

Constipation • FREQUENCY 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 anti-depressants, 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 9-5). To prevent constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be

TABLE 9-5 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.

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.

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, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT3) 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 either metoclopramide, a serotonin type 3 (5-HT3) receptor antagonist like 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-HT3 receptor antagonists or neurokinin-1 antagonists, such as aprepitant or fosaprepitant, is recommended. Clinicians should attempt prevention of post-chemotherapy 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 non-pharmacological interventions, such as biofeedback and hypnosis. The most common pharmacological 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-HT3 receptor antagonists and can sometimes even cause cannabis hyperemesis syndrome. Older patients—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, congestive heart failure 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 P_{O_2} , P_{CO_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 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 9-6**). 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 neither be used as first-line therapy nor 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.

Fatigue • FREQUENCY Fatigue is one of the most commonly reported symptoms of not only 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 among the most distressing symptoms.

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's question "How much of the day does the patient spend in bed?" may be the best measure. In this 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 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 per 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

TABLE 9-6 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
Oxycodone	30–50% of baseline opioid dose q4h	
Hydromorphone	5–10 mg PO q4h	
Anxiolytics	1–2 mg PO q4h	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	

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

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. Palliative sedation is used in distressing situations that cannot be addressed in other ways. It can be abused if done to hasten death (which it usually does not), when at the request of the family, rather than 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, 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 per 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, most studies suggest that depressed patients at the end of life are neither diagnosed, nor even properly treated if diagnosed. 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*) 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 either with the Patient Health Questionnaire-9 (PHQ-9) or the PHQ-2 which asks "Over the past two 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: Not at all, Several days, More than half the days, Nearly every day. There are other possible diagnostic tools such as the short form of the Beck Depression Index or a visual analog 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.

INTERVENTIONS Under-treatment 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.

While there are no 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 per 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. 24) • 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 9-7**). 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 anti-nausea, 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 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

TABLE 9-7 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

drugs are used to treat terminal delirium. If patients develop dystonic reactions, benzotropine should be administered. Neuroleptics may be combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

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 the patient's violence is threatening to himself 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 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, steroids, 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 which 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. In the United States, which has one of the least comprehensive health insurance systems among developed 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 >\$8,000. 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, recent 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 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 last 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' home. 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 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. In 2014, 14.5% of hospice care was provided 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. Finally, the Center for Medicare and Medicaid Innovation (CMMI) is testing concurrent hospice and palliative care services with curative treatment with ~120 providers.

By 2014, the mean length of enrollment in a hospice was 71 days, with the median being 17 days. 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, as leading physicians have increasingly emphasized the need to introduce palliative care much earlier in patients' illness, 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. For instance, some companies and home health agencies are offering non-hospice palliative care services in patients' homes in an effort to increase quality of life and forestall hospitalizations. Similarly, palliative care services are increasingly available via consultation, rather than present only in hospital, day care, outpatient, and nursing home settings. Palliative care consultations for non-hospice 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 developed countries is that patients have a moral as well as constitutional or common law 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 in which 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 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 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; 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 $\mu\text{g}/\text{h}$ 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 AMA 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 DNR" 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 Futility 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 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, about 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 transferred to other facilities or died while awaiting transfer.

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
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, Switzerland, Oregon, Washington, Montana, Vermont, California

EUTHANASIA AND PHYSICIAN-ASSISTED SUICIDE

Euthanasia and PAS are defined in **Table 9-8**. 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.

LEGAL ASPECTS Euthanasia and PAS are legal in the Netherlands, Belgium, Luxembourg, Colombia, and Canada. 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 can legally elect assisted suicide. In the United States, PAS is legal in five states: Oregon, Washington State, Montana, Vermont, and California. 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 Australia, 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 relatively rare. In all countries, even the Netherlands and Belgium where these practices have been tolerated and legal for many years, fewer than 5% of deaths occur by euthanasia or PAS. As of the most recent data, the share of deaths attributable to euthanasia or PAS was 2.9% in the Netherlands (2010) and 4.6% in Belgium (2013). In 2015, 0.39% of all deaths in Oregon and 0.31% of all deaths in Washington State were reported to be by PAS, although these may be underestimates.

In the Netherlands, Belgium, 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.

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. Only 25% 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 family members appear to be more important factors motivating a desire for euthanasia or PAS. Losing autonomy (91% Oregon, 90% Washington), not being able to enjoy activities (89% OR, 89% WA), or fear of losing dignity (68% OR, 76% WA) are the most cited end of life concerns in both states. Over a third of patients seeking PAS note being a burden on family (41% OR, 53% 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 indicate that between 1998 and 2015, 53% of cases had no complications, 44% of patients had no data on complications, and 2.4% of cases had regurgitation after taking the prescribed medicine as the only complication. In addition, six patients awakened and the reported range of time to death extended to 104 h. In Washington State between 2014 and 2015, 1.4% of cases had regurgitation, 1 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 empathetic, 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 controversial 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 9-9**). 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.

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.

TABLE 9-9 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 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.

(Continued)

TABLE 9-9 Managing Changes in the Patient's Condition during the Final Days and Hours (Continued)

CHANGES IN THE PATIENT'S CONDITION	POTENTIAL COMPLICATION	FAMILY'S POSSIBLE REACTION AND CONCERN	ADVICE AND INTERVENTION
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 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 modify 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.

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.
- KELLEY AS, MEIER DE: Palliative care —A shifting paradigm. *N Engl J Med* 363:781, 2010.
- KELLEY AS et al: Hospice enrollment saves money for medicare and improves care quality across a number of different lengths-of-stay. *Health Affairs* 32:552, 2012.
- KELLEY AS et al: Palliative care for the seriously ill. *N Engl J Med* 373:747, 2015.
- 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.
- MURRAY SA et al: Illness trajectories and palliative care. *BMJ* 330:1007, 2005.
- NEUMAN P et al: Medicare per capita spending by age and service: New data highlights oldest beneficiaries. *Health Aff (Millwood)* 34:335, 2015.
- NICHOLAS LH et al: Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA* 306:1447, 2011.

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.

VAN DEN BEUKEN-VAN EVERDINGEN MH et al: Update on prevalence of pain in patients with cancer: Systematic review and meta-analysis. *J Pain Symptom Manage* 51:1070, 2016.

WEBSITES

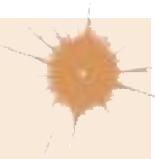
- American Academy of Hospice and Palliative Medicine: www.aahpm.org
 Center to Advance Palliative Care: <http://www.capc.org>
 Education in Palliative and End of Life Care (EPEC): <http://www.epec.net>
 End of Life—Palliative Education Resource Center: <http://www.eperc.mcw.edu>
 Family Caregiver Alliance: <http://www.caregiver.org>
 The Medical Directive: <http://www.medicaldirective.org>
 National Family Caregivers Association: <http://www.nfcacares.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>

Section 1 Pain

10

Pain: Pathophysiology and Management

James P. Rathmell, Howard L. Fields



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. 10-1). 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 (A β), 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 (A δ) and the unmyelinated (C) axons (Fig. 10-1). 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 A δ and C fiber afferents. Most A δ and C fiber afferents respond maximally only to intense (painful) stimuli and produce the subjective experience of pain when they are electrically stimulated; this defines them as *primary afferent nociceptors (pain receptors)*. The ability to detect painful stimuli is completely abolished when conduction in A δ and C fiber axons is blocked.

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 acidic pH, endogenous mediators and by 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 on the back 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

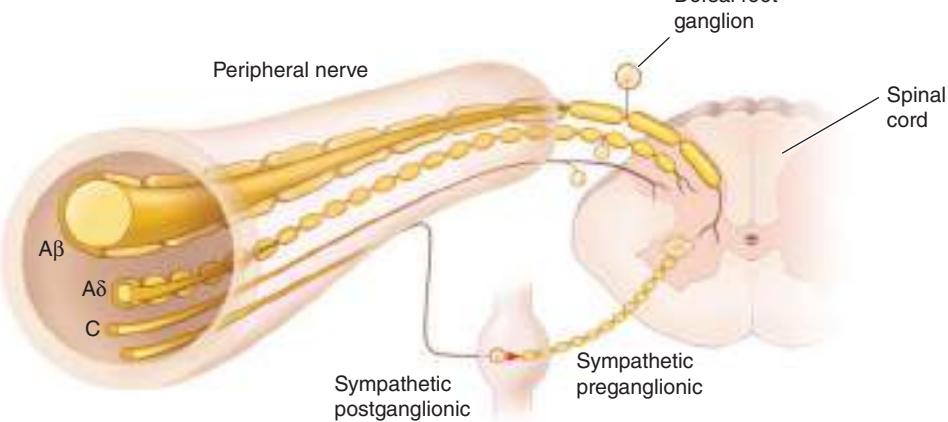


FIGURE 10-1 Components of a typical cutaneous nerve. There are two distinct functional categories of axons: primary afferents with cell bodies in the dorsal root ganglion and sympathetic postganglionic fibers with cell bodies in the sympathetic ganglion. Primary afferents include those with large-diameter myelinated (A β), small-diameter myelinated (A δ), and unmyelinated (C) axons. All sympathetic postganglionic fibers are unmyelinated.

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 A δ 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 also have a neuroeffector function. Most nociceptors contain polypeptide mediators that are released from their peripheral terminals when they are activated (Fig. 10-2). An example is substance P,

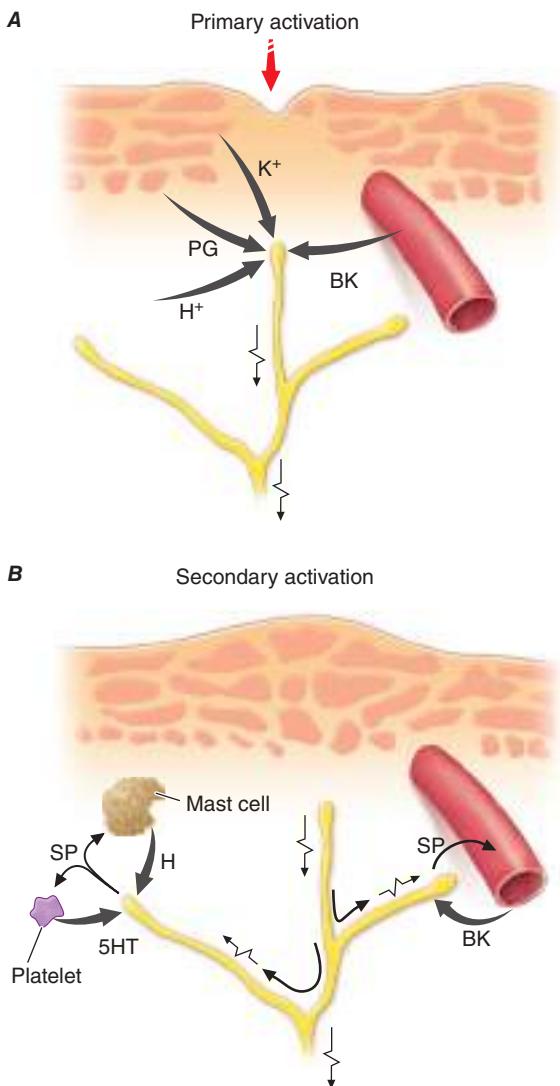


FIGURE 10-2 Events leading to activation, sensitization, and spread of sensitization of primary afferent nociceptor terminals. **A.** Direct activation by intense pressure and consequent cell damage. Cell damage induces lower pH (H^+) and leads to release of potassium (K^+) and to synthesis of prostaglandins (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.

an 11-amino-acid peptide. Substance P is released 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. Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through these neuroeffector functions.

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. 10-3). 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 peptides, including substance P and calcitonin gene-related peptide, 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.

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 mislocalized by the patient to a place 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

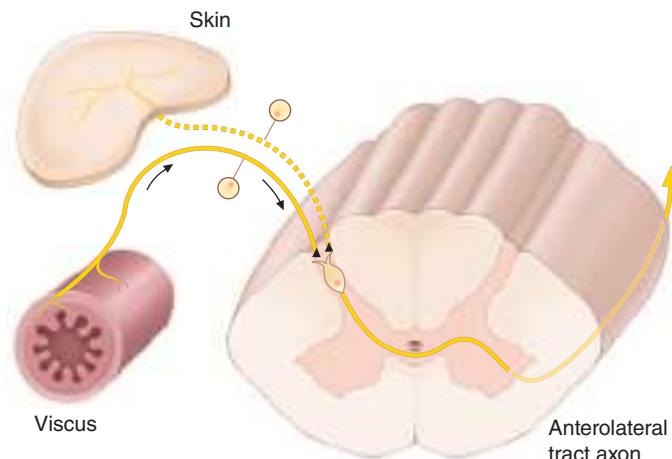


FIGURE 10-3 The convergence-projection hypothesis of referred pain. According to this hypothesis, visceral afferent nociceptors converge on the same pain-projection neurons as the afferents from the somatic structures in which the pain is perceived. The brain has no way of knowing the actual source of input and mistakenly “projects” the sensation to the somatic structure.

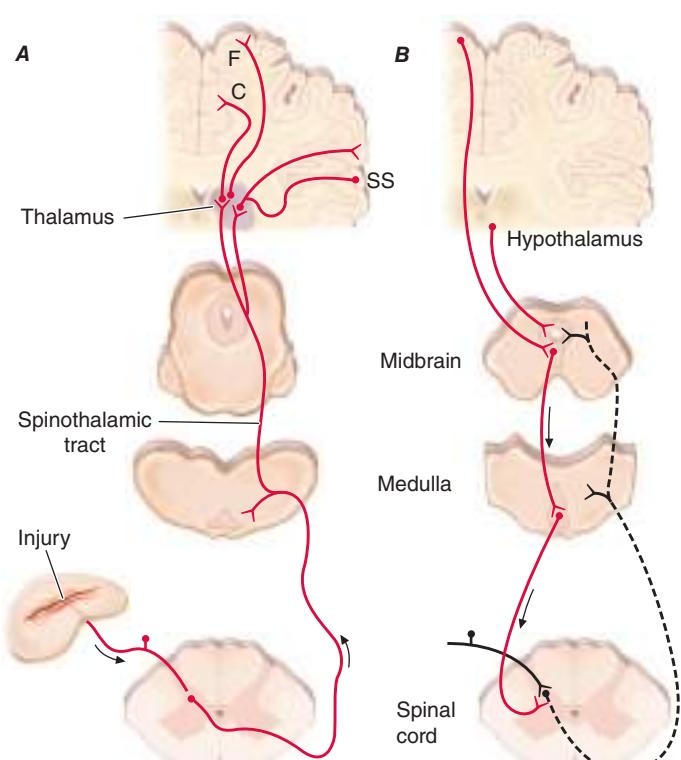


FIGURE 10-4 Pain-transmission and modulatory pathways. **A.** Transmission system for nociceptive messages. Noxious stimuli activate the sensitive peripheral ending of the primary afferent nociceptor by the process of transduction. The message is then transmitted over the peripheral nerve to the spinal 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.

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. 10-4). One of the thalamic projections is to the somatosensory cortex. This projection mediates the purely sensory 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 gyrus and other areas of the frontal lobes, including the 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.

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 spinal pain-transmission neurons through a descending pathway (Fig. 10-4).

Human brain-imaging studies have implicated this pain-modulating circuit in the pain-relieving effect of attention, suggestion, and opioid analgesic medications (Fig. 10-5). 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.

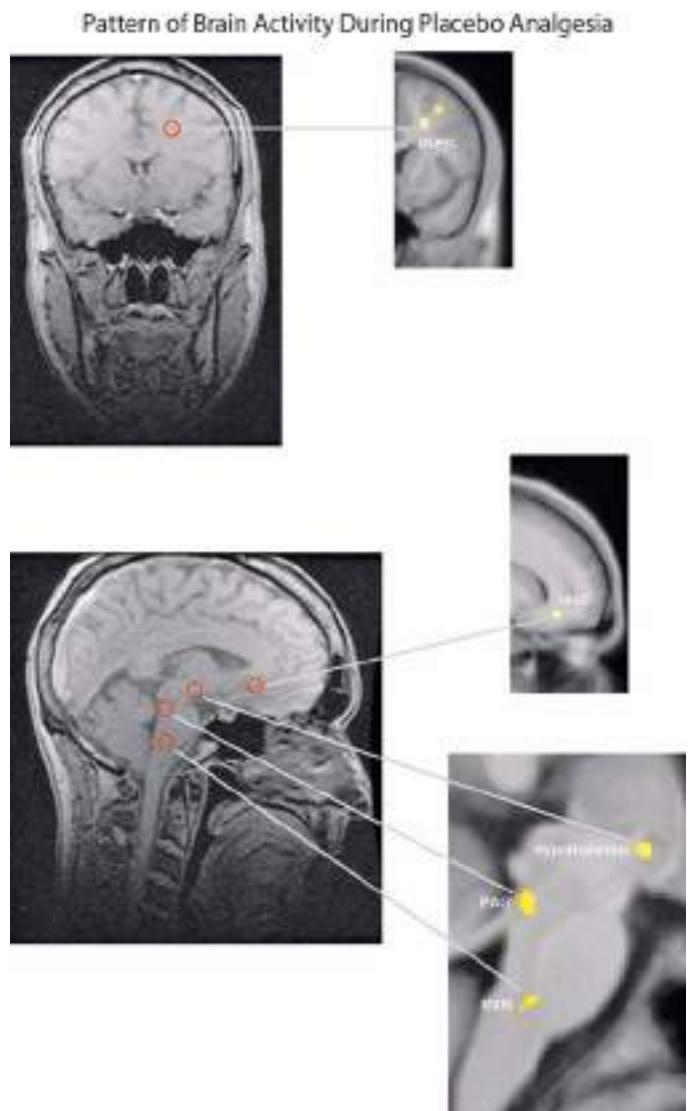


FIGURE 10-5 Functional magnetic resonance imaging (fMRI) demonstrates placebo-enhanced brain activity in anatomic regions correlating with the opioidergic descending pain control system. Top panel: Frontal fMRI image shows placebo-enhanced brain activity in the dorsal lateral prefrontal cortex (DLPFC). 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 et al: *Neuron* 63:533, 2009.)

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 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 are typically 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 present in 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 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. 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 post-traumatic 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 10-1**). 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 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 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

TABLE 10-1 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					
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					
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 5-HT	NE	SEDATIVE POTENCY	ANTICHOLINERGIC POTENCY	ORTHOSTATIC HYPOTENSION	CARDIAC ARRHYTHMIA	AVE. DOSE, mg/d	RANGE, mg/d
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	GENERIC NAME	PO DOSE, mg	INTERVAL			
Anticonvulsants and Antiarrhythmics ^a								
Phenytoin	300	daily/qhs	Clonazepam	1	q6h			
Carbamazepine	200–300	q6h	Gabapentin ^b	600–1200	q8h			
Oxcarbazepine	300	bid	Pregabalin	150–600	bid			

^aAntidepressants, anticonvulsants, and antiarrhythmics have not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of pain. ^bGabapentin in doses up to 1800 mg/d is FDA approved for postherpetic neuralgia.

Abbreviations: 5-HT, serotonin; NE, norepinephrine; NSAID, nonsteroidal anti-inflammatory agent.

coagulation. Nonselective COX inhibitors 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 method for rapid pain relief. Although side effects are common, most are reversible: nausea, vomiting, pruritus, 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 10-1 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 typically accompanied by sedation and a reduction in respiratory rate. 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, most commonly the benzodiazepines. Because of this, 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 frequent 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 μ g 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 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.

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 usually 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 organic factors before initiating therapy. Addressing these issues together, rather than waiting to address emotional issues after organic 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 an organic 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 looked for 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 is unresponsive to more conservative treatment, placement of electrodes within the spinal canal 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 10-1), 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 10-2** 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 10-1; [Chap. 444](#)). 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.

TABLE 10-2 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

^aControlled trials demonstrate analgesia. ^bControlled studies indicate benefit but not analgesia.

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. 433**). 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.

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 non-malignant 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 physically dependent on other opioid analgesics.

With long-term outpatient use of orally administered opioids, it may be desirable to use long-acting compounds such as levorphanol, methadone, sustained-release morphine, or transdermal fentanyl (Table 10-1). 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. 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 above 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 analgesics.

Soon after the introduction of a controlled-release oxycodone formulation (OxyContin) in the late 1990s, a dramatic rise in emergency department visits and deaths associated with oxycodone ingestion appeared, focusing public attention on misuse of prescription pain medications. The magnitude of prescription opioid abuse has grown over the last decade, leading the Centers for Disease Control and Prevention to classify prescription opioid analgesic abuse as an epidemic. This appears to be due in large part to individuals using a prescription drug nonmedically, most often an opioid analgesic. 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 (CDC) released the *CDC Guideline for Prescribing Opioids for Chronic Pain*, with recommendations for primary care clinicians who

TABLE 10-3 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 et al: J Pain 10:113, 2009.

are prescribing opioids for chronic noncancer. The guideline is based on the best available scientific evidence and addresses (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 10-3**; a checklist for primary care clinicians prescribing opioids for noncancer pain is shown in **Table 10-4**.

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. 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

TABLE 10-4 Centers for Disease Control 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	
<ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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	
<ul style="list-style-type: none"> • Check that return visit is scheduled ≤3 months from last visit. 	
WHEN REASSESSING AT A PATIENT VISIT	
<ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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). <ul style="list-style-type: none"> • 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, Available at: <https://stacks.cdc.gov/view/cdc/38025>, accessed May 25, 2017 (Public Domain).

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 very common. 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. 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.

FURTHER READING

- DOWELL D et al: CDC guideline for prescribing opioids for chronic pain—United States, 2016. *JAMA* 315:1624, 2016.
 FINNERUP NB et al: Pharmacotherapy for neuropathic pain in adults: A systematic review and meta-analysis. *Lancet Neurol* 14:162, 2015.
 SUN EC et al: Incidence of and risk factors for chronic opioid use among opioid-naïve patients in the postoperative period. *JAMA Intern Med* 176:1286, 2016.

11

Chest Discomfort

David A. Morrow



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 (pericardial disease, aortic emergencies, and pulmonary conditions); and (3) non-cardiopulmonary 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 the third 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 the remainder undergo additional investigation in the ED. As few as 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. 11-1), and fewer than 10% 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, a disconcerting 2% to 6% of patients with chest discomfort of presumed non-ischemic etiology who are discharged from the ED are 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.

The natural histories of ACS, acute pericardial diseases, pulmonary embolism, and aortic emergencies are discussed in Chaps. 265, 268, 269, 273, and 274, respectively. In a study of >350,000 patients with unspecified presumed non-cardiopulmonary 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.

CAUSES OF CHEST DISCOMFORT

The major etiologies of chest discomfort are discussed in this section and summarized in Table 11-1. 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”).

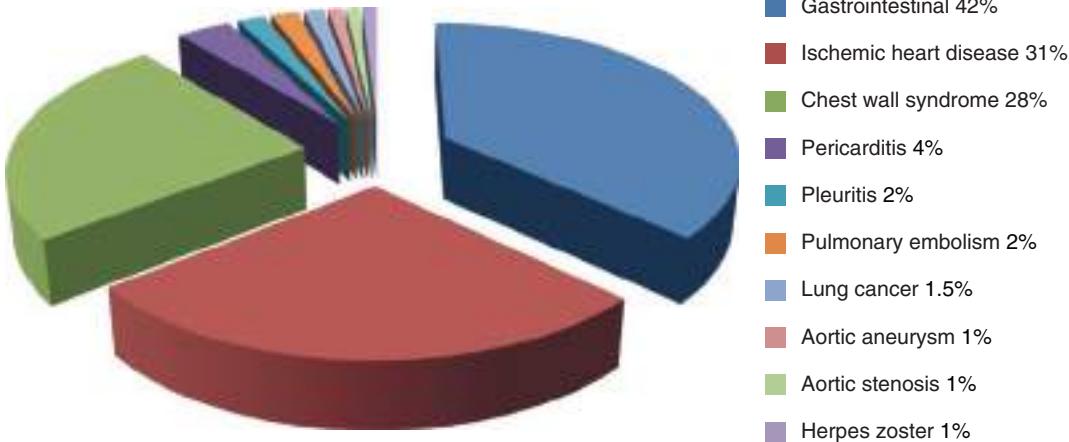


FIGURE 11-1 Distribution of final discharge diagnoses in patients with nontraumatic acute chest pain. (Figure prepared from data in P Fruergaard et al: Eur Heart J 17:1028, 1996.)

TABLE 11-1 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 <i>Unstable angina:</i> Increasing pattern or at rest <i>Myocardial infarction:</i> 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					
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
Non-cardiopulmonary					
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 atherosomatous plaque that obstructs one or more of the epicardial coronary arteries. Stable ischemic heart disease (Chap. 267) 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 most commonly when rupture or erosion of one or more atherosclerotic lesions triggers coronary thrombosis. Unstable ischemic heart disease is classified clinically by the presence or absence of detectable 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. Marked by ischemic symptoms at rest, with minimal activity, or in an accelerating pattern, unstable ischemic heart disease is classified as *unstable angina* when there is no detectable myocardial injury and as *non-ST elevation MI* (NSTEMI) when there is evidence of myocardial necrosis (Chap. 268). 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. 269).

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) from MI occurring secondary to other imbalances of myocardial oxygen supply and demand (type 2; see Chap. 268).

Other contributors to stable and unstable ischemic heart disease, such as endothelial dysfunction, microvascular disease, and vasoconstriction, may exist alone or in combination with coronary atherosclerosis and may be the dominant cause of myocardial ischemia in some patients. Moreover, non-atherosclerotic 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. 274), hypertrophic cardiomyopathy, or idiopathic dilated cardiomyopathy (Chap. 254), 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 11-1). 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 typically 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 (See also Chap. 265) Inflammation of the pericardium due to infectious or noninfectious causes can be responsible for acute or chronic chest discomfort. 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 pericarditis 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 non-ischemic myocardial diseases can also produce chest discomfort. 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. 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.

Diseases of the Aorta (See also Chap. 274) Acute aortic dissection (Fig. 11-1) 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. 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 atherosomatous 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 (SEE ALSO CHAP. 273) Pulmonary emboli (annual incidence, ~1 per 1000) can produce dyspnea and chest discomfort that is sudden in onset. 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”).

PNEUMOTHORAX (SEE ALSO CHAP. 289) 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. 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 (See also Chaps. 277, 278, and 288) Most pulmonary

diseases that produce chest pain, including pneumonia and malignancy, do so because of involvement of the pleura or surrounding structures. 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.

■ NON-CARDIOPULMONARY CAUSES

Gastrointestinal Conditions (See also Chap. 314) 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. 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. 11-1 and Table 11-1). Acid reflux often causes a burning discomfort. The pain of esophageal spasm, in contrast, 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 syndrome) caused by severe vomiting. Peptic ulcer disease is most commonly epigastric in location but can radiate into the chest (Table 11-1).

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 (See also Chap. 363)

Chest discomfort can be produced by any musculoskeletal disorder involving the chest wall or the nerves of the chest wall, neck, or upper limbs. 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. 188).

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 11-1). The symptoms may include chest tightness or aching that is associated with a sense of anxiety and difficulty in breathing. The symptoms may be prolonged or fleeting.

APPROACH TO THE PATIENT

Chest Discomfort

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. Among non-cardiopulmonary causes

TABLE 11-2 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*.

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 non-cardiopulmonary conditions has a more favorable prognosis during completion of the diagnostic work-up. 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 11-2**.

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 11-1), including ACS in particular (Fig. 11-2). 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 severe, 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—for example, 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. However, 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 “warm-up angina” in which some patients experience relief from angina as they continue at the same or even a greater level of exertion (Chap. 267). 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 by 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

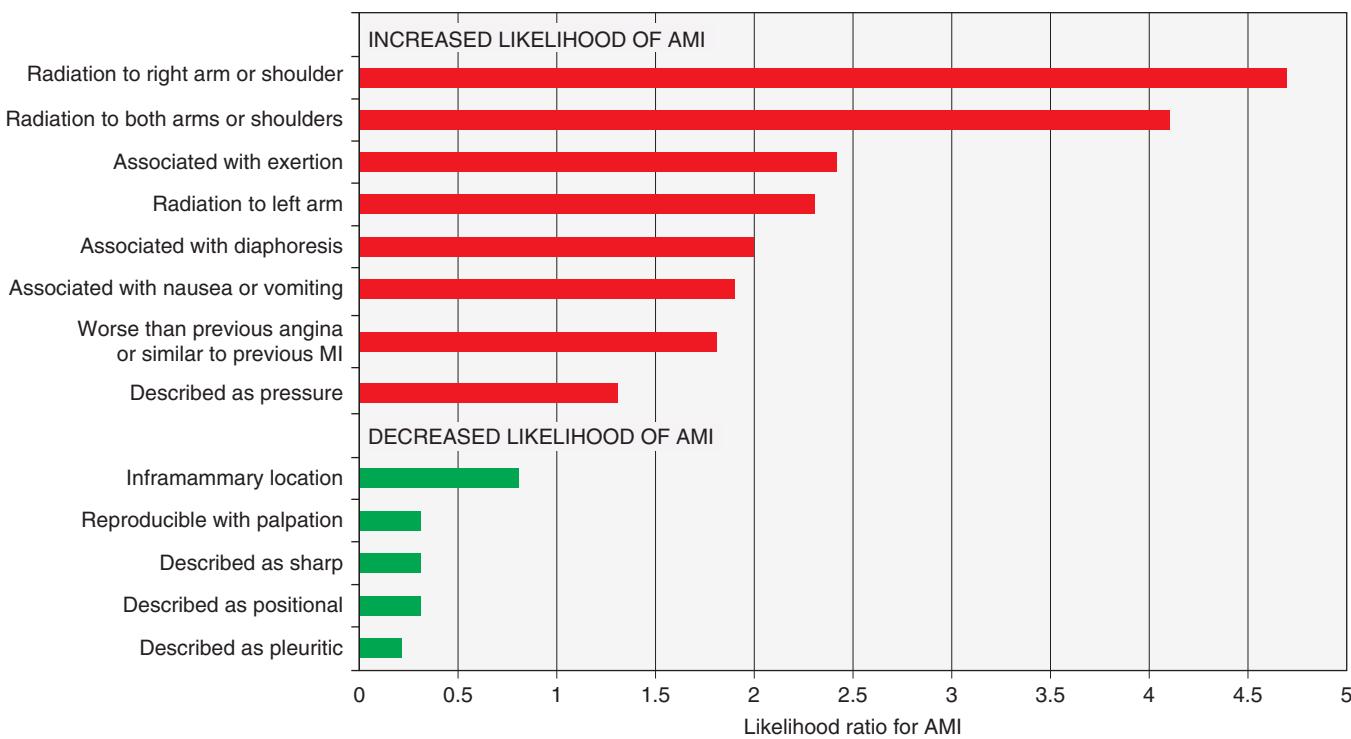


FIGURE 11-2 Association of chest pain characteristics with the probability of acute myocardial infarction (AMI). Note that a subsequent larger study showed a non-significant association with radiation to the right arm. (Figure prepared from data in CJ Swap, JT Nagurney: JAMA 294:2623, 2005.)

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 pre-syncope 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. 273) 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 examination 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—for example, 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 non-specific 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. 234 and 265). 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 proximal 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 examination 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 underlying chronic cardiopulmonary disease or severe right ventricular dysfunction leading to hepatic congestion.

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, though 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. 235 and 273**). 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 patients with

pulmonary embolism (**Chaps. 273 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 myocardial cells. Owing to the time necessary for this release, initial biomarkers of injury may be in the normal range, even in patients with STEMI. Because of superior cardiac tissue-specificity compared with creatine kinase MB, cardiac troponin is the preferred biomarker for the diagnosis of MI and should be measured in all patients with suspected ACS at presentation and repeated in 3–6 h. Testing after 6 h is required only when there is uncertainty regarding the onset of pain or when stuttering symptoms have occurred. 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, enhances the overall accuracy of a diagnosis of MI, and improves risk stratification in suspected ACS. The greater negative predictive value of a negative troponin result with current-generation 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 promising and have been adopted in some centers where high-sensitivity assays for troponin are used routinely. In patients presenting >2 h after symptom onset, a concentration of cardiac troponin below the limit of detection using a high-sensitivity assay may be sufficient to exclude MI with a negative predictive value >99% at the time of hospital presentation. However, with these advantages has come a trade-off: myocardial injury is detected in a larger proportion of patients who have non-ACS cardiopulmonary conditions than with previous, less sensitive assays. This evolution in testing for myocardial necrosis has rendered other aspects of the clinical evaluation 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 useful in discriminating acute causes of myocardial injury from chronic elevation due to underlying structural heart disease, end-stage renal disease, or 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 non-ischemic insults, such as myocarditis, may result in myocardial injury but should not be labeled MI.

Other laboratory assessments may include the D-dimer test to aid in exclusion of pulmonary embolism (**Chap. 273**). Measurement of a B-type natriuretic peptide is useful when considered in conjunction with the clinical history and examination 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 either for early provocative testing for ischemia or for discharge from the ED. 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



FIGURE 11-3 Examples of decision-aids used in conjunction with serial measurement of cardiac troponin for evaluation of acute chest pain. (Figure prepared from data in SA Mahler et al: *Int J Cardiol* 168:795, 2013.)

ranging from a <1% to a >16% probability of a major cardiovascular complication. The Acute Cardiac Ischemia Time-Insensitive Predictive Instrument (ACI-TIPI) combines age, sex, chest pain presence, and ST-segment abnormalities to define a probability of ACS. More recently developed decision-aids are shown in Fig. 11-3. Elements common to each of these 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 levels. 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), they can help identify patients with a very low probability of ACS (e.g., <1%). Nevertheless, no such decision-aid (or single clinical factor) is sufficiently sensitive and well validated to use as a sole tool for clinical decision-making.

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. 269) in patients who already have an established diagnosis of ACS. The latter risk scores were not designed to be used for diagnostic assessment.

PROVOCATIVE TESTING FOR ISCHEMIA

Exercise electrocardiography ("stress testing") is 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 them as being at low or selectively intermediate risk of ACS. Early exercise testing is safe in patients without high-risk findings after 8–12 h of observation and can assist in refining their prognostic assessment. For example, of low-risk patients who underwent exercise testing in the first 48 h after presentation, those without evidence of ischemia had a 2% rate of cardiac events through 6 months, whereas the rate was 15% among patients with either clear evidence of ischemia or an equivocal result. Patients who are unable to exercise may undergo pharmacological stress

testing with either nuclear perfusion imaging or echocardiography. Notably, some experts have deemed the routine use of stress testing for low-risk patients unsupported by direct clinical evidence and a potentially unnecessary source of cost.

Professional society guidelines identify ongoing chest pain as a contraindication to stress testing. 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 some centers, early myocardial perfusion imaging is performed as part of a routine strategy for evaluating patients at low or intermediate risk of ACS in parallel with other testing. Management of patients with normal perfusion images can be expedited with earlier discharge and outpatient stress testing, if indicated. Those with abnormal rest perfusion imaging, which cannot discriminate between old or new myocardial defects, usually warrant additional in-hospital evaluation.

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 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.

CT Angiography (See Chap. 236) CT angiography is emerging as a modality for the evaluation of patients with acute chest discomfort. Coronary CT angiography is a sensitive technique for detection of obstructive coronary disease, particularly in the proximal third of the major epicardial coronary arteries. CT appears to enhance the speed to disposition of patients with a low-intermediate probability for ACS; its major strength being the negative predictive value of a finding of no significant disease. 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. Balancing factors in the consideration of the emerging role of coronary CT angiography in low-risk patients are radiation exposure and additional testing prompted by nondiagnostic abnormal results.

MRI (See Chap. 236) Cardiac magnetic resonance (CMR) imaging is an evolving, versatile technique for structural and functional evaluation of the heart and the vasculature of the chest. 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 non-ischemic 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.

■ 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 exercise 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.

■ 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.

FANAROFF AC et al: Does this patient with chest pain have acute coronary syndrome? *JAMA* 314:1955, 2015.

HERMANN LK et al: Yield of routine provocative cardiac testing among patients in an emergency department-based chest pain unit. *JAMA Int Med* 173:1128, 2013.

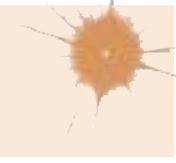
MAHLER SA et al: The HEART Pathway randomized trial: Identifying emergency department patients with acute chest pain for early discharge. *Circulation Cardiovasc Qual Outcomes* 8:195, 2015.

SHAH AS et al: High-sensitivity cardiac troponin I at presentation in patients with suspected acute coronary syndrome: A cohort study. *Lancet* 386:2481, 2016.

12

Abdominal Pain

Danny O. Jacobs



Correctly interpreting 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 12-1**).

The etiologic classification in **Table 12-2**, although not complete, provides a useful framework for evaluating patients with abdominal pain.

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 intra-abdominal 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

TABLE 12-1 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

TABLE 12-2 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
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	

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 a significant amount 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 its 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 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. 10.**

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 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. 362**).

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. It 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. 14**). 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, are either relieved 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. 320**) 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 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. 44). 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. 320). The location of the pain can assist in narrowing the differential diagnosis (Table 12-3); 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.

TABLE 12-3 Differential Diagnoses of Abdominal Pain by Location

Right Upper Quadrant	Epigastric	Left Upper Quadrant
Cholecystitis	Peptic ulcer disease	Splenic infarct
Cholangitis	Gastritis	Splenic rupture
Pancreatitis	GERD	Splenic abscess
Pneumonia/empyema	Pancreatitis	Gastritis
Pleurisy/pleurodynia	Myocardial infarction	Gastric ulcer
Subdiaphragmatic abscess	Pericarditis	Pancreatitis
Hepatitis	Ruptured aortic aneurysm	Subdiaphragmatic abscess
Budd-Chiari syndrome	Esophagitis	
Right Lower Quadrant	Perumbilical	Left Lower Quadrant
Appendicitis	Early appendicitis	Diverticulitis
Salpingitis	Gastroenteritis	Salpingitis
Inguinal hernia	Bowel obstruction	Inguinal hernia
Ectopic pregnancy	Ruptured aortic aneurysm	Ectopic pregnancy
Nephrolithiasis		Nephrolithiasis
Inflammatory bowel disease		Irritable bowel syndrome
Mesenteric lymphadenitis		Inflammatory bowel disease
Typhlitis		
Diffuse Nonlocalized Pain		
Gastroenteritis	Malaria	
Mesenteric ischemia	Familial Mediterranean fever	
Bowel obstruction	Metabolic diseases	
Irritable bowel syndrome	Psychiatric disease	
Peritonitis		
Diabetes		

Abbreviation: GERD, gastroesophageal reflux disease.

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 briskly, 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 $>20,000/\mu\text{L}$ 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 intervention are not clear, low dose

computed tomography is preferred to abdominal radiography when evaluating non-traumatic acute abdominal pain.

Very rarely, barium or water-soluble contrast study of the upper part of the gastrointestinal tract are appropriate radiographic investigations 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, and 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.

ACKNOWLEDGMENT

We gratefully acknowledge the enormous contribution to this chapter and the approach it espouses to 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.
- 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 with abdominal pain. Available from <http://dx.doi.org/10.1016/j.jem.2016.11.027>. Accessed November 2016.
- NAYOR J et al: Tracing the cause of abdominal pain. *N Engl J Med* 375:e8, 2016.
- 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.

TABLE 13-1 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.

focus on the general approach to a patient with headache; migraine and other primary headache disorders are discussed in **Chap. 422**.

GENERAL PRINCIPLES

A classification system developed by the International Headache Society (www.ihf-headache.org/ichd-guidelines) characterizes headache as primary or secondary (**Table 13-1**). 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.

ANATOMY AND PHYSIOLOGY OF HEADACHE

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors (**Chap. 10**). 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 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 appear to be 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 trigeminal nociceptors at all levels of the pain-processing pathways and influence vegetative functions, such as hypothalamus and brainstem structures

The innervation of the large intracranial vessels and dura mater by the trigeminal nerve is known as the *trigeminovascular system*. Cranial autonomic symptoms, such as *lacrimation*, *conjunctival injection*, *nasal congestion*, *rhinorrhea*, *periorbital swelling*, *aural fullness*, and *ptosis*, are prominent in the trigeminal autonomic cephalgias (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. Moreover, they can often be mistaken for symptoms or signs of cranial sinus inflammation, which is thus overdiagnosed and inappropriately managed. Migraine and other primary headache types are

13

Headache

Peter J. Goadsby



Headache is among the most common reasons patients seek medical attention, on a global basis being responsible 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

TABLE 13-2 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, cough
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

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 13-2**), rapid diagnosis and management are critical.

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 funduscopic, intraocular pressure measurement, and refraction.

The psychological state of the patient 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

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. 133 and 134.

■ 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. 302, and intracranial hemorrhage in Chap. 421.

■ 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 the 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, or both. 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. 86.

■ TEMPORAL ARTERITIS

(See also Chaps. 28 and 356) Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation. 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. 356), 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 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. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headaches 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. 28.

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. 422**.

■ 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 not a single entity; it encompasses a number of different headache syndromes, both primary and secondary (**Table 13-3**). In aggregate, this group presents considerable disability and is thus specially dealt with here. Population-based estimates suggest that about 4% of adults have daily or near-daily headache.

APPROACH TO THE PATIENT

Chronic Daily Headache

The first step in the management of patients with CDH is to diagnose any secondary headache and treat that problem (**Table 13-3**). This can sometimes be a challenge where the underlying cause triggers a 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 excess morning

sleepiness. Medicines including topiramate, valproate, propranolol, flunarizine (not available in the United States), and candesartan are also useful in migraine.

MANAGEMENT OF MEDICALLY INTRACTABLE DISABLING PRIMARY CHRONIC DAILY HEADACHE

The management of medically intractable headache is difficult, although developments in therapy are at hand. Monoclonal antibodies to calcitonin gene-related peptide (CGRP) or its receptor have been reported to be effective and well-tolerated in chronic migraine in phase II/III randomized placebo-controlled trials. Non-invasive neuromodulatory approaches, such as single pulse transcranial magnetic stimulation and non-invasive vagal nerve stimulation, which appear to modulate thalamic processing or brainstem mechanisms, respectively, in migraine have, or are, entering clinical practice, respectively. Non-invasive vagal nerve stimulation has also shown promise in chronic cluster headache, chronic paroxysmal hemicrania, short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA), short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT), and hemicrania continua (**Chap. 422**). Other modalities are discussed in **Chap. 422**.

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 medications, it is often helpful that analgesic use be reduced and eliminated. 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 the 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 a NSAID with a longer half-life is taken once or twice daily; however, overuse problems may develop with more frequent dosing schedules or shorter acting NSAIDs. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced, although another equally widely used approach is to commence the preventive at the same time as the analgesic reduction is started. It must be emphasized that *preventives often do not work in the presence of analgesic overuse*. 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 pain 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, acute 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

TABLE 13-3 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 New daily persistent headache ^a	SUNCT/SUNA Hypnic headache	Chronic CNS infection Medication-overuse headache ^a

^aMay be complicated by medication overuse. ^bSome patients may have headache >4 h/d.

Abbreviations: CNS, central nervous system; SUNA, short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms; SUNCT, short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing.

opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in 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, can induce a significant remission that allows a preventive treatment to be established. Serotonin 5-HT₃ 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 13-4**.

Clinical Presentation The patient with 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. 302**).

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 1 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. 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. 432**) can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration in this setting.

TABLE 13-4 Differential Diagnosis of New Daily Persistent Headache

PRIMARY	SECONDARY
Migrainous-type Featureless (tension-type)	Subarachnoid hemorrhage Low cerebrospinal fluid (CSF) volume headache Raised CSF pressure headache Posttraumatic headache ^a Chronic meningitis

^aIncludes postinfectious forms.

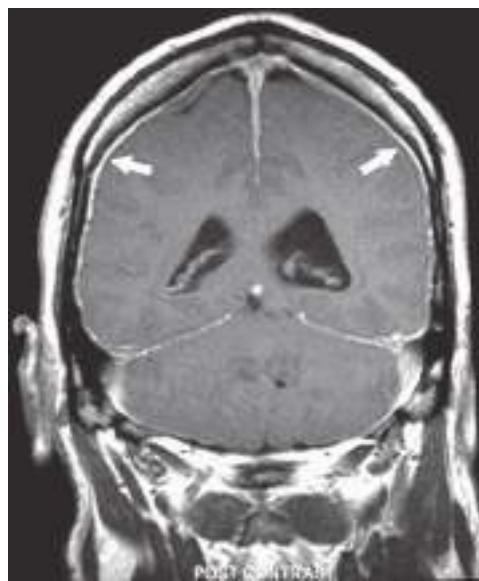


FIGURE 13-1 Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low cerebrospinal fluid (CSF) volume headache.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice (**Fig. 13-1**). 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, by CT, or increasingly by MR myelography. Less used now, ¹¹¹In-DTPA CSF studies in the absence of a directly identified site of leakage, may demonstrate early emptying of ¹¹¹In-DTPA tracer into the bladder or slow progress of tracer across the brain suggesting a CSF leak.

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; however, its effect is less rapid than caffeine.

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. NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathic intracranial hypertension (pseudotumor cerebri) without visual problems, particularly when the fundi are normal. Persistently raised intracranial pressure can trigger chronic migraine. 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 worse with recumbency. Visual obscurations are frequent. 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. Headache on rising in the morning or nocturnal headache is also characteristic of 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.

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, 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.

PRIMARY CARE AND HEADACHE MANAGEMENT

Most patients with headache will be seen first in a primary care setting. The task of the primary care physician is to identify the very few worrisome secondary headaches from the very great majority of primary and less troublesome secondary headaches (Table 13-2).

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 gaining 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 with a complaint of headache will have migraine (**Chap. 422**).

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. Cephalgia 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 D: *Wolff's Headache and Other Head Pain*, 8th ed. New York, Oxford, 2008.

14

Back and Neck Pain

John W. Engstrom



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 \$177 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 held together 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. 14-1 and 14-2**). Desiccation of the nucleus pulposus and degeneration of the annulus fibrosus increase 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.

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 that results in flexion, extension, and lateral bending movements of the spine.

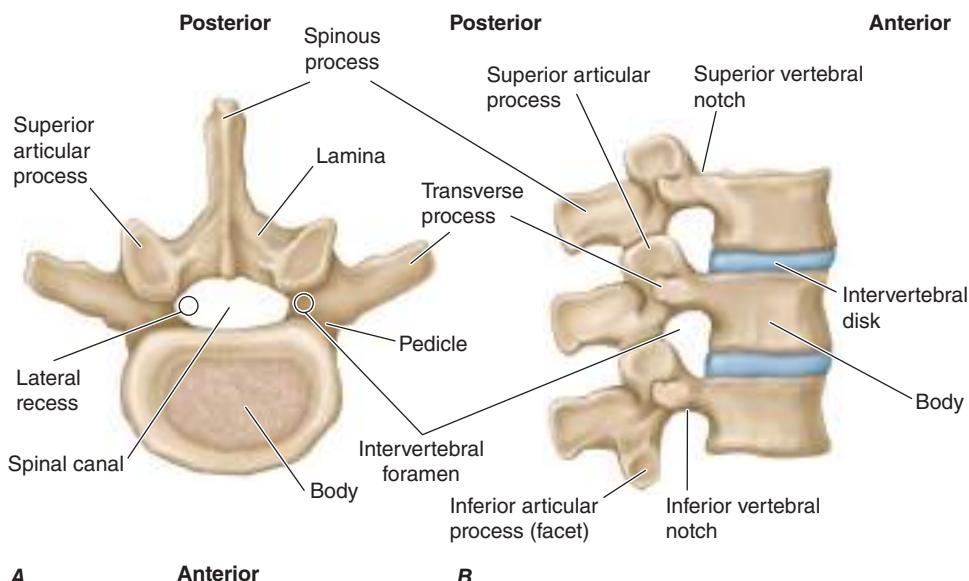


FIGURE 14-1 Vertebral anatomy. (From A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998; with permission.)

Nerve root injury (*radiculopathy*) is a common cause of neck and arm, or low back and buttock or leg, pain (see **dermatomes in Figs. 22-2 and 22-3**). The nerve roots exit at a level above their respective vertebral bodies in the cervical region (e.g., the C7 nerve root exits at the C6-C7 level) and below their respective vertebral bodies in the thoracic and lumbar regions (e.g., the T1 nerve root exits at the T1-T2 level). The cervical nerve roots follow a short intraspinal course before exiting. By contrast, because the spinal cord ends at the vertebral L1 or L2 level, the lumbar nerve roots follow a long intraspinal course and can be injured anywhere from the upper lumbar spine to the intervertebral foramen or extraforaminal space. 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 (Fig. 14-3). 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. 14-2 and 14-3). Neuroimaging of the spine must include both sagittal and axial views to assess possible compression in either the lateral recess or intervertebral foramen.

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 radiculopathy in the cervical spine.

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 identification of 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 14-1).

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 can explain 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,”

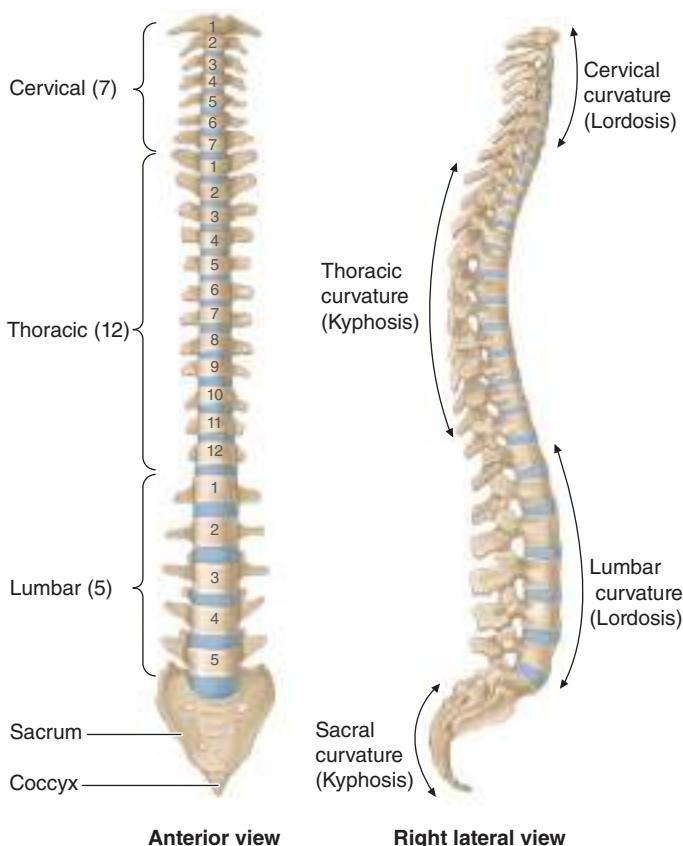


FIGURE 14-2 Spinal column. (From A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998; with permission.)

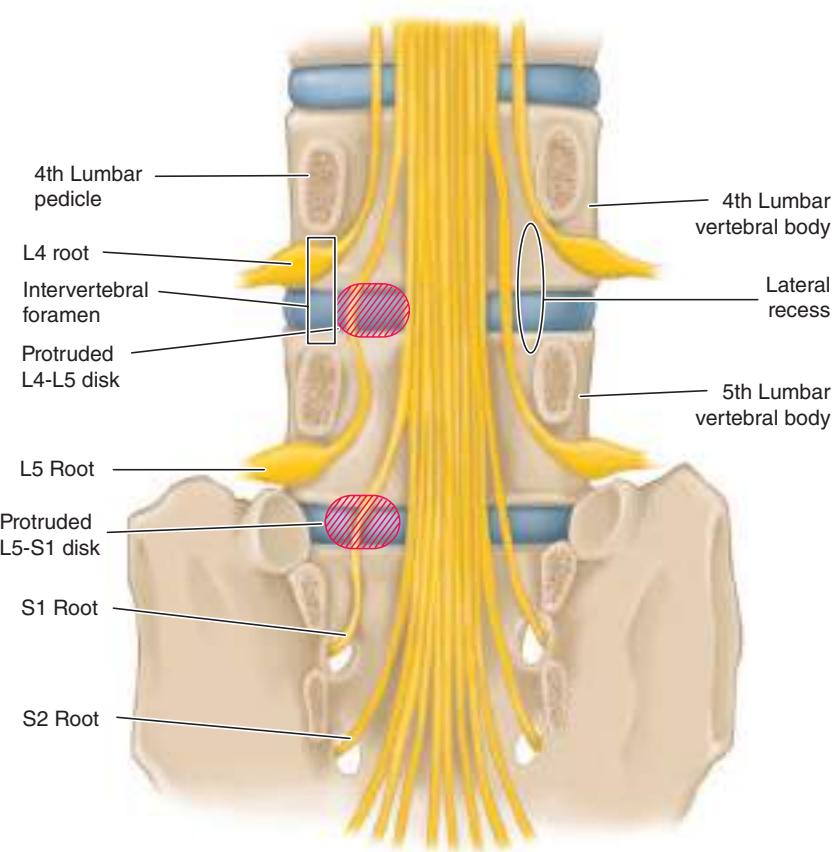


FIGURE 14-3 Compression of L5 and S1 roots by herniated disks. (From AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009; with permission.)

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 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 physical examination that includes the abdomen and rectum 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 may reveal a lateral curvature of the spine (scoliosis). An asymmetry in the prominence of the paraspinal muscles suggests muscle spasm. Spine pain reproduced by palpation over the spinous process reflects injury of the affected vertebrae or adjacent pain-sensitive structures.

Forward bending is often limited by paraspinal muscle spasm; the latter may flatten the usual lumbar lordosis. 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

TABLE 14-1 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)
- 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
- Palpation/percussion tenderness over the midline spine
- Abdominal, rectal, or pelvic mass
- Internal/external rotation of the leg at the hip; heel percussion sign
- Straight leg- or reverse straight leg-raising signs
- Progressive focal neurologic deficit

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.

The neurologic examination includes a search for focal weakness or muscle atrophy, focal 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 14-2** and are discussed below.

LABORATORY, IMAGING, AND EMG STUDIES

Laboratory studies are rarely needed for the initial evaluation of non-specific acute (<3 months in 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 14-1), 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).

Computed tomography (CT) scanning 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. CT scans are increasingly used as a primary screening modality for moderate to severe acute trauma. Magnetic resonance imaging (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.

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 reported, 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. Based in part on such evidence, the American College of Physicians and the North American Spine Society have partnered to make parsimonious use of spine imaging a high priority in the "Choosing Wisely" campaign, aimed at reducing unnecessary spine care. Successful efforts to reduce unnecessary imaging have typically been multifaceted. Some include physician education and computerized decision support to identify prior imaging examinations and to require specific indications for approval of imaging tests. Other strategies have included audit and feedback of individual practitioners' rates of ordering, and more rapid access to physical therapy or expert consultation for patients without imaging indications.

For example, educational tools 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

TABLE 14-2 Lumbosacral Radiculopathy: Neurologic Features

LUMBOSACRAL NERVE ROOTS	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
L2 ^a	—	Upper anterior thigh	Psoas (hip flexors)	Anterior thigh
L3 ^a	—	Lower anterior thigh Anterior knee	Psoas (hip flexors) Quadriceps (knee extensors) Thigh adductors	Anterior thigh, knee
L4 ^a	Quadriceps (knee)	Medial calf	Quadriceps (knee extensors) ^b Thigh adductors	Knee, medial calf Anterolateral thigh
L5 ^c	—	Dorsal surface—foot Lateral calf	Peronei (foot evertors) ^b Tibialis anterior (foot dorsiflexors) Gluteus medius (leg abductors) Toe dorsiflexors	Lateral calf, dorsal foot, posterior lateral thigh, buttocks
S1 ^c	Gastrocnemius/soleus (ankle)	Plantar surface—foot Lateral aspect—foot	Gastrocnemius/soleus (foot plantar flexors) ^b Abductor hallucis (toe flexors) ^b Gluteus maximus (leg extensors)	Bottom foot, posterior calf, posterior thigh, buttocks

^aReverse straight leg-raising sign present—see "Examination of the Back." ^bThese muscles receive the majority of innervation from this root. ^cStraight leg-raising sign present—see "Examination of the Back."

spine fusion surgery. (4) Do not use EMG and nerve conduction studies (NCSs) to determine the cause of axial 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. 438). 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.

CAUSES OF BACK PAIN (TABLE 14-3)

LUMBAR DISK DISEASE

This is a common cause of acute, chronic, or recurrent low back and leg pain (Figs. 14-3 and 14-4). Disk disease is most likely to occur at the

TABLE 14-3 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) and spondylolisthesis
Spine Infection
Vertebral osteomyelitis
Spinal epidural abscess
Septic disk (diskitis)
Meningitis
Lumbar arachnoiditis
Neoplasms—Metastatic, Hematologic, Primary Bone Tumors, Fractures
Trauma/falls, motor vehicle accidents
Atraumatic fractures: osteoporosis, neoplastic infiltration, osteomyelitis
Minor Trauma
Strain or sprain
Whiplash injury
Metabolic Spine Disease
Osteoporosis—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

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.

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 at the 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 compress multiple roots or cause inflammation of nerve roots within the spinal canal. Clinical manifestations of specific nerve root lesions are summarized in Table 14-2.

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 spinal cord disease suggest an etiology other than lumbar disk disease. Absence of ankle reflexes can be a normal finding in persons >60 years or a sign of bilateral S1 radiculopathy. 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. For example, an absent knee reflex may be due to a femoral neuropathy or an L4 nerve root injury, and a 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 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 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

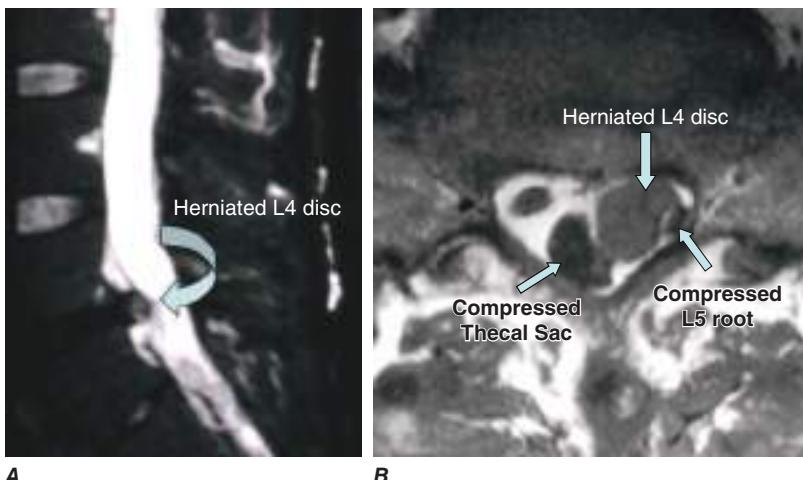


FIGURE 14-4 Left L5 radiculopathy. **A.** Sagittal T2-weighted image on the left reveals disk herniation at the L4-L5 level. **B.** Axial T1-weighted image shows paracentral disk herniation with displacement of the thecal sac medially and the left L5 nerve root posteriorly in the left lateral recess.

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. 434), and Guillain-Barré syndrome (Chap. 439). 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 surgical decompression, sometimes on an urgent basis in an attempt to restore or preserve motor or sphincter function, or radiotherapy for metastatic tumors (Chap. 86).

■ DEGENERATIVE CONDITIONS

Lumbar spinal stenosis (LSS) describes a narrowed lumbar spinal canal. *Neurogenic claudication* consists of pain, typically in the back and buttock or leg, that is brought on by walking or standing and relieved by sitting. Symptoms in the legs are usually bilateral. 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 frequently asymptomatic. The correlation between the severity of symptoms and the degree of spinal canal stenosis is variable. LSS is most often acquired (75%), but can also be congenital or due to a mixture of both. 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. 14-5).

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. A recent trial for LSS accompanied by leg pain did not show an overall benefit for epidural glucocorticoids plus lidocaine, but subgroup analysis showed a small improvement in disability scores at 6 weeks of uncertain clinical significance.

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. Surgical therapy 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 can produce significant relief of exertional back and leg pain, leading to less disability and improved functional outcome at 4 years. 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 spinal level or at an adjacent level within 7–10 years; recurrent symptoms usually respond to a second surgical decompression.

Neural foraminal narrowing with radiculopathy is a common consequence of osteoarthritic processes that cause LSS (Figs. 14-1 and 14-6), 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 other less common causes. Most common is bony foraminal narrowing leading to nerve root ischemia and persistent symptoms, in contrast to the inflammation associated with a 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 specific 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

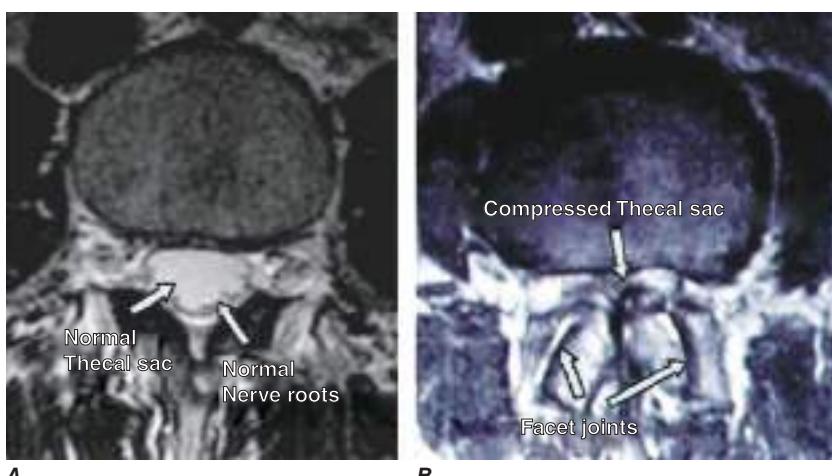


FIGURE 14-5 Axial T2-weighted images of the lumbar spine. **A.** The image shows a normal thecal sac within the lumbar spinal canal. The thecal sac is bright. The lumbar roots are dark punctate dots in the posterior thecal sac with the patient supine. **B.** The thecal sac is not well visualized due to severe lumbar spinal canal stenosis, partially the result of hypertrophic facet joints.

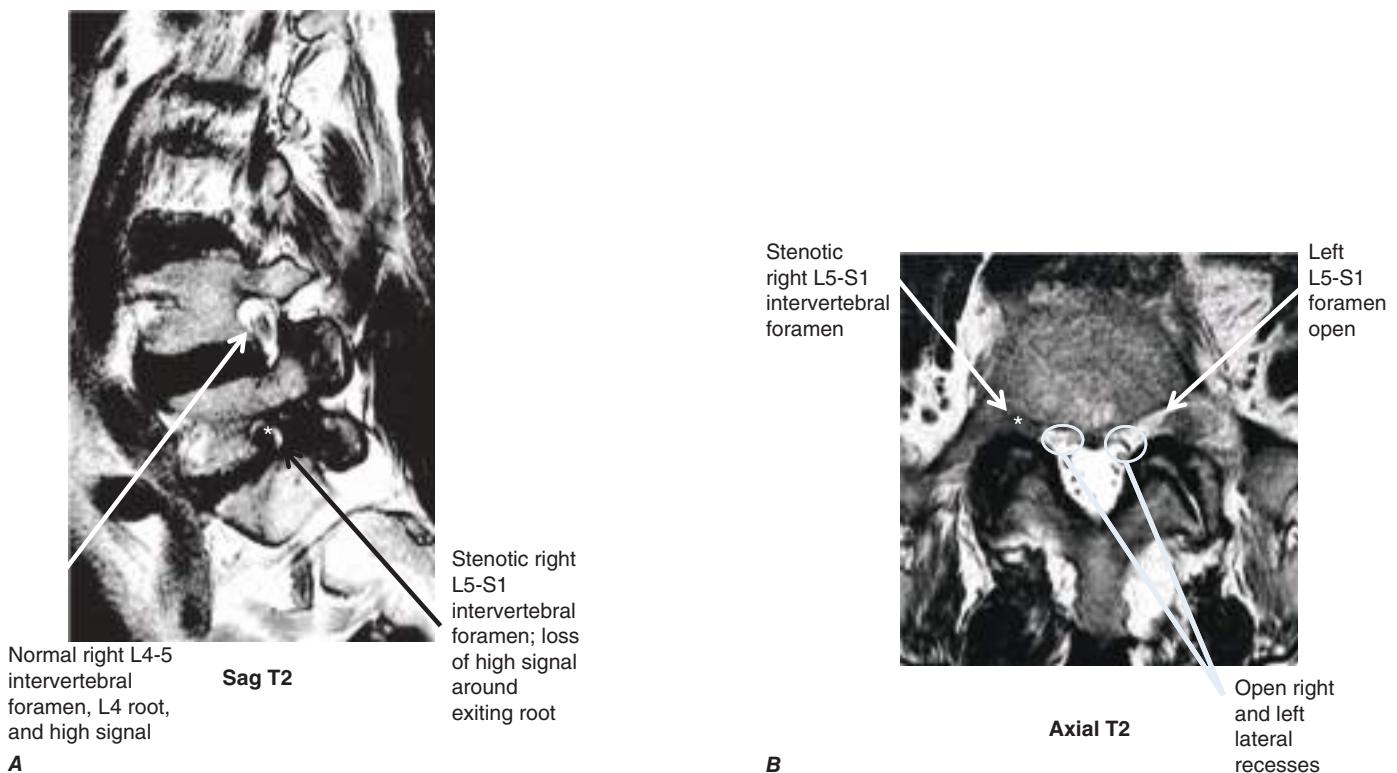


FIGURE 14-6 Right L5 radiculopathy. **A.** Sagittal T2-weighted image. There is normal high signal around the exiting right L4 nerve root in the right neural foramen at L4-L5; effacement of the high signal in the right L5-S1 foramen is present one level caudal on the right at L5-S1. **B.** Axial T2-weighted image. The lateral recesses are normal bilaterally; the intervertebral foramen is normal on the left, but severely stenotic on the right. *Severe right L5-S1 foraminal stenosis.

can help direct the attention of the radiologist to specific nerve roots, especially on axial images. For *facet joint hypertrophy*, surgical foraminotomy produces long-term relief of leg and back pain in 80–90% of patients. Facet joint 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 anatomic spine disease include infections (e.g., herpes zoster and Lyme disease), carcinomatous meningitis, and root avulsion or traction (trauma).

■ 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 x-ray, CT, or MRI findings are minimal, and prominent degenerative spine disease can be seen in asymptomatic patients. Osteophytes, combined disk-osteophytes, or 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, prior surgery). The slippage may be asymptomatic or may cause LBP and hamstring tightness, nerve root injury (the L5 root most frequently), symptomatic spinal stenosis, or CES in severe cases. A “step-off” on palpation or tenderness may be elicited near the segment that has “slipped” forward (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 neck or low back 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 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 the 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. 86.**

■ 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; IV drug use, poor dentition, endocarditis, pulmonary disease, IV catheters, or post-operative wound sites may also be responsible. Back pain at rest, tenderness over the involved vertebra, and an elevated ESR or 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 vertebra can produce a spinal epidural abscess.

Spinal epidural abscess (Chap. 434) presents with back pain (aggravated by movement or spinous process palpation), fever, radiculopathy,

or signs of spinal cord compression. The subacute development of two or more of these findings should increase the index of 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 alone are not diagnostic and may also occur 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; with fractures above L1 the spinal cord is 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 clearly describe a specific anatomic lesion. 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 excluded patients aged >65—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 postmenopausal or senile osteoporosis (Chap. 404). 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 senile 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 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 is mixed, but meta-analyses do not support their utility.

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. 403, 404, and 405.

■ 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. 351), ankylosing spondylitis, reactive arthritis, psoriatic arthritis, or inflammatory bowel disease (Chaps. 319 and 355).

■ CONGENITAL ANOMALIES OF THE LUMBAR SPINE

Spondylolisthesis is a bony defect in the vertebral pars interarticularis (a segment near the junction of the pedicle with the lamina); the cause is usually a stress microfracture in a congenitally abnormal segment. It occurs in up to 6% of adolescents. The defect (usually bilateral) is best visualized on plain x-rays or CT scan and is frequently asymptomatic. Symptoms may occur in the setting of a single injury, repeated minor injuries, or during a growth spurt. Spondylolisthesis 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 occasionally progressive due to 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 an evaluation 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.

■ 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. 76 and 317), 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 suggests 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) is an important physical finding. Patients with suspected AAA should be evaluated with abdominal ultrasound, CT, or MRI ([Chap. 274](#)).

Sacral Pain with Gynecologic and Urologic Disease Pelvic organs rarely cause 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. 83](#)), 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.

Psychiatric Disease CLBP may be encountered in patients who seek financial compensation; in malingering; 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. Pre-operative 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 low back pain 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 2010 reported that LBP ranked #6 overall as a cause of disability-related life years (DALYs), and was the #1 cause overall for total years lived with disability (YLD). These numbers increased substantially from 1990 estimates, and with the aging of the population worldwide, the numbers of individuals suffering from low back pain are expected to increase further in the future. Although rankings for low back pain generally were higher in developed regions of the world, this was not uniformly the case; for example, in North Africa and the Middle East low back pain ranked #2 for DALYs. Another area of uncertainty is the extent 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 prior spine infection, but in developed countries is multiple lumbar spine surgeries. The longstanding history and acceptance of acupuncture in China may also explain the large number of studies from China regarding the efficacy of acupuncture in many pain settings.

TREATMENT

Back Pain

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. 10](#)). Safety may be improved with automated notices for high doses, early refills, prescriptions from multiple pharmacies, and overlapping opioid and benzodiazepine prescriptions. Greater access to alternative treatments for chronic pain, such as tailored exercise programs and cognitive-behavioral therapy (CBT), may also reduce opioid prescribing. Public concern in the United States resulted in passage of the Comprehensive Addiction and Recovery Act of 2016.

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 low back pain without radiculopathy. Finally, educating patients and the public about the risks of overtreatment may be necessary.

ALBP WITHOUT RADICULOPATHY

ALBP is defined as pain of <3 months in 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 excludes 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 14-1. 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 being substantially improved after 7 weeks. Spontaneous improvement can mislead clinicians and patients alike about the efficacy of treatment interventions unless subjected to 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, treatment methods, activity modifications, and strategies to prevent future exacerbations. 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 kept to a day or two at most. Several randomized trials suggest that bed rest does not hasten the pace of recovery. In general, the best activity recommendation is for early resumption of normal physical activity, avoiding only strenuous manual labor. Possible advantages of early ambulation for ALBP 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. Use of heating pads or blankets is sometimes helpful.

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 accompany 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 a NSAID together in hopes of a more rapid benefit. 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.

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. 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. Clinical efficacy of opioids for chronic pain beyond 16 weeks of use is unproven.

There is no evidence to support use of oral or injected glucocorticoids, antiepileptics, antidepressants, therapies for neuropathic pain such as gabapentin or herbal therapies. Commonly used non-pharmacologic 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, 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, 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.

CLBP WITHOUT RADICULOPATHY

CLBP is defined as pain lasting >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, activity tolerance is the primary goal, while pain relief is secondary.

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. A long-term benefit of spinal manipulation or massage for CLBP is unproven.

Medications for CLBP may include short courses of NSAIDs or acetaminophen. Tricyclic antidepressants can provide modest pain relief for some patients without evidence of depression. Trials do not support the efficacy of selective serotonin reuptake inhibitors (SSRIs) for CLBP. However, depression is common among patients with chronic pain and should be appropriately treated.

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.

Back pain is the most frequent reason for seeking complementary and alternative treatments, most commonly spinal manipulation, acupuncture, and massage. The value of these approaches remains unclear, however. Biofeedback has not been studied rigorously. There is no convincing evidence that either spinal manipulation, 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. Some trials of massage therapy have been encouraging for short-term relief only.

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 effective.

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 is commonly 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 both types of 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 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). A few small trials have produced conflicting results for facet joint and diskogenic pain. A trial in patients with chronic radicular pain found no difference between radiofrequency denervation of the dorsal root ganglia and sham treatment. These interventional therapies have not been studied in sufficient detail to draw firm conclusions regarding their value for CLBP.

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 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 be considered for people 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.

Lumbar disk replacement with prosthetic disks is U.S. 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." Serious complications are somewhat more likely with the artificial disk. This treatment remains controversial for CLBP.

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 suggest that the evidence is limited and benefits are limited.

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, ultrasoundradiofrequency facet joint denervation, intradiscal electrothermal therapy, or intradiscal radiofrequency thermocoagulation. On the other hand, exercise therapy and treatment of depression appear to be useful and underused.

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, there are several important treatment options that 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 of opioid analgesics. Opioids are superior for acute pain relief in the emergency room.

Epidural glucocorticoid injections have a role in providing symptom relief for acute lumbar radiculopathy due to a herniated disk. However, there does not appear to be a benefit in terms of reducing subsequent surgical interventions. A brief course of high dose oral glucocorticoids for 5 days followed by a rapid taper >5 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 manifest 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.

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 indicate that, compared to nonsurgical treatment, surgery results in more rapid pain relief. 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, 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 (discectomy). Minimally invasive techniques have gained in popularity in recent years, but preliminary 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 low back pain despite an initial discectomy fared no better with spine fusion than with a conservative regimen of cognitive intervention and exercise. Artificial disks 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. Neck pain arising from the cervical spine is typically precipitated by movement and may be accompanied by focal tenderness and limitation of motion. Many of the prior comments made regarding causes of low back pain 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 14-4**), 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 low back pain 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.

■ TRAUMA TO THE CERVICAL SPINE

Trauma to the cervical spine (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. 434**). Immediate immobilization of the neck is essential to minimize further spinal cord injury from movement of unstable cervical spine segments. The decision to obtain imaging should be based on the nature 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 feet 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. A CT scan is the diagnostic procedure of choice for detection of acute fractures following severe trauma; plain x-rays can be used for lesser degrees of trauma. When traumatic injury to the vertebral arteries or cervical spinal cord is suspected, visualization by MRI with magnetic resonance angiography is preferred.

Whiplash injury is due to rapid flexion and extension of the neck, usually from automobile accidents. The exact mechanism of injury is unclear. 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.

TABLE 14-4 Cervical Radiculopathy: Neurologic Features

CERVICAL NERVE ROOTS	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
C5	Biceps	Lateral deltoid	Rhomboids ^a (elbow extends backward with hand on hip) Infraspinatus ^a (arm rotates externally with elbow flexed at the side) Deltoid ^a (arm raised laterally 30–45° from the side)	Lateral arm, medial scapula
C6	Biceps	Thumb/index finger; Dorsal hand/lateral forearm	Biceps ^a (arm flexed at the elbow in supination) Pronator teres (forearm pronated)	Lateral forearm, thumb/index fingers
C7	Triceps	Middle fingers	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 and medial arm	Abductor pollicis brevis (abduction of thumb) First dorsal interosseous (abduction of index finger) Abductor digiti minimi (abduction of little finger)	Medial arm, axilla

^aThese muscles receive the majority of innervation from this root.

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 14-4. 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 SPONDYLOYSIS

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. 429), multiple sclerosis (Chap. 436), spinal cord tumors, or syringomyelia (Chap. 434). 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, or spinal canal stenosis. With 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 MR imaging, 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 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

RA (Chap. 351) of the cervical facet joints produces neck pain, stiffness, and limitation of motion. Synovitis of the atlantoaxial joint (C1-C2; Fig. 14-2) 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).

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 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 does 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. 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. 74*) is another cause and should be considered, especially when a concurrent Horner's syndrome is present. *Suprascapular neuropathy* may produce severe shoulder pain, weakness, and wasting of the supraspinatus and infraspinatus muscles. *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.

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. 438*.

■ 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 (tendonitis, 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. Neck pain was ranked #21 as a cause of DALYs in the Global Burden of Diseases Study 2010,

accounting for ~40% of the total global DALYs due to LBP. 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 low back pain, but the approach is remarkably similar in many respects. As with low back pain, 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 a combination of NSAIDs, acetaminophen, cold packs, or heat while awaiting spontaneous recovery. 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, electromagnetic therapy, 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.

TREATMENT

Neck Pain with Radiculopathy

The natural history of neck pain with acute 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 radiculopathy, but have not been subjected to rigorous trials. The risk of injection complications is higher in the neck than the low back; vertebral artery dissection, dural puncture, and embolism from injection particles in the vertebral arteries have all been reported. Opioid analgesics can be used in the emergency room 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 whether long-term 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.



Surgical treatments include anterior cervical discectomy alone, laminectomy with discectomy, or discectomy 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): Non-invasive 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>.
- BENZON HT et al: Improving the safety of epidural steroid injections. *JAMA* 313:1713, 2015.
- FRIEDLY JL et al: A randomized trial of epidural glucocorticoid injections for spinal stenosis. *N Engl J Med* 371:11, 2014.
- GOLDBERG H et al: Oral steroids for acute radiculopathy due to a herniated lumbar disk. *JAMA* 313:1915, 2015.
- HOY DG et al: Reflecting on the global burden of musculoskeletal conditions: Lessons learnt from the global burden of disease 2010 study and the next steps forward. *Ann Rheum Dis* 74:4, 2015.
- KATZ JN, HARRIS MB: Clinical practice. Lumbar spinal stenosis. *N Engl J Med* 358:818, 2008.
- LAMB SE et al: Group cognitive behavioural treatment for low-back pain in primary care: A randomised controlled trial and cost-effectiveness analysis. *Lancet* 375:916, 2010.
- MALMIVAARA A et al: The treatment of acute low back pain—Bed rest, exercises, or ordinary activity? *N Engl J Med* 332:351, 1995.
- MELANICA J et al: Spinal stenosis. *Handb Clin Neurol* 109:541, 2014.
- SERINKEN M et al: Comparison of intravenous morphine versus paracetamol in sciatica: A randomized placebo controlled trial. *Acad Emerg Med* 23:674, 2016.
- ZYGOURAKIS CC et al: Geographic and hospital variation in cost of lumbar laminectomy and lumbar fusion for degenerative conditions. *Neurosurgery* 81:331, 2017.

Section 2 Alterations in Body Temperature

15

Fever

Charles A. Dinarello, Reuven Porat



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 studies of healthy individuals 18–40 years of age, the mean oral temperature is $36.8^\circ \pm 0.4^\circ\text{C}$ ($98.2^\circ \pm 0.7^\circ\text{F}$), with low levels at 6 A.M. and higher levels at 4–6 P.M. The maximal normal oral temperature is 37.2°C (98.9°F) at 6 A.M. and 37.7°C (99.9°F) at 4 P.M.; these values define the 99th percentile for healthy individuals. In

light of these studies, an A.M. temperature of $>37.2^\circ\text{C}$ ($>98.9^\circ\text{F}$) or a P.M. temperature of $>37.7^\circ\text{C}$ ($>99.9^\circ\text{F}$) would define a fever. The normal daily temperature variation, also called the *circadian rhythm*, is typically 0.5°C (0.9°F). However, in some individuals recovering from a febrile illness, this daily variation can be as great as 1.0°C . During a febrile illness, the diurnal variation is usually maintained, but at higher, febrile levels. The daily temperature variation appears to be fixed in early childhood; in contrast, elderly individuals can exhibit a reduced ability to develop fever, with only a modest fever even in severe infections.

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 $\sim 0.6^\circ\text{C}$ (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 HYPERTERMIA

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 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 $>41.5^\circ\text{C}$ ($>106.7^\circ\text{F}$) 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 *subnormal*, 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. 455.**

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.

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 E₂ (PGE₂) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE₂ 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. 15-1. 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 PGE₂, they also induce PGE₂ in peripheral tissues. The increase in PGE₂ in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. It is thought that some systemic PGE₂ escapes destruction by the lung and gains access to the hypothalamus via the internal carotid. However, it is the elevation of PGE₂ in the brain that starts the process of raising the hypothalamic set point for core temperature.

There are four receptors for PGE₂, 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 PGE₂ receptor genes leaves the fever mechanism intact. Although PGE₂ is essential for fever, it is not a neurotransmitter. Rather, the release of PGE₂ from the brain side of the hypothalamic endothelium triggers the PGE₂ 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. 15-1, 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 PGE₂ production and fever.

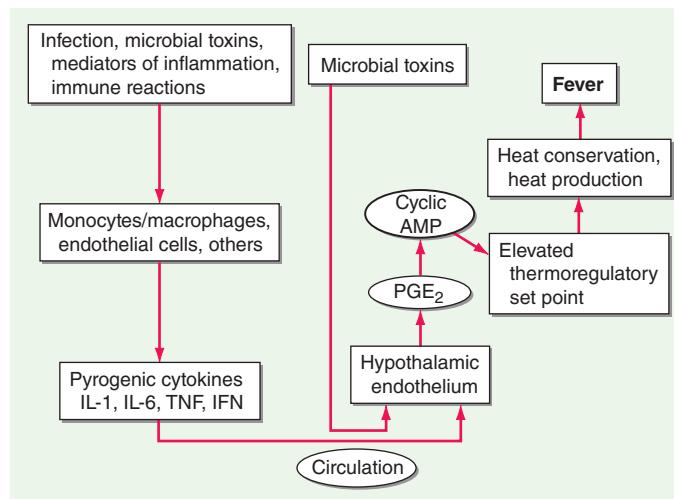


FIGURE 15-1 Chronology of events required for the induction of fever. AMP, adenosine 5'-monophosphate; IFN, interferon; IL, interleukin; PGE₂, prostaglandin E₂; TNF, tumor necrosis factor.

■ 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

PHYSICAL EXAMINATION

The chronology of events preceding fever, including exposure to other infected 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 infection 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. 297.**

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).

TREATMENT

Fever

THE DECISION TO TREAT FEVER

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 is no periodicity of fever in patients with familial Mediterranean fever. However, these patterns have limited or no diagnostic value compared with specific and rapid laboratory tests.

ANTICYTOKINE THERAPY TO REDUCE FEVER IN AUTOIMMUNE AND AUTOINFLAMMATORY DISEASES

Recurrent fever is documented at some point in most autoimmune diseases and nearly all autoinflammatory diseases. Although fever can be a manifestation of autoimmune diseases, recurrent fevers are characteristic of autoinflammatory diseases (Table 15-1), 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 these illnesses are dramatically reduced by blocking of IL-1 activity with anakinra or canakinumab. Anticytokines therefore reduce fever in

TABLE 15-1 Autoinflammatory Diseases in Which Fever Is Characteristic

Adult and juvenile Still's disease
Cryopyrin-associated periodic syndromes (CAPS)
Familial Mediterranean fever
Hyper-IgD syndrome
Behçet's syndrome
Macrophage activation syndrome
Normocomplementemic urticarial vasculitis
Antisynthetase myositis
PAPA ^a syndrome
Blau syndrome
Gouty arthritis

^aPyogenic arthritis, pyoderma gangrenosum, and acne.

autoimmune and autoinflammatory diseases. Although fevers in autoinflammatory diseases are mediated by IL-1 β , 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 PGE₂ level in the thermoregulatory center. The synthesis of PGE₂ 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 PGE₂. 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, PGE₂ appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE₂ synthesis by inhibiting the activity of phospholipase A₂, 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. 455.

FURTHER READING

- DINARELLO CA et al: Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nature Rev* 11:633, 2012.
KULLENBERG T et al: Long-term safety profile of anakinra in patients with severe cryopyrin-associated periodic syndromes. *Rheumatology* 55:1499, 2016.

16

Fever and Rash

Elaine T. Kaye, Kenneth M. Kaye



The acutely ill patient with fever and rash often presents a diagnostic challenge for physicians, yet the distinctive appearance of an eruption in concert with a clinical syndrome can facilitate a prompt diagnosis and the institution of life-saving therapy or critical infection-control interventions. **Representative images of many of the rashes discussed in this chapter are included in Chap. A1.**

APPROACH TO THE PATIENT

Fever and Rash

A thorough history of patients with fever and rash includes the following relevant information: immune status, medications taken within the previous month, specific travel history, immunization status, exposure to domestic pets and other animals, history of animal (including arthropod) bites, recent dietary exposures, existence of cardiac abnormalities, presence of prosthetic material, recent exposure to ill individuals, and sexual exposures. The history should also include the site of onset of the rash and its direction and rate of spread.

A thorough physical examination entails close attention to the rash, with an assessment and precise definition of its salient features. First, it is critical to determine what *type* of lesions make up the eruption. *Macules* are flat lesions defined by an area of changed color (i.e., a blanchable erythema). *Papules* are raised, solid lesions <5 mm in diameter; *plaques* are lesions >5 mm in diameter with a flat, plateau-like surface; and *nodules* are lesions >5 mm in diameter with a more rounded configuration. *Wheals* (urticaria, hives) are papules or plaques that are pale pink and may appear annular (ringlike) as they enlarge; classic (nonvasculitic) wheals are transient, lasting only 24 h in any defined area. *Vesicles* (<5 mm) and *bullae* (>5 mm) are circumscribed, elevated lesions containing fluid. *Pustules* are raised lesions containing purulent exudate; vesicular processes such as varicella or herpes simplex may evolve to pustules. *Nonpalpable purpura* is a flat lesion that is due to bleeding into the skin. If <3 mm in diameter, the purpuric lesions are termed *petechiae*; if >3 mm, they are termed *ecchymoses*. *Palpable purpura* is a raised lesion that is due to inflammation of the vessel wall (vasculitis) with subsequent hemorrhage. An *ulcer* is a defect in the skin extending at least into the upper layer of the dermis, and an *eschar* (tâche noire) is a necrotic lesion covered with a black crust.

Other pertinent features of rashes include their *configuration* (i.e., annular or target), the *arrangement* of their lesions, and their *distribution* (i.e., central or peripheral).

For further discussion, see Chaps. 52, 54, 117, and 124.

CLASSIFICATION OF RASH

This chapter reviews rashes that reflect systemic disease, but it does not include localized skin eruptions (i.e., cellulitis, impetigo) that may also be associated with fever (Chap. 124). The chapter is not intended to be all-inclusive, but it covers the most important and most common diseases associated with fever and rash. Rashes are classified herein on the basis of lesion morphology and distribution. For practical purposes, this classification system is based on the most typical disease presentations. However, morphology may vary as rashes evolve, and the presentation of diseases with rashes is subject to many variations (Chap. 54). For instance, the classic petechial rash of Rocky Mountain spotted fever (Chap. 182) may initially consist of blanchable erythematous macules distributed peripherally; at times, however, the rash associated with this disease may not be predominantly acral, or no rash may develop at all.

Diseases with fever and rash may be classified by type of eruption: centrally distributed maculopapular, peripheral, confluent desquamative erythematous, vesiculobullous, urticaria-like, nodular, purpuric, ulcerated, or with eschars. Diseases are listed by these categories in **Table 16-1**, and many are highlighted in the text. However, for a more detailed discussion of each disease associated with a rash, the reader is referred to the chapter dealing with that specific disease. (**Reference chapters are cited in the text and listed in Table 16-1.**)

CENTRALLY DISTRIBUTED MACULOPAPULAR ERUPTIONS

Centrally distributed rashes, in which lesions are primarily truncal, are the most common type of eruption. The rash of *rubeola* (measles) starts at the hairline 2–3 days into the illness and moves down the body, typically sparing the palms and soles (**Chap. 200**). It begins as discrete erythematous lesions, which become confluent as the rash spreads. Koplik's spots (1- to 2-mm white or bluish lesions with an erythematous halo on the buccal mucosa) are pathognomonic for measles and are generally seen during the first 2 days of symptoms. They should not be confused with Fordyce's spots (ectopic sebaceous glands), which have no erythematous halos and are found in the mouth of healthy individuals. Koplik's spots may briefly overlap with the measles exanthem.

Rubella (German measles) also spreads from the hairline downward; unlike that of measles, however, the rash of rubella tends to clear from originally affected areas as it migrates, and it may be pruritic (**Chap. 201**). Forchheimer spots (palatal petechiae) may develop but are nonspecific because they also develop in *infectious mononucleosis* (**Chap. 189**), *scarlet fever* (**Chap. 143**), and *Zika virus infection* (**Chap. 204**). Postauricular and suboccipital adenopathy and arthritis are common among adults with rubella. Exposure of pregnant women to ill individuals should be avoided, as rubella causes severe congenital abnormalities. Numerous strains of *enteroviruses* (**Chap. 199**), primarily echoviruses and coxsackieviruses, cause nonspecific syndromes of fever and eruptions that may mimic rubella or measles. Patients with *infectious mononucleosis* caused by Epstein-Barr virus (**Chap. 189**) or with *primary HIV infection* (**Chap. 197**) may exhibit pharyngitis, lymphadenopathy, and a nonspecific maculopapular exanthem.

The rash of *erythema infectiosum* (fifth disease), which is caused by human parvovirus B19, primarily affects children 3–12 years old; it develops after fever has resolved as a bright blanchable erythema on the cheeks ("slapped cheeks") with perioral pallor (**Chap. 192**). A more diffuse rash (often pruritic) appears the next day on the trunk and extremities and then rapidly develops into a lacy reticular eruption that may wax and wane (especially with temperature change) over 3 weeks. Adults with fifth disease often have arthritis, and fetal hydrops can develop in association with this condition in pregnant women.

Exanthem subitum (roseola) is caused by human herpesvirus 6 and is most common among children <3 years of age (**Chap. 190**). As in *erythema infectiosum*, the rash usually appears after fever has subsided. It consists of 2- to 3-mm rose-pink macules and papules that coalesce only rarely, occur initially on the trunk and sometimes on the extremities (sparing the face), and fade within 2 days.

Although drug reactions have many manifestations, including urticaria, exanthematic drug-induced eruptions (**Chap. 56**) are most common and are often difficult to distinguish from viral exanthems. Eruptions elicited by drugs are usually more intensely erythematous and pruritic than viral exanthems, but this distinction is not reliable. A history of new medications and an absence of prostration may help to distinguish a drug-related rash from an eruption of another etiology. Rashes may persist for up to 2 weeks after administration of the offending agent is discontinued. Certain populations are more prone than others to drug rashes. Of HIV-infected patients, 50–60% develop a rash in response to sulfa drugs; 30–90% of patients with mononucleosis due to Epstein-Barr virus develop a rash when given ampicillin.

Rickettsial illnesses (**Chap. 182**) should be considered in the evaluation of individuals with centrally distributed maculopapular eruptions. The usual setting for *epidemic typhus* is a site of war or natural disaster in which people are exposed to body lice. Endemic typhus or

leptospirosis (the latter caused by a spirochete) (**Chap. 179**) may be seen in urban environments where rodents proliferate. Outside the United States, other rickettsial diseases cause a spotted-fever syndrome and should be considered in residents of or travelers to endemic areas. Similarly, *typhoid fever*, a nonrickettsial disease caused by *Salmonella typhi* (**Chap. 160**), is usually acquired during travel outside the United States. *Dengue fever*, caused by a mosquito-transmitted flavivirus, occurs in tropical and subtropical regions of the world (**Chap. 204**).

Some centrally distributed maculopapular eruptions have distinctive features. *Erythema migrans*, the rash of *Lyme disease* (**Chap. 181**), typically manifests as single or multiple annular lesions. Untreated *erythema migrans* lesions usually fade within a month but may persist for more than a year. *Southern tick-associated rash illness* (STARI) (**Chap. 181**) has an *erythema migrans*-like rash, but is less severe than Lyme disease and often occurs in regions where Lyme is not endemic. *Erythema marginatum*, the rash of *acute rheumatic fever* (**Chap. 352**), has a distinctive pattern of enlarging and shifting transient annular lesions.

Collagen vascular diseases may cause fever and rash. Patients with *systemic lupus erythematosus* (**Chap. 349**) typically develop a sharply defined, erythematous eruption in a butterfly distribution on the cheeks (malar rash) as well as many other skin manifestations. *Still's disease* presents as an evanescent, salmon-colored rash on the trunk and proximal extremities that coincides with fever spikes.

Zika virus is a mosquito-transmitted flavivirus that is associated with severe birth defects (**Chap. 204**). Zika is rapidly spreading among tropical and subtropical regions of the world. The eruption of Zika virus infection is typically pruritic and often accompanied by conjunctival injection.

PERIPHERAL ERUPTIONS

These rashes are alike in that they are most prominent peripherally or begin in peripheral (acral) areas before spreading centripetally. Early diagnosis and therapy are critical in *Rocky Mountain spotted fever* (**Chap. 182**) because of its grave prognosis if untreated. Lesions evolve from macular to petechial, start on the wrists and ankles, spread centripetally, and appear on the palms and soles only later in the disease. The rash of *secondary syphilis* (**Chap. 177**), which may be generalized but is prominent on the palms and soles, should be considered in the differential diagnosis of pityriasis rosea, especially in sexually active patients. *Chikungunya fever* (**Chap. 204**), which is transmitted by mosquito bite in tropical and subtropical regions, is associated with a maculopapular eruption and severe polyarticular small-joint arthralgias. *Hand-foot-and-mouth disease* (**Chap. 199**), most commonly caused by coxsackievirus A16 or enterovirus 71, is distinguished by tender vesicles distributed on the hands and feet and in the mouth; coxsackievirus A6 causes an atypical syndrome with more extensive lesions. The classic target lesions of *erythema multiforme* appear symmetrically on the elbows, knees, palms, soles, and face. In severe cases, these lesions spread diffusely and involve mucosal surfaces. Lesions may develop on the hands and feet in *endocarditis* (**Chap. 123**).

CONFLUENT DESQUAMATIVE ERYTHEMAS

These eruptions consist of diffuse erythema frequently followed by desquamation. The eruptions caused by group A *Streptococcus* or *Staphylococcus aureus* are toxin-mediated. *Scarlet fever* (**Chap. 143**) usually follows pharyngitis; patients have a facial flush, a "strawberry" tongue, and accentuated petechiae in body folds (Pastia's lines). *Kawasaki disease* (**Chaps. 54 and 356**) presents in the pediatric population as fissuring of the lips, a strawberry tongue, conjunctivitis, adenopathy, and sometimes cardiac abnormalities. *Streptococcal toxic shock syndrome* (**Chap. 143**) manifests with hypotension, multiorgan failure, and, often, a severe group A streptococcal infection (e.g., necrotizing fasciitis). *Staphylococcal toxic shock syndrome* (**Chap. 142**) also presents with hypotension and multiorgan failure, but usually only *S. aureus* colonization—not a severe *S. aureus* infection—is documented. *Staphylococcal scalded-skin syndrome* (**Chap. 142**) is seen primarily in children and in immunocompromised adults. Generalized erythema is often evident during the prodrome of fever and malaise; profound tenderness of the skin is distinctive. In the exfoliative stage, the skin can be

TABLE 16-1 Diseases Associated with Fever and Rash

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Centrally Distributed Maculopapular Eruptions					
Acute meningococcemia ^a	—	—	—	—	150
Drug reaction with eosinophilia and systemic symptoms (DRESS); also termed drug-induced hypersensitivity syndrome (DIHS) ^b	—	—	—	—	56
Rubeola (measles, first disease)	Paramyxovirus	Discrete lesions that become confluent as rash spreads from hairline downward, usually sparing palms and soles; lasts ≥3 days; Koplik's spots	Nonimmune individuals	Cough, conjunctivitis, coryza, severe prostration	200
Rubella (German measles, third disease)	Togavirus	Spreads from hairline downward, clearing as it spreads; Forschheimer spots	Nonimmune individuals	Adenopathy, arthritis	201
Erythema infectiosum (fifth disease)	Human parvovirus B19	Bright-red "slapped-cheeks" appearance followed by lacy reticular rash that waxes and wanes over 3 weeks; rarely, papular-purpuric "gloves-and-socks" syndrome on hands and feet	Most common among children 3–12 years old; occurs in winter and spring	Mild fever; arthritis in adults; rash following resolution of fever	192
Exanthem subitum (roseola, sixth disease)	Human herpesvirus 6	Diffuse maculopapular eruption over trunk and neck; resolves within 2 days	Usually affects children <3 years old	Rash following resolution of fever; similar to Boston exanthem (echovirus 16); febrile seizures may occur	190
Primary HIV infection	HIV	Nonspecific diffuse macules and papules; less commonly, urticarial or vesicular oral or genital ulcers	Individuals recently infected with HIV	Pharyngitis, adenopathy, arthralgias	197
Infectious mononucleosis	Epstein-Barr virus	Diffuse maculopapular eruption (5% of cases; 30–90% if ampicillin is given); urticaria, petechiae in some cases; periorbital edema (50%); palatal petechiae (25%)	Adolescents, young adults	Hepatosplenomegaly, pharyngitis, cervical lymphadenopathy, atypical lymphocytosis, heterophile antibody	189
Other viral exanthems	Echoviruses 2, 4, 9, 11, 16, 19, 25; coxsackieviruses A9, B1, B5; etc.	Wide range of skin findings that may mimic rubella or measles	Affect children more commonly than adults	Nonspecific viral syndromes	199
Exanthematous drug-induced eruption	Drugs (antibiotics, anticonvulsants, diuretics, etc.)	Intensely pruritic, bright-red macules and papules, symmetric on trunk and extremities; may become confluent	Occurs 2–3 days after exposure in previously sensitized individuals; otherwise, after 2–3 weeks (but can occur anytime, even shortly after drug is discontinued)	Variable findings: fever and eosinophilia	56
Epidemic typhus	<i>Rickettsia prowazekii</i>	Maculopapular eruption appearing in axillae, spreading to trunk and later to extremities; usually spares face, palms, soles; evolves from blanchable macules to confluent eruption with petechiae; rash evanescent in recrudescent typhus (Brill-Zinsser disease)	Exposure to body lice; occurrence of recrudescent typhus as relapse after 30–50 years	Headache, myalgias; mortality rates 10–40% if untreated; milder clinical presentation in recrudescent form	182
Endemic (murine) typhus	<i>Rickettsia typhi</i>	Maculopapular eruption, usually sparing palms, soles	Exposure to rat or cat fleas	Headache, myalgias	182
Scrub typhus	<i>Orientia tsutsugamushi</i>	Diffuse macular rash starting on trunk; eschar at site of mite bite	Endemic in South Pacific, Australia, Asia; transmitted by mites	Headache, myalgias, regional adenopathy; mortality rates up to 30% if untreated	182
Rickettsial spotted fevers	<i>Rickettsia conorii</i> (boutonneuse fever), <i>Rickettsia australis</i> (North Queensland tick typhus), <i>Rickettsia sibirica</i> (Siberian tick typhus), and others	Eschar common at bite site; maculopapular (rarely, vesicular and petechial) eruption on proximal extremities, spreading to trunk and face	Exposure to ticks; <i>R. conorii</i> in Mediterranean region, India, Africa; <i>R. australis</i> in Australia; <i>R. sibirica</i> in Siberia, Mongolia	Headache, myalgias, regional adenopathy	182

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Human monocytotropic ehrlichiosis ^c	<i>Ehrlichia chaffeensis</i>	Maculopapular eruption (40% of cases), involves trunk and extremities; may be petechial	Tick-borne; most common in U.S. Southeast, southern Midwest, and mid-Atlantic regions	Headache, myalgias, leukopenia	182
Leptospirosis	<i>Leptospira interrogans</i> and other <i>Leptospira</i> species	Maculopapular eruption; conjunctivitis; scleral hemorrhage in some cases	Exposure to water contaminated with animal urine	Myalgias; aseptic meningitis; fulminant form: icterohemorrhagic fever (Weil's disease)	179
Lyme disease	<i>Borrelia burgdorferi</i> (sole cause in U.S.), <i>Borrelia afzelii</i> , <i>Borrelia garinii</i>	Papule expanding to erythematous annular lesion with central clearing (erythema migrans; average diameter, 15 cm), sometimes with concentric rings, sometimes with indurated or vesicular center; multiple secondary erythema migrans lesions in some cases	Bite of <i>Ixodes</i> tick vector	Headache, myalgias, chills, photophobia occurring acutely; CNS disease, myocardial disease, arthritis weeks to months later in some cases	181
Southern tick-associated rash illness (STARI, Master's disease)	Unknown (possibly <i>Borrelia lonestari</i> or other <i>Borrelia</i> spirochetes)	Similar to erythema migrans of Lyme disease with several differences, including: multiple secondary lesions less likely; lesions tending to be smaller (average diameter, ~8 cm); central clearing more likely	Bite of tick vector <i>Amblyomma americanum</i> (Lone Star tick); often found in regions where Lyme disease is uncommon, including southern United States	Compared with Lyme disease: fewer constitutional symptoms, tick bite more likely to be recalled; other Lyme disease sequelae lacking	181
Typhoid fever	<i>Salmonella typhi</i>	Transient, blanchable erythematous macules and papules, 2–4 mm, usually on trunk (rose spots)	Ingestion of contaminated food or water (rare in U.S.)	Variable abdominal pain and diarrhea; headache, myalgias, hepatosplenomegaly	160
Dengue fever ^d	Dengue virus (4 serotypes; flaviviruses)	Rash in 50% of cases; initially diffuse flushing; midway through illness, onset of maculopapular rash, which begins on trunk and spreads centrifugally to extremities and face; pruritus, hyperesthesia in some cases; after defervescence, petechiae on extremities may occur	Occurs in tropics and subtropics; transmitted by mosquito	Headache; musculoskeletal pain ("breakbone fever"); leukopenia; occasionally biphasic ("saddleback") fever	204
Rat-bite fever (sodoku)	<i>Spirillum minus</i>	Eschar at bite site; then blotchy violaceous or red-brown rash involving trunk and extremities	Rat bite; primarily found in Asia; rare in U.S.	Regional adenopathy; recurrent fevers if untreated	136
Relapsing fever	<i>Borrelia</i> species	Central rash at end of febrile episode; petechiae in some cases	Exposure to ticks or body lice	Recurrent fever, headache, myalgias, hepatosplenomegaly	180
Erythema marginatum (rheumatic fever)	Group A Streptococcus	Erythematous annular papules and plaques occurring as polycyclic lesions in waves over trunk, proximal extremities; evolving and resolving within hours	Patients with rheumatic fever	Pharyngitis preceding polyarthritis, carditis, subcutaneous nodules, chorea	381
Systemic lupus erythematosus (SLE)	Autoimmune disease	Macular and papular erythema, often in sun-exposed areas; discoid lupus lesions (local atrophy, scale, pigmentary changes); periungual telangiectasis; malar rash; vasculitis sometimes causing urticaria, palpable purpura; oral erosions in some cases	Most common in young to middle-aged women; flares precipitated by sun exposure	Arthritis; cardiac, pulmonary, renal, hematologic, and vasculitic disease	352
Still's disease	Autoimmune disease	Transient 2- to 5-mm erythematous papules appearing at height of fever on trunk, proximal extremities; lesions evanescent	Children and young adults	High spiking fever, polyarthritis, splenomegaly; erythrocyte sedimentation rate, >100 mm/h	—
African trypanosomiasis	<i>Trypanosoma brucei rhodesiense/gambiense</i>	Blotchy or annular erythematous macular and papular rash (trypanid), primarily on trunk; pruritus; chancre at site of tsetse fly bite may precede rash by several weeks	Tsetse fly bite in eastern (<i>T. brucei rhodesiense</i>) or western (<i>T. brucei gambiense</i>) Africa	Hemolympathic disease followed by meningoencephalitis; Winterbottom's sign (posterior cervical lymphadenopathy) (<i>T. brucei gambiense</i>)	222

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Arcanobacterial pharyngitis	<i>Arcanobacterium (Corynebacterium) haemolyticum</i>	Diffuse, erythematous, maculopapular eruption involving trunk and proximal extremities; may desquamate	Children and young adults	Exudative pharyngitis, lymphadenopathy	145
West Nile fever	West Nile virus	Maculopapular eruption involving the trunk, extremities, and head or neck; rash in 20–50% of cases	Mosquito bite; rarely, blood transfusion or transplanted organ	Headache, weakness, malaise, myalgia, neuroinvasive disease (encephalitis, meningitis, flaccid paralysis)	204
Zika virus infection	Zika virus	Pruritic macular and papular erythema; rash may begin on trunk and descend to lower body; conjunctival injection; palatal petechiae may occur	Mosquito bite; sexual transmission or blood transfusion less common	Arthralgia (especially of small joints), myalgia, lymphadenopathy, headache, low-grade fever; illness in pregnancy may cause severe birth defects, including microcephaly; neurologic complications, including Guillain-Barré, may occur	204
Peripheral Eruptions					
Chronic meningococcemia, disseminated gonococcal infection, ^a human parvovirus B19 infection ^e	—	—	—	—	150, 151, 192
Rocky Mountain spotted fever	<i>Rickettsia rickettsii</i>	Rash beginning on wrists and ankles and spreading centripetally; appears on palms and soles later in disease; lesion evolution from blanchable macules to petechiae	Tick vector; widespread but more common in southeastern and southwest-central U.S.	Headache, myalgias, abdominal pain; mortality rates up to 40% if untreated	182
Secondary syphilis	<i>Treponema pallidum</i>	Coincident primary chancre in 10% of cases; copper-colored, scaly papular eruption, diffuse but prominent on palms and soles; rash never vesicular in adults; condyloma latum, mucous patches, and alopecia in some cases	Sexually transmitted	Fever, constitutional symptoms	177
Chikungunya fever	Chikungunya virus	Maculopapular eruption; typically occurs on trunk, but also occurs on extremities and face	<i>Aedes aegypti</i> and <i>A. albopictus</i> mosquito bites; tropical and subtropical regions	Severe polyarticular, migratory arthralgias, especially involving small joints (e.g., hands, wrists, ankles)	204
Hand-foot-and-mouth disease	Coxsackievirus A16 and enterovirus 71 most common causes; coxsackievirus A6 associated with atypical syndrome	Tender vesicles, erosions in mouth; 0.25-cm papules on hands and feet with rim of erythema evolving into tender vesicles; shedding of nails can occur 1–2 months after acute illness; coxsackievirus A6 lesions extend to perioral area, extremities, trunk, buttocks, genitals, and areas affected by eczema	Summer and fall; primarily children <10 years old; multiple family members; coxsackievirus A6 infection also occurs in young adults	Transient fever; enterovirus 71 can be associated with brain stem encephalitis, flaccid paralysis resembling polio, or aseptic meningitis	199
Erythema multiforme (EM)	Infection, drugs, idiopathic causes	Target lesions (central erythema surrounded by area of clearing and another rim of erythema) up to 2 cm; symmetric on knees, elbows, palms, soles; spreads centripetally; papular, sometimes vesicular; when extensive and involving mucous membranes, termed EM major	Herpes simplex virus or <i>Mycoplasma pneumoniae</i> infection; drug intake (i.e., sulfa, phenytoin, penicillin)	50% of patients <20 years old; fever more common in most severe form, EM major, which can be confused with Stevens-Johnson syndrome (but EM major lacks prominent skin sloughing)	— ^f
Rat-bite fever (Haverhill fever)	<i>Streptobacillus moniliformis</i>	Maculopapular eruption over palms, soles, and extremities; tends to be more severe at joints; eruption sometimes becoming generalized; may be purpuric; may desquamate	Rat bite, ingestion of contaminated food	Myalgias; arthritis (50%); fever recurrence in some cases	136

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLIC FACTORS	CLINICAL SYNDROME	CHAPTER
Peripheral Eruptions (Continued)					
Bacterial endocarditis	Streptococcus, Staphylococcus, etc.	Subacute course (e.g., viridans streptococci): Osler's nodes (tender pink nodules on finger or toe pads); petechiae on skin and mucosa; splinter hemorrhages. Acute course (e.g., <i>Staphylococcus aureus</i>): Janeway lesions (painless erythematous or hemorrhagic macules, usually on palms and soles)	Abnormal heart valve (e.g., viridans streptococci), intravenous drug use	New or changing heart murmur	123
Confluent Desquamative Erythemas					
Scarlet fever (second disease)	Group A Streptococcus (pyrogenic exotoxins A, B, C)	Diffuse blanchable erythema beginning on face and spreading to trunk and extremities; circumoral pallor; "sandpaper" texture to skin; accentuation of linear erythema in skin folds (Pastia's lines); enanthem of white evolving into red "strawberry" tongue; desquamation in second week	Most common among children 2–10 years old; usually follows group A streptococcal pharyngitis	Fever, pharyngitis, headache	143
Kawasaki disease	Idiopathic causes	Rash similar to scarlet fever (scarlatiniform) or EM; fissuring of lips, strawberry tongue; conjunctivitis; edema of hands, feet; desquamation later in disease	Children <8 years old	Cervical adenopathy, pharyngitis, coronary artery vasculitis	54, 356
Streptococcal toxic shock syndrome	Group A Streptococcus (associated with pyrogenic exotoxin A and/or B or certain M types)	When present, rash often scarlatiniform	May occur in setting of severe group A streptococcal infections (e.g., necrotizing fasciitis, bacteremia, pneumonia)	Multiorgan failure, hypotension; mortality rate 30%	143
Staphylococcal toxic shock syndrome	<i>S. aureus</i> (toxic shock syndrome toxin 1, enterotoxins B and others)	Diffuse erythema involving palms; pronounced erythema of mucosal surfaces; conjunctivitis; desquamation 7–10 days into illness	Colonization with toxin-producing <i>S. aureus</i>	Fever >39°C (>102°F), hypotension, multiorgan dysfunction	142
Staphylococcal scalded-skin syndrome	<i>S. aureus</i> , phage group II	Diffuse tender erythema, often with bullae and desquamation; Nikolsky's sign	Colonization with toxin-producing <i>S. aureus</i> ; occurs in children <10 years old (termed Ritter's disease in neonates) or adults with renal dysfunction	Irritability; nasal or conjunctival secretions	142
Exfoliative erythroderma syndrome	Underlying psoriasis, eczema, drug eruption, mycosis fungoides	Diffuse erythema (often scaling) interspersed with lesions of underlying condition	Usually occurs in adults over age 50; more common among men	Fever, chills (i.e., difficulty with thermoregulation); lymphadenopathy	54, 56
DRESS [drug-induced hypersensitivity syndrome (DIHS)]	Aromatic anticonvulsants; other drugs, including sulfonamides, minocycline	Maculopapular eruption (mimicking exanthematous drug rash), sometimes progressing to exfoliative erythroderma; profound edema, especially facial; pustules may occur	Individuals genetically unable to detoxify arene oxides (anticonvulsant metabolites), patients with slow N-acetylating capacity (sulfonamides)	Lymphadenopathy, multiorgan failure (especially hepatic), eosinophilia, atypical lymphocytes; mimics sepsis	56
Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN)	Drugs (80% of cases; often allopurinol, anticonvulsants, antibiotics), infection, idiopathic factors	Erythematous and purpuric macules, sometimes targetoid, or diffuse erythema progressing to bullae, with sloughing and necrosis of entire epidermis; Nikolsky's sign; involves mucosal surfaces; TEN (>30% epidermal necrosis) is maximal form; SJS involves <10% of epidermis; SJS/TEN overlap involves 10–30% of epidermis	Uncommon among children; more common among patients with HIV infection, systemic lupus erythematosus, certain HLA types, or slow acetylators	Dehydration, sepsis sometimes resulting from lack of normal skin integrity; mortality rates up to 30%	56
Vesiculobullous or Pustular Eruptions					
Hand-foot-and-mouth syndrome ^e ; staphylococcal scalded-skin syndrome; TEN ^b ;DRESS ^b	—	—	—	—	— ^f

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Varicella (chickenpox)	VZV	Macules (2–3 mm) evolving into papules, then vesicles (sometimes umbilicated), on an erythematous base (“dewdrops on a rose petal”); pustules then forming and crusting; lesions appearing in crops; may involve scalp, mouth; intensely pruritic	Usually affects children; 10% of adults susceptible; most common in late winter and spring; incidence down by 90% in U.S. as a result of varicella vaccination	Malaise; generally mild disease in healthy children; more severe disease with complications in adults and immunocompromised children	188
Pseudomonas “hot-tub” folliculitis	Pseudomonas aeruginosa	Pruritic erythematous follicular, papular, vesicular, or pustular lesions that may involve axillae, buttocks, abdomen, and especially areas occluded by bathing suits; can manifest as tender isolated nodules on palmar or plantar surfaces (the latter designated “Pseudomonas hot-foot syndrome”)	Bathers in hot tubs or swimming pools; occurs in outbreaks	Earache, sore eyes and/or throat; fever may be absent; generally self-limited	159
Variola (smallpox)	Variola major virus	Red macules on tongue and palate evolving to papules and vesicles; skin macules evolving to papules, then vesicles, then pustules over 1 week, with subsequent lesion crusting; lesions initially appearing on face and spreading centrifugally from trunk to extremities; differs from varicella in that (1) skin lesions in any given area are at same stage of development and (2) there is a prominent distribution of lesions on face and extremities (including palms, soles)	Nonimmune individuals exposed to smallpox	Prodrome of fever, headache, backache, myalgias; vomiting in 50% of cases	S2
Primary herpes simplex virus (HSV) infection	HSV	Erythema rapidly followed by hallmark painful grouped vesicles that may evolve into pustules that ulcerate, especially on mucosal surfaces; lesions at site of inoculation: commonly gingivostomatitis for HSV-1 and genital lesions for HSV-2; recurrent disease milder (e.g., herpes labialis does not involve oral mucosa)	Primary infection most common among children and young adults for HSV-1 and among sexually active young adults for HSV-2; no fever in recurrent infection	Regional lymphadenopathy	187
Disseminated herpesvirus infection	Varicella-zoster virus (VZV) or HSV	Generalized vesicles that can evolve to pustules and ulcerations; individual lesions similar for VZV and HSV. <i>Zoster cutaneous dissemination</i> : >25 lesions extending outside involved dermatome. <i>HSV</i> : extensive, progressive mucocutaneous lesions that may occur in absence of dissemination, sometimes disseminate in eczematous skin (<i>eczema herpeticum</i>); <i>HSV</i> visceral dissemination may occur with only localized mucocutaneous disease; in disseminated neonatal disease, skin lesions diagnostically helpful when present, but rash absent in a substantial minority of cases	Patients with immunosuppression, eczema; neonates	Visceral organ involvement (e.g., liver, lungs) in some cases; neonatal disease particularly severe	133, 187, 188
Rickettsialpox	<i>Rickettsia akari</i>	Eschar found at site of mite bite; generalized rash involving face, trunk, extremities; may involve palms and soles; <100 papules and plaques (2–10 mm); tops of lesions developing vesicles that may evolve into pustules	Seen in urban settings; transmitted by mouse mites	Headache, myalgias, regional adenopathy; mild disease	182

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Vesiculobullous or Pustular Eruptions (Continued)					
Acute generalized exanthematous pustulosis	Drugs (mostly anticonvulsants or antimicrobials); also viral	Tiny sterile nonfollicular pustules on erythematous, edematous skin; begins on face and in body folds, then becomes generalized	Appears 2–21 days after start of drug therapy, depending on whether patient has been sensitized	Acute fever, pruritus, leukocytosis	56
Disseminated <i>Vibrio vulnificus</i> infection	<i>V. vulnificus</i>	Erythematous lesions evolving into hemorrhagic bullae and then into necrotic ulcers	Patients with cirrhosis, diabetes, renal failure; exposure by ingestion of contaminated saltwater, seafood	Hypotension; mortality rate 50%	163
Ecthyma gangrenosum	<i>P. aeruginosa</i> , other gram-negative rods, fungi	Indurated plaque evolving into hemorrhagic bulla or pustule that sloughs, resulting in eschar formation; erythematous halo; most common in axillary, groin, perianal regions	Usually affects neutropenic patients; occurs in up to 28% of individuals with <i>Pseudomonas</i> bacteremia	Clinical signs of sepsis	159
Urticaria-Like Eruptions					
Urticular vasculitis	Serum sickness, often due to infection (including hepatitis B viral, enteroviral, parasitic), drugs; connective tissue disease	Erythematous, edematous “urticaria-like” plaques, pruritic or burning; unlike urticaria: typical lesion duration >24 h (up to 5 days) and lack of complete lesion blanching with compression due to hemorrhage	Patients with serum sickness (including hepatitis B), connective tissue disease	Fever variable; arthralgias/ arthritis	356 ^f
Nodular Eruptions					
Disseminated infection	Fungal infections (e.g., candidiasis, histoplasmosis, cryptococcosis, sporotrichosis, coccidioidomycosis); mycobacteria	Subcutaneous nodules (up to 3 cm); fluctuance, draining common with mycobacteria; necrotic nodules (extremities, periorbital or nasal regions) common with <i>Aspergillus</i> , <i>Mucor</i>	Immunocompromised hosts (e.g., bone marrow transplant recipients, patients undergoing chemotherapy, HIV-infected patients)	Features vary with organism	— ^f
Erythema nodosum (septal panniculitis)	Infections (e.g., streptococcal, fungal, mycobacterial, yersinial); drugs (e.g., sulfas, penicillins, oral contraceptives); sarcoidosis; idiopathic causes	Large, violaceous, nonulcerative, subcutaneous nodules; exquisitely tender; usually on lower legs but also on upper extremities	More common among girls and women 15–30 years old	Arthralgias (50%); features vary with associated condition	— ^f
Sweet syndrome (acute febrile neutrophilic dermatosis)	<i>Yersinia</i> infection; upper respiratory infection; inflammatory bowel disease; pregnancy; malignancy (usually hematologic); drugs (G-CSF)	Tender red or blue edematous nodules giving impression of vesiculation; usually on face, neck, upper extremities; when on lower extremities, may mimic erythema nodosum	More common among women and among persons 30–60 years old; 20% of cases associated with malignancy (men and women equally affected in this group)	Headache, arthralgias, leukocytosis	54
Bacillary angiomatosis	<i>Bartonella henselae</i> , <i>B. quintana</i>	Many forms, including erythematous, smooth vascular nodules; friable, exophytic lesions; erythematous plaques (may be dry, scaly); subcutaneous nodules (may be erythematous)	Immunosuppressed individuals, especially those with advanced HIV infection	Peliosis of liver and spleen in some cases; lesions sometimes involving multiple organs; bacteremia	167
Purpuric Eruptions					
Rocky Mountain spotted fever, rat-bite fever, endocarditis ^g ; epidemic typhus ^h ; dengue fever ^{i,j,k} ; human parvovirus B19 infection ^e	—	—	—	—	— ^f
Acute meningococcemia	<i>Neisseria meningitidis</i>	Initially pink maculopapular lesions evolving into petechiae; petechiae rapidly becoming numerous, sometimes enlarging and becoming vesicular; trunk, extremities most commonly involved; may appear on face, hands, feet; may include purpura fulminans (see below) reflecting DIC	Most common among children, individuals with asplenia or terminal complement component deficiency (C5–C8)	Hypotension, meningitis (sometimes preceded by upper respiratory infection)	150

(Continued)

TABLE 16-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Purpura fulminans	Severe DIC	Large ecchymoses with sharply irregular shapes evolving into hemorrhagic bullae and then into black necrotic lesions	Individuals with sepsis (e.g., involving <i>N. meningitidis</i>), malignancy, or massive trauma; asplenic patients at high risk for sepsis	Hypotension	150, 297
Chronic meningococcemia	<i>N. meningitidis</i>	Variety of recurrent eruptions, including pink maculopapular; nodular (usually on lower extremities); petechial (sometimes developing vesicular centers); purpuric areas with pale blue-gray centers	Individuals with complement deficiencies	Fevers, sometimes intermittent; arthritis, myalgias, headache	150
Disseminated gonococcal infection	<i>Neisseria gonorrhoeae</i>	Papules (1–5 mm) evolving over 1–2 days into hemorrhagic pustules with gray necrotic centers; hemorrhagic bullae occurring rarely; lesions (usually <40) distributed peripherally near joints (more commonly on upper extremities)	Sexually active individuals (more often females), some with complement deficiency	Low-grade fever, tenosynovitis, arthritis	151
Enteroviral petechial rash	Usually echovirus 9 or coxsackievirus A9	Disseminated petechial lesions (may also be maculopapular, vesicular, or urticarial)	Often occurs in outbreaks	Pharyngitis, headache; aseptic meningitis with echovirus 9	199
Viral hemorrhagic fever	Arboviruses (including dengue) and arenaviruses	Petechiae	Residence in or travel to endemic areas, other virus exposure	Triad of fever, shock, hemorrhage from mucosa or gastrointestinal tract	204, 205
Thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome	Idiopathic, bloody diarrhea caused by Shiga toxin-generating bacteria (e.g., <i>Escherichia coli</i> O157:H7), deficiency in ADAMTS13 (cleaves von Willebrand factor), drugs (e.g., quinine, chemotherapy, immunosuppression)	Petechiae	Individuals with <i>E. coli</i> O157:H7 gastroenteritis (especially children), cancer chemotherapy, HIV infection, autoimmune diseases, pregnant/postpartum women	Fever (not always present), microangiopathic hemolytic anemia, thrombocytopenia, renal dysfunction, neurologic dysfunction; coagulation studies normal	54, 96, 111, 156, 161
Cutaneous small-vessel vasculitis (leukocytoclastic vasculitis)	Infections (including group A streptococcal infection, hepatitis B or C), drugs, idiopathic factors	Palpable purpuric lesions appearing in crops on legs or other dependent areas; may become vesicular or ulcerative	Occurs in a wide spectrum of diseases, including connective tissue disease, cryoglobulinemia, malignancy, Henoch-Schönlein purpura (HSP); more common among children	Fever (not always present), malaise, arthralgias, myalgias; systemic vasculitis in some cases; renal, joint, and gastrointestinal involvement common in HSP	54
Eruptions with Ulcers and/or Eschars					
Scrub typhus, rickettsial spotted fevers, rat-bite fever ^a ; rickettsialpox, ecthyma gangrenosum ^b	—	—	—	—	— ^c
Tularemia	<i>Francisella tularensis</i>	Ulceroglandular form: erythematous, tender papule evolves into necrotic, tender ulcer with raised borders; in 35% of cases, eruptions (maculopapular, vesiculopapular, acneiform, or urticarial; erythema nodosum; or EM) may occur	Exposure to ticks, biting flies, infected animals	Fever, headache, lymphadenopathy	165
Anthrax	<i>Bacillus anthracis</i>	Pruritic papule enlarging and evolving into a 1- by 3-cm painless ulcer surrounded by vesicles and then developing a central eschar with edema; residual scar	Exposure to infected animals or animal products, other exposure to anthrax spores	Lymphadenopathy, headache	S2

^aSee “Purpuric Eruptions.” ^bSee “Confluent Desquamative Erythemas.” ^cRash is rare in human granulocytotropic ehrlichiosis or anaplasmosis (caused by *Anaplasma phagocytophila*; most common in the upper midwestern and northeastern United States). ^dSee “Viral hemorrhagic fever” under “Purpuric Eruptions” for dengue hemorrhagic fever/dengue shock syndrome. ^eSee “Centrally Distributed Maculopapular Eruptions.” ^fSee etiology-specific chapters. ^gSee “Peripheral Eruptions.” ^hSee “Vesiculobullous or Pustular Eruptions.”

Abbreviations: CNS, central nervous system; DIC, disseminated intravascular coagulation; G-CSF, granulocyte colony-stimulating factor; HLA, human leukocyte antigen.

induced to form bullae with light lateral pressure (Nikolsky's sign). In a mild form, a scarlatiniform eruption mimics scarlet fever, but the patient does not exhibit a strawberry tongue or circumoral pallor. In contrast to the staphylococcal scalded-skin syndrome, in which the cleavage plane is superficial in the epidermis, *toxic epidermal necrolysis* (Chap. 56), a maximal variant of *Stevens-Johnson syndrome*, involves sloughing of the entire epidermis, resulting in severe disease. *Exfoliative erythroderma syndrome* (Chaps. 54 and 56) is a serious reaction associated with systemic toxicity that is often due to eczema, psoriasis, a drug reaction, or mycosis fungoides. *Drug rash with eosinophilia and systemic symptoms* (DRESS), often due to antiepileptic and antibiotic agents (Chap. 56), initially appears similar to an exanthematous drug reaction but may progress to exfoliative erythroderma; it is accompanied by multiorgan failure and has an associated mortality rate of ~10%.

VESICULOBULLOUS OR PUSTULAR ERUPTIONS

Varicella (Chap. 188) is highly contagious, often occurring in winter or spring, and is characterized by pruritic lesions that, within a given region of the body, are in different stages of development at any point in time. In immunocompromised hosts, varicella vesicles may lack the characteristic erythematous base or may appear hemorrhagic. Lesions of *Pseudomonas* "hot-tub" folliculitis (Chap. 159) are also pruritic and may appear similar to those of varicella. However, hot-tub folliculitis generally occurs in outbreaks after bathing in hot tubs or swimming pools, and lesions occur in regions occluded by bathing suits. Lesions of *variola* (smallpox) (Chap. S2) also appear similar to those of varicella but are all at the same stage of development in a given region of the body. Variola lesions are most prominent on the face and extremities, while varicella lesions are most prominent on the trunk. *Herpes simplex virus infection* (Chap. 187) is characterized by hallmark grouped vesicles on an erythematous base. Primary herpes infection is accompanied by fever and toxicity, while recurrent disease is milder. *Rickettsialpox* (Chap. 182) is often documented in urban settings and is characterized by vesicles followed by pustules. It can be distinguished from varicella by an eschar at the site of the mouse-mite bite and the papule/plaque base of each vesicle. *Acute generalized exanthematous pustulosis* should be considered in individuals who are acutely febrile and are taking new medications, especially anticonvulsant or antimicrobial agents (Chap. 56). Disseminated *Vibrio vulnificus* infection (Chap. 163) or *ecthyma gangrenosum* due to *Pseudomonas aeruginosa* (Chap. 159) should be considered in immunosuppressed individuals with sepsis and hemorrhagic bullae.

URTICARIA-LIKE ERUPTIONS

Individuals with classic urticaria ("hives") usually have a hypersensitivity reaction without associated fever. In the presence of fever, urticaria-like eruptions are most often due to *urticular vasculitis* (Chap. 356). Unlike individual lesions of classic urticaria, which last up to 24 h, these lesions may last 3–5 days. Etiologies include serum sickness (often induced by drugs such as penicillins, sulfas, salicylates, or barbiturates), connective-tissue disease (e.g., systemic lupus erythematosus or Sjögren's syndrome), and infection (e.g., with hepatitis B virus, enteroviruses, or parasites). Malignancy, especially lymphoma, may be associated with fever and chronic urticaria (Chap. 54).

NODULAR ERUPTIONS

In immunocompromised hosts, nodular lesions often represent disseminated infection. Patients with disseminated *candidiasis* (often due to *Candida tropicalis*) may have a triad of fever, myalgias, and eruptive nodules (Chap. 211). Disseminated *cryptococcosis* lesions (Chap. 210) may resemble molluscum contagiosum (Chap. 191). Necrosis of nodules should raise the suspicion of *aspergillosis* (Chap. 212) or *mucomycosis* (Chap. 213). *Erythema nodosum* presents with exquisitely tender nodules on the lower extremities. *Sweet syndrome* (Chap. 54) should be considered in individuals with multiple nodules and plaques, often so edematous that they give the appearance of vesicles or bullae. Sweet syndrome may occur in individuals with infection, inflammatory bowel disease, or malignancy and can also be induced by drugs.

PURPURIC ERUPTIONS

Acute meningococcemia (Chap. 150) classically presents in children as a petechial eruption, but initial lesions may appear as blanchable macules or urticaria. Rocky Mountain spotted fever should be considered in the differential diagnosis of acute meningococcemia. *Echovirus 9 infection* (Chap. 199) may mimic acute meningococcemia; patients should be treated as if they have bacterial sepsis because prompt differentiation of these conditions may be impossible. Large ecchymotic areas of *purpura fulminans* (Chaps. 150 and 297) reflect severe underlying disseminated intravascular coagulation, which may be due to infectious or noninfectious causes. The lesions of *chronic meningococcemia* (Chap. 150) may have a variety of morphologies, including petechial. Purpuric nodules may develop on the legs and resemble erythema nodosum but lack its exquisite tenderness. Lesions of *disseminated gonococcemia* (Chap. 151) are distinctive, sparse, countable hemorrhagic pustules, usually located near joints. The lesions of chronic meningococcemia and those of gonococcemia may be indistinguishable in terms of appearance and distribution. *Viral hemorrhagic fever* (Chaps. 204 and 205) should be considered in patients with an appropriate travel history and a petechial rash. *Thrombotic thrombocytopenic purpura* (Chaps. 54, 96, and 111) and *hemolytic-uremic syndrome* (Chaps. 111, 156, and 161) are closely related and are noninfectious causes of fever and petechiae. *Cutaneous small-vessel vasculitis* (*leukocytoclastic vasculitis*) typically manifests as palpable purpura and has a wide variety of causes (Chap. 54).

ERUPTIONS WITH ULCERS OR ESCHARS

The presence of an ulcer or eschar in the setting of a more widespread eruption can provide an important diagnostic clue. For example, an eschar may suggest the diagnosis of *scrub typhus* or *rickettsialpox* (Chap. 182) in the appropriate setting. In other illnesses (e.g., anthrax) (Chap. S2), an ulcer or eschar may be the only skin manifestation.

FURTHER READING

- CHERRY JD: Cutaneous manifestations of systemic infections, in *Feigin and Cherry's Textbook of Pediatric Infectious Diseases*, 7th ed. JD Cherry et al (eds). Houston, Elsevier Saunders, 2014, pp 741–768.
 WEBER DJ et al: The acutely ill patient with fever and rash, in *Principles and Practice of Infectious Diseases*, vol 1, 8th ed. JI Bennett et al (eds). Philadelphia, Elsevier Saunders, 2015, pp 732–747.
 WOLFF K et al: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 7th ed. New York, McGraw-Hill, 2013.
 WOLFF K et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 8th ed. New York, McGraw-Hill, 2012.

17

Fever of Unknown Origin

Chantal P. Bleeker-Rovers,
 Jos W. M. van der Meer



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 classic 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 $\geq 38.3^{\circ}\text{C}$ ($\geq 101^{\circ}\text{F}$) 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 $\geq 38.3^{\circ}\text{C}$ ($\geq 101^{\circ}\text{F}$) 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).

■ ETIOLOGY AND EPIDEMIOLOGY

The range of FUO etiologies has evolved over time 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 four decades ago.



Table 17-1 summarizes the findings of large studies on FUO conducted over the past 25 years. In general, infection accounts for about one-fifth of cases of FUO in Western countries; next in frequency are noninfectious inflammatory diseases (NIIDs, which include “collagen or rheumatic diseases,” vasculitis syndromes, granulomatous disorders, and autoinflammatory syndromes) and neoplasms. In geographic areas outside the West, infections are a much more common cause of FUO (43% vs 17%), while the proportions of cases due to NIIDs and neoplasms are similar. Up to 50% of cases caused by infections in patients with FUO outside Western nations are due to tuberculosis, which is a less common cause in the United States and Western Europe. The number of FUO patients diagnosed with NIIDs probably will not decrease in the near future, as fever may precede more typical manifestations or serologic evidence of these diseases by months. Moreover, many NIIDs can be diagnosed only after prolonged observation and exclusion of other diseases.

In the West, the proportion of patients who remain undiagnosed is higher than in non-Western populations and has been increasing over figures reported in studies before the 1990s. An important factor contributing to the seemingly high diagnostic failure rate is that a diagnosis is more often being established before 3 weeks have elapsed, given that patients with fever tend to seek medical advice earlier and that better diagnostic techniques, such as CT and MRI, are available; therefore, only the cases that are most difficult to diagnose continue to

TABLE 17-1 Etiology of Fever of Unknown Origin (FUO) Over the Past 25 Years: Findings from Large FUO Studies

FIRST AUTHOR (COUNTRY, YEAR OF PUBLICATION)	NO. OF PATIENTS (RECRUITMENT PERIOD)	PERCENTAGE OF CASES DUE TO INDICATED CAUSE				
		INFECTIONS	NONINFECTIOUS INFLAMMATORY DISEASES	NEOPLASMS	MISCELLANEOUS	UNKNOWN
Western Countries						
De Kleijn et al. (Netherlands, 1997)	167 (1992–1994)	26	24	13	8	30
Vanderschueren et al. (Belgium, 2003)	185 (1990–1999)	11	18	10	8	53
Hot et al. (France, 2005)	280 (1995–2005)	11	20	27	9	33
Zenone et al. (France, 2006)	144 (1999–2005)	23	26	10	15	26
Bleeker-Rovers (Netherlands, 2007)	73 (2003–2005)	16	22	7	4	51
Mansueto et al. (Italy, 2008)	91 (1991–2002)	32	12	14	10	32
Vanderschueren et al. (Belgium, 2009)	114 (2003–2007)	15	22	13	10	40
Efstathiou et al. (Greece, 2010)	112 (2001–2007)	30	33	11	5	21
Pedersen et al. (Denmark, 2012)	52 (2005–2010)	19	33	8	0	40
Robine et al. (France, 2014)	103 (2002–2012)	12	30	3	5	51
Vanderschueren et al. (Belgium, 2014)	436 (2000–2010)	17	24	11	10	39
Total	1757	19	24	12	8	38
Other Geographic Locations						
Tabak et al. (Turkey, 2003)	117 (1984–2001)	34	29	19	4	14
Saltoglu et al. (Turkey, 2004)	87 (1994–2002)	59	18	14	2	7

(Continued)

TABLE 17-1 Etiology of Fever of Unknown Origin (FUO) Over the Past 25 Years: Findings from Large FUO Studies (Continued)

FIRST AUTHOR (COUNTRY, YEAR OF PUBLICATION)	NO. OF PATIENTS (RECRUITMENT PERIOD)	PERCENTAGE OF CASES DUE TO INDICATED CAUSE				
		INFECTIONS	NONINFECTIOUS INFLAMMATORY DISEASES	NEOPLASMS	MISCELLANEOUS	UNKNOWN
Ergonul et al. (Turkey, 2005)	80 (1993–1999)	52	16	18	3	11
Brahim et al. (Turkey, 2005)	97 (1990–2005)	36	8	16	5	35
Chin et al. (Taiwan, 2006)	94 (2001–2002)	57	7	9	9	18
Colpan et al. (Turkey, 2007)	71 (2001–2004)	45	27	14	6	9
Hu et al. (China, 2008)	142 (2002–2003)	36	32	13	5	14
Kucukardali et al. (Turkey, 2008)	154 (2003–2004)	34	31	14	5	16
Ali-Eldin et al. (Egypt, 2011)	93 (2009–2010)	42	15	30	0	12
Bandyopadhyay et al. (India, 2011)	164 (2008–2009)	55	11	22	0	12
Mete et al. (Turkey, 2012)	100 (2001–2009)	26	38	14	2	20
Ma et al. (China, 2012)	397 (2000–2009)	49	18	16	7	10
Ryuko et al. (Japan, 2013)	174 (2004–2010)	41	27	7	6	19
Mahmood et al. (Pakistan, 2013)	205 (2006–2011)	49	20	13	2	17
Alvi et al. (Iran, 2013)	106 (2007–2011)	44	18	12	10	15
Naito et al. (Japan, 2013)	121 (2011)	23	31	11	12	23
Yamanouchi et al. (Japan, 2014)	256 (1994–2012)	28	18	10	15	29
Moawad et al. (Turkey, 2014)	98 (1995–2008)	33	14	18	18	17
Yu et al. (China, 2014)	107 (2010–2011)	30	17	18	14	22
Mir et al. (India, 2014)	91 (2010–2012)	44	12	12	4	27
Kabapy et al. (Egypt, 2015)	979 (2009–2010)	79	17	1	1	2
Montasser et al. (Egypt, 2015)	217 (unknown)	66	7	7	12	8
Popvska-Jovicic et al. (Serbia, 2016)	74	38	26	15	18	4
Total	4024	43	20	14	7	16

meet the criteria for FUO. Furthermore, most patients who have FUO without a diagnosis currently do well, and thus a less aggressive diagnostic approach may be used in clinically stable patients once diseases with immediate therapeutic or prognostic consequences have been ruled out to a reasonable extent. This factor may be especially relevant to patients with recurrent fever who are asymptomatic between febrile episodes. 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

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 17-2** 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 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 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

TABLE 17-2 All Reported Causes of 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, spondylositis, xanthogranulomatous urinary tract infection
Bacterial, specific	Actinomycosis, atypical mycobacterial infection, bartonellosis, brucellosis, <i>Campylobacter</i> infection, <i>Chlamydia pneumoniae</i> infection, chronic meningococcemia, ehrlichiosis, gonococcemia, legionellosis, leptospirosis, listeriosis, louse-borne relapsing fever (<i>Borrelia recurrentis</i>), Lyme disease, melioidosis (<i>Pseudomonas pseudomallei</i>), <i>Mycoplasma</i> infection, nocardiosis, psittacosis, Q fever (<i>Coxiella burnetii</i>), rickettsiosis, <i>Spirillum minor</i> infection, <i>Streptobacillus moniliformis</i> infection, syphilis, tick-borne relapsing fever (<i>Borrelia duttonii</i>), tuberculosis, tularemia, typhoid fever and other salmonelloses, Whipple's disease (<i>Tropheryma whipplei</i>), yersiniosis
Fungal	Aspergillosis, blastomycosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, <i>Malassezia furfur</i> infection, paracoccidioidomycosis, <i>Pneumocystis jirovecii</i> 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, Behcet'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 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

^aThis table includes all causes of FUO that have been described in the literature. ^bCAPS 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.

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 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. paraprophilus*, *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.

Of the NIIDs, large-vessel vasculitis, polymyalgia rheumatica, sarcoidosis, familial Mediterranean fever, and adult-onset Still's disease are rather common diagnoses in patients with FUO. The hereditary autoinflammatory syndromes are very rare 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 that commencing 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. 15 and 455**) 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; typically these patients 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 a dissociation between pulse rate and temperature.

Previous studies of FUO have shown that a diagnosis is more likely 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 17-1 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 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/ μ L.

Microbiologic serology should not be included in the diagnostic workup of patients without PDCs for specific infections. A TST 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 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, funduscopic examination by an ophthalmologist may be useful in the early stage of the diagnostic workup. When the first-stage diagnostic tests do not lead to a diagnosis, 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

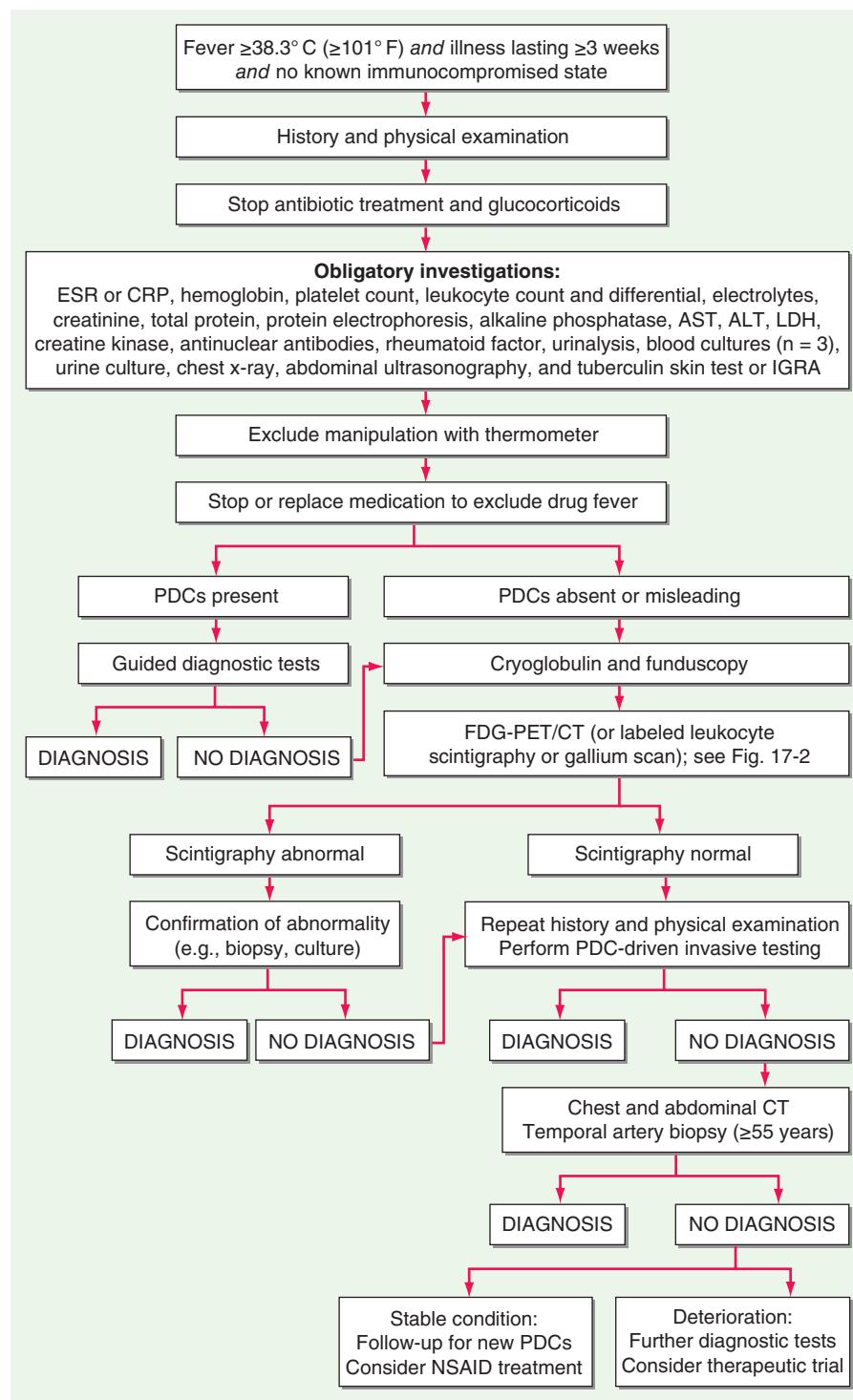


FIGURE 17-1 Structured approach to patients with FUO. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FDG-PET/CT, ¹⁸F-fluorodeoxyglucose positron emission tomography combined with low-dose CT; IGRA, interferon γ release assay; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; PDCs, potentially diagnostic clues (all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis).

examination, and obligatory tests. The search for PDCs should be directed toward clues matching known recurrent syndromes (Table 17-3). 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 scintigraphic imaging (see below), should be performed only during a febrile episode 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.

Scintigraphy Scintigraphic imaging is a noninvasive method allowing delineation of foci in all parts of the body on the basis of functional changes in tissues. This procedure plays an important role in the diagnosis of patients with FUO in clinical practice. Conventional scintigraphic methods used in clinical practice are ⁶⁷Ga-citrate scintigraphy and ¹¹¹In- or ^{99m}Tc-labeled leukocyte scintigraphy. Focal infectious and inflammatory processes can also be

TABLE 17-3 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 (<i>Borrelia recurrentis</i>), melioidosis (<i>Pseudomonas pseudomallei</i>), Q fever (<i>Coxiella burnetii</i>), salmonellosis, <i>Spirillum minor</i> infection, <i>Streptobacillus moniliformis</i> infection, syphilis, tick-borne relapsing fever (<i>Borrelia duttonii</i>), tularemia, Whipple's disease (<i>Tropheryma whipplei</i>), 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, Behcet'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 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

^aThis table includes all causes of recurrent fever that have been described in the literature. ^bCAPS 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.

detected by several radiologic techniques, such as CT, MRI, and ultrasound. However, because of the lack of substantial pathologic changes in the early phase, infectious and inflammatory foci cannot be detected at this time. Furthermore, distinguishing active infectious or inflammatory lesions from residual changes due to cured processes or surgery remains critical. Finally, CT and MRI routinely provide information on only part of the body, while scintigraphy readily allows whole-body imaging.

Fluorodeoxyglucose Positron Emission Tomography ¹⁸F-Fluorodeoxyglucose (FDG) positron emission tomography (PET) combined with 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. Normal uptake 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 non-specific way due to cytokine activation, which upregulates glucose transporters in bone marrow cells. Compared with conventional scintigraphy, 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. 17-2). 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.

In recent years, many cohort studies and several meta-analyses have focused on the diagnostic yield of PET and PET/CT in FUO. Although these studies are highly variable in terms of the selection of patients, follow-up, and the selection of a gold-standard reference point, all 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. In one study, FDG-PET was never helpful in diagnosing FUO in patients who had a normal CRP level and a normal ESR. In a meta-analysis on the performance, diagnostic yield, and management decision impact of nuclear imaging tests in patients with FUO, the diagnostic yield of gallium scintigraphy ranged from 21% to 54%, and, on average, the location of a source of fever was correctly localized in approximately one-third of patients. Moreover, in gallium scintigraphy, results do not become available for days, whereas FDG-PET/CT yields results within hours. In this meta-analysis, estimates of the diagnostic yield of labeled leukocyte scintigraphy ranged from 8% to 31%, and overall the cause of FUO was correctly identified on the basis of the scan results in only one-fifth of patients. 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



FIGURE 17-2 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.

FDG-PET/CT was more likely than alternative tests to correctly identify the cause of FUO.

Although scintigraphic techniques do not directly provide a definitive diagnosis, they often identify the anatomic location of a particular ongoing metabolic process and, with the help of other techniques such as biopsy and culture, facilitate timely diagnosis and treatment. 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. Results reported in the literature and the advantages offered by FDG-PET/CT indicate that conventional scintigraphic techniques should be replaced by FDG-PET/CT in the investigation of patients with FUO at institutions where this technique is available. 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.

LATER-STAGE DIAGNOSTIC TESTS

In some cases, more invasive tests are appropriate. Abnormalities found with scintigraphic techniques 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. In one study, pulmonary wedge excision, histologic examination of an excised tonsil, and biopsy of the peritoneum were performed in light of PDCs or abnormal FDG-PET results and yielded a diagnosis.

If no diagnosis is reached despite scintigraphic and PDC-driven histologic investigations or culture, second-stage screening diagnostic tests should be considered (Fig. 17-1). 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 have often been performed as a screening procedure in patients with FUO. In each of two recent 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.

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 but is not always effective once an attack is well under way. When familial Mediterranean fever 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 familial Mediterranean fever, 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.

ANAKINRA

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-1 α and IL-1 β . Anakinra is extremely effective in the treatment of many autoinflammatory syndromes, such as familial Mediterranean fever, cryopyrin-associated periodic syndrome, tumor necrosis factor receptor-associated periodic syndrome, mevalonate kinase deficiency (hyper IgD syndrome), and Schnitzler syndrome. 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

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. In a study by our group (Table 17-1), none of 37 FUO patients without a diagnosis died during a follow-up period of at least 6 months; 4 of 36 patients with a diagnosis died during follow-up as a result of infection ($n = 1$) or malignancy ($n = 3$). A large study on the prognosis of FUO (Vanderschueren et al, 2014; Table 17-1) included 436 patients and documented a mortality rate of 10%, of which 68% was related to the febrile illness—malignancy in most cases. In this study, only 4 of 168 patients in whom no diagnosis could be made died, all during their first admission. In two of these patients, diagnosis (lymphoma and pneumonia) was made during autopsy. Other studies have also shown that malignancy accounts for most FUO-related deaths. Non-Hodgkin's lymphoma carries a disproportionately high death toll. In nonmalignant FUO, fatality rates are very low. The good outcome in patients without a diagnosis confirms that potentially lethal occult diseases are very unusual and that empirical therapy with antibiotics, antituberculous agents, or glucocorticoids is rarely required in stable patients. In less affluent regions, infectious diseases are still a major cause of FUO, and outcomes may be different.

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.
- KNOCKAERT DC et al: Fever of unknown origin in adults: 40 years on. J Intern Med 253:263, 2003.
- MULDERS-MANDERS C et al: Fever of unknown origin. Clin Med 15:280, 2015.
- TAKEUCHI M et al: Nuclear imaging for classical fever of unknown origin: Meta-analysis. J Nucl Med 57:1913, 2016.
- VANDERSCHUEREN S et al: Mortality in patients presenting with fever of unknown origin. Acta Clin Belg 69:12, 2014.

Section 3 Nervous System Dysfunction

18

Syncope

Roy Freeman



Syncope is a transient, self-limited loss of consciousness due to acute global impairment of cerebral blood flow. The onset is rapid, duration brief, and recovery spontaneous and complete. Other causes of transient loss of consciousness need to be distinguished from syncope; these include seizures, vertebrobasilar ischemia, hypoxemia, and hypoglycemia. A syncopal prodrome (*presyncope*) is common, although loss of consciousness may occur without any warning symptoms. Typical presyncopal symptoms include dizziness, lightheadedness or faintness, weakness, fatigue, and visual and auditory disturbances. The causes of syncope can be divided into three general categories: (1) neurally mediated syncope (also called *reflex* or *vasovagal syncope*), (2) orthostatic hypotension, and (3) cardiac syncope.

Neurally mediated syncope comprises a heterogeneous group of functional disorders that are characterized by a transient change in the reflexes responsible for maintaining cardiovascular homeostasis. Episodic vasodilation (or loss of vasoconstrictor tone) and bradycardia occur in varying combinations, resulting in temporary failure of blood

pressure control. In contrast, in patients with orthostatic hypotension due to autonomic failure, these cardiovascular homeostatic reflexes are chronically impaired. Cardiac syncope may be due to arrhythmias or structural cardiac diseases that cause a decrease in cardiac output. The clinical features, underlying pathophysiologic mechanisms, therapeutic interventions, and prognoses differ markedly among these three causes.

■ EPIDEMIOLOGY AND NATURAL HISTORY

Syncope is a common presenting problem, accounting for ~3% of all emergency room visits and 1% of all hospital admissions. The annual cost for syncope-related hospitalization in the United States is ~\$2.4 billion. Syncope has a lifetime cumulative incidence of up to 35% in the general population. The peak incidence in the young occurs between ages 10 and 30 years, with a median peak around 15 years. Neurally mediated syncope is the etiology in the vast majority of these cases. In elderly adults, there is a sharp rise in the incidence of syncope after 70 years.

In population-based studies, neurally mediated syncope is the most common cause of syncope. The incidence is slightly higher in females than males. In young subjects, there is often a family history in first-degree relatives. Cardiovascular disease due to structural disease or arrhythmias is the next most common cause in most series, particularly in emergency room settings and in older patients. Orthostatic hypotension also increases in prevalence with age because of the reduced baroreflex responsiveness, decreased cardiac compliance, and attenuation of the vestibulosympathetic reflex associated with aging. In the elderly, orthostatic hypotension is substantially more common in institutionalized (54–68%) than community-dwelling (6%) individuals, an observation most likely explained by the greater prevalence of predisposing neurologic disorders, physiologic impairment, and vasoactive medication use among institutionalized patients.

The prognosis after a single syncopal event for all age groups is generally benign. In particular, syncope of noncardiac and unexplained origin in younger individuals has an excellent prognosis; life expectancy is unaffected. By contrast, syncope due to a cardiac cause, either structural heart disease or primary arrhythmic disease, is associated with an increased risk of sudden cardiac death and mortality from other causes. Similarly, mortality rate is increased in individuals with syncope due to orthostatic hypotension related to age and the associated comorbid conditions ([Table 18-1](#)).

■ PATHOPHYSIOLOGY

The upright posture imposes a unique physiologic stress upon humans; most, although not all, syncopal episodes occur from a standing position. Standing results in pooling of 500–1000 mL of blood in the lower

extremities and splanchnic circulation. There is a decrease in venous return to the heart and reduced ventricular filling that result in diminished cardiac output and blood pressure. These hemodynamic changes provoke a compensatory reflex response, initiated by the baroreceptors in the carotid sinus and aortic arch, resulting in increased sympathetic outflow and decreased vagal nerve activity ([Fig. 18-1](#)). The reflex increases peripheral resistance, venous return to the heart, and cardiac output and thus limits the fall in blood pressure. If this response fails, as is the case chronically in orthostatic hypotension and transiently in neurally mediated syncope, cerebral hypoperfusion occurs.

Syncope is a consequence of global cerebral hypoperfusion and thus represents a failure of cerebral blood flow autoregulatory mechanisms. Myogenic factors, local metabolites, and to a lesser extent autonomic neurovascular control are responsible for the autoregulation of cerebral blood flow ([Chap. 301](#)). The latency of the autoregulatory response is 5–10 s. Typically cerebral blood flow ranges from 50 to 60 mL/min per 100 g brain tissue and remains relatively constant over perfusion pressures ranging from 50 to 150 mmHg. Cessation of blood flow for 6–8 s will result in loss of consciousness, while impairment of consciousness ensues when blood flow decreases to 25 mL/min per 100 g brain tissue.

From the clinical standpoint, a fall in systemic systolic blood pressure to ~50 mmHg or lower will result in syncope. A decrease in cardiac output and/or systemic vascular resistance—the determinants of blood pressure—thus underlies the pathophysiology of syncope. Common causes of impaired cardiac output include decreased effective circulating blood volume; increased thoracic pressure; massive pulmonary embolus; cardiac brady- and tachyarrhythmias; valvular heart disease; and myocardial dysfunction. Systemic vascular resistance may be decreased by central and peripheral autonomic nervous system diseases, sympatholytic medications, and transiently during neurally mediated syncope. Increased cerebral vascular resistance, most frequently due to hypocapnia induced by hyperventilation, may also contribute to the pathophysiology of syncope.

Two patterns of electroencephalographic (EEG) changes occur in syncopal subjects. The first is a “slow-flat-slow” pattern ([Fig. 18-2](#)) in which normal background activity is replaced with high-amplitude slow delta waves. This is followed by sudden flattening of the EEG—a cessation or attenuation of cortical activity—followed by the return of slow waves, and then normal activity. A second pattern, the “slow pattern,” is characterized by increasing and decreasing slow wave activity only. The EEG flattening that occurs in the slow-flat-slow pattern is a marker of more severe cerebral hypoperfusion. Despite the presence of myoclonic movements and other motor activity during some syncopal events, EEG seizure discharges are not detected.

CLASSIFICATION

■ NEURALLY MEDIED SYNCOPES

Neurally mediated (reflex; vasovagal) syncope is the final pathway of a complex central and peripheral nervous system reflex arc. There is a sudden, transient change in autonomic efferent activity with increased parasympathetic outflow, plus sympathoinhibition (the vasodepressor response), resulting in bradycardia, vasodilation, and/or reduced vasoconstrictor tone. The resulting fall in systemic blood pressure can then reduce cerebral blood flow to below the compensatory limits of autoregulation ([Fig. 18-3](#)). In order to elicit neurally mediated syncope, a functioning autonomic nervous system is necessary, in contrast to syncope resulting from autonomic failure (discussed below).

Multiple triggers of the afferent limb of the reflex arc can result in neurally mediated syncope. In some situations, these can be clearly defined, e.g., the carotid sinus, the gastrointestinal tract, or the bladder. Often, however, the trigger is less easily recognized and the cause is multifactorial. Under these circumstances, it is likely that different afferent pathways converge on the central autonomic network within the medulla that integrates the neural impulses and mediates the vasodepressor-bradycardic response.

Classification of Neurally Mediated Syncope Neurally mediated syncope may be subdivided based on the afferent pathway

TABLE 18-1 High-Risk Features Indicating Hospitalization or Intensive Evaluation of Syncope

Chest pain suggesting coronary ischemia
Features of congestive heart failure
Moderate or severe valvular disease
Moderate or severe structural cardiac disease
Electrocardiographic features of ischemia
History of ventricular arrhythmias
Prolonged QT interval (>500 ms)
Repetitive sinoatrial block or sinus pauses
Persistent sinus bradycardia
Bi- or trifascicular block or intraventricular conduction delay with QRS duration ≥120 ms
Atrial fibrillation
Nonsustained ventricular tachycardia
Family history of sudden death
Preexcitation syndromes
Brugada pattern on ECG
Palpitations at time of syncope
Syncope at rest or during exercise

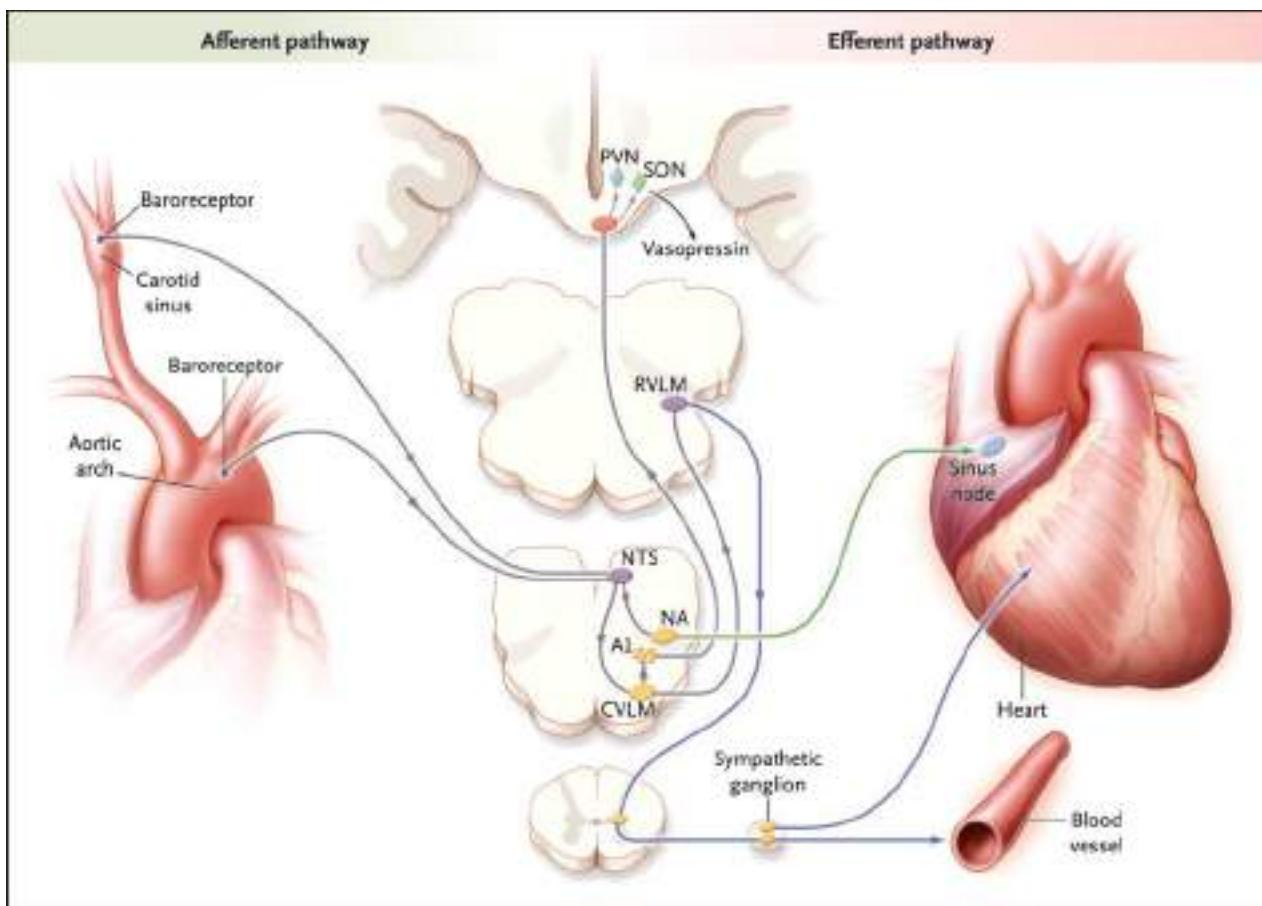


FIGURE 18-1 The baroreflex. A decrease in arterial pressure unloads the baroreceptors—the terminals of afferent fibers of the glossopharyngeal and vagus nerves—that are situated in the carotid sinus and aortic arch. This leads to a reduction in the afferent impulses that are relayed from these mechanoreceptors through the glossopharyngeal and vagus nerves to the nucleus of the tractus solitarius (NTS) in the dorsomedial medulla. The reduced baroreceptor afferent activity produces a decrease in vagal nerve input to the sinus node that is mediated via connections of the NTS to the nucleus ambiguus (NA). There is an increase in sympathetic efferent activity that is mediated by the NTS projections to the caudal ventrolateral medulla (CVLM) (an excitatory pathway) and from there to the rostral ventrolateral medulla (RVLM) (an inhibitory pathway). The activation of RVLM presynaptic neurons in response to hypotension is thus predominantly due to disinhibition. In response to a sustained fall in blood pressure, vasopressin release is mediated by projections from the A1 noradrenergic cell group in the ventrolateral medulla. This projection activates vasopressin-synthesizing neurons in the magnocellular portion of the paraventricular nucleus (PVN) and the supraoptic nucleus (SON) of the hypothalamus. Blue denotes sympathetic neurons, and green denotes parasympathetic neurons. (From R Freeman: *N Engl J Med* 358:615, 2008.)

and provocative trigger. Vasovagal syncope (the common faint) is provoked by intense emotion, pain, and/or orthostatic stress, whereas the situational reflex syncopes have specific localized stimuli that provoke the reflex vasodilation and bradycardia that leads to syncope. The underlying mechanisms have been identified and pathophysiology delineated for most of these situational reflex syncopes. The afferent trigger may originate in the pulmonary system, gastrointestinal system, urogenital system, heart, and carotid artery (Table 18-2). Hyper-ventilation leading to hypocapnia and cerebral vasoconstriction, and raised intrathoracic pressure that impairs venous return to the heart, play a central role in many of the situational reflex syncopes. The afferent pathway of the reflex arc differs among these disorders, but the efferent response via the vagus and sympathetic pathways is similar.

Alternately, neurally mediated syncope may be subdivided based on the predominant efferent pathway. Vasodepressor syncope describes syncope predominantly due to efferent, sympathetic, vasoconstrictor failure; cardioinhibitory syncope describes syncope predominantly associated with bradycardia or asystole due to increased vagal outflow; and mixed syncope describes syncope in which there are both vagal and sympathetic reflex changes.

Features of Neurally Mediated Syncope In addition to symptoms of orthostatic intolerance such as dizziness, lightheadedness, and fatigue, premonitory features of autonomic activation may be present in patients with neurally mediated syncope. These include diaphoresis, pallor, palpitations, nausea, hyperventilation, and yawning. During the syncopal event, proximal and distal myoclonus (typically arrhythmic

and multifocal) may occur, raising the possibility of epilepsy. The eyes typically remain open and usually deviate upward. Pupils are usually dilated. Roving eye movements may occur. Grunting, moaning, snorting, and stertorous breathing may be present. Urinary incontinence may occur. Fecal incontinence is very rare. Postictal confusion is also rare, although visual and auditory hallucinations and near death and out-of-body experiences are sometimes reported.

Although some predisposing factors and provocative stimuli are well established (for example, motionless upright posture, warm ambient temperature, intravascular volume depletion, alcohol ingestion, hypoxemia, anemia, pain, the sight of blood, venipuncture, and intense emotion), the underlying basis for the widely different thresholds for syncope among individuals exposed to the same provocative stimulus is not known. A genetic basis for neurally mediated syncope may exist; several studies have reported an increased incidence of syncope in first-degree relatives of fainters, but no gene or genetic marker has been identified, and environmental, social, and cultural factors have not been excluded by these studies.

TREATMENT

Neurally Mediated Syncope

Reassurance, avoidance of provocative stimuli, and plasma volume expansion with fluid and salt are the cornerstones of the management of neurally mediated syncope. Isometric counterpressure maneuvers of the limbs (leg crossing or handgrip and arm tensing)

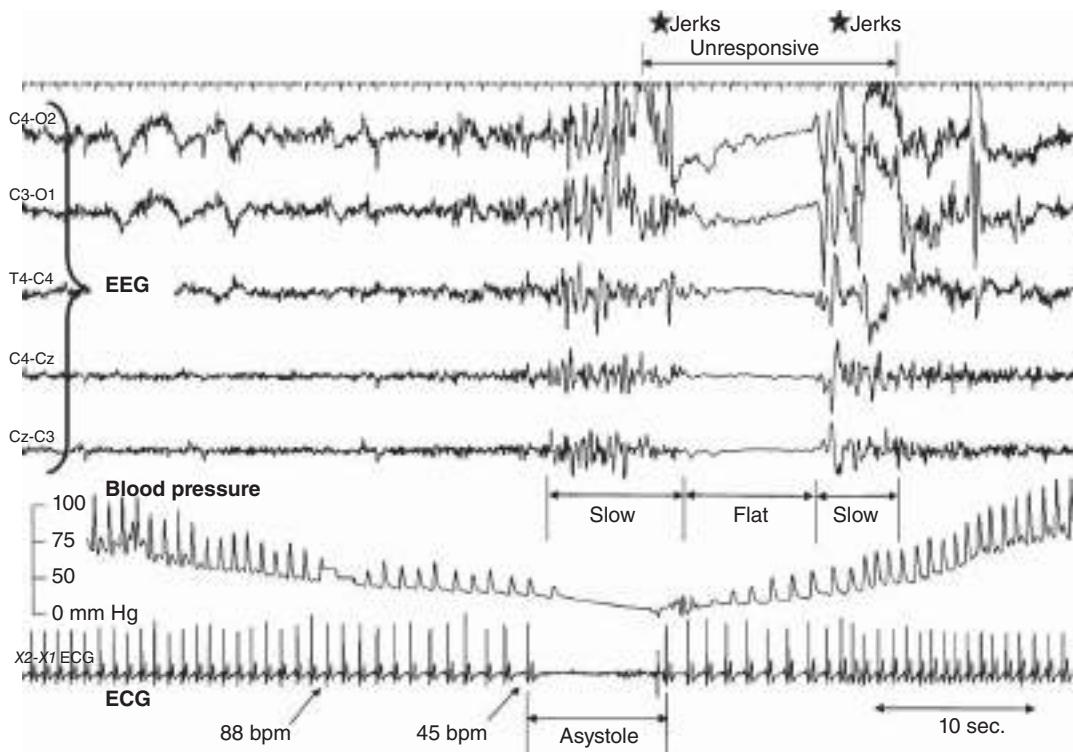


FIGURE 18-2 The electroencephalogram (EEG) in vasovagal syncope. A 1-min segment of a tilt-table test with typical vasovagal syncope demonstrating the “slow-flat-slow” EEG pattern. Finger beat-to-beat blood pressure, electrocardiogram (ECG), and selected EEG channels are shown. EEG slowing starts when systolic blood pressure drops to ~50 mmHg; heart rate is then ~45 beats/min (bpm). Asystole occurred, lasting about 8 s. The EEG flattens for a similar period, but with a delay. A transient loss of consciousness, lasting 14 s, was observed. There were muscle jerks just before and just after the flat period of the EEG. (Figure reproduced with permission from W Wieling et al: *Brain* 132:2630, 2009.)

may raise blood pressure by increasing central blood volume and cardiac output. By maintaining pressure in the autoregulatory zone, these maneuvers avoid or delay the onset of syncope. Randomized controlled trials support this intervention.

Fludrocortisone, vasoconstricting agents, and β -adrenoreceptor antagonists are widely used by experts to treat refractory patients, although there is no consistent evidence from randomized controlled trials for any pharmacotherapy to treat neurally mediated

syncope. Because vasodilation is the dominant pathophysiologic syncopal mechanism in most patients, use of a cardiac pacemaker is rarely beneficial. Possible exceptions are older patients (>40 years) in whom syncope is associated with asystole or severe bradycardia and patients with prominent cardioinhibition due to carotid sinus syndrome. In these patients, dual-chamber pacing may be helpful although this continues to be an area of uncertainty.

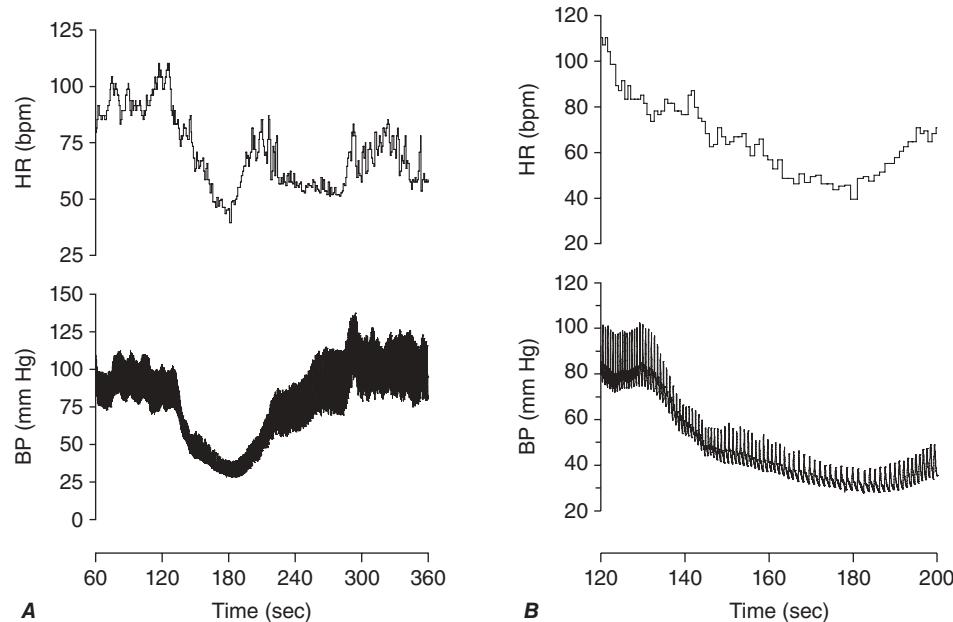


FIGURE 18-3 **A.** The paroxysmal hypotensive-bradycardic response that is characteristic of neurally mediated syncope. Noninvasive beat-to-beat blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. **B.** The same tracing expanded to show 80 s of the episode (from 80 to 200 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

TABLE 18-2 Causes of Syncope**A. Neurally Mediated Syncope**

Vasovagal syncope
Provoked fear, pain, anxiety, intense emotion, sight of blood, unpleasant sights and odors, orthostatic stress
Situational reflex syncope
Pulmonary
Cough syncope, wind instrument player's syncope, weightlifter's syncope, "mess trick" ^a and "fainting lark," ^b sneeze syncope, airway instrumentation
Urogenital
Postmicturition syncope, urogenital tract instrumentation, prostatic massage
Gastrointestinal
Swallow syncope, glossopharyngeal neuralgia, esophageal stimulation, gastrointestinal tract instrumentation, rectal examination, defecation syncope
Cardiac
Bezold-Jarisch reflex, cardiac outflow obstruction
Carotid sinus
Carotid sinus sensitivity, carotid sinus massage
Ocular
Ocular pressure, ocular examination, ocular surgery

B. Orthostatic Hypotension

Primary autonomic failure due to idiopathic central and peripheral neurodegenerative diseases—the "synucleinopathies"
Lewy body diseases
Parkinson's disease
Lewy body dementia
Pure autonomic failure
Multiple system atrophy (Shy-Drager syndrome)
Secondary autonomic failure due to autonomic peripheral neuropathies
Diabetes
Hereditary amyloidosis (familial amyloid polyneuropathy)
Primary amyloidosis (AL amyloidosis; immunoglobulin light chain associated)
Hereditary sensory and autonomic neuropathies (HSAN) (especially type III—familial dysautonomia)
Idiopathic immune-mediated autonomic neuropathy
Autoimmune autonomic ganglionopathy
Sjögren's syndrome
Paraneoplastic autonomic neuropathy
HIV neuropathy
Postprandial hypotension
Iatrogenic (drug-induced)
Volume depletion

C. Cardiac Syncope

Arrhythmias
Sinus node dysfunction
Atrioventricular dysfunction
Supraventricular tachycardias
Ventricular tachycardias
Inherited channelopathies
Cardiac structural disease
Valvular disease
Myocardial ischemia
Obstructive and other cardiomyopathies
Atrial myxoma
Pericardial effusions and tamponade

^aHyperventilation for ~1 min, followed by sudden chest compression. ^bHyperventilation (~20 breaths) in a squatting position, rapid rise to standing, then Valsalva.

■ ORTHOSTATIC HYPOTENSION

Orthostatic hypotension, defined as a reduction in systolic blood pressure of at least 20 mmHg or diastolic blood pressure of at least 10 mmHg within 3 min of standing or head-up tilt on a tilt table, is a manifestation of sympathetic vasoconstrictor (autonomic) failure (Fig. 18-4). In many (but not all) cases, there is no compensatory increase in heart rate despite hypotension; with partial autonomic failure, heart rate may increase to some degree but is insufficient to maintain cardiac output. A variant of orthostatic hypotension is "delayed" orthostatic hypotension, which occurs beyond 3 min of standing; this may reflect a mild or early form of sympathetic adrenergic dysfunction. In some cases, orthostatic hypotension occurs within 15 s of standing (so-called "initial" orthostatic hypotension), a finding that may reflect a transient mismatch between cardiac output and peripheral vascular resistance and does not represent autonomic failure.

Characteristic symptoms of orthostatic hypotension include light-headedness, dizziness, and presyncope (near-faintness) occurring in response to sudden postural change. However, symptoms may be absent or nonspecific, such as generalized weakness, fatigue, cognitive slowing, leg buckling, or headache. Visual blurring may occur, likely due to retinal or occipital lobe ischemia. Neck pain, typically in the suboccipital, posterior cervical, and shoulder region (the "coat-hanger headache"), most likely due to neck muscle ischemia, may be the only symptom. Patients may report orthostatic dyspnea (thought to reflect ventilation-perfusion mismatch due to inadequate perfusion of ventilated lung apices) or angina (attributed to impaired myocardial perfusion even with normal coronary arteries). Symptoms may be exacerbated by exertion, prolonged standing, increased ambient temperature, or meals. Syncope is usually preceded by warning symptoms, but may occur suddenly, suggesting the possibility of a seizure or cardiac cause.

Supine hypertension is common in patients with orthostatic hypotension due to autonomic failure, affecting >50% of patients in some series. Orthostatic hypotension may present after initiation of therapy for hypertension, and supine hypertension may follow treatment of orthostatic hypotension. However, in other cases, the association of the two conditions is unrelated to therapy; it may in part be explained by baroreflex dysfunction in the presence of residual sympathetic outflow, particularly in patients with central autonomic degeneration.

Causes of Neurogenic Orthostatic Hypotension Causes of neurogenic orthostatic hypotension include central and peripheral autonomic nervous system dysfunction (Chap. 432). Autonomic dysfunction of other organ systems (including the bladder, bowels, sexual organs, and sudomotor system) of varying severity frequently accompanies orthostatic hypotension in these disorders (Table 18-2).

The primary autonomic degenerative disorders are multiple system atrophy (Shy-Drager syndrome; Chap. 432), Parkinson's disease (Chap. 427), dementia with Lewy bodies (Chap. 426), and pure autonomic failure (Chap. 432). These are often grouped together as "synucleinopathies" due to the presence of α -synuclein, a small protein that aggregates predominantly in the cytoplasm of neurons in the Lewy body disorders (Parkinson's disease, dementia with Lewy bodies, and pure autonomic failure) and in the glia in multiple system atrophy.

Peripheral autonomic dysfunction may also accompany small-fiber peripheral neuropathies such as those seen in diabetes mellitus, amyloid, immune-mediated neuropathies, hereditary sensory and autonomic neuropathies (HSAN; particularly HSAN type III, familial dysautonomia) (Chaps. 438 and 439). Less frequently, orthostatic hypotension is associated with the peripheral neuropathies that accompany vitamin B₁₂ deficiency, neurotoxic exposure, HIV and other infections, and porphyria.

Patients with autonomic failure and the elderly are susceptible to falls in blood pressure associated with meals. The magnitude of the blood pressure fall is exacerbated by large meals, meals high in carbohydrate, and alcohol intake. The mechanism of postprandial syncope is not fully elucidated.

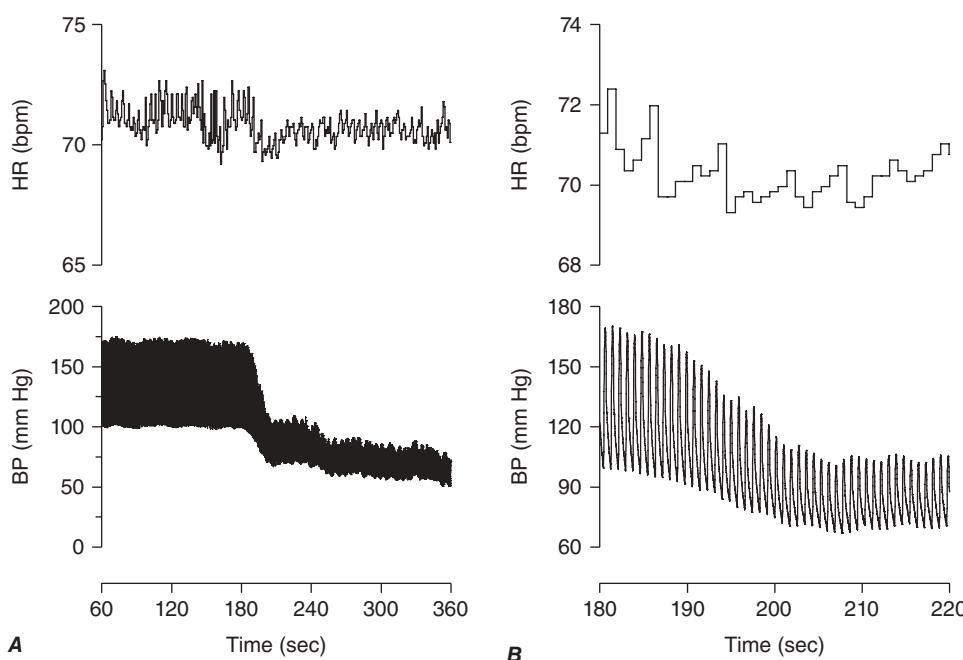


FIGURE 18-4 **A.** The gradual fall in blood pressure without a compensatory heart rate increase that is characteristic of orthostatic hypotension due to autonomic failure. Blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. **B.** The same tracing expanded to show 40 s of the episode (from 180 to 220 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

Orthostatic hypotension is often iatrogenic. Drugs from several classes may lower peripheral resistance (e.g., α -adrenoreceptor antagonists used to treat hypertension and prostatic hypertrophy; antihypertensive agents of several classes; nitrates and other vasodilators; tricyclic agents and phenothiazines). Iatrogenic volume depletion due to diuresis and volume depletion due to medical causes (hemorrhage, vomiting, diarrhea, or decreased fluid intake) may also result in decreased effective circulatory volume, orthostatic hypotension, and syncope.

TREATMENT

Orthostatic Hypotension

The first step is to remove reversible causes—usually vasoactive medications (Table 432-6). Next, nonpharmacologic interventions should be introduced. These interventions include patient education regarding staged moves from supine to upright; warnings about the hypotensive effects of large meals; instructions about the isometric counterpressure maneuvers that increase intravascular pressure (see above); and raising the head of the bed to reduce supine hypertension. Intravascular volume should be expanded by increasing dietary fluid and salt. If these nonpharmacologic measures fail, pharmacologic intervention with fludrocortisone acetate and vasoconstricting agents such as midodrine, L-dihydroxyphenylserine, and pseudoephedrine should be introduced. Some patients with intractable symptoms require additional therapy with supplementary agents that include pyridostigmine, atomoxetine, yohimbine, desmopressin acetate (DDAVP), and erythropoietin (Chap. 432).

CARDIAC SYNCOPES

Cardiac (or cardiovascular) syncope is caused by arrhythmias and structural heart disease. These may occur in combination because structural disease renders the heart more vulnerable to abnormal electrical activity.

Arrhythmias Bradyarrhythmias that cause syncope include those due to severe sinus node dysfunction (e.g., sinus arrest or sinoatrial block) and atrioventricular (AV) block (e.g., Mobitz type II, high-grade, and complete AV block). The bradyarrhythmias due to sinus node dysfunction are often associated with an atrial tachyarrhythmia, a disorder known as the tachycardia-bradycardia syndrome. A prolonged

pause following the termination of a tachycardic episode is a frequent cause of syncope in patients with the tachycardia-bradycardia syndrome. Medications of several classes may also cause bradyarrhythmias of sufficient severity to cause syncope. Syncope due to bradycardia or asystole is referred to as a Stokes-Adams attack.

Ventricular tachyarrhythmias frequently cause syncope. The likelihood of syncope with ventricular tachycardia is in part dependent on the ventricular rate; rates <200 beats/min are less likely to cause syncope. The compromised hemodynamic function during ventricular tachycardia is caused by ineffective ventricular contraction, reduced diastolic filling due to abbreviated filling periods, loss of AV synchrony, and concurrent myocardial ischemia.

Several disorders associated with cardiac electrophysiologic instability and arrhythmogenesis are due to mutations in ion channel subunit genes. These include the long QT syndrome, Brugada syndrome, and catecholaminergic polymorphic ventricular tachycardia. The long QT syndrome is a genetically heterogeneous disorder associated with prolonged cardiac repolarization and a predisposition to ventricular arrhythmias. Syncope and sudden death in patients with long QT syndrome result from a unique polymorphic ventricular tachycardia called *torsades des pointes* that degenerates into ventricular fibrillation. The long QT syndrome has been linked to genes encoding K⁺ channel α -subunits, K⁺ channel β -subunits, voltage-gated Na⁺ channel, and a scaffolding protein, ankyrin B (ANK2).

Brugada syndrome is characterized by idiopathic ventricular fibrillation in association with right ventricular electrocardiogram (ECG) abnormalities without structural heart disease. This disorder is also genetically heterogeneous, although it is most frequently linked to mutations in the Na⁺ channel α -subunit, SCN5A. Catecholaminergic polymorphic tachycardia is an inherited, genetically heterogeneous disorder associated with exercise- or stress-induced ventricular arrhythmias, syncope, or sudden death. Acquired QT interval prolongation, most commonly due to drugs, may also result in ventricular arrhythmias and syncope. **These disorders are discussed in detail in Chap. 249.**

Structural Disease Structural heart disease (e.g., valvular disease, myocardial ischemia, hypertrophic and other cardiomyopathies, cardiac masses such as atrial myxoma, and pericardial effusions) may lead to syncope by compromising cardiac output. Structural disease may also contribute to other pathophysiologic mechanisms of syncope. For example, cardiac structural disease may predispose to arrhythmogenesis; aggressive treatment of cardiac failure with diuretics and/or vasodilators may lead to orthostatic hypotension; and inappropriate reflex vasodilation may occur with structural disorders such as aortic stenosis and hypertrophic cardiomyopathy, possibly provoked by increased ventricular contractility.

TREATMENT

Cardiac Syncope

Treatment of cardiac disease depends on the underlying disorder. Therapies for arrhythmias include cardiac pacing for sinus node disease and AV block, and ablation, antiarrhythmic drugs, and cardioverter-defibrillators for atrial and ventricular tachyarrhythmias. These disorders are best managed by physicians with specialized skills in this area.

APPROACH TO THE PATIENT

Syncope

DIFFERENTIAL DIAGNOSIS

Syncope is easily diagnosed when the characteristic features are present; however, several disorders with transient real or apparent loss of consciousness may create diagnostic confusion.

Generalized and partial seizures may be confused with syncope; however, there are a number of differentiating features. Whereas tonic-clonic movements are the hallmark of a generalized seizure, myoclonic and other movements also may occur in up to 90% of syncopal episodes. Myoclonic jerks associated with syncope may be multifocal or generalized. They are typically arrhythmic and of short duration (<30 s). Mild flexor and extensor posturing also may occur. Partial or partial-complex seizures with secondary generalization are usually preceded by an aura, commonly an unpleasant smell; fear; anxiety; abdominal discomfort; or other visceral sensations. These phenomena should be differentiated from the premonitory features of syncope.

Autonomic manifestations of seizures (autonomic epilepsy) may provide a more difficult diagnostic challenge. Autonomic seizures have cardiovascular, gastrointestinal, pulmonary, urogenital, pupillary, and cutaneous manifestations that are similar to the premonitory features of syncope. Furthermore, the cardiovascular manifestations of autonomic epilepsy include clinically significant tachycardias and bradycardias that may be of sufficient magnitude to cause loss of consciousness. The presence of accompanying nonautonomic auras may help differentiate these episodes from syncope.

Loss of consciousness associated with a seizure usually lasts >5 min and is associated with prolonged postictal drowsiness and disorientation, whereas reorientation occurs almost immediately after a syncopal event. Muscle aches may occur after both syncope and seizures, although they tend to last longer and be more severe following a seizure. Seizures, unlike syncope, are rarely provoked by emotions or pain. Incontinence of urine may occur with both seizures and syncope; however, fecal incontinence occurs very rarely with syncope.

Hypoglycemia may cause transient loss of consciousness, typically in individuals with type 1 or type 2 diabetes treated with insulin. The clinical features associated with impending or actual hypoglycemia include tremor, palpitations, anxiety, diaphoresis, hunger, and paresthesias. These symptoms are due to autonomic activation to counter the falling blood glucose. Hunger, in particular, is not a typical premonitory feature of syncope. Hypoglycemia also impairs neuronal function, leading to fatigue, weakness, dizziness, and cognitive and behavioral symptoms. Diagnostic difficulties may occur in individuals in strict glycemic control; repeated hypoglycemia impairs the counterregulatory response and leads to a loss of the characteristic warning symptoms that are the hallmark of hypoglycemia.

Patients with cataplexy experience an abrupt partial or complete loss of muscular tone triggered by strong emotions, typically anger or laughter. Unlike syncope, consciousness is maintained throughout the attacks, which typically last between 30 s and 2 min. There are no premonitory symptoms. Cataplexy occurs in 60–75% of patients with narcolepsy.

The clinical interview and interrogation of eyewitnesses usually allow differentiation of syncope from falls due to vestibular dysfunction, cerebellar disease, extrapyramidal system dysfunction, and other gait disorders. A diagnosis of syncope can be particularly challenging in patients with dementia who experience repeated falls and are unable to provide a clear history of the episodes. If the fall is accompanied by head trauma, a postconcussive syndrome, amnesia for the precipitating events, and/or the presence of loss of consciousness may also contribute to diagnostic difficulty.

Apparent loss of consciousness can be a manifestation of psychiatric disorders such as generalized anxiety, panic disorders, major depression, and somatization disorder. These possibilities should be

considered in individuals who faint frequently without prodromal symptoms. Such patients are rarely injured despite numerous falls. There are no clinically significant hemodynamic changes concurrent with these episodes. In contrast, transient loss of consciousness due to vasovagal syncope precipitated by fear, stress, anxiety, and emotional distress is accompanied by hypotension, bradycardia, or both.

INITIAL EVALUATION

The goals of the initial evaluation are to determine whether the transient loss of consciousness was due to syncope; to identify the cause; and to assess risk for future episodes and serious harm (Table 18-1). The initial evaluation should include a detailed history, thorough questioning of eyewitnesses, and a complete physical and neurologic examination. Blood pressure and heart rate should be measured in the supine position and after 3 min of standing to determine whether orthostatic hypotension is present. An ECG should be performed if there is suspicion of syncope due to an arrhythmia or underlying cardiac disease. Relevant electrocardiographic abnormalities include bradyarrhythmias or tachyarrhythmias, AV block, ischemia, old myocardial infarction, long QT syndrome, and bundle branch block. This initial assessment will lead to the identification of a cause of syncope in ~50% of patients and also allows stratification of patients at risk for cardiac mortality.

Laboratory Tests Baseline laboratory blood tests are rarely helpful in identifying the cause of syncope. Blood tests should be performed when specific disorders, e.g., myocardial infarction, anemia, and secondary autonomic failure, are suspected (Table 18-2).

Autonomic Nervous System Testing (Chap. 432) Autonomic testing, including tilt-table testing, can be performed in specialized centers. Autonomic testing is helpful to uncover objective evidence of autonomic failure and also to demonstrate a predisposition to neurally mediated syncope. Autonomic testing includes assessments of parasympathetic autonomic nervous system function (e.g., heart rate variability to deep respiration and a Valsalva maneuver), sympathetic cholinergic function (e.g., thermoregulatory sweat response and quantitative sudomotor axon reflex test), and sympathetic adrenergic function (e.g., blood pressure response to a Valsalva maneuver and a tilt-table test with beat-to-beat blood pressure measurement). The hemodynamic abnormalities demonstrated on the tilt-table test (Figs. 18-3 and 18-4) may be useful in distinguishing orthostatic hypotension due to autonomic failure from the hypotensive bradycardic response of neurally mediated syncope. Similarly, the tilt-table test may help identify patients with syncope due to immediate or delayed orthostatic hypotension.

Carotid sinus massage should be considered in patients with symptoms suggestive of carotid sinus syncope and in patients >50 years with recurrent syncope of unknown etiology. This test should only be carried out under continuous ECG and blood pressure monitoring and should be avoided in patients with carotid bruits, plaques, or stenosis.

Cardiac Evaluation ECG monitoring is indicated for patients with a high pretest probability of arrhythmia causing syncope. Patients should be monitored in hospital if the likelihood of a life-threatening arrhythmia is high, e.g., patients with severe structural or coronary artery disease, nonsustained ventricular tachycardia, trifascicular heart block, prolonged QT interval, Brugada syndrome ECG pattern, or family history of sudden cardiac death (Table 18-1). Outpatient Holter monitoring is recommended for patients who experience frequent syncopal episodes (one or more per week), whereas loop recorders, which continually record and erase cardiac rhythm, are indicated for patients with suspected arrhythmias with low risk of sudden cardiac death. Loop recorders may be external (recommended for evaluation of episodes that occur at a frequency of >1 per month) or implantable (if syncope occurs less frequently).

Echocardiography should be performed in patients with a history of cardiac disease or if abnormalities are found on physical

examination or the ECG. Echocardiographic diagnoses that may be responsible for syncope include aortic stenosis, hypertrophic cardiomyopathy, cardiac tumors, aortic dissection, and pericardial tamponade. Echocardiography also has a role in risk stratification based on the left ventricular ejection fraction.

Treadmill exercise testing with ECG and blood pressure monitoring should be performed in patients who have experienced syncope during or shortly after exercise. Treadmill testing may help identify exercise-induced arrhythmias (e.g., tachycardia-related AV block) and exercise-induced exaggerated vasodilation.

Electrophysiologic studies are indicated in patients with structural heart disease and ECG abnormalities in whom noninvasive investigations have failed to yield a diagnosis. Electrophysiologic studies have low sensitivity and specificity and should only be performed when a high pretest probability exists. Currently, this test is rarely performed to evaluate patients with syncope.

Psychiatric Evaluation Screening for psychiatric disorders may be appropriate in patients with recurrent unexplained syncope episodes. Tilt-table testing, with demonstration of symptoms in the absence of hemodynamic change, may be useful in reproducing syncope in patients with suspected psychogenic syncope.

FURTHER READING

AL-KHATIB SM et al: Risk stratification for arrhythmic events in patients with asymptomatic pre-excitation: A systematic review for the 2015 ACC/AHA/HRS guideline for the management of adult patients with supraventricular tachycardia: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 67:1624, 2016.

FREEMAN R et al: Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome. *Auton Neurosci* 161:46, 2011.

GIBBONS CH et al: The recommendations of a consensus panel for the screening, diagnosis, and treatment of neurogenic orthostatic hypotension and associated supine hypertension. *J Neurol* 264:1567, 2017.

SHELDON RS, RAJ SR: Pacing and vasovagal syncope: Back to our physiologic roots. *Clin Auton Res* 27:213, 2017.

VAROSY PD et al: Pacing as a treatment for reflex-mediated (vasovagal, situational, or carotid sinus hypersensitivity) syncope: A systematic review for the 2017 ACC/AHA/HRS guideline for the evaluation and management of patients with syncope: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 70:664, 2017.

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. 18, 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 non-vestibular imbalance and 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 asymmetry of vestibular inputs from the two labyrinths or in their central pathway that is usually acute. 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 body position. 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 a number of 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 aural symptoms (ear pain, pressure, fullness) 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. Particular focus should be given to assessment of eye movements, vestibular function, and hearing.

19

Dizziness and Vertigo

Mark F. Walker, Robert B. Daroff



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.

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) can aid in the detection of peripheral vestibular nystagmus, because they reduce the patient's ability to use visual fixation to suppress nystagmus. **Table 19-1** 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 vestibuloocular reflex (VOR) is assessed with small-amplitude (~20 degrees) rapid head rotations. While the patient fixates on a target, the head is rotated 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 vestibular hypofunction (catch-up saccades after rotations in both directions).

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 upbeat-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

TABLE 19-1 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.

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. Electronystagmography or videonystagmography includes recordings of spontaneous nystagmus (if present) and measurement of positional nystagmus. Caloric testing assesses the responses of the two horizontal semicircular canals. The test battery often includes recording of saccades and pursuit to assess 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 magnetic resonance imaging (MRI) of the internal auditory canals, including administration of gadolinium, to rule out a schwannoma.

■ 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 19-1). 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, but glucocorticoids can improve outcome if administered within 3 days of symptom onset. 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, rolling over in bed, rising from a supine position, and extending the head to look upward. 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

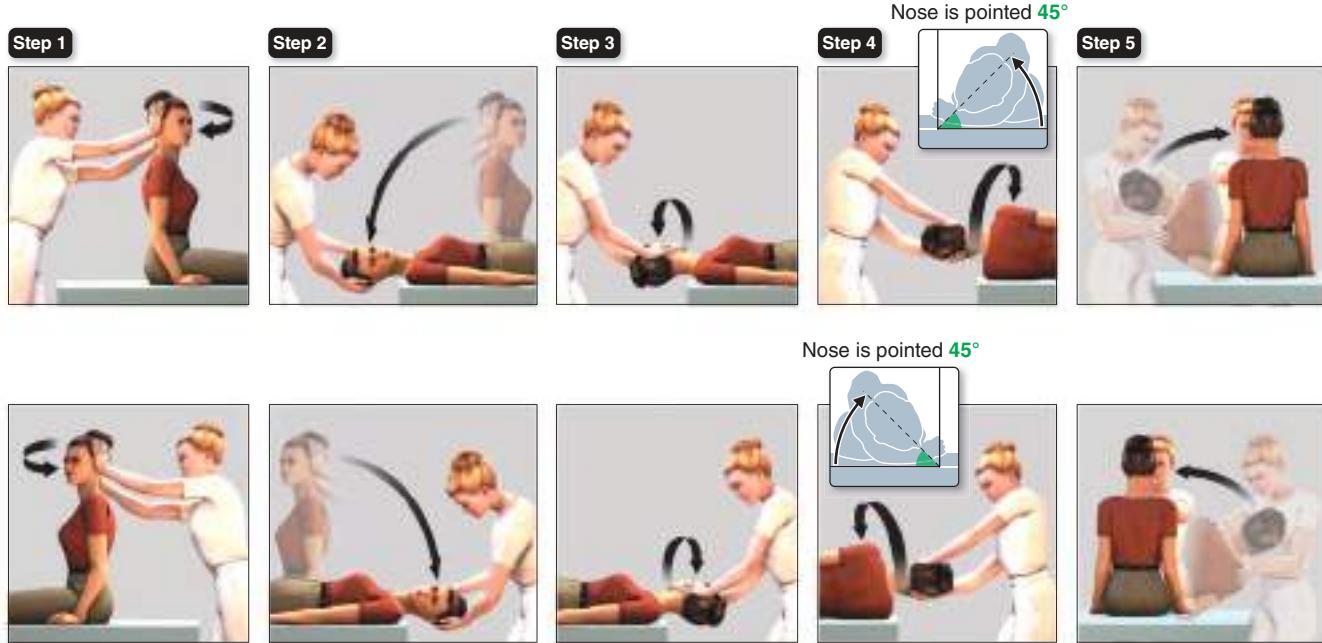


FIGURE 19-1 Modified Epley maneuver for treatment of benign paroxysmal positional vertigo of the right (top panels) and left (bottom panels) posterior semicircular canals. **Step 1.** With the patient seated, turn the head 45 degrees toward the affected ear. **Step 2.** Keeping the head turned, lower 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>.)

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. 19-1) 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>).

■ VESTIBULAR MIGRAINE

Vestibular migraine is a very 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. 422). 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. The 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 thought to be due to excess fluid (endolymph) in the inner ear; hence the term *endolymphatic hydrops*. 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. Non-ablative 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 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. 419–421), demyelination (Chap. 436), or tumors (Chap. 86), or they may be due to neurodegenerative conditions that include the vestibulocerebellum (Chaps. 423–426). Subacute cerebellar degeneration may be due to immune, including paraneoplastic, processes (Chaps. 90 and 431). Table 19-1 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. 443). 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*, *psychophysiological 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 neurootologic 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 19-2 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.

TABLE 19-2 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	
Methylprednisolone ^g	100 mg daily days 1–3; 80 mg daily days 4–6; 60 mg daily days 7–9; 40 mg daily days 10–12; 20 mg daily days 13–15; 10 mg daily days 16–18, 20, 22
Selective serotonin reuptake inhibitors ^h	

^aAll listed drugs are approved by the U.S. Food and Drug Administration, but most are not approved for the treatment of vertigo. ^bUsual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase. ^cFor motion sickness only. ^dFor benign paroxysmal positional vertigo. ^eFor Ménière's disease. ^fFor vestibular migraine. ^gFor acute vestibular neuritis (started within 3 days of onset). ^hFor persistent postural-perceptual vertigo and anxiety.

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.

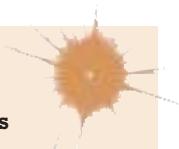
FURTHER READING

- DIETERICH M, STAAB JP: Functional dizziness: From phobic postural vertigo and chronic subjective dizziness to persistent postural-perceptual dizziness. *Curr Opin Neurol* 30:107, 2017.
 KIM JS, ZEE DS: Benign paroxysmal positional vertigo. *N Engl J Med* 370:1138, 2014.
 VON BREVERN M, LEMPERT T: Vestibular Migraine. *Handb Clin Neurol* 137:301, 2016.

20

Fatigue

Jeffrey M. Gelfand, Vanja C. Douglas



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 human experience of physical and mental weariness, sluggishness, low energy, and exhaustion. In the context of clinical medicine, fatigue is most typically and 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. 21); 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. 27), and from *dyspnea on exertion*, although patients may use the word fatigue to describe those symptoms. The task facing clinicians when a patient presents with fatigue is to identify the underlying cause and to 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, as defined by the U.S. Centers for Disease Control and Prevention, is low (Chap. 442).

■ DIFFERENTIAL DIAGNOSIS

Psychiatric Disease Fatigue is a common somatic manifestation of many major psychiatric syndromes, including depression, anxiety, and somatoform disorders. Psychiatric symptoms are reported in more than three-quarters of patients with unexplained chronic fatigue. Even in patients with systemic or neurologic syndromes in which fatigue is independently recognized as a manifestation of disease, comorbid psychiatric symptoms or disease may still be an important source of interaction.

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. 21). **Fatigability** of muscle power is a cardinal manifestation of some neuromuscular disorders such as myasthenia gravis and is distinguished from *fatigue* by finding clinically apparent diminution of the amount of force that a muscle generates upon repeated contraction (Chap. 440). Fatigue is one of the most common and bothersome symptoms reported in multiple sclerosis (MS) (Chap. 436), 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, central dysautonomias, and amyotrophic lateral sclerosis. Fatigue after stroke is a well-described but poorly understood entity with a widely varying prevalence. Episodic fatigue can be a premonitory symptom of migraine. Fatigue is also a frequent result of traumatic brain injury, 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. 291). Whether the cumulative sleep deprivation that is common in modern society contributes to clinically apparent fatigue is not known (Chap. 27).

Endocrine Disorders Fatigue, sometimes in association with true muscle weakness, can be a heralding symptom of hypothyroidism,

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. Adrenal insufficiency 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. 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. 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.

Obesity Obesity 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, 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.

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 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.

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 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.

Systemic Inflammatory/Rheumatologic Disorders Fatigue is a prominent complaint in many chronic inflammatory disorders, including systemic lupus erythematosus, polymyalgia

rheumatics, rheumatoid arthritis, inflammatory bowel disease, anti-neutrophil 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 Chronic fatigue syndrome (**Chap. 442**) and fibromyalgia (**Chap. 366**) incorporate chronic fatigue as part of the syndromic definition when present in association with a number of other inclusion and exclusion 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. 56**). 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 chronic fatigue syndrome.

APPROACH TO THE PATIENT

Fatigue

A detailed history focusing on the quality, pattern, time-course, associated symptoms, and alleviating factors of fatigue is critical 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, 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 can 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 conduction 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 more likely to lead to false-positive results 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 of treatment 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. Antidepressant treatment (**Chap. 444**) may be helpful for treatment of chronic fatigue when symptoms of depression are present and may be 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 the context of chronic fatigue syndrome 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 1 α and 1 β , and

tumor necrosis factor α , might mediate fatigue in some patients. Preliminary data suggests that biological therapies that inhibit IL-1 or other cytokines can help to ameliorate fatigue in some patients with inflammatory conditions in addition to, or as part of, their disease modifying effect; thus, cytokine antagonists represent one possible future approach.

■ PROGNOSIS

Acute fatigue significant enough to require medical evaluation is more likely to lead to an identifiable medical, neurologic, or psychiatric cause than 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 on longitudinal follow-up in patients with unexplained chronic fatigue. Complete resolution of unexplained chronic fatigue is uncommon, at least over the short term, but multidisciplinary treatment approaches can lead to symptomatic improvements that can substantially improve quality of life.

■ FURTHER READING

- DAVID A et al: Tired, weak, or in need of rest: Fatigue among general practice attenders. *BMJ* 301:1199, 1990.
- KROENKE K et al: Chronic fatigue in primary care. Prevalence, patient characteristics, and outcome. *JAMA* 260:929–934, 1988.
- ROERINK ME et al: Interleukin-1 as a mediator of fatigue in disease: A narrative review. *J Neuroinflammation* 14:16, 2017.
- SHARPE M et al: Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. *Lancet Psychiatry* 2:1067, 2015.
- 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.

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. 26](#)).

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 21-1](#)).

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 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. 415](#)) 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. 440](#)).

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 twitch within a muscle due to the

21

Neurologic Causes of Weakness and Paralysis

Michael J. Aminoff



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. 431](#)) or abnormal movements ([Chap. 428](#)). *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

TABLE 21-1 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

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. 21-1) produce weakness through decreased activation of lower motor neurons.

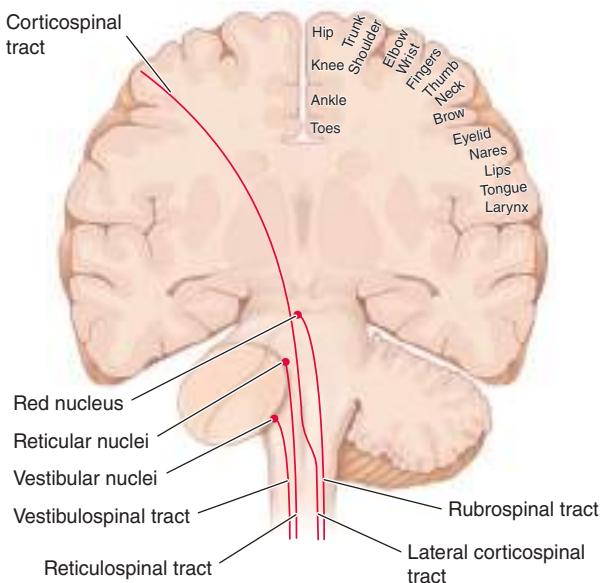


FIGURE 21-1 The corticospinal and bulbospinal upper motor neuron pathways. Upper motor neurons have their cell bodies in layer V of the primary motor cortex (the precentral gyrus, or Brodmann's area 4) and in the premotor and supplemental motor cortex (area 6). The upper motor neurons in the primary 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.

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. 438) shows that with weakness of the upper motor neuron type, motor units have a diminished maximal discharge frequency.

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. 21-2). 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.

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.

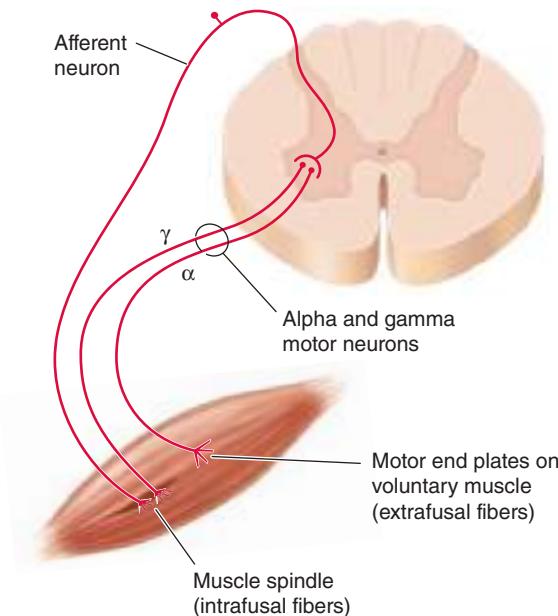
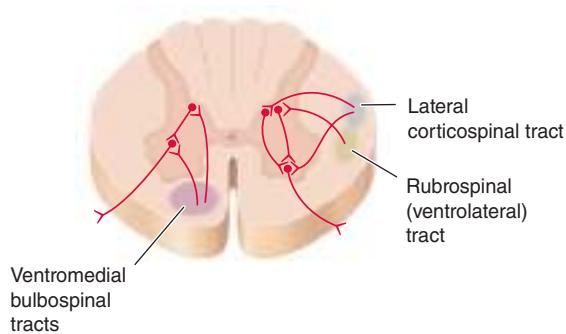


FIGURE 21-2 Lower motor neurons are divided into α and γ types. The larger α motor neurons are more numerous and innervate the extrafusal muscle fibers of the motor unit. Loss of α motor neurons or disruption of their axons produces lower motor neuron weakness. The smaller, less numerous γ motor neurons 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 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.

Neuromuscular Junction Weakness Disorders of the neuromuscular junctions 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 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. 419). 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 the Brown-Séquard syndrome (Chap. 434).

Acute or episodic hemiparesis usually results from focal structural lesions, particularly 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. 21-3) of acute origin starts with a computed tomography (CT) scan of the brain and laboratory studies. If the CT is normal, or in subacute or chronic cases of hemiparesis, magnetic resonance imaging (MRI) of the brain and/or cervical spine (including the foramen magnum) is performed, depending on the clinical accompaniments.

Paraparesis *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 suggestive of a central cord syndrome (Chap. 434), or hyperreflexia in the legs with normal reflexes in the arms. Imaging the spinal cord (Fig. 21-3) may reveal compressive lesions, infarction (proprioception usually is spared), arteriovenous fistulas or other vascular anomalies, or transverse myelitis (Chap. 434).

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 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. 439), or myopathy (Chap. 441).

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. 434). 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

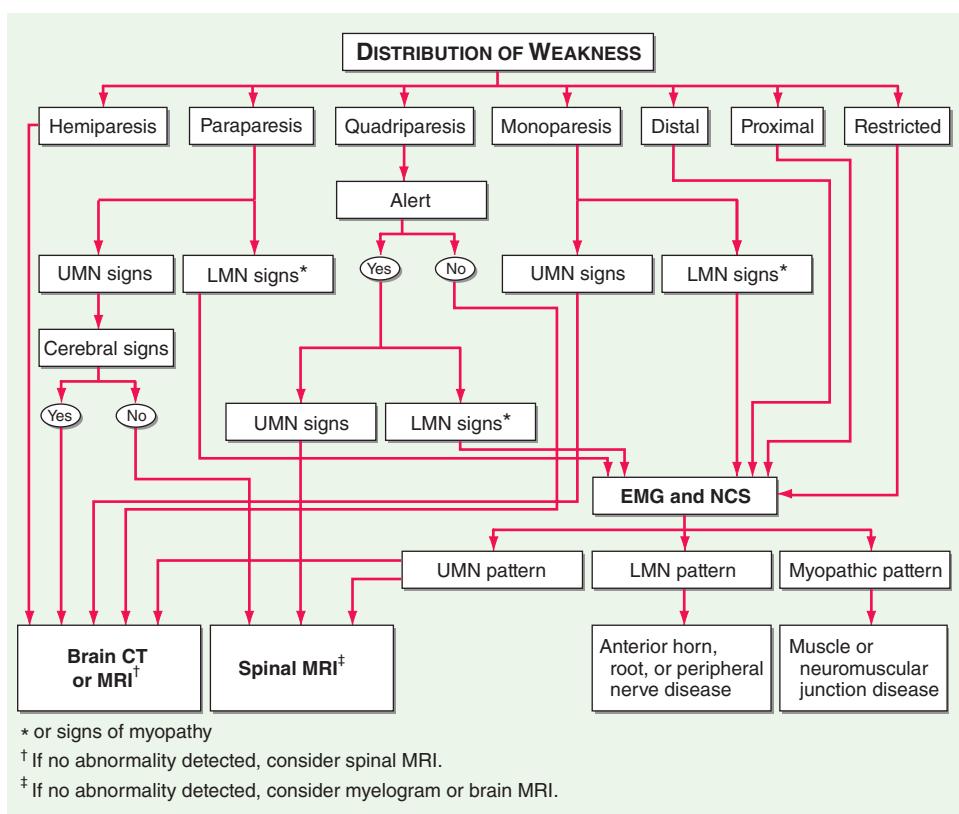


FIGURE 21-3 An algorithm for the initial workup of a patient with weakness. CT, computed tomography; EMG, electromyography; LMN, lower motor neuron; MRI, magnetic resonance imaging; NCS, nerve conduction studies; UMN, upper motor neuron.

TABLE 21-2 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

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.

Quadriplegia or Generalized Weakness Generalized weakness may be due to disorders of the CNS or the motor unit. Although the terms often are used interchangeably, *quadriplegia* 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 21-2**. A patient with generalized fatigability without objective weakness may have the chronic fatigue syndrome (**Chap. 442**).

ACUTE QUADRIPLAESIS Quadriplegia 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. 439**).

In obtunded patients, evaluation begins with a CT 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 QUADRIPLAESIS Quadriplegia 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 *quadriplegia* 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 to determine

the levels of muscle enzymes and electrolytes, and EMG and nerve conduction studies help to localize the pathologic process.

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 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. 420**); 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. 438**). Anterior horn cell disease may begin distally but is typically asymmetric and without accompanying numbness (**Chap. 429**). Rarely, myopathies present with distal weakness (**Chap. 441**). Electrodiagnostic studies help localize the disorder (Fig. 21-3).

Proximal Weakness Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles (**Chap. 441**). Diseases of the neuromuscular junction, such as myasthenia gravis (**Chap. 440**), may present with symmetric proximal weakness often associated with ptosis, diplopia, or bulbar weakness and fluctuating in severity during the day. In anterior horn cell disease, proximal weakness is usually asymmetric, but it may be symmetric if familial. 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. 441**) or neuromuscular junction disorder (**Chap. 440**). Bilateral facial palsy with areflexia suggests Guillain-Barré syndrome (**Chap. 439**). 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. 358**).

FURTHER READING

- BRAZIS P, MASDEU JC, BILLER J: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.
- CAMPBELL WW: *DeJong's The Neurological Examination*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2012.
- GUARANTORS OF BRAIN: *Aids to the Examination of the Peripheral Nervous System*, 4th ed. Edinburgh, Saunders, 2000.



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. 10.**

■ 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 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 *dysesthesia*s 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.

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). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities.

Afferent fibers in peripheral nerve trunks traverse the dorsal roots and enter the dorsal horn of the spinal cord (Fig. 22-1). From there, 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. 10). 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.

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.

■ 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 22-1). 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.

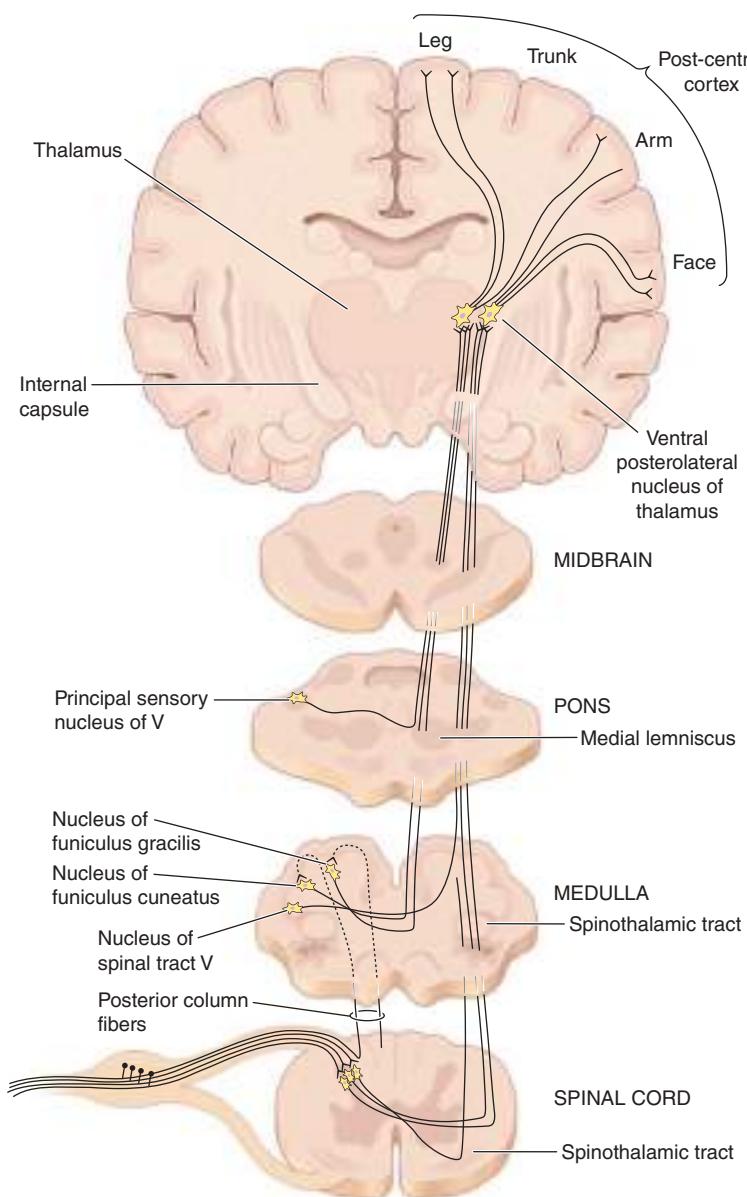


FIGURE 22-1 The main somatosensory pathways. The spinothalamic tract (pain, thermal sense) and the posterior column-lemniscal system (touch, pressure, joint position) are shown. Offshoots from the ascending anterolateral fasciculus (spinothalamic tract) to nuclei in the medulla, pons, and mesencephalon and nuclear terminations of the tract are indicated. (From AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)

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. 22-2 and 22-3).

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 whether the sensory symptoms are a disguised request for help with psychologic or situational problems. 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 or a fine camel hair brush, 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 indicate 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.

TABLE 22-1 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.

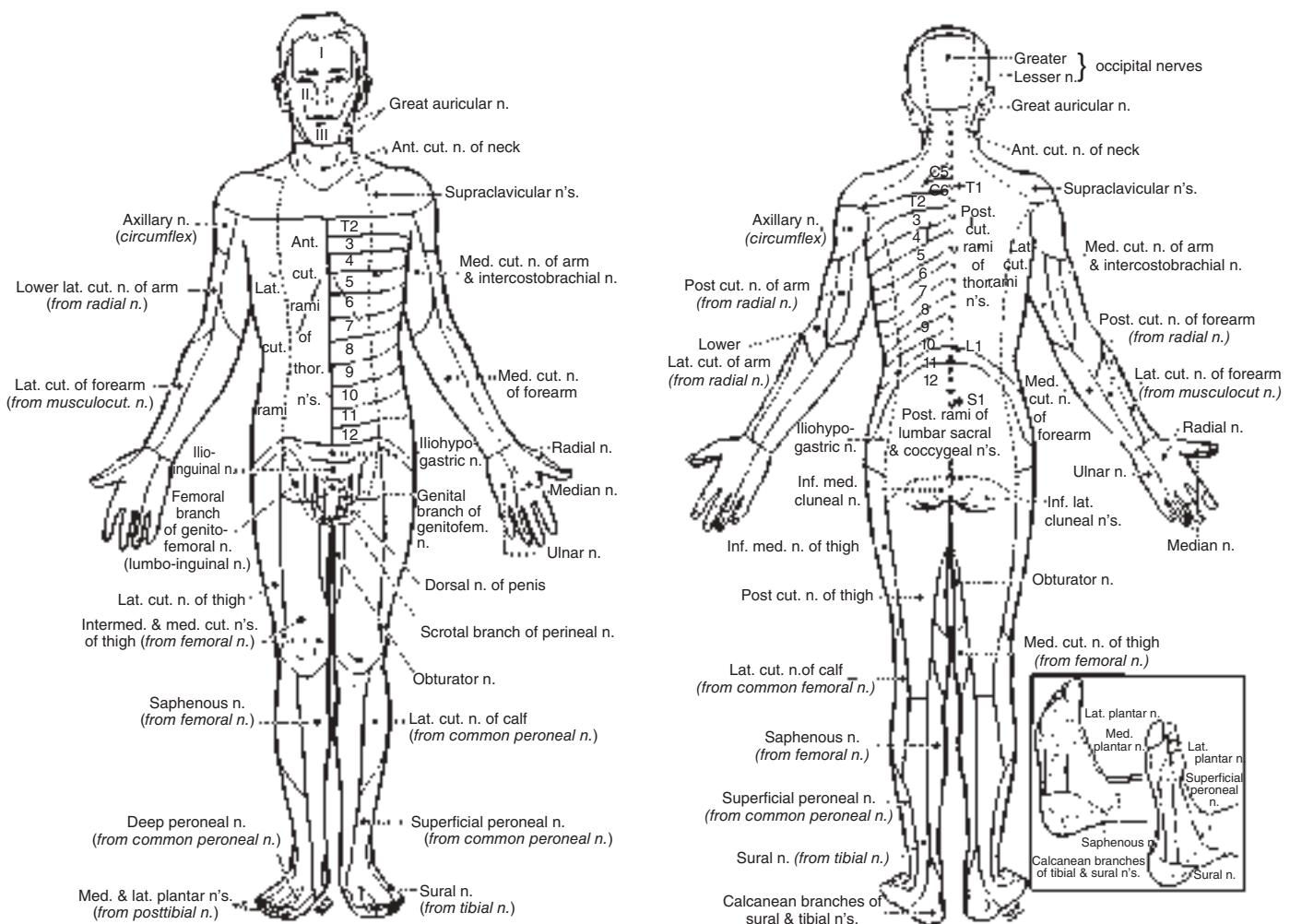


FIGURE 22-2 The cutaneous fields of peripheral nerves. (Reproduced by permission from W Haymaker, B Woodhall: *Peripheral Nerve Injuries*, 2nd ed. Philadelphia, Saunders, 1953.)

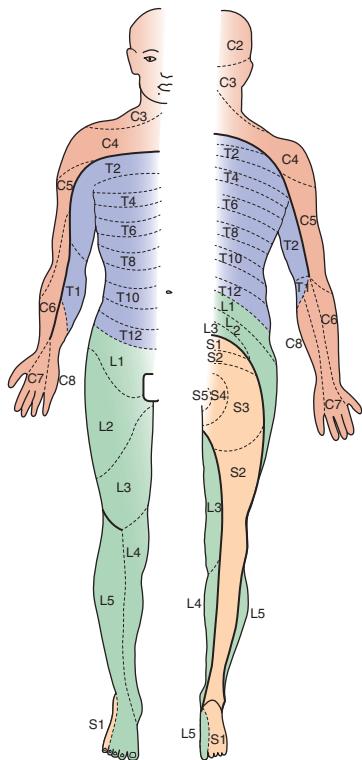


FIGURE 22-3 Distribution of the sensory spinal roots on the surface of the body (dermatomes). (From D Sinclair: *Mechanisms of Cutaneous Sensation*. Oxford, UK, Oxford University Press, 1981; with permission from Dr. David Sinclair.)

Quantitative Sensory Testing 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.

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 is tested with special 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.

■ 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 B₁₂ 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. 22-2 and 22-3). 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. 14). 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 appropriate 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. 438). 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. 21). Although most polyneuropathies are pансensory 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. 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 with time. This condition is usually paraneoplastic or idiopathic in origin (Chaps. 90 and 438) or related to an autoimmune disease, particularly Sjögren's syndrome.

Spinal Cord (See also Chap. 434) 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. 22-1 and 434-1); 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. 21) 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 recent 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. 419-7). A lesion in the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, causes pансensory 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.

Cortex With lesions of the parietal lobe involving either the cortex or the 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.

FURTHER READING

BRAZIS P, MASDEU JC, BILLER J: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.

CAMPBELL WW: *DeJong's The Neurological Examination*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2012.

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 neurological and non-neurological causes, though significant overlap often exists. The *antalginic 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 non-neurological 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 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 23-1).

23

Gait Disorders, Imbalance, and Falls

Jessica M. Baker, Lewis R. Sudarsky

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

TABLE 23-1 Etiology of Gait Disorders

ETIOLOGY	NO. OF CASES	PERCENT
Sensory deficits	22	18.3
Myelopathy	20	16.7
Multiple infarcts	18	15.0
Parkinsonism	14	11.7
Cerebellar degeneration	8	6.7
Hydrocephalus	8	6.7
Toxic/metabolic causes	3	2.5
Psychogenic causes	4	3.3
Other	6	5.0
Unknown causes	17	14.2
Total	120	100

Source: Reproduced with permission from J Masdeu et al: *Gait Disorders of Aging*. Lippincott Raven, 1997.

■ 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. 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. 434.](#)

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. 428](#)) and stiff-person syndrome ([Chap. 90](#)). 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 DISORDERS

Parkinson's disease ([Chap. 427](#)) is common, affecting 1% of the population >55 years of age. The stooped posture and shuffling gait 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 fall risk. 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, an 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. 427](#)).

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.

Hyperkinetic movement disorders also produce characteristic and recognizable disturbances in gait. In Huntington's disease ([Chap. 428](#)), 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 D₂ 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. 26](#)), 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. 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 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. 432](#)) and various forms of hereditary cerebellar degeneration ([Chap. 431](#)). A short expansion at the site of the fragile X mutation (*fragile X pre-mutation*) 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.

TABLE 23-2 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

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 B₁₂ 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 23-2** compares sensory ataxia with cerebellar ataxia and frontal gait disorder.

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 disorders (formerly “psychogenic”) are common in neurologic practice, and the presentation often involves gait. The hallmark of a functional gait disorder is an internal inconsistency of deficits that may be incompatible with a neurological deficit. For example, odd gyrations of posture with wastage of muscular energy (astasia-abasia) appear superficially unsteady, yet in reality require significant postural control. Falls are rare, and there are often discrepancies between examination findings and the patient’s functional status. Extreme slow motion, an inappropriately overcautious gait, and dramatic fluctuations over time may improve with distraction, keeping in mind that numerous organic neurological diseases are also paroxysmal in nature. Preceding stress or trauma is variably present, and its absence no longer precludes the diagnosis of a functional neurological disorder. Functional gait disorders are among the most dramatic encountered, and should be differentiated from the slowness and psychomotor retardation seen in certain patients with major depression.

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, though neurologic and orthopedic problems may coexist. Characteristic patterns of abnormality are sometimes seen, though, 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, 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, though 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. 19**) 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. 90**). 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 nursing homes and hospitals. Elderly people are not only at higher risk for falls, but are 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 last 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, and environmental factors); some risk factors are modifiable. The presence of multiple risk factors is associated with a substantially increased risk of falls. (**Table 23-3**) 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.

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

TABLE 23-3 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.

adults should be asked about falls at least annually. The Timed Up and Go ("TUG") test involves timing a patient as they stand up from a chair, walk 10 ft, turn, then sit down. Patients with a history of falls, or those requiring >12 s to complete the TUG test, are 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, pre-syncopal 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 neuroleptics, benzodiazepines, anti-depressants, anti-arrhythmics, 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 neurological evaluation and intervention (**Chaps. 18 and 418**), though 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 neurological 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. Neurological 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 PSP, 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 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 etiology of the gait disorder and the mechanism underlying the falls by a given patient. Orthostatic changes in blood pressure and pulse should be recorded. Rising from a chair and walking should be evaluated for safety. Specific treatment may be possible once a diagnosis is established. Therapeutic intervention is often recommended for older patients at substantial risk for falls, even if no neurologic disease is identified. A home visit to look for environmental hazards can be helpful. A variety of modifications may be recommended to improve safety, including improved lighting and the installation of grab bars and nonslip surfaces.

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. A Tai Chi exercise program has been demonstrated to reduce the risk of falls and injury in patients with Parkinson's disease.

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.
- NUTT JC: Classification of Gait and Balance Disorders. *Adv Neurol* 87:135, 2001.
- PIRKER W, KATZENSCHLAGER R: Gait disorders in adults and the elderly. *Wien Klin Wochenschr* 129:81, 2017.

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 largely unknown. 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 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 re-experiencing 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, including those preparing for elective surgery or being admitted to the hospital. 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. 25) 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

24

Confusion and Delirium

S. Andrew Josephson, Bruce L. Miller

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

least two-thirds of cases of delirium occur in patients with coexisting underlying dementia. A form of dementia with parkinsonism, *dementia with Lewy bodies*, 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, 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 above, patients with preexisting dementia are particularly susceptible to episodes of delirium. Alzheimer's disease, dementia with Lewy bodies, and Parkinson's disease dementia 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 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.

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 24-1**). 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

TABLE 24-1 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)?

^aInformation is usually obtained from a reliable reporter, such as a family member, caregiver, or nurse.

Source: Modified from SK Inouye et al: Clarifying confusion: The Confusion Assessment Method. A new method for detection of delirium. Ann Intern Med 113:941, 1990.

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 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. 25); 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 24-2.

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 recent rise in availability of "bath salts," synthetic cannabis, 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. 447). 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. 445). Alcohol and benzodiazepine withdrawal should be considered in all cases of delirium because even patients who drink only a few servings of alcohol every day can experience relatively severe withdrawal symptoms upon hospitalization.

Metabolic abnormalities such as electrolyte disturbances of sodium, calcium, magnesium, or glucose can cause delirium, and

TABLE 24-2 Common Etiologies 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

Hypothermia and hyperthermia

Pulmonary failure: hypoxemia and hypercarbia

Liver failure/hepatic encephalopathy

Renal failure/uremia

Cardiac failure

Vitamin deficiencies: B₁₂, 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

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.

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 B₁₂, 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 lead 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, sometimes described as *terminal restlessness*, 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 24-3. 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

TABLE 24-3 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
 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: B₁₂, folate, thiamine
 Endocrinologic laboratories: thyroid-stimulating hormone (TSH) and free T₄; 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 reagent (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.

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 examination can also be useful in identifying 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.

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 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. 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. 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; however, there is little evidence that these medications are effective in delirium, and therefore they should be reserved for patients who display severe agitation and significant potential to harm themselves or staff. The recent 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 still use benzodiazepines to treat acute confusion, their use should be limited to cases in which delirium is caused by alcohol or benzodiazepine withdrawal.

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 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

- 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.
- HATTA K et al: Preventive effects of ramelteon on delirium: A randomized placebo-controlled trial. *JAMA Psychiatry* 71:397, 2014.
- NEUFELD KJ et al: Antipsychotic medication for prevention and treatment of delirium in hospitalized adults: A systematic review and meta-analysis. *J Am Geriatr Soc* 64:705, 2016.

25

Dementia

William W. Seeley, Bruce L. Miller



Dementia, a syndrome with many causes, affects >5 million people in the United States and results in a total annual health care cost in excess of \$250 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 aged >70 years and 20–40% of individuals aged >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, although in other dementias, such as *frontotemporal dementia (FTD)*, memory loss is not typically a presenting feature. **Focal cerebral disorders are discussed in Chap. 26 and illustrated in a video library in Chap. V2; detailed discussions of AD can be found in Chap. 423; FTD and related disorders in Chap. 424; vascular dementia in Chap. 425; DLB in Chap. 426; Huntington's disease (HD) in Chap. 428; and prion diseases in Chap. 430.**

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. 26). 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 begins in the entorhinal region of the medial temporal lobe, spreads to the hippocampus, and then moves to 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 disconnect 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-striatal¹ 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.

THE CAUSES OF DEMENTIA

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade for those aged >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.

The many causes of dementia are listed in Table 25-1. 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

TABLE 25-1 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)	

Less Common Causes of Dementia

Vitamin deficiencies	Toxic disorders
Thiamine (B ₁): Wernicke's encephalopathy ^a	Drug, medication, and narcotic poisoning ^a
B ₁₂ (subacute combined degeneration) ^a	Heavy metal intoxication ^a
Nicotinic acid (pellagra) ^a	Organic toxins
Endocrine and other organ failure	Psychiatric
Hypothyroidism ^a	Depression (pseudodementia) ^a
Adrenal insufficiency and Cushing's syndrome ^a	Schizophrenia ^a
Hypo- and hyperparathyroidism ^a	Conversion disorder ^a
Renal failure ^a	Degenerative disorders
Liver failure ^a	Huntington's disease
Pulmonary failure ^a	Multisystem atrophy
Chronic infections	Hereditary ataxias (some forms)
HIV	Frontotemporal lobar degeneration spectrum
Neurosphilis ^a	Multiple sclerosis
Papovavirus (JC virus) (progressive multifocal leukoencephalopathy)	Adult Down's syndrome with Alzheimer's disease
Tuberculosis, fungal, and protozoal ^a	ALS-parkinsonism-dementia complex of Guam
Whipple's disease ^a	Prion (Creutzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases)
Head trauma and diffuse brain damage	Miscellaneous
Chronic traumatic encephalopathy	Sarcoidosis ^a
Chronic subdural hematoma ^a	Vasculitis ^a
Postanoxia	CADASIL, etc.
Postencephalitis	Acute intermittent porphyria ^a
Normal-pressure hydrocephalus ^a	Recurrent nonconvulsive seizures ^a
Intracranial hypotension	Additional conditions in children or adolescents
Neoplastic	Pantothenate kinase-associated neurodegeneration
Primary brain tumor ^a	Subacute sclerosing panencephalitis
Metastatic brain tumor ^a	Metabolic disorders (e.g., Wilson's and Leigh's diseases, leukodystrophies, lipid storage diseases, mitochondrial mutations)
Paraneoplastic/autoimmune limbic encephalitis ^a	

^aPotentially reversible dementia.

Abbreviations: ALS, amyotrophic lateral sclerosis; CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LBD, Lewy body disease; PDD, Parkinson's disease dementia.

¹The striatum comprises the caudate/putamen/nucleus accumbens.

background. AD is the most common cause of dementia in Western countries, accounting for more than half of all patients. Vascular disease is considered 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, 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 can occur concurrently with or preceding the motor syndrome, as in DLB. A mixed pathology is common, especially in very old individuals. In patients aged <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 25-1 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.

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 25-1).

Subtle cumulative decline in episodic memory is a common part of aging. This frustrating experience, often the source of jokes and humor, is often referred to as *benign forgetfulness of the elderly*. Benign means that it is not so progressive or serious that it impairs reasonably successful and productive daily functioning, although the distinction between benign and more significant memory loss can be difficult to make. 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. A measurable cognitive problem that does not seriously disrupt daily activities is often referred to as *mild cognitive impairment* (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 deposition on positron emission tomography (PET) imaging.

The major degenerative dementias include AD, DLB, FTD and related disorders, HD, and prion diseases, including Creutzfeldt-Jakob disease (CJD). These disorders are all associated with the abnormal aggregation of a specific protein: Aβ₄₂ and tau in AD; α-synuclein in DLB; tau, TAR DNA-binding protein of 43 kDa (TDP-43), or fused in sarcoma (FUS) in FTD; huntingtin in HD; and misfolded prion protein (PrP^{Sc}) in CJD (Table 25-2).

APPROACH TO THE PATIENT

Dementias

Three major issues should be kept at the forefront: (1) What is the best fit for a 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 25-3. The major degenerative dementias can usually be distinguished by the initial symptoms; neuropsychological, neuropsychiatric, and neurologic findings; and neuroimaging features (Table 25-4).

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. 24) and should trigger the 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 difficulty with 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 empathy for others, or progressive loss of speech fluency or single-word comprehension and by a 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); the loss of skeletal muscle paralysis during dreaming); 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, 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 putative 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. 430). 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, chronic traumatic encephalopathy (a progressive dementia best characterized in contact sport athletes such as

TABLE 25-2 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
	FUS	Very rare FUS		FUS neuronal and glial inclusions varying in morphology and distribution
DLB	α-Synuclein	Very rare SNCA (4)	Unknown	α-Synuclein neuronal inclusions (Lewy bodies)
CJD	PrP ^{Sc}	PRNP (20) (up to 15% of patients carry these dominant mutations)	Codon 129 homozygosity for methionine or valine	PrP ^{Sc} deposition, panlaminar spongiosis

Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia.

TABLE 25-3 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 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
		Lab screen for 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).

boxers and American football players), intracranial hypotension, or NPH. Subacute onset of severe amnesia and psychosis with mesial temporal T2/fluid-attenuated inversion recovery (FLAIR) hyperintensities on magnetic resonance imaging (MRI) should raise concern

for paraneoplastic limbic encephalitis, especially in a long-term smoker or other patients at risk for cancer. Related autoimmune conditions, such as voltage-gated potassium channel (VGKC)- or N-methyl-D-aspartate (NMDA)-receptor antibody-mediated encephalopathy, can present with a similar tempo and imaging signature with or without characteristic motor manifestations such as myokymia (anti-VGKC) and faciobrachial dystonic seizures (anti-NMDA) (Chap. 90). Alcohol abuse creates risk for malnutrition and thiamine deficiency. Veganism, bowel irradiation, an autoimmune diathesis, a remote history of gastric surgery, and chronic antihistamine therapy for dyspepsia or gastroesophageal reflux predispose to B₁₂ 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 disorders, the recent death of a loved one, or depressive signs, such as insomnia or weight loss, raise the possibility of depression-related cognitive impairments.

PHYSICAL AND NEUROLOGIC EXAMINATION

A thorough general and neurologic examination is essential to document dementia, to look for other signs of nervous system involvement, and to search for clues suggesting a systemic disease that might be responsible for the cognitive disorder. Typical AD spares motor systems until later in the course. In contrast, FTD patients 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 the new onset of a parkinsonian syndrome (resting tremor, cogwheel rigidity, bradykinesia, festinating gait), but DLB often starts with visual hallucinations or dementia. 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 B₁₂ deficiency. Peripheral neuropathy could also indicate another vitamin deficiency, heavy metal

TABLE 25-4 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; PSP, progressive supranuclear palsy; REM, rapid eye movement.

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), and Cognistat can be used to capture dementia and follow progression. None of these tests is highly sensitive to early-stage dementia or 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 in the areas 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 min 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) function, but some patients lack either finding despite profound social-emotional deficits. PDD or DLB patients have more severe deficits in visuospatial function but do better on episodic memory tasks than patients with AD. Patients with vascular dementia often demonstrate a mixture of executive control 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, euphoria, 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,

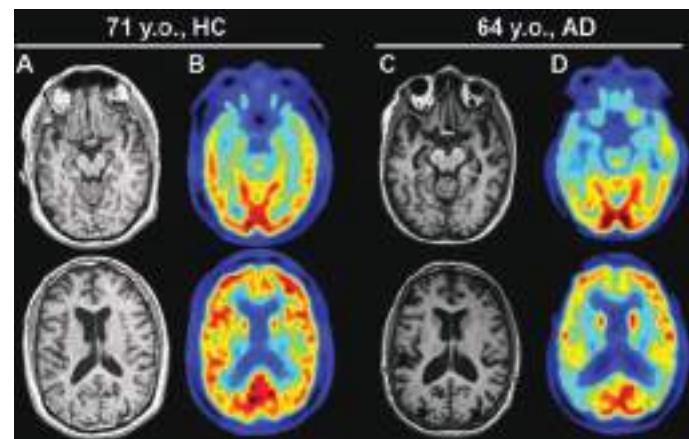


FIGURE 25-1 **Alzheimer’s disease (AD).** Axial T1-weighted magnetic resonance images of a healthy 71-year-old (**A**) and a 64-year-old with AD (**C**). Note the reduction in medial temporal lobe volume in the patient with AD. Fluorodeoxyglucose positron emission tomography scans of the same individuals (**B** and **D**) demonstrate reduced glucose metabolism in the posterior temporoparietal regions bilaterally in AD, a typical finding in this condition. HC, healthy control. (Images courtesy of Gil Rabinovici, University of California, San Francisco and William Jagust, University of California, Berkeley.)

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 25-3 lists most screening tests for dementia. The American Academy of Neurology recommends the routine measurement of a complete blood count, electrolytes, renal and thyroid function, a vitamin B₁₂ level, and a neuroimaging study (computed tomography [CT] or MRI).

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. 25-1). Focal frontal, insular, and/or anterior temporal atrophy suggests FTD (Chap. 424). DLB often features less prominent atrophy, with greater involvement of amygdala than 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. 25-2). Communicating hydrocephalus with



FIGURE 25-2 **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.

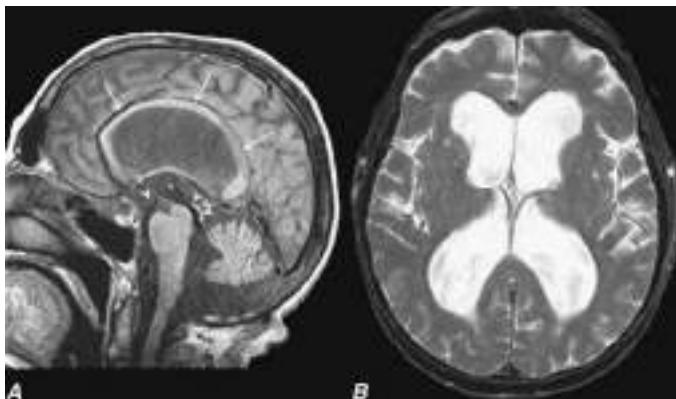


FIGURE 25-3 Normal-pressure hydrocephalus. **A.** Sagittal T1-weighted magnetic resonance image (MRI) demonstrates dilation of the lateral ventricle and stretching of the corpus callosum (arrows), depression of the floor of the third ventricle (single arrowhead), and enlargement of the aqueduct (double 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.

vertex effacement (crowding of dorsal convexity gyri/sulci), gaping Sylvian fissures despite minimal cortical atrophy, and additional features shown in Fig. 25-3 suggest NPH. Single-photon emission computed tomography (SPECT) and PET scanning show temporal-parietal hypoperfusion or hypometabolism in AD and frontotemporal deficits in FTD, but these changes often reflect atrophy and can therefore be detected with MRI alone in many patients. Recently, amyloid imaging has shown promise for the diagnosis of AD, and Pittsburgh Compound-B (PiB) (not available outside of research settings) and ¹⁸F-AV-45 (florbetapir; approved by the U.S. Food and Drug Administration in 2013) are reliable radioligands for detecting brain amyloid associated with amyloid angiopathy or neuritic plaques of AD (Fig. 25-4). Because these abnormalities can be seen in cognitively normal older persons (~25% of individuals at age 65), however, amyloid imaging may also detect preclinical or incidental AD in patients lacking an AD-like dementia syndrome. Currently, the main clinical value of amyloid imaging is to exclude AD as the likely cause of dementia in patients who have negative scans. Once disease-modifying therapies become available, use of these biomarkers may help to identify treatment candidates before irreversible brain injury has occurred. In the meantime, the prognostic value of detecting brain amyloid in an asymptomatic elder remains a topic of vigorous investigation. Similarly, MRI perfusion and structural/functional connectivity methods are being explored as potential treatment-monitoring strategies.

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_β₄₂ and tau proteins show differing patterns with the various dementias, and the presence of low A_β₄₂ and mildly elevated CSF tau is highly suggestive of AD. The routine use of lumbar puncture in the diagnosis of dementia is debated, but the sensitivity and specificity of AD diagnostic measures are not yet high enough to warrant routine use. Formal psychometric testing helps to document the severity of cognitive disturbance, suggest psychogenic causes, and provide 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, potentially treatable neoplasms, or unusual infections when the diagnosis is uncertain. Systemic disorders with CNS manifestations, such as sarcoidosis, can usually 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.

■ GLOBAL CONSIDERATIONS



Vascular dementia (Chap. 425) is more common in Asian countries, 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, particularly with HIV (and associated opportunistic infections), syphilis, and tuberculosis, likewise represent major contributors to dementia in the developing world. 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 to provide comfort and support to the patient and caregivers. Treatment of underlying causes includes thyroid replacement for hypothyroidism; vitamin therapy for thiamine or B₁₂ deficiency or for elevated serum homocysteine; antimicrobials for opportunistic infections or antiretrovirals for HIV; ventricular shunting for NPH; or appropriate surgical, radiation, and/or chemotherapeutic treatment for CNS neoplasms. Removal of cognition-imparing

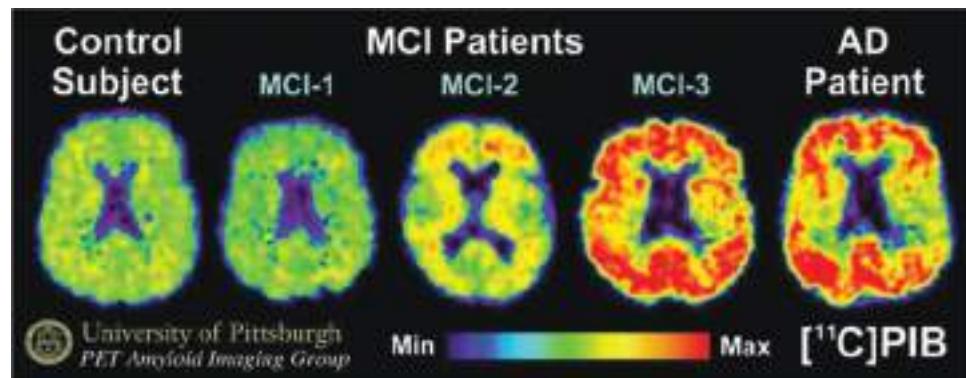


FIGURE 25-4 Positron emission tomography (PET) images obtained with the amyloid-imaging agent Pittsburgh Compound-B ([¹¹C]PiB) in a normal control (left); three different patients with mild cognitive impairment (MCI; center); and a patient with mild Alzheimer's disease (AD; right). Some MCI patients have control-like levels of amyloid, some have AD-like levels of amyloid, and some have intermediate levels. (Images courtesy of William Klunk and Chester Mathis, University of Pittsburgh.)

drugs or medications is frequently useful. If the patient's cognitive complaints stem from a psychiatric disorder, vigorous treatment of this condition should seek to eliminate the cognitive complaint or 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. Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) (**Chap. 443**), which feature anxiolytic properties but few cognitive side effects, provide the mainstay of treatment when necessary. Anticonvulsants are used to control seizures. Levetiracetam may be particularly useful, but there have as yet been no randomized trials for treatment of 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, dyskinesia, and occasionally paradoxical disinhibition (benzodiazepines). 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. When patients do not respond to treatment, it is usually a mistake to advance to higher doses or to use anticholinergic drugs 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 the 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). Recent work has focused on developing antibodies against $\text{A}\beta_{42}$ as a treatment for AD. Although the initial randomized controlled trials failed, there was some evidence for efficacy in the mildest patient groups. Therefore, researchers have begun to focus on patients with very mild disease and asymptomatic individuals at risk for AD, such as those who carry autosomal dominantly inherited genetic mutations or healthy elders with CSF or amyloid imaging biomarker evidence supporting presymptomatic AD. Memantine proves useful when treating 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.

A proactive strategy has been shown to reduce the occurrence of delirium in hospitalized patients. This strategy includes frequent orientation, cognitive activities, sleep-enhancement measures, vision and hearing aids, and correction of dehydration.

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 activities such as walking, bowling, dancing, singing, bingo, and golf. Demented patients often object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help or take over 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 in this setting. 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 be of considerable benefit.

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 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.
- GRIEM J et al: Psychologic/functional forms of memory disorder. *Handb Clin Neurol* 139:407, 2017.

26

Aphasia, Memory Loss, Hemispatial Neglect, Frontal Syndromes, and Other Cerebral Disorders

M.-Marsel Mesulam



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. 26-1**). 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 explicit episodic memory, 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.

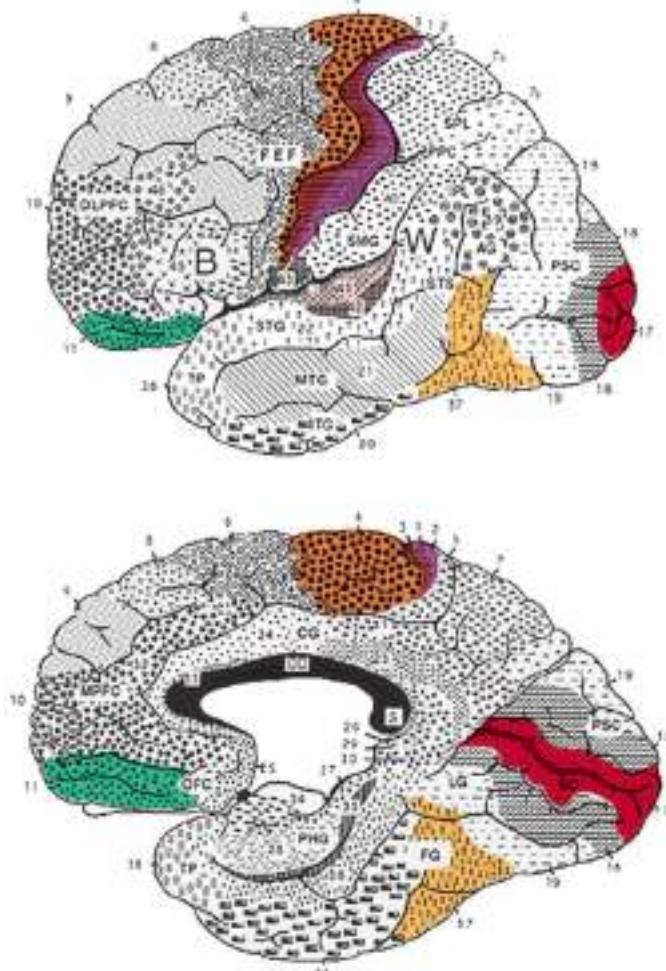


FIGURE 26-1 Lateral (top) and medial (bottom) views of the cerebral hemispheres. The numbers refer to the Brodmann cyto-architectonic designations. Area 17 corresponds to the primary visual cortex, 41–42 to the primary auditory cortex, 1–3 to the primary somatosensory cortex, and 4 to the primary motor 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.

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 rather than 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.

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 palilalia 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 26-1 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.

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 down-stairs was my teeth-tick ... a ... den ... dentith ... my dentist. And they

TABLE 26-1 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

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. 419). 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. 26-1) and surrounding anterior perisylvian and insular cortex due to occlusion of the superior division of the MCA (Chap. 419). 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 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. 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 Agraphia *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 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 *apraxia* 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 26-1. 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 26-1.

Several neurodegenerative syndromes, such as typical Alzheimer-type (amnestic; *Chap. 423*) and frontotemporal (behavioral; *Chap. 424*) 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 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. 430*).

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 26-1 but usually has longer and more frequent word-finding pauses. When repetition is impaired the aphasia resembles the *conduction aphasia* in Table 26-1. 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 26-1. 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. 26-2*). 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.

■ 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 hemisphere. 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 lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemisphere 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,

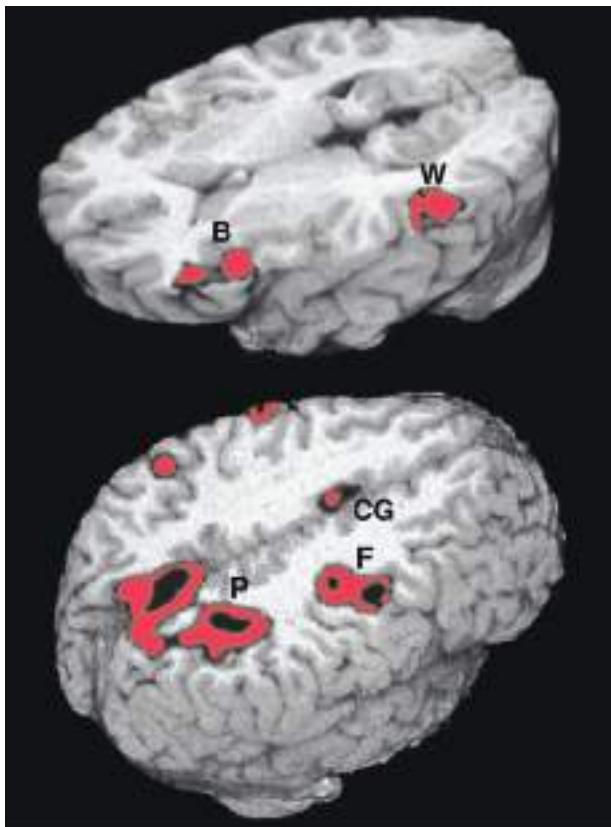


FIGURE 26-2 Functional magnetic resonance imaging of language and spatial attention in neurologically intact subjects. The red and black areas show regions of task-related significant activation. (Top) The subjects were asked to determine if two words were synonymous. This language task led to the simultaneous 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; with permission.)

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. 26-3A). 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*.

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 Alzheimer's disease.

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. 26-3B). 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. 26-3B is not by itself sufficient to diagnose simultanagnosia because some patients with a frontal network syndrome may omit the strange looking large letters, perhaps because they lack the mental flexibility needed 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 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. 26-4). 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.

THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION

A patient with *prosopagnosia* cannot recognize familiar faces, including, sometimes, the reflection of his or her 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

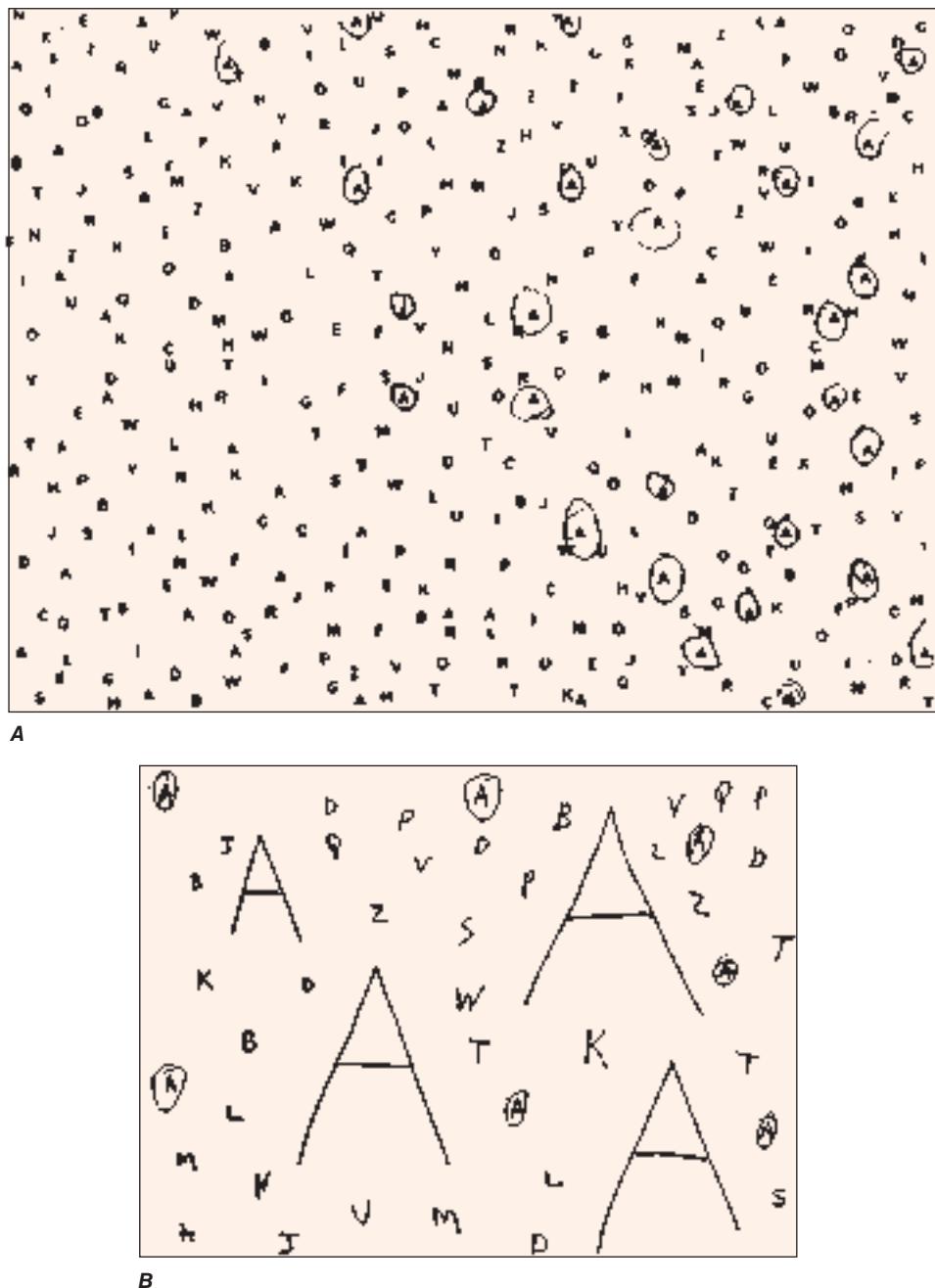


FIGURE 26-3 **A.** A 47-year-old man with a large frontoparietal lesion in the right hemisphere was asked to circle all the 'A's. Only targets on the right are circled. This is a manifestation of left hemispatial neglect. **B.** A 70-year-old woman with a 2-year history of degenerative dementia was able to circle most of the small targets but ignored the larger ones. This is a manifestation of simultanagnosia.

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 understanding words that denote objects and in naming faces and objects but a relative preservation of face

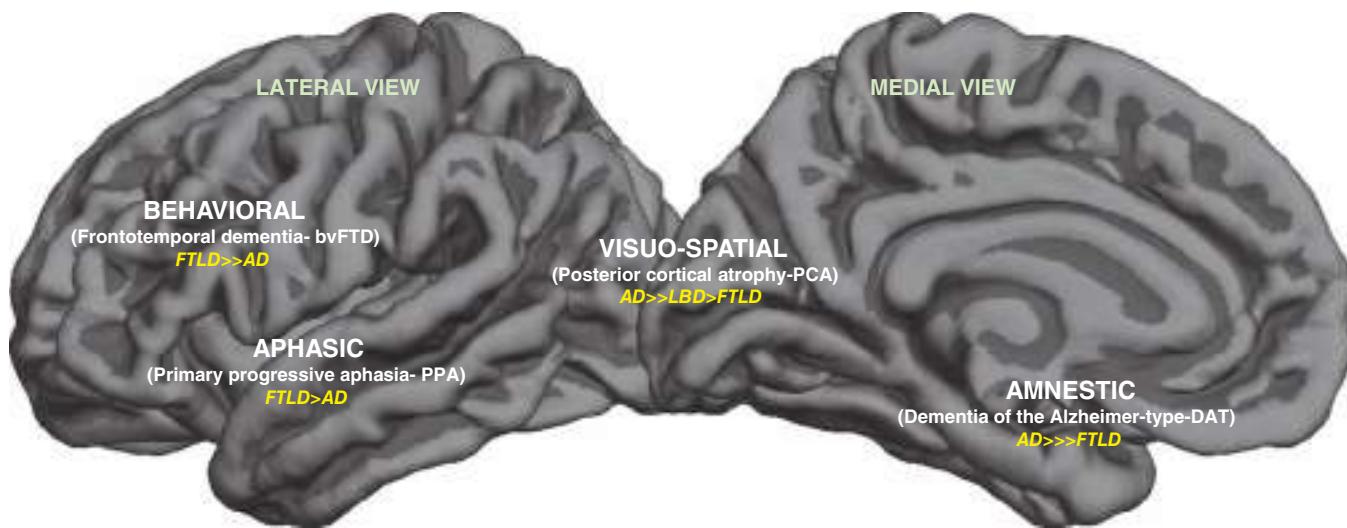


FIGURE 26-4 Four focal dementia syndromes and their most likely neuropathologic correlates. AD, Alzheimer's disease; bvFTD, behavioral variant frontotemporal dementia; DAT, amnestic dementia of the Alzheimer-type; FTLD, frontotemporal lobar degeneration (tau or TDP-43 type); LBD, Lewy body disease; PCA, posterior cortical atrophy syndrome; PPA, primary progressive aphasia.

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 and paralimbic areas (such as the hippocampus, amygdala, and entorhinal cortex), 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 include the coordination of emotion, motivation, autonomic tone, and endocrine function. 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. A disturbance in this function 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 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 term *salience network* has been introduced to designate parts of the frontal network and their interactions with adjacent paralimbic cortices of the insula and cingulate gyrus. Impairments of social conduct and empathy seen in neurodegenerative frontal dementias are attributed to pathology of the salience network.

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 abulia 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 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 known as the behavioral variant of frontotemporal dementia (bvFTD). In many patients with bvFTD the atrophy extends into the anterior temporal lobes. 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 B₁₂ 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 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 Non-pharmacologic 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 non-pharmacologic interventions have been used to treat higher cortical deficits. These include speech therapy for aphasias, behavioral modification for comportmental 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 CSF biomarkers now makes it possible to address this question with reasonable success and to make specific diagnoses of AD, LBD, CJD, FTLD. A specific etiological 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

- 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*, M-M Mesulam (ed). New York, Oxford University Press, 2000, pp 1-120.
- MESULAM M-M et al: Case 1-2017: A 70-year-old woman with gradually progressive loss of language. *N Engl J Med* 376:158, 2017.
- MILLER BL, BOEVE BF (eds): *The Behavioral Neurology of Dementia*, 2nd ed. Cambridge University Press, 2017.
- TEICHMANN M et al: Direct current stimulation over the anterior temporal areas boosts semantic processing in primary progressive aphasia. *Ann Neurol* 80:693, 2016.

27

Sleep Disorders

Thomas E. Scammell, Clifford B. Saper,
Charles A. Czeisler



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 20 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

Adults need at least 7 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.

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 increasing arousal threshold and slowing of the cortical EEG. REM sleep is characterized by a low-amplitude, mixed-frequency EEG similar to that of NREM stage N1 sleep, and the EOG shows REMs which 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. 27-1). 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.

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. 27-1). N3 sleep is most intense and prominent during childhood, decreasing with puberty and across the second and third decades of life. N3 sleep declines during adulthood to the point where it may be completely absent in older adults. The remaining NREM sleep becomes more fragmented, with many more 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 to the neurobehavioral performance impairment induced by acute sleep deprivation than young adults, 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 is generally recovered 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, plus 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. 27-2, 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 GABA to disinhibit cortical inhibitory interneurons, thus promoting arousal. Additional wake-promoting neurons in the hypothalamus use the peptide neurotransmitter orexin (also known as hypocretin, shown in blue) to reinforce activity in the other arousal cell groups.

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). Damage to the thalamus causes loss of the content of wakefulness, 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. 27-2 in red. These neurons in the preoptic area and pons use γ -aminobutyric acid (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

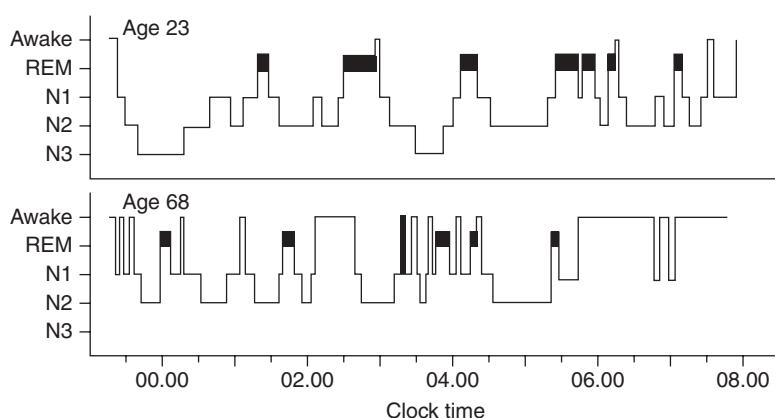


FIGURE 27-1 Wake-sleep architecture. Alternating stages of wakefulness, the three stages of non-rapid eye movement sleep (N1–N3), and rapid eye movement (REM) sleep (solid bars) occur over the course of the night for representative young and older adult men. Characteristic features of sleep in older people include reduction of N3 slow-wave sleep, frequent spontaneous awakenings, early sleep onset, and early morning awakening.

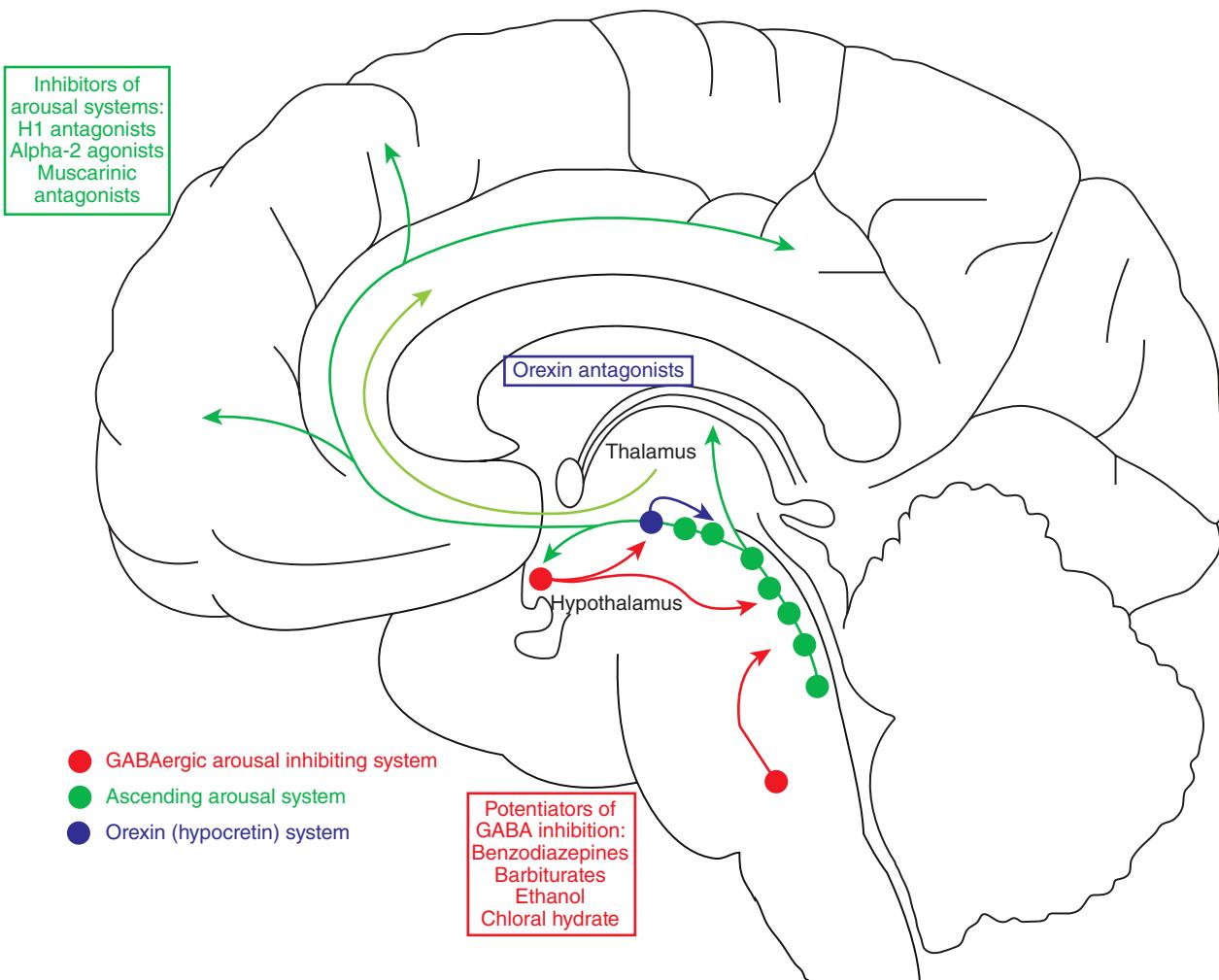


FIGURE 27-2 Relationship of drugs for insomnia with wake-sleep systems. The arousal system in the brain (green) includes monoaminergic, glutamatergic, and cholinergic neurons in the brainstem that activate neurons in the hypothalamus, thalamus, basal forebrain, and cerebral cortex. Orexin neurons (blue) 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).

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 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 D₂, have also been implicated in this process. Both adenosine and prostaglandin D₂ 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 targets in the limbic system 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. 27-2) or potentiate the output of the sleep-promoting system (red in Fig. 27-2). 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. At the molecular level, endogenous circadian rhythmicity is driven by self-sustaining transcriptional/translational feedback loops. 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).

While it is now recognized that most cells in the body have circadian clocks that regulate diverse physiologic processes, most of these disparate clocks 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 neurons in the SCN 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. They also receive visual input to synchronize them with the external world and 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. For example, young adults typically have long intrinsic circadian periods and consequently go to bed late and rise late, whereas others have short periods and go to bed and rise earlier. 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. 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 (see “Shift-Work Disorder,” below).

Subjects woken 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 PCO_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 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 adrenocorticotrophic 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. 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 patients 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.

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 needn't fall asleep to have an accident, as the inattention and slowed responses of drowsy drivers are a major contributor. 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. 436*), fibromyalgia (*Chap. 366*), chronic fatigue syndrome (*Chap. 442*), or endocrine deficiencies such as hypothyroidism (*Chap. 376*) or Addison's disease (*Chap. 379*). 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 reflecting poor motivation or bad sleep habits.

Table 27-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

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

TABLE 27-1 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 (<i>Chap. 291</i>)	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); 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; 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

episodes while driving or in other safety-related settings, sleepiness while at work or school (and the relationship of sleepiness to work and school 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 the nighttime sleep episode, and the occurrence of REM sleep in two or more of the MSLT 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 from state to state, 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 lead to impairment of daytime alertness, increased risk of sleep-related motor vehicle crashes, depression, hypertension, myocardial infarction, diabetes, stroke, and increased 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. 291 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. In addition, they often experience symptoms related to an intrusion of REM sleep characteristics. 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 27-1); (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.

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 cerebrospinal fluid, 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. 27-3). Patients with narcolepsy but no cataplexy

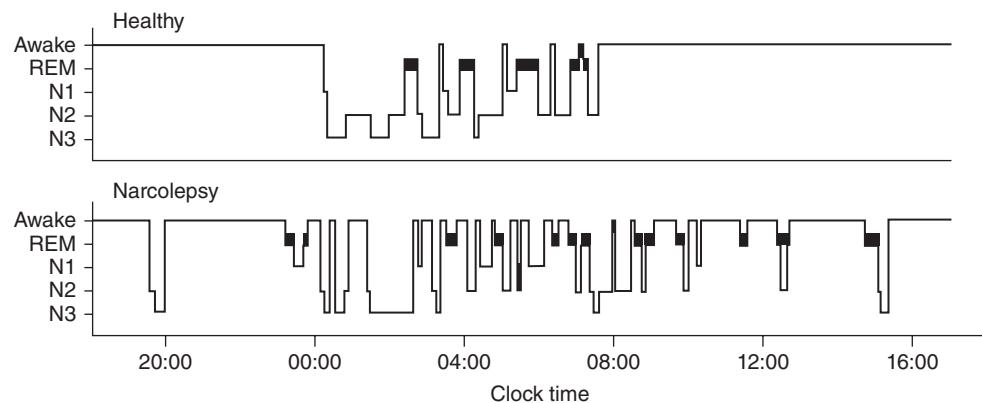


FIGURE 27-3 Polysomnographic recordings of a healthy individual and a patient with narcolepsy. The healthy individual has a long period of NREM sleep before entering REM sleep, but the individual with narcolepsy enters rapid eye movement (REM) sleep quickly at night and has moderately fragmented sleep. During the day, the healthy subject stays awake from 8:00 AM until midnight, but the patient with narcolepsy dozes off frequently, with many daytime naps that include REM sleep.

(type 2 narcolepsy) usually have normal orexin levels and may have other yet uncharacterized causes of their excessive daytime sleepiness.

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. 343), 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). Traumatic brain injury can also damage orexin-containing neurons, inducing type 2 narcolepsy.

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 heartfelt mirth when laughing at a great 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 was not sleep deprived 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 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 used quite often 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 often 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. Sodium oxybate (gamma hydroxybutyrate) is given twice each night and 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, given at bedtime and 3–4 h later, is also very helpful in reducing cataplexy.

¹No antidepressant has been approved by the U.S. 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 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., TV, 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 benzodiazepines such as alprazolam.

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 in bed due to circadian variations in airway resistance and postural changes 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. 25) 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, and although insomnia is a common early symptom, most patients present with other obvious neurologic signs such as dementia, myoclonus, dysarthria, or autonomic dysfunction.

TREATMENT

Insomnia

Treatment of insomnia improves quality of life and can promote long-term health. With improved sleep, patients often report less daytime fatigue, improved cognition, and more energy. Treating the insomnia can also improve the comorbid disease. For example, management of insomnia at the time of diagnosis of major depression often improves the response to antidepressants and reduces the risk of relapse. Sleep loss can heighten the perception of pain, so a similar approach is warranted in acute and chronic pain management.

The treatment plan should target all putative contributing factors: establish good sleep hygiene, treat medical disorders, use behavioral therapies for anxiety and negative conditioning, and use pharmacotherapy and/or psychotherapy for psychiatric disorders. Behavioral therapies should be the first-line treatment, followed by judicious use of sleep-promoting medications if needed.

TREATMENT OF MEDICAL AND PSYCHIATRIC DISEASE

If the history suggests that a medical or psychiatric disease contributes to the insomnia, then it should be addressed by, for example, treating the pain, improving breathing, and switching or adjusting the timing of medications.

IMPROVE SLEEP HYGIENE

Attention should be paid to improving sleep hygiene and avoiding counterproductive, arousing behaviors before bedtime. Patients should establish a regular bedtime and wake time, even on weekends, to help synchronize their circadian rhythms and sleep patterns. The amount of time allocated for sleep should not be more than their actual total amount of sleep. In the 30 min before bedtime, patients should establish a relaxing "wind-down" routine that can include a warm bath, listening to music, meditation, or other relaxation techniques. The bedroom should be off-limits to computers, televisions, radios, smartphones, videogames, and tablets. Once in bed, patients should try to avoid thinking about anything stressful or arousing such as problems with relationships or work. If they cannot fall asleep within 20 min, it often helps to get out of bed and read or listen to relaxing music in dim light as a form of distraction from any anxiety, but artificial light, including light from a television, cell phone, or computer, should be avoided, because light itself suppresses melatonin secretion and is arousing.

Table 27-2 outlines some of the key aspects of good sleep hygiene to improve insomnia.

COGNITIVE BEHAVIORAL THERAPY

CBT uses a combination of the techniques above plus additional methods to improve insomnia. A trained therapist may use cognitive psychology techniques to reduce excessive worrying about sleep and to reframe faulty beliefs about the insomnia and its daytime consequences. The therapist may also teach the patient relaxation techniques, such as progressive muscle relaxation or meditation, to reduce autonomic arousal, intrusive thoughts, and anxiety.

MEDICATIONS FOR INSOMNIA

If insomnia persists after treatment of these contributing factors, pharmacotherapy is often used on a nightly or intermittent basis. A variety of sedatives can improve sleep.

Antihistamines, such as diphenhydramine, are the primary active ingredient in most over-the-counter sleep aids. These may be of benefit when used intermittently, but can produce tolerance and anticholinergic side effects such as dry mouth and constipation, which limit their use, particularly in the elderly.

Benzodiazepine receptor agonists (BzRAs) are an effective and well-tolerated class of medications for insomnia. BzRAs bind to the GABA_A receptor and potentiate the postsynaptic response to GABA. GABA_A receptors are found throughout the brain, and BzRAs may

TABLE 27-2 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

globally reduce neural activity and may enhance the activity of specific sleep-promoting GABAergic pathways. Classic BzRAs include lorazepam, triazolam, and clonazepam, whereas newer agents such as zolpidem and zaleplon have more selective affinity for the α_1 subunit of the GABA_A receptor.

Specific BzRAs are often chosen based on the desired duration of action. The most commonly prescribed agents in this family are zaleplon (5–20 mg), with a half-life of 1–2 h; zolpidem (5–10 mg) and triazolam (0.125–0.25 mg), with half-lives of 2–4 h; eszopiclone (1–3 mg), with a half-life of 5–8 h; and temazepam (15–30 mg), with a half-life of 8–20 h. Generally, side effects are minimal when the dose is kept low and the serum concentration is minimized during the waking hours (by using the shortest-acting effective agent). For chronic insomnia, intermittent use is recommended, unless the consequences of untreated insomnia outweigh concerns regarding chronic use.

The heterocyclic *antidepressants* (trazodone, amitriptyline,² and doxepin) are the most commonly prescribed alternatives to BzRAs due to their lack of abuse potential and lower cost. Trazodone (25–100 mg) is used more commonly than the tricyclic antidepressants, because it has a much shorter half-life (5–9 h) and less anticholinergic activity.

The orexin receptor antagonist suvorexant (10–20 mg) can also improve insomnia by blocking the wake-promoting effects of the orexin neuropeptides. It has a long half-life and can produce morning sedation, and as it reduces orexin signaling, it can rarely produce hypnagogic hallucinations and sleep paralysis (see narcolepsy section above).

Medications for insomnia are now among the most commonly prescribed medications, but they should be used cautiously. All sedatives increase the risk of injurious falls and confusion in the elderly, and therefore if needed, these medications should be used at the lowest effective dose. Morning sedation can interfere with driving and judgment, and when selecting a medication, one should consider the duration of action. Benzodiazepines carry a risk of addiction and abuse, especially in patients with a history of alcohol or sedative abuse. In patients with depression, all sedatives can worsen the depression. Like alcohol, some sleep-promoting medications can worsen sleep apnea. Sedatives can also produce complex behaviors during sleep, such as sleep walking and sleep eating, although this seems more likely at higher doses.

²Trazodone and amitriptyline have not been approved by the FDA for treating insomnia.

■ 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 behavior such as gambling and sex. 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 do not 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

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 as used for RLS, including dopamine agonists. Recent 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 occasional 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 insufficient sleep, which subsequently causes an increase in deep NREM sleep; alcohol; and stress. 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

the parents that the condition is self-limited and benign, and like sleep-walking, 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.

REM Sleep Behavior Disorder (RBD) RBD ([Video 27-2](#)) 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 non-respiratory 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.

The prevalence of RBD increases with age, afflicting about 2% of adults aged >70, and is about twice as common in men. Most already have or will develop a neurodegenerative disorder. Within 12 years of disease onset, half of RBD patients develop a synucleinopathy such as Parkinson's disease ([Chap. 427](#)) or dementia with Lewy bodies ([Chap. 426](#)), or occasionally multiple system atrophy ([Chap. 432](#)), 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.

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 an appropriate phase.

Delayed Sleep-Wake Phase Disorder DSWPD is characterized by: (1) reported 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). 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 rate of 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 DSWD from other forms of sleep-onset insomnia. DSWD 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 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 (N24SWRD) 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 non-photic 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 N24SWRD is restricted to the nighttime hours due to social or occupational demands. Despite this regular sleep-wake schedule, affected patients with N24SWRD 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 N24SWRD, 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,

³No medications have been approved by the FDA for the treatment of RBD.

in some cases, induce synchronization of the circadian pacemaker. In sighted patients, N24SWRD is usually 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 N24SWRD.

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 and to 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 produce decreased alertness and performance, increased reaction time, and increased 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 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.

From 5 to 15% of individuals scheduled to work at night or in the early morning hours 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 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 following 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

- DING F et al: Changes in the composition of brain interstitial ions control the sleep-wake cycle. *Science* 352:550, 2016.
- JU YE et al: Sleep and Alzheimer disease pathology—A bidirectional relationship. *Nat Rev Neurol* 10:115, 2014.
- LEE ML et al: High risk of near-crash driving events following night-shift work. *Proc Natl Acad Sci USA* 113:176, 2016.
- 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.
- LIU Y et al: Prevalence of healthy sleep duration among adults—United States, 2014. *MMWR Morb Mortal Wkly Rep* 65:137, 2016.
- RIEMANN D et al: The neurobiology, investigation, and treatment of chronic insomnia. *Lancet Neurol* 14:547, 2015.
- SCAMMELL TE: Narcolepsy. *N Engl J Med* 373:2654, 2015.
- SCAMMELL TE et al: Neural circuitry of wakefulness and sleep. *Neuron* 93:747, 2017.
- STOTHARD ER et al: Circadian entrainment to the natural light-dark cycle across seasons and the weekend. *Curr Biol* 27:508, 2017.
- XIE L et al: Sleep drives metabolite clearance from the adult brain. *Science* 342:373, 2013.

VIDEO 27-1 A typical episode of severe cataplexy. The patient is joking and then falls to the ground with an abrupt loss of muscle tone. The electromyogram recordings (four lower traces on the right) show reductions in muscle activity during the period of paralysis. The electroencephalogram (top two traces) shows wakefulness throughout the episode. (Video courtesy of Giuseppe Plazzi, University of Bologna.)

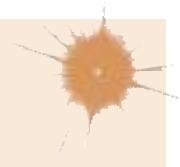
VIDEO 27-2 Typical aggressive movements in rapid eye movement (REM) sleep behavior disorder. (Video courtesy of Dr. Carlos Schenck, University of Minnesota Medical School.)

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

28

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.



FIGURE 28-1 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.

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.

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. 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 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

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 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. 28-1). 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.

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 (*Papilledema'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. 28-2). It is an extremely useful sign in retrobulbar optic neuritis and other optic nerve diseases, in which

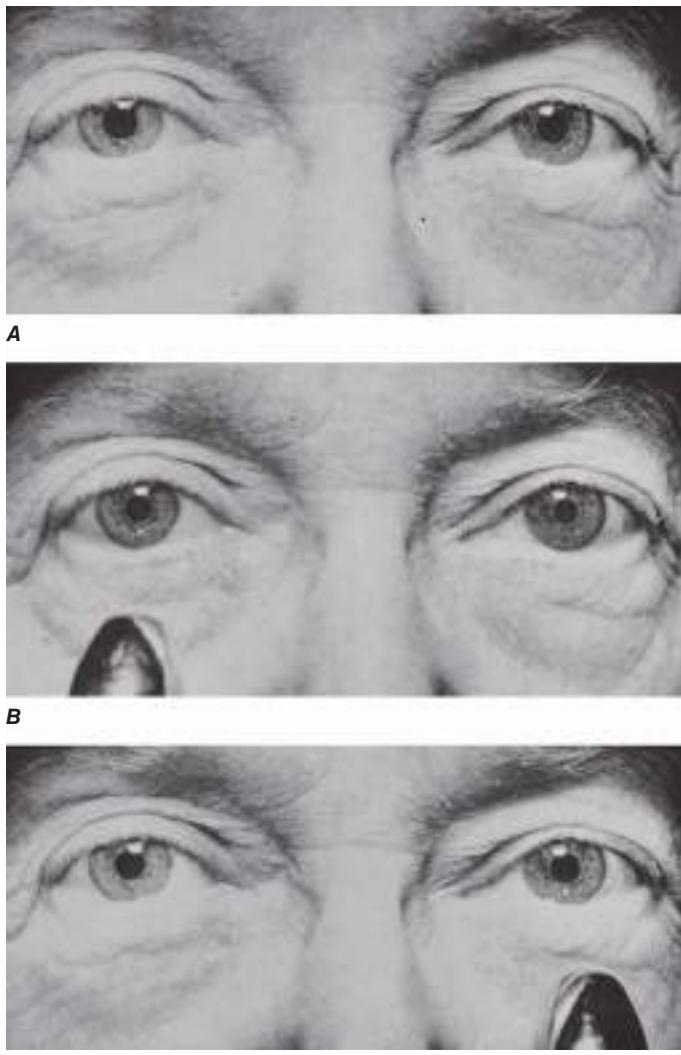


FIGURE 28-2 Demonstration of a relative afferent pupillary defect (Marcus Gunn pupil) in the left eye, done with the patient fixating on a distant target. **A.** With dim background lighting, the pupils are equal and relatively large. **B.** Shining a flashlight into the right eye evokes equal, strong 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.)

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

Stereoacluity 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 stereoacluity is 40 s of arc. If a patient achieves this level of stereoacluity, 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 visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worth 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). Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors but cannot name them.

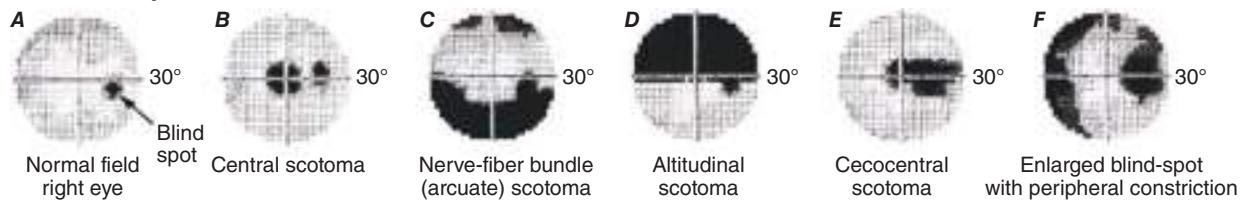
VISUAL FIELDS

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. 28-3**). 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. 28-3A**). 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.

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

Monocular prechiasmal field defects:



Binocular chiasmal or postchiasmal field defects:

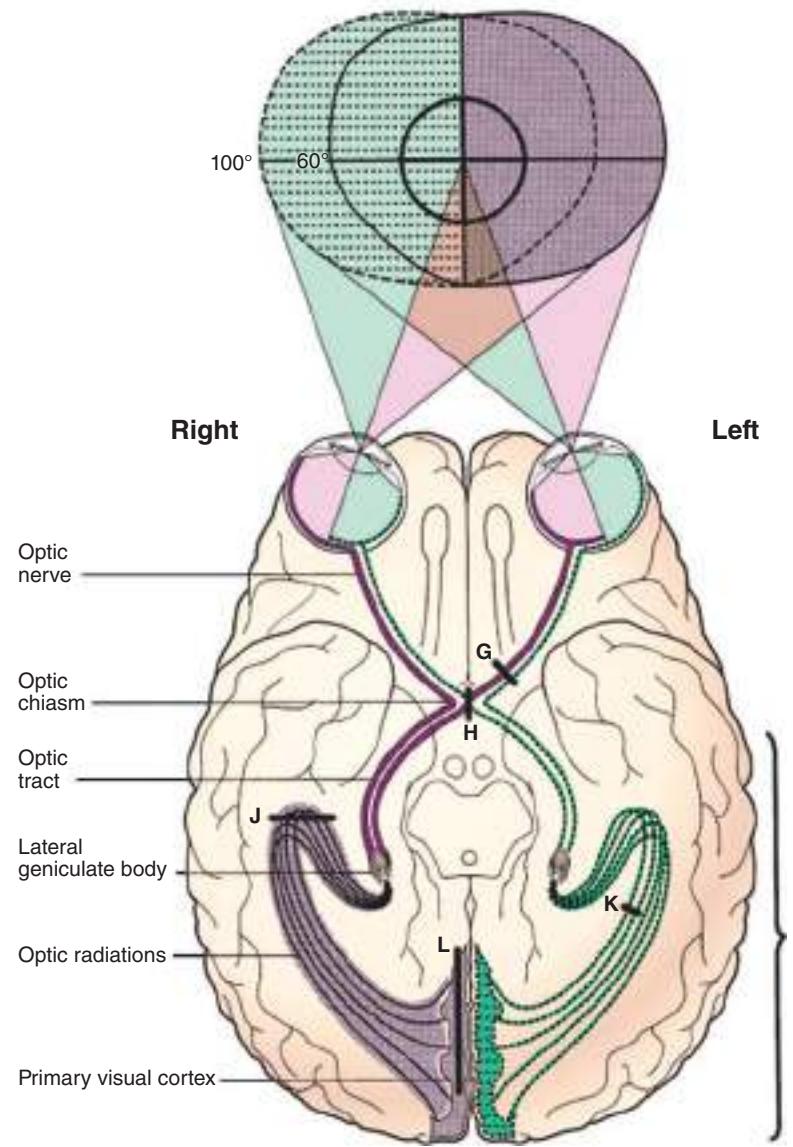
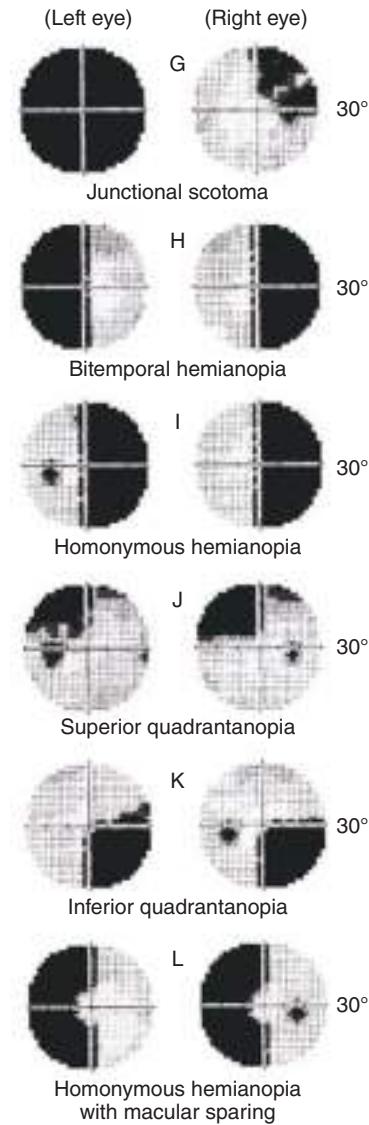


FIGURE 28-3 Ventral view of the brain, correlating patterns of visual field loss with the sites of lesions in the visual pathway. The visual fields overlap partially, creating 120° of central binocular field flanked by a 40° monocular crescent on either side. The visual field maps in this figure were 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.

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. 28-3B**).

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. 28-3C**). 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. 28-3D**). 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. 28-3E**). 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. 28-3G**). More symmetric compression of the optic chiasm by a pituitary adenoma (see **Fig. 373-1**), meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia (**Fig. 28-3H**). 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. 28-3I**). 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. 28-3J**), whereas injury to the optic radiations in the parietal lobe results in an inferior quadrantic homonymous hemianopia (**Fig. 28-3K**). 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 with hemianopia after occipital stroke have macular sparing, because the macular representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery (**Fig. 28-3L**). 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 real improvement of the visual fields can be demonstrated. No effective treatment has been devised for homonymous hemianopia caused by loss of visual cortex.

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 exposed basement membrane underlying the 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 should 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 frequent 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 a common problem.

Blepharitis 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, using 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 hypertrophic 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 (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 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, 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. 187). 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. A dendritic pattern of corneal epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but is seen in only a minority of primary infections. Recurrent ocular infection arises from reactivation of the latent herpesvirus. Viral eruption in the corneal epithelium may result in the characteristic herpes dendrite. Involvement of the corneal stroma produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with cycloplegia, and either a topical antiviral (trifluridine, ganciclovir) or an oral antiviral (acyclovir, ganciclovir) agent. Topical glucocorticoids are effective in mitigating corneal scarring but are generally reserved for cases involving stromal damage, because of the danger of corneal melting and perforation. Topical glucocorticoids also carry the risk of prolonging infection and inducing glaucoma.

Herpes Zoster Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis (Chap. 188). 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.

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.

Uveitis Involving the anterior structures of the eye, uveitis also is 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 rheumatoid arthritis, inflammatory bowel disease, psoriasis, reactive arthritis, and Behcet'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 This is diagnosed 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, Behcet'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. 190-1); 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. Monoclonal antibodies which 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. 28-4) are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis also 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.

■ TRANSIENT OR SUDDEN VISUAL LOSS

Amaurosis Fugax This term refers to a transient ischemic attack of the retina (Chap. 420). Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a



FIGURE 28-4 Roth's spot, cotton-wool spot, and retinal hemorrhages in a 48-year-old liver transplant patient with candidemia from immunosuppression.

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. 28-5). 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. 28-6). 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.

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,



FIGURE 28-5 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.



FIGURE 28-6 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.

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), and leakage of lipid and fluid (hard exudate) into the macula (**Fig. 28-7**). In hypertensive crisis, sudden visual loss can result from vasospasm of retinal arterioles and retinal ischemia. In addition, acute hypertension may produce visual loss from ischemic swelling of the optic disc. 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.

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. 28-8**). 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.

Anterior Ischemic Optic Neuropathy (AION) This is caused by insufficient blood flow through the posterior ciliary arteries that

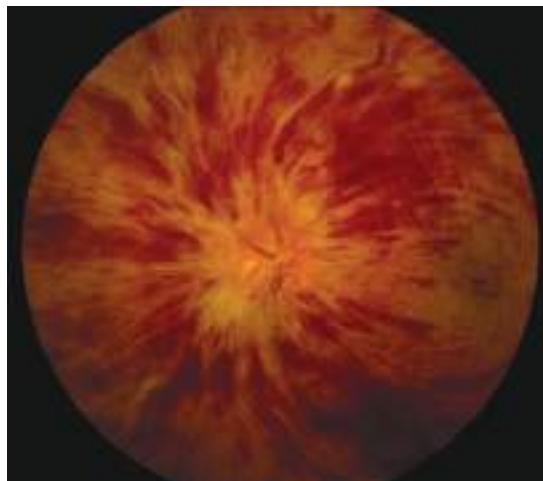


FIGURE 28-8 Central retinal vein occlusion can produce massive retinal hemorrhage ("blood and thunder"), ischemia, and vision loss.

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. 28-9**). 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 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.

About 5% of patients, especially Caucasian females aged >60, develop the arteritic form of AION in conjunction with giant-cell (temporal) arteritis (**Chap. 356**). 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 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



FIGURE 28-7 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.



FIGURE 28-9 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.

examine a sufficient number of tissue sections. The histological 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. 28-10), although optic disc pallor slowly developed over subsequent months.

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.

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. 436).

A particularly severe form of optic neuritis occurs in neuromyelitis optica (NMO); it is typically longitudinally extensive, and may be bilateral or 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 is diagnostic. Treatment for acute episodes consists of glucocorticoids and, in resistant cases, plasma exchange. **Neuromyelitis optica** is discussed in detail in Chap. 437.

■ LEBER'S HEREDITARY OPTIC NEUROPATHY

This disease usually affects young men, causing gradual, 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 inherited from the mother by all her children, but for unknown reasons, daughters are rarely affected. Early stage clinical trials of gene therapy for this condition are in progress.

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. 28-11) 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, or alcoholism can lead to insidious visual loss. Thiamine, vitamin B₁₂, and folate levels should be checked in any patient with unexplained bilateral central scotomas and optic pallor.

Papilledema This connotes bilateral optic disc swelling from raised intracranial pressure (Fig. 28-12). 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 can 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. 28-3F). With unremitting papilledema, peripheral visual field loss



FIGURE 28-10 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.

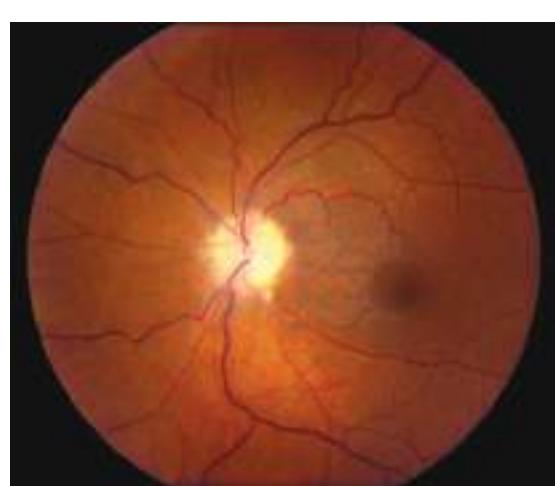


FIGURE 28-11 Optic atrophy is not a specific diagnosis but refers to the combination of optic disc pallor, arteriolar narrowing, and nerve fiber layer destruction produced by a host of eye diseases, especially optic neuropathies.



FIGURE 28-12 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.

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.

Evaluation of papilledema requires neuroimaging to exclude an intracranial lesion. MR angiography is appropriate in selected cases to search for a dural venous sinus occlusion 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. Optic nerve sheath fenestration is less efficacious, and does not address other neurological symptoms. Occasionally, fulminant papilledema produces rapid onset of blindness. In such patients, emergency surgery should be performed to install a shunt without delay.

Optic Disc Drusen These are refractile deposits within the substance of the optic nerve head (**Fig. 28-13**). They are unrelated to



FIGURE 28-13 Optic disc drusen are calcified, mulberry-like deposits of unknown etiology within the optic disc, giving rise to “pseudopapilledema.”

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 as glittering particles 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.

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. 28-14**). 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

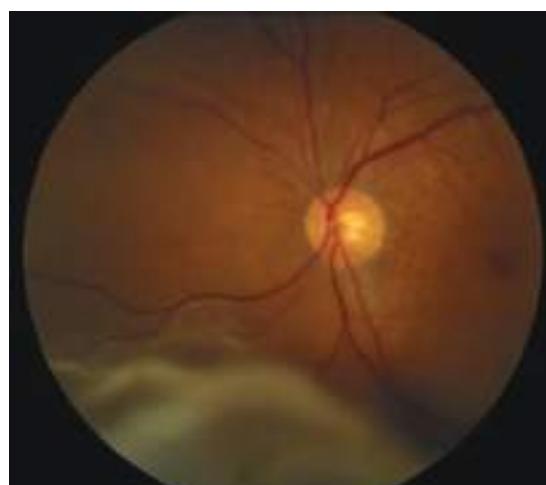


FIGURE 28-14 Retinal detachment appears as an elevated sheet of retinal tissue with folds. In this patient, the fovea was spared, so acuity was normal, but an inferior detachment produced a superior scotoma.

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.

Classic Migraine (See also Chap. 422) 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 Vertebralbasilar 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 brain-stem 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 vertebralbasilar 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 malingers. 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

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.

Glucoma 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 or arcuate scotomas on perimetric testing. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges (Fig. 28-15). 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.

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



FIGURE 28-15 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.

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 non-exudative (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. 28-16). 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.

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, or aflibercept 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

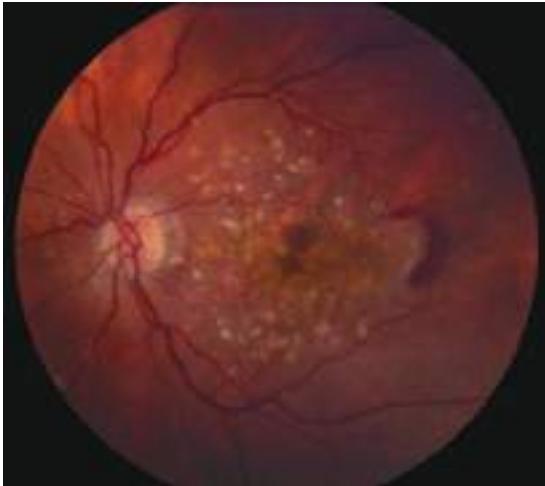


FIGURE 28-16 Age-related macular degeneration consisting of scattered yellow drusen in the macula (dry form) and a crescent of fresh hemorrhage temporal to the fovea from a subretinal neovascular membrane (wet form).

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. 28-17). 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. 396–398.](#)

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. 28-18). The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Most cases are due to 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.

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

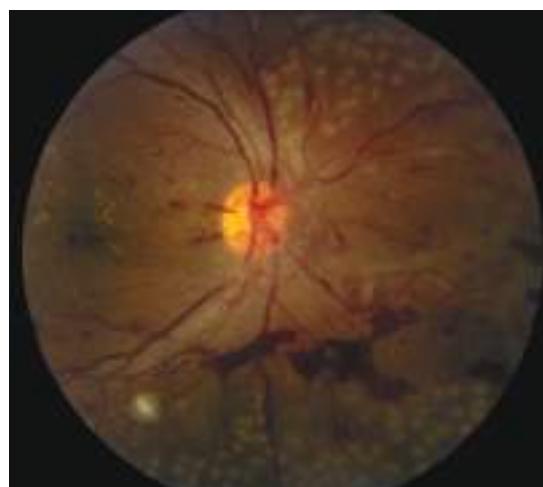


FIGURE 28-17 Proliferative diabetic retinopathy in a 25-year-old man with an 18-year history of diabetes, showing neovascular vessels emanating from the optic disc, retinal and vitreous hemorrhage, cotton-wool spots, and macular exudate. Round spots in the periphery represent recently applied panretinal photocoagulation.



FIGURE 28-18 Retinitis pigmentosa with black clumps of pigment known as “bone spicules.” The patient had peripheral visual field loss with sparing of central (macular) vision.

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. 28-19). 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. Treatment of



FIGURE 28-19 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.

melanoma is controversial. Options include enucleation, local resection, and irradiation. *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.

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 opto-ciliary 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. 118). 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 a 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. 375). 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. Graves’ ophthalmopathy can be treated with oral prednisone (60 mg/d) for 1 month, followed by a taper over several months. Worsening of symptoms upon glucocorticoid withdrawal is common. Topical lubricants, taping the eyelids closed at night, moisture chambers, and eyelid surgery are helpful to limit exposure of ocular tissues. 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. 433**) 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.

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. 440**) 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. In the above example, a cover test with the head turned to the right 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.

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. 440) 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 measurements of a variable, fatigable ptosis, often accompanied by diplopia, are helpful to establish the diagnosis. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. The diagnosis can be confirmed by an IV edrophonium injection, which produces a transient reversal of eyelid or eye muscle weakness. 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. With improvement in the resolution of these non-invasive techniques, 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.

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. The vertical diplopia also 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.

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 is thought to impinge 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., chorodoma, 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 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. 28-20). 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.

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. 19). 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

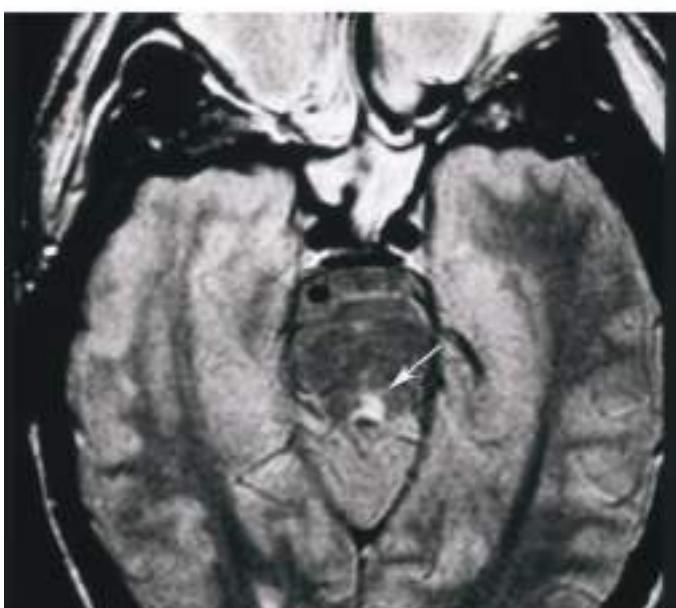
**A****B****C****D**

FIGURE 28-20 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).

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.

DOWNSBEAT 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.

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

- BAINBRIDGE JW et al: Long-term effect of gene therapy on Leber's congenital amaurosis. *N Engl J Med* 372:1887, 2015.
- BUTTIGEREI TF et al: Polymyalgia rheumatica and giant cell arteritis. *JAMA* 315:2442, 2016.
- CAMPOCHIARO PA et al: Anti-vascular endothelial growth factor agents in the treatment of retinal disease. *Ophthalmology* 123:S78, 2016.
- GROSS JG et al: Panretinal photocoagulation vs intravitreous ranibizumab for proliferative diabetic retinopathy. *JAMA* 314:2137, 2015
- JAFFE GJ et al: Adalimumab in patients with active noninfectious uveitis. *N Engl J Med* 375:932, 2016
- PEARSON RA et al: Donor and host photoreceptors engage in material transfer following transplantation of post-mitotic photoreceptor precursors. *Nat Commun* 7:13029, 2016.
- STONE JH et al: Trial of tocilizumab in giant-cell arteritis. *N Engl J Med* 377:317, 2017.
- WALL M et al: Effect of acetazolamide on visual function in patients with idiopathic intracranial hypertension and mild visual loss: The idiopathic intracranial hypertension treatment trial. *JAMA* 311:1641, 2014.
- WILLIAMS PA et al: Vitamin B3 modulates mitochondrial vulnerability and prevents glaucoma in aged mice. *Science* 355:756, 2017.
- YANOFF M, DUKER J: *Ophthalmology*, 4th ed. Atlanta, Georgia, Saunders, 2014.

Richard L. Doty, Steven M. Bromley

All environmental chemicals necessary for life enter the body by the nose and mouth. The senses of smell (olfaction) and taste (gustation) monitor such chemicals, determine the flavor and palatability of foods and beverages, and warn of dangerous environmental conditions, including fire, air pollution, leaking natural gas, and bacteria-laden foodstuffs. These senses contribute significantly to quality of life and, when dysfunctional, can have untoward physical and psychological consequences. Indeed, a recent longitudinal study of 1162 non-demented elderly persons found, even after controlling for confounders, that those with the lowest baseline olfactory test scores had a 45% mortality rate over a 4-year period, compared to an 18% mortality rate for those with the highest olfactory test scores. A basic understanding of these senses in health and disease is critical for the physician, because thousands of patients present to doctors' offices each year with complaints of chemosensory dysfunction. Among the more important recent developments in neurology is the discovery that decreased smell function is among the first signs, if not the first sign, of such neurodegenerative diseases as Parkinson's disease (PD) and Alzheimer's disease (AD), signifying their "presymptomatic" phase.

ANATOMY AND PHYSIOLOGY

Olfactory System Odorous chemicals enter the front of nose during inhalation and active sniffing, as well as the back of the nose (nasopharynx) during deglutition. After reaching the highest recesses of the nasal cavity, they dissolve in the olfactory mucus and diffuse or are actively transported by specialized proteins to receptors located on the cilia of olfactory receptor cells. The cilia, dendrites, cell bodies, and proximal axonal segments of these bipolar cells are located within a unique neuroepithelium covering the cribriform plate, the superior nasal septum, superior turbinate, and sectors of the middle turbinate (Fig. 29-1). Nearly 400 types of G-protein-coupled odor receptors (GPCRs) are expressed on the cilia of the receptor cells, with only one type of GPCR receptor being expressed on a given cell. Other receptors, including trace amine-associated receptors and members of the non-GPCR membrane-spanning 4-domain family, subfamily A (MS4A) protein family, are also present on some receptor cells. Such a plethora of receptor cell types does not exist in any other sensory system. Importantly, when damaged, the receptor cells can be replaced by stem

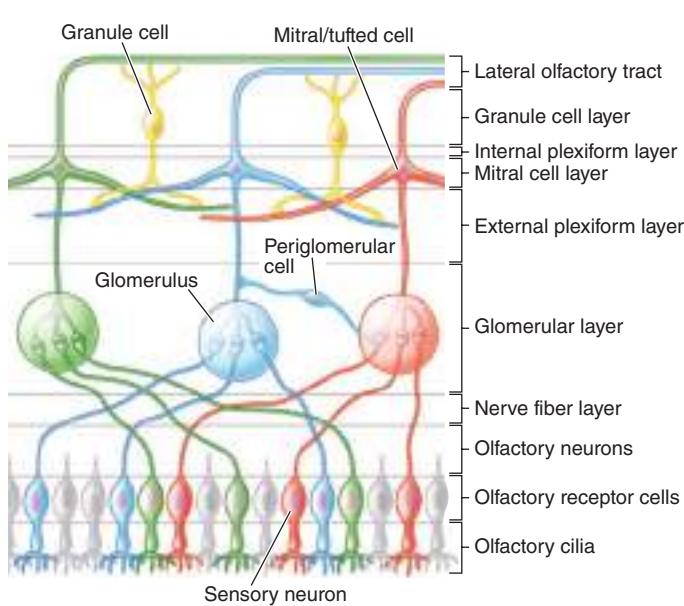


FIGURE 29-2 Schematic of the layers and wiring of the olfactory bulb. Each receptor type (red, green, blue) projects to a common glomerulus. The neural activity within each glomerulus is modulated by periglomerular cells. The activity of the primary projection cells, the mitral and tufted cells, is modulated by granule cells, periglomerular cells, and secondary dendrites from adjacent mitral and tufted cells. (From www.med.yale.edu/neurosurg/treloar/index.html.)

cells near the basement membrane, although such replacement is often incomplete.

After coalescing into bundles surrounded by glia-like ensheathing cells (termed fila), the receptor cell axons pass through the cribriform plate to the olfactory bulbs, where they synapse with dendrites of other cell types within the glomeruli (Fig. 29-2). These spherical structures, which make up a distinct layer of the olfactory bulb, are a site of convergence of information, because many more fibers enter than leave them. Receptor cells that express the same type of receptor project to the same glomeruli, effectively making each glomerulus a functional unit. The major projection neurons of the olfactory system—the mitral and tufted cells—send primary dendrites into the glomeruli, connecting not only with the incoming receptor cell axons, but with dendrites of periglomerular cells. The activity of the mitral/tufted cells is modulated by the periglomerular cells, secondary dendrites from other mitral/tufted cells, and granule cells, the most numerous cells of the bulb. The latter cells, which are largely GABAergic, receive inputs from central brain structures and modulate the output of the mitral/

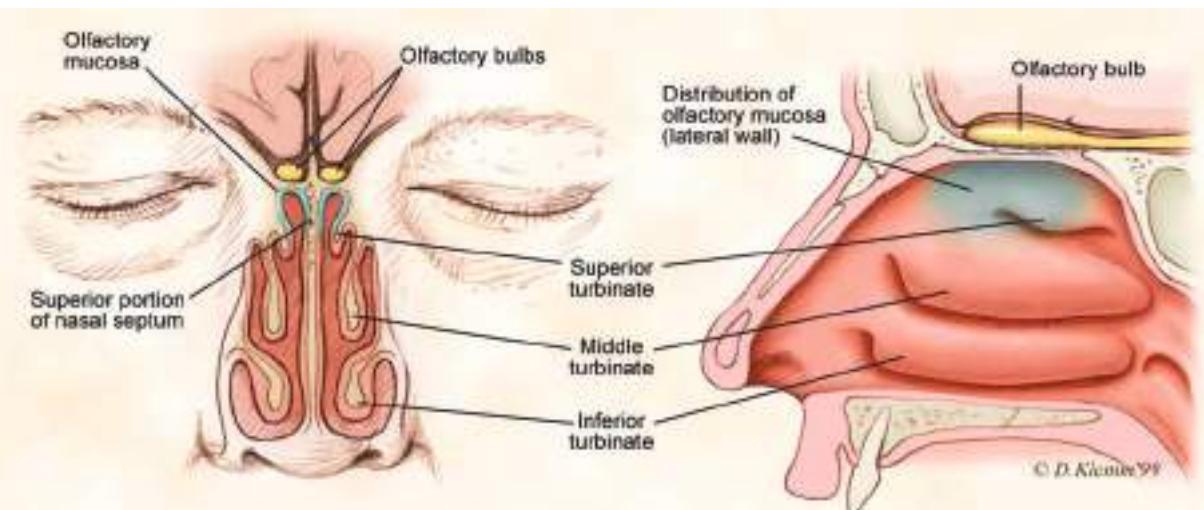


FIGURE 29-1 Anatomy of the nose, showing the distribution of olfactory receptors in the roof of the nasal cavity. (Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center; used with permission.)

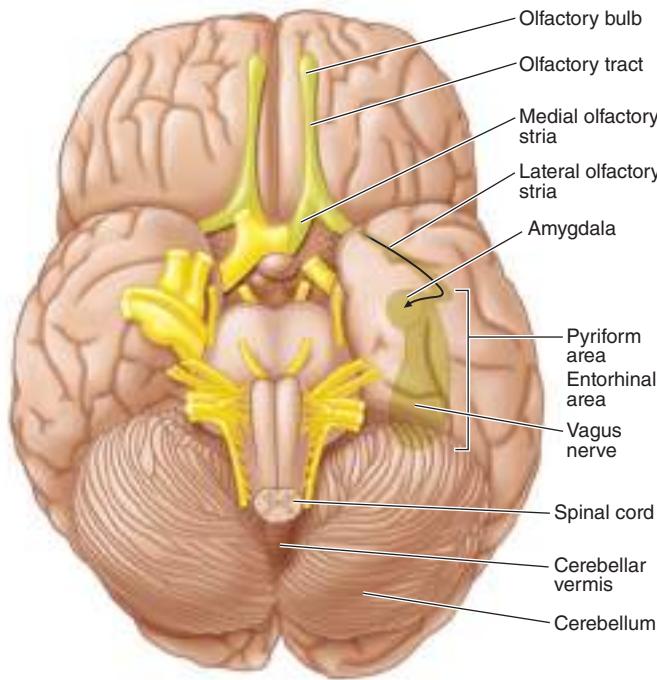


FIGURE 29-3 Anatomy of the base of the brain showing the primary olfactory cortex.

tufted cells. Interestingly, like the olfactory receptor cells, some cells within the bulb undergo replacement. Thus, neuroblasts formed within the anterior subventricular zone of the brain migrate along the rostral migratory stream, ultimately becoming granule and periglomerular cells.

The axons of the mitral and tufted cells synapse within secondary olfactory structures which largely comprise the primary olfactory cortex (POC) (Fig. 29-3). The POC is defined as those cortical structures that receive direct projections from the olfactory bulb, most notably the piriform and entorhinal cortices. Although olfaction is unique in

that its initial afferent projections bypass the thalamus, persons with damage to the thalamus can exhibit olfactory deficits, particularly ones of odor identification. Such deficits likely reflect the involvement of thalamic connections between the POC and the orbitofrontal cortex (OFC), where odor identification largely occurs. The close anatomic ties between the olfactory system and the amygdala, hippocampus, and hypothalamus help to explain the intimate associations between odor perception and cognitive functions such as memory, motivation, arousal, autonomic activity, digestion, and sex.

Taste System Tastants are sensed by specialized receptor cells present within taste buds—small grapefruit-like segmented structures located on the lateral margins and dorsum of the tongue, roof of the mouth, pharynx, larynx, and superior esophagus (Fig. 29-4). Lingual taste buds are embedded in well-defined protuberances, termed fungiform, foliate, and circumvallate papillae. After dissolving in a liquid, tastants enter the opening of the taste bud—the taste pore—and bind to receptors on microvilli, small extensions of receptor cells within each taste bud. Such binding changes the electrical potential across the taste cell, resulting in neurotransmitter release onto the first-order taste neurons. Although humans have ~7500 taste buds, not all harbor taste-sensitive cells; some contain only one class of receptor (e.g., cells responsive only to sugars), whereas others contain cells sensitive to more than one class. The number of taste receptor cells per taste bud ranges from zero to well over 100. A small family of three G-protein-coupled receptors (GPCRs), namely T1R1, T1R2, and T1R3, mediate sweet and umami taste sensations. Bitter sensations, on the other hand, depend on T2R receptors, a family of ~30 GPCRs expressed on cells different from those that express the sweet and umami receptors. T2Rs sense a wide range of bitter substances but do not distinguish among them. Sour tastants are sensed by the PKD2L1 receptor, a member of the transient receptor potential protein (TRP) family. Perception of salty sensations, such as induced by sodium chloride, arises from the entry of Na^+ ions into the cells via specialized membrane channels, such as the amiloride-sensitive Na^+ channel.

It is now well established that both bitter and sweet taste-related receptors are also present elsewhere in the body, most notably in the alimentary and respiratory tracts. This important discovery generalizes the concept of taste-related chemoreception to areas of the body

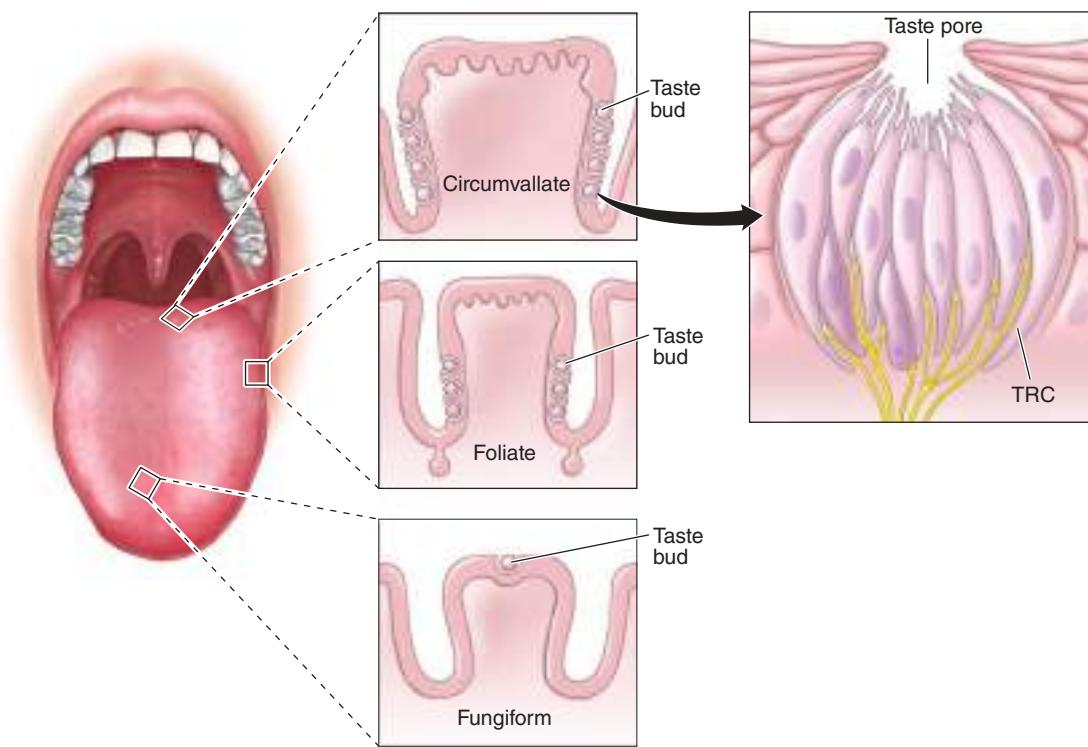


FIGURE 29-4 Schematic of the taste bud and its opening (pore), as well as the location of buds on the three major types of papillae: Fungiform (anterior), foliate (lateral), and circumvallate (posterior).

beyond the mouth and throat, with α -gustducin, the taste-specific G-protein α -subunit, expressed in so-called brush cells found specifically within the human trachea, lung, pancreas, and gallbladder. These brush cells are rich in nitric oxide (NO) synthase, known to defend against xenobiotic organisms, protect the mucosa from acid-induced lesions, and, in the case of the gastrointestinal tract, stimulate vagal and splanchnic afferent neurons. NO further acts on nearby cells, including enteroendocrine cells, absorptive or secretory epithelial cells, mucosal blood vessels, and cells of the immune system. Members of the T2R family of bitter receptors and the sweet receptors of the T1R family have been identified within the gastrointestinal tract and in enteroendocrine cell lines. In some cases, these receptors are important for metabolism, with the T1R3 receptors and gustducin playing decisive roles in the sensing and transport of dietary sugars from the intestinal lumen into absorptive enterocytes via a sodium-dependent glucose transporter and in regulation of hormone release from gut enteroendocrine cells. In other cases, these receptors may be important for airway protection, with a number of T2R bitter receptors in the motile cilia of the human airway that responded to bitter compounds by increasing their beat frequency. One specific T2R38 taste receptor is expressed in human upper respiratory epithelia and responds to acyl-monoserine lactone quorum-sensing molecules secreted by *Pseudomonas aeruginosa* and other gram-negative bacteria. Differences in T2R38 functionality, as related to TAS2R38 genotype, correlate with susceptibility to upper respiratory infections in humans.

Taste information is sent to the brain via three cranial nerves (CNs): CN VII (the *facial nerve*, which involves the intermediate nerve with its branches, the greater petrosal and chorda tympani nerves), CN IX (the *glossopharyngeal nerve*), and CN X (the *vagus nerve*) (Fig. 29-5). CN VII innervates the anterior tongue and all of the soft palate, CN IX innervates the posterior tongue, and CN X innervates the laryngeal surface of the epiglottis, larynx, and proximal portion of the esophagus. The mandibular branch of CN V (V_3) conveys somatosensory information (e.g., touch, burning, cooling, irritation) to the brain. Although not technically a gustatory nerve, CN V shares primary nerve routes with many of the gustatory nerve fibers and adds temperature, texture, pungency, and spiciness to the taste experience. The chorda tympani nerve is famous for taking a recurrent course through the facial canal in the petrosal portion of the temporal bone, passing through the middle ear, and then exiting the skull via the petrotympanic fissure, where it joins the lingual nerve (a division of CN V) near the tongue. This nerve also

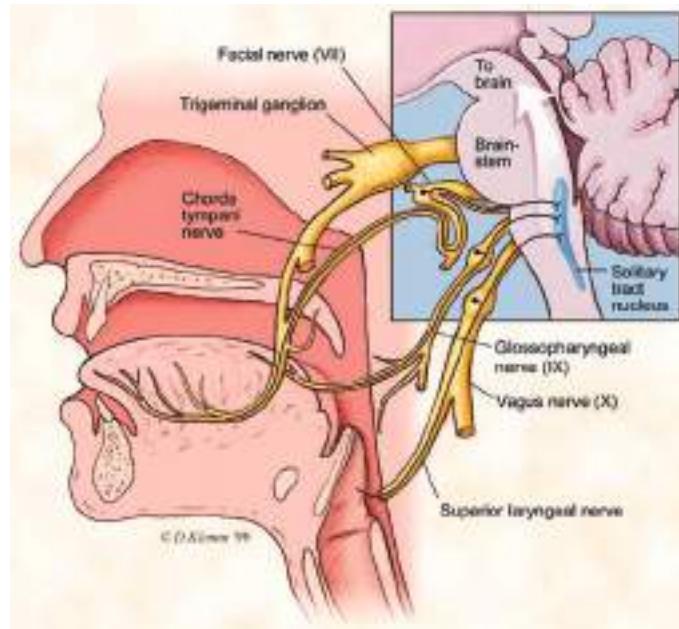


FIGURE 29-5 Schematic of the cranial nerves (CNs) that mediate taste function, including the chorda tympani nerve (CN VII), the glossopharyngeal nerve (CN IX), and the vagus nerve (CN X). (Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center; used with permission.)

carries parasympathetic fibers to the submandibular and sublingual glands, whereas the greater petrosal nerve supplies the palatine glands, thereby influencing saliva production.

The axons of the projection cells, which synapse with taste buds, enter the rostral portion of the nucleus of the solitary tract (NTS) within the medulla of the brainstem (Fig. 29-5). From the NTS, neurons then project to a division of the ventroposteromedial thalamic nucleus (VPM) via the medial lemniscus. From here, projections are made to the rostral part of the frontal operculum and adjoining insula, a brain region considered the *primary taste cortex* (PTC). Projections from the PTC then go to the *secondary taste cortex*, namely the caudolateral OFC. This brain region is involved in the conscious recognition of taste qualities. Moreover, because it contains cells that are activated by several sensory modalities, it is likely a center for establishing “flavor.”

DISORDERS OF OLFACTION

The ability to smell is influenced, in everyday life, by such factors as age, gender, general health, nutrition, smoking, and reproductive state. Women typically outperform men on tests of olfactory function and retain normal smell function to a later age than do men.

Estimates of the prevalence of olfactory dysfunction in the general population vary; a recent cross-sectional analysis from the National Health and Nutrition Examination Survey (NHANES 2013–2014) found an overall prevalence of 13.5%. However, it is apparent that significant decrements in the ability to smell are present in >50% of the population between 65 and 80 years of age and in 75% of those aged ≥ 80 years (Fig. 29-6). Such presbyosmia helps to explain why many elderly report that food has little flavor, a problem that can result in nutritional disturbances. This also helps to explain why a disproportionate number of elderly die in accidental gas poisonings. A relatively complete listing of conditions and disorders that have been associated with olfactory dysfunction is presented in Table 29-1.

Aside from aging, the three most common identifiable causes of long-lasting or permanent smell loss seen in the clinic are, in order of frequency, severe upper respiratory infections, head trauma, and chronic rhinosinusitis. The physiologic basis for most head trauma-related losses is the shearing and subsequent scarring of the olfactory fila as they pass from the nasal cavity into the brain cavity. The cribriform plate does not have to be fractured or show pathology for smell loss to be present. Severity of trauma, as indexed by a poor Glasgow Coma Scale score on presentation and the length of posttraumatic amnesia, is associated with higher risk of olfactory impairment. Less than 10% of posttraumatic anosmic patients will recover age-related normal function over time. This increases to nearly 25% of those with less-than-total loss. Upper respiratory infections, such as those

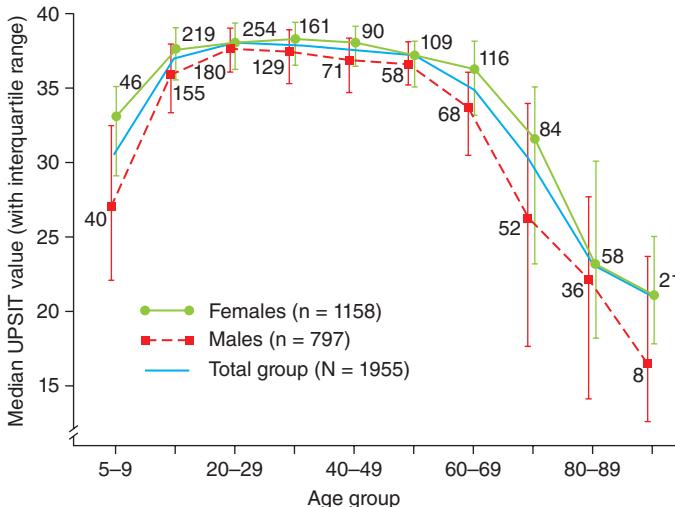


FIGURE 29-6 Scores on the University of Pennsylvania Smell Identification Test (UPSIT) as a function of subject age and sex. Numbers by each data point indicate sample sizes. Note that women identify odorants better than men at all ages. (From RL Doty et al: Science 226:1421, 1984. Copyright © 1984 American Association for the Advancement of Science.)

TABLE 29-1 Disorders and Conditions Associated with Compromised Olfactory Function, as Measured by Olfactory Testing

22q11 deletion syndrome	Korsakoff's psychosis
AIDS/HIV infection	Laryngopharyngeal reflux disease
Adenoid hypertrophy	Legionnaires' disease
Adrenal cortical insufficiency	Leprosy
Age	Liver disease
Alcoholism	Lubag disease
Allergies	Medications
Alzheimer's disease	Migraine
Amyotrophic lateral sclerosis (ALS)	Multiple sclerosis
Anorexia nervosa	Multi-infarct dementia
Asperger's syndrome	Myasthenia gravis
Ataxias	Narcolepsy with cataplexy
Attention deficit/hyperactivity disorder	Neoplasms, cranial/nasal
Behcet's disease	Nutritional deficiencies
Bardet-Biedl syndrome	Obstructive pulmonary disease
Chagas' disease	Obesity
Chemical exposure	Obsessive compulsive disorder
Chronic obstructive pulmonary disease	Orthostatic tremor
Congenital	Panic disorder
Cushing's syndrome	Parkinson's disease (PD)
Cystic fibrosis	Pick's disease
Degenerative ataxias	Posttraumatic stress disorder
Depression	Pregnancy
Diabetes	Pseudohypoparathyroidism
Down's syndrome	Psychopathy
Epilepsy	Radiation (therapeutic, cranial)
Facial paralysis	REM behavior disorder
Fibromyalgia	Refsum's disease
Frontotemporal lobe degeneration	Renal failure/end-stage kidney disease
Gonadal dysgenesis (Turner's syndrome)	Restless leg syndrome
Granulomatosis with Polyangiitis (Wegener's)	Rhinosinusitis/polyposis
Guamanian ALS/PD/dementia syndrome	Schizophrenia
Head trauma	Seasonal affective disorder
Herpes simplex encephalitis	Sjögren's syndrome
Hypothyroidism	Stroke
Huntington's disease	Systemic sclerosis
Iatrogenesis	Tobacco smoking
Idiopathic inflammatory myopathies	Toxic chemical exposure
Kallmann's syndrome	Upper respiratory infections
	Usher syndrome
	Vitamin B ₁₂ deficiency

associated with the common cold, influenza, pneumonia, or HIV, can directly and permanently harm the olfactory epithelium by decreasing receptor cell number, damaging cilia on remaining receptor cells, and inducing the replacement of sensory epithelium with respiratory epithelium. The smell loss associated with chronic rhinosinusitis is related to disease severity, with most loss occurring in cases where rhinosinusitis and polyposis are both present. Although systemic glucocorticoid therapy can usually induce short-term functional improvement, it does not, on average, return smell test scores to normal, implying that chronic permanent neural loss is present and/or that short-term administration of systemic glucocorticoids does not completely mitigate the inflammation. It is well established that microinflammation in an otherwise seemingly normal epithelium can influence smell function.

A number of neurodegenerative diseases are accompanied by olfactory impairment, including PD, AD, Huntington's disease, parkinsonism-dementia complex of Guam, dementia with Lewy bodies (DLB), multiple system atrophy, corticobasal degeneration, frontotemporal

dementia, and Down's syndrome; smell loss can also occur in idiopathic rapid eye movement (REM) behavioral sleep disorder (iRBD), as well as in multiple sclerosis (MS) related to lesions within olfaction-related structures. Olfactory impairment in PD often predates the clinical diagnosis by a number of years. In staged cases, studies of the sequence of formation of abnormal α -synuclein aggregates and Lewy bodies suggest that the olfactory bulbs may be, along with the dorsomotor nucleus of the vagus, the first site of neural damage in PD. In postmortem studies of patients with very mild "presymptomatic" signs of AD, poorer smell function has been associated with higher levels of AD-related pathology. Smell loss is more marked in patients with early clinical manifestations of DLB than in those with mild AD. Interestingly, smell loss is minimal or nonexistent in progressive supranuclear palsy and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism. The relative contributions of disease-specific pathology or differential damage to forebrain neuromodulator/neurotransmitter systems in explaining different degrees of olfactory dysfunction among the various neurodegenerative diseases are presently unknown.

The smell loss seen in iRBD is of the same magnitude as that found in PD. This is of particular interest because patients with iRBD frequently develop PD and hyposmia. REM behavior disorder is not only seen in its idiopathic form, but can also be associated with narcolepsy (**Chap. 27**). A study of narcoleptic patients with and without REM behavior disorder demonstrated that narcolepsy, independent of REM behavior disorder, was associated with impairments in olfactory function. Loss of hypothalamic neurons expressing orexin (also known as hypocretin) neuropeptides is believed to be responsible for narcolepsy and cataplexy. Orexin-containing neurons project throughout the entire olfactory system (from the olfactory epithelium to the olfactory cortex), and damage to these projections may be one underlying mechanism for impaired olfactory performance in narcoleptic patients. Administration of intranasal orexin A (hypocretin-1) improved olfactory function, supporting the notion that mild olfactory impairment is not only a primary feature of narcolepsy with cataplexy, but that orexin deficiency may be directly responsible for the loss of smell in this condition.

■ DISORDERS OF TASTE

The majority of patients who present with taste dysfunction exhibit olfactory, not taste, loss. This is because most flavors attributed to taste actually depend on retronasal stimulation of the olfactory receptors during deglutition. As noted earlier, taste buds only mediate basic tastes such as sweet, sour, bitter, salty, and umami. Significant impairment of whole-mouth gustatory function is rare outside of generalized metabolic disturbances or systemic use of some medications, because taste bud regeneration occurs and peripheral damage alone would require the involvement of multiple CN pathways. Taste function can be influenced by age, diet, smoking behavior, use of medications, and other subject-related factors including (1) the release of foul-tasting materials from the oral cavity from oral medical conditions (e.g., gingivitis, purulent sialadenitis) or appliances; (2) transport problems of tastants to the taste buds (e.g., drying, infections, or inflammatory conditions of the orolinguinal mucosa), (3) damage to the taste buds themselves (e.g., local trauma, invasive carcinomas), (4) damage to the neural pathways innervating the taste buds (e.g., middle ear infections), (5) damage to central structures (e.g., multiple sclerosis, tumor, epilepsy, stroke), and (6) systemic disturbances of metabolism (e.g., diabetes, thyroid disease, medications). Unlike CN VII, CN IX is relatively protected along its path, although iatrogenic interventions such as tonsillectomy, bronchoscopy, laryngoscopy, endotracheal intubation, and radiation therapy can result in selective injury. CN VII damage commonly results from mastoidectomy, tympanoplasty, and stapedectomy, in some cases inducing persistent metallic sensations. Bell's palsy (**Chap. 433**) is one of the most common causes of CN VII injury that results in taste disturbance. On rare occasions, migraine (**Chap. 422**) is associated with a gustatory prodrome or aura, and in some cases tastants can trigger a migraine attack. Interestingly, dysgeusia occurs in some cases of *burning mouth syndrome* (also termed *glossodynia* or *glossalgia*), as does dry mouth and thirst. Burning mouth syndrome is likely associated with dysfunction of the trigeminal nerve (CN V).

Some of the etiologies suggested for this poorly understood syndrome are amenable to treatment, including (1) nutritional deficiencies (e.g., iron, folic acid, B vitamins, zinc), (2) diabetes mellitus (possibly predisposing to oral candidiasis), (3) denture allergy, (4) mechanical irritation from dentures or oral devices, (5) repetitive movements of the mouth (e.g., tongue thrusting, teeth grinding, jaw clenching), (6) tongue ischemia as a result of temporal arteritis, (7) periodontal disease, (8) reflux esophagitis, and (9) geographic tongue.

Although both taste and smell can be adversely influenced by drugs, taste alterations are more common. Indeed, over 250 medications have been reported to alter the ability to taste. Major offenders include antineoplastic agents, antirheumatic drugs, antibiotics, and blood pressure medications. Terbinafine, a commonly used antifungal, has been linked to taste disturbance lasting up to 3 years. In a recent controlled trial, nearly two-thirds of individuals taking eszopiclone (Lunesta) experienced a bitter dysgeusia that was stronger in women, systematically related to the time since drug administration, and positively correlated with both blood and saliva levels of the drug. Intranasal use of nasal gels and sprays containing zinc, which are common over-the-counter prophylactics for upper respiratory viral infections, has been implicated in loss of smell function. Whether their efficacy in preventing such infections, which are the most common cause of anosmia and hyposmia, outweighs their potential detriment to smell function requires study. Dysgeusia occurs commonly in the context of drugs used to treat or minimize symptoms of cancer, with a weighted prevalence from 56 to 76% depending on the type of cancer treatment. Attempts to prevent taste problems from such drugs using prophylactic zinc sulfate or amifostine have proven to be minimally beneficial. Although antiepileptic medications are occasionally used to treat smell or taste disturbances, the use of topiramate has been reported to result in a reversible loss of an ability to detect and recognize tastes and odors during treatment.

As with olfaction, a number of systemic disorders can affect taste. These include, but are not limited to, chronic renal failure, end-stage liver disease, vitamin and mineral deficiencies, diabetes mellitus, and hypothyroidism. In diabetes, there appears to be a progressive loss of taste beginning with glucose and then extending to other sweeteners, salty stimuli, and then all stimuli. Psychiatric conditions can be associated with chemosensory alterations (e.g., depression, schizophrenia, bulimia). A recent review of tactile, gustatory, and olfactory hallucinations demonstrated that no one type of hallucinatory experience is pathognomonic to any given diagnosis.

Pregnancy is a unique condition with regard to taste function. There appears to be an increase in dislike and intensity of bitter tastes during the first trimester that may help to ensure that pregnant women avoid poisons during a critical phase of fetal development. Similarly, a relative increase in the preference for salt and bitter in the second and third trimesters may support the ingestion of much needed electrolytes to expand fluid volume and support a varied diet.

CLINICAL EVALUATION

In most cases, a careful clinical history will establish the probable etiology of a chemosensory problem, including questions about its nature, onset, duration, and pattern of fluctuations. *Sudden loss* suggests the possibility of head trauma, ischemia, infection, or a psychiatric condition. *Gradual loss* can reflect the development of a progressive obstructive lesion, although gradual loss can also follow head trauma. *Intermittent loss* suggests the likelihood of an inflammatory process. The patient should be asked about potential precipitating events, such as cold or flu infections prior to symptom onset, because these often go underappreciated. Information regarding head trauma, smoking habits, drug and alcohol abuse (e.g., intranasal cocaine, chronic alcoholism), exposures to pesticides and other toxic agents, and medical interventions is also informative. A determination of all the medications that the patient was taking before and at the time of symptom onset is important, because many can cause chemosensory disturbances. Comorbid medical conditions associated with smell impairment, such as renal failure, liver disease, hypothyroidism, diabetes, or dementia, should be assessed. Delayed puberty in association with anosmia (with

or without midline craniofacial abnormalities, deafness, and renal anomalies) suggests the possibility of Kallmann's syndrome. Recollection of epistaxis, discharge (clear, purulent, or bloody), nasal obstruction, allergies, and somatic symptoms, including headache or irritation, may have localizing value. Questions related to memory, parkinsonian symptoms, and seizure activity (e.g., automatisms, blackouts, auras, *déjà vu*) should be posed. Pending litigation and the possibility of malingering should be considered. Modern forced-choice olfactory tests can detect malingering from improbable responses.

Neurologic and otorhinolaryngologic (ORL) examinations, along with appropriate brain and nasosinus imaging, aid in the evaluation of patients with olfactory or gustatory complaints. The neural evaluation should focus on CN function, with particular attention to possible skull base and intracranial lesions. Visual acuity, field, and optic disc examinations aid in the detection of intracranial mass lesions that produce raised intracranial pressure (papilledema) and optic atrophy. Foster Kennedy syndrome refers to raised intracranial pressure plus a compressive optic neuropathy; typical causes are olfactory groove meningiomas or other frontal lobe tumors. The ORL examination should thoroughly assess the intranasal architecture and mucosal surfaces. Polyps, masses, and adhesions of the turbinates to the septum may compromise the flow of air to the olfactory receptors, because less than a fifth of the inspired air traverses the olfactory cleft in the unobstructed state. Blood tests may be helpful to identify such conditions as diabetes, infection, heavy metal exposure, nutritional deficiency (e.g., vitamin B₆ or B₁₂), allergy, and thyroid, liver, and kidney disease.

As with other sensory disorders, quantitative sensory testing is advised. Self-reports of patients can be misleading, and a number of patients who complain of chemosensory dysfunction have normal function for their age and gender. Quantitative smell and taste testing provides objective information for worker's compensation and other legal claims, as well as a way to accurately assess the effects of treatment interventions. A number of standardized olfactory and taste tests are commercially available. The most widely used of these tests, the 40-item University of Pennsylvania Smell Identification Test (UPSIT), uses norms based on nearly 4000 normal subjects. A determination is made of both absolute dysfunction (i.e., mild loss, moderate loss, severe loss, total loss, probable malingering) and relative dysfunction (percentile rank for age and gender). Although electrophysiologic testing is available at some smell and taste centers (e.g., odor event-related potentials), they require complex stimulus presentation and recording equipment and rarely provide additional diagnostic information. With the exception of electrogustometers, commercially available taste tests have only recently become available. Most use filter paper strips impregnated with tastants, so no stimulus preparation is required.

TREATMENT AND MANAGEMENT

Given the various mechanisms by which olfactory and gustatory disturbance can occur, management of patients tends to be condition-specific. For example, patients with hypothyroidism, diabetes, or infections often benefit from specific treatments to correct the underlying disease process that is adversely influencing chemoreception. For most patients who present primarily with obstructive/transport loss affecting the nasal and paranasal regions (e.g., allergic rhinitis, polyposis, intranasal neoplasms, nasal deviations), medical and/or surgical intervention is often beneficial. Antifungal and antibiotic treatments may reverse taste problems secondary to candidiasis or other oral infections. Chlorhexidine mouthwash mitigates some salty or bitter dysgeusias, conceivably as a result of its strong positive charge. Excessive dryness of the oral mucosa is a problem with many medications and conditions, and artificial saliva (e.g., Xerolube) or oral pilocarpine treatments may prove beneficial. Other methods to improve salivary flow include the use of mints, lozenges, or sugarless gum. Flavor enhancers may make food more palatable (e.g., monosodium glutamate), but caution is advised to avoid overusing ingredients containing sodium or sugar, particularly in circumstances when a patient also has underlying hypertension or diabetes. Medications that induce distortions of taste can often be discontinued and replaced with other types of medications or modes of therapy. As mentioned earlier, pharmacologic agents result

in taste disturbances much more frequently than smell disturbances. It is important to note, however, that many drug-related effects are long lasting and not reversed by short-term drug discontinuance.

A recent study of endoscopic sinus surgery in patients with chronic rhinosinusitis and hyposmia revealed that patients with severe olfactory dysfunction prior to the surgery had a more dramatic and sustained improvement over time compared to patients with more mild olfactory dysfunction prior to intervention. In the case of intranasal and sinus-related inflammatory conditions, such as seen with allergy, viruses, and traumas, the use of intranasal or systemic glucocorticoids may also be helpful. One common approach is to use a tapering course of oral prednisone. Topical intranasal administration of glucocorticoids was found to be less effective in general than systemic administration, however the effects of different nasal administration techniques were not analyzed; for example, intranasal glucocorticoids are more effective if administered in the Moffett's position (head in the inverted position such as over the edge of the bed with the bridge of the nose perpendicular to the floor). After head trauma, an initial trial of glucocorticoids may help to reduce local edema and the potential deleterious deposition of scar tissue around olfactory fila at the level of the cribriform plate.

Treatments are limited for patients with chemosensory loss or primary injury to neural pathways. Nonetheless, spontaneous recovery can occur. In a follow-up study of 542 patients presenting to our center with smell loss from a variety of causes, modest improvement occurred over an average time period of 4 years in about half of the participants. However, only 11% of the anosmic and 23% of the hyposmic patients regained normal age-related function. Interestingly, the amount of dysfunction present at the time of presentation, not etiology, was the best predictor of prognosis. Other predictors were age and the duration of dysfunction prior to initial testing.

Several studies have reported that patients with hyposmia may benefit from repeated smelling of odors over the course of weeks or months. The usual paradigm is to smell odors such as eucalyptol, citronella, eugenol, and phenyl ethyl alcohol before going to bed and immediately upon awakening each day. The rationale for such an approach comes from animal studies demonstrating that prolonged exposure to odorants can induce increased neural activity within the olfactory bulb. There is also limited evidence that α -lipoic acid (400 mg/d), an essential cofactor for many enzyme complexes with possible antioxidant effects, may be beneficial in mitigating smell loss following viral infection of the upper respiratory tract. However, double-blind studies are needed to confirm this observation. α -lipoic acid has also been suggested to be useful in some cases of hypogeusia and burning mouth syndrome.

The use of zinc and vitamin A in treating olfactory disturbances is controversial, and there does not appear to be much benefit beyond replenishing established deficiencies. However, zinc has been shown to improve taste function secondary to hepatic deficiencies, and retinoids (bioactive vitamin A derivatives) are known to play an essential role in the survival of olfactory neurons. One protocol in which zinc was infused with chemotherapy treatments suggested a possible protective effect against developing taste impairment. Diseases of the alimentary tract can not only influence chemoreceptive function, but also occasionally influence vitamin B_{12} absorption. This can result in a relative deficiency of vitamin B_{12} , theoretically contributing to olfactory nerve disturbance. Vitamin B_2 (riboflavin) and magnesium supplements are reported in the alternative literature to aid in the management of migraine that, in turn, may be associated with smell dysfunction. Because vitamin D deficiency is a cofactor of chemotherapy-induced mucocutaneous toxicity and dysgeusia, adding vitamin D₃ 1000–2000 units per day, may benefit some patients with smell and taste complaints during or following chemotherapy.

A number of medications have reportedly been used with success in ameliorating olfactory symptoms, although strong scientific evidence for efficacy is generally lacking. A report that theophylline improved smell function was uncontrolled and failed to account for the fact that some meaningful improvement occurs without treatment; indeed, the

percentage of responders was about the same (~50%) as that noted by others to show spontaneous improvement over a similar time period. Antiepileptics and some antidepressants (e.g., amitriptyline) have been used to treat dysosmias and smell distortions, particularly following head trauma. Ironically, amitriptyline is also frequently on the list of medications that can ultimately distort smell and taste function, possibly from its anticholinergic effects. One study suggested that the centrally acting acetylcholinesterase inhibitor donepezil in AD resulted in improvements on smell identification measures that correlated with overall clinician-based impressions of change in dementia severity scores.

Alternative therapies, such as acupuncture, meditation, cognitive-behavioral therapy, and yoga, can help patients manage uncomfortable experiences associated with chemosensory disturbance and oral pain syndromes and to cope with the psychosocial stressors surrounding the impairment. Additionally, modification of diet and eating habits is also important. By accentuating the other sensory experiences of a meal, such as food texture, aroma, temperature, and color, one can optimize the overall eating experience for a patient. In some cases, a flavor enhancer like monosodium glutamate (MSG) can be added to foods to increase palatability and encourage intake.

Proper oral and nasal hygiene and routine dental care are extremely important ways for patients to protect themselves from disorders of the mouth and nose that can ultimately result in chemosensory disturbance. Patients should be warned not to overcompensate for their taste loss by adding excessive amounts of sugar or salt. Smoking cessation and the discontinuance of oral tobacco use are essential in the management of any patient with smell and/or taste disturbance and should be repeatedly emphasized.

A major and often overlooked element of therapy comes from chemosensory testing itself. Confirmation or lack of conformation of loss is beneficial to patients who come to believe, in light of unsupportive family members and medical providers, that they may be "crazy." In cases where the loss is minor, patients can be informed of the likelihood of a more positive prognosis. Importantly, quantitative testing places the patient's problem into overall perspective. Thus, it is often therapeutic for an older person to know that, while his or her smell function is not what it used to be, it still falls above the average of his or her peer group. Without testing, many such patients are simply told that they are getting old and nothing can be done for them, leading in some cases to depression and decreased self-esteem.

FURTHER READING

- DEVANAND DP et al: Olfactory identification deficits are associated with increased mortality in a multiethnic urban community. *Ann Neurol* 78:401, 2015.
- DORY RL: Olfaction in Parkinson's disease and related disorders. *Neurobiol Dis* 46:527, 2012.
- DORY RL: Neurotoxic exposure and alterations in olfaction and gustation. *Handbook Clin Neurol* 131:299, 2015.
- DORY RL (ed): *Handbook of Olfaction and Gustation*, 3rd ed. Hoboken, Wiley-Liss, 2015.
- DORY RL et al: Influences of hormone replacement therapy on olfactory and cognitive function in the menopause. *Neurobiol Aging* 36:2053, 2015.
- DORY RL et al: Taste function in early stage treated and untreated Parkinson's disease. *J Neurol* 262:547, 2015.
- KOHLI P et al: The association between olfaction and depression: A systematic review. *Chem Senses* 41:479, 2016.
- LIU G et al: Prevalence and risk factors of taste and smell impairment in a nationwide sample of the US population: A cross-sectional study. *BMJ Open* 6:e013246, 2016.
- LONDON B et al: Predictors of prognosis in patients with olfactory disturbance. *Ann Neurol* 63:159, 2008.
- PEKALA K et al: Efficacy of olfactory training in patients with olfactory loss: A systematic review and meta-analysis. *Int Forum Allergy Rhinol* 6:299, 2016.
- PERRICONE C et al: Smell and autoimmunity: A comprehensive review. *Clin Rev Allergy Immunol* 45:87, 2013.



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. 30-1). 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 ear drum and the ossicles boost the sound energy nearly 200-fold by the time it reaches the inner ear.

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 acoustical 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.

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. 30-2). 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.

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 acoustical 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

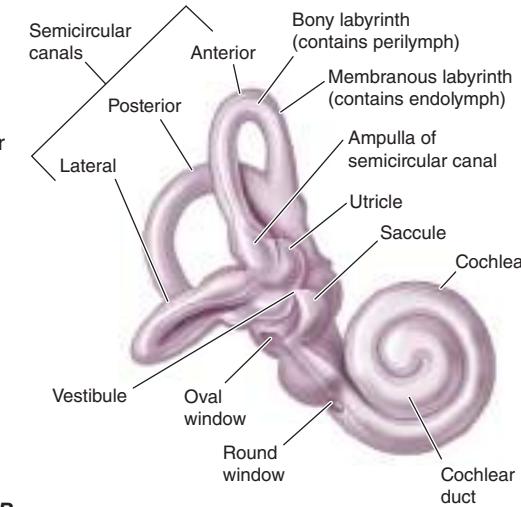
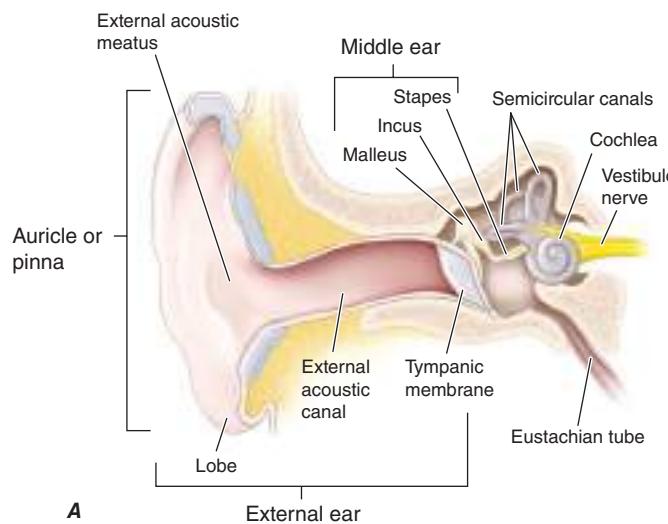


FIGURE 30-1 Ear anatomy. A. Drawing of modified coronal section through external ear and temporal bone, with structures of the middle and inner ear demonstrated. **B.** High-resolution view of inner ear.

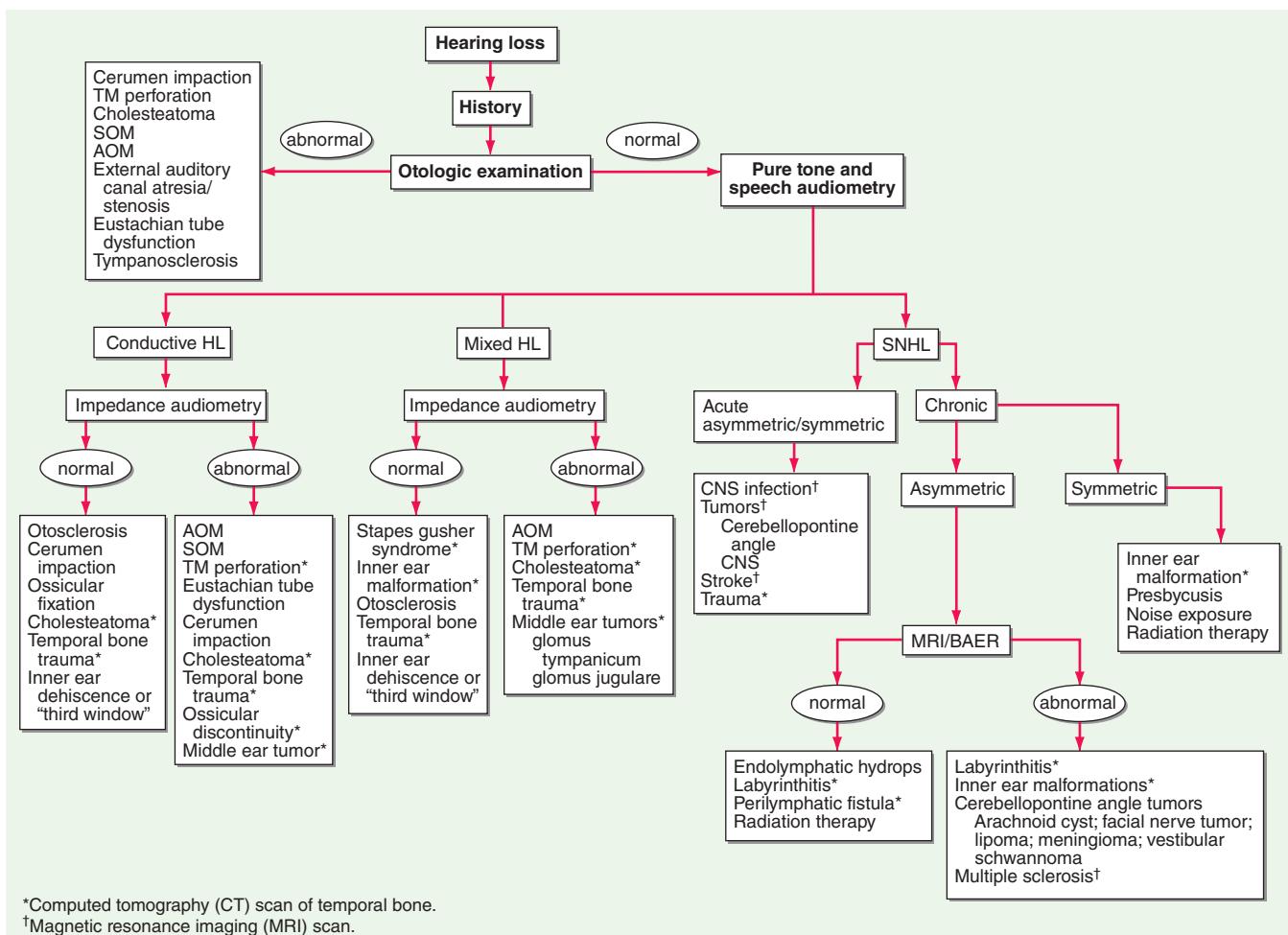


FIGURE 30-2 An algorithm for the approach to hearing loss. AOM, acute otitis media; BAER, brainstem auditory-evoked response; CNS, central nervous system; HL, hearing loss; SNHL, sensorineural hearing loss; SOM, serous otitis media; TM, tympanic membrane.

ear malformations or pathologies, 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.

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). 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 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. 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 also highly suggestive of an underlying cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Bony destruction visualized on computerized tomography (CT) of the temporal bone is

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 "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. 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 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).

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 ear plugs 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 the adults aged >75 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. 30-3). 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 and rate of cognitive decline. 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. 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).

Ménière’s disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but it invariably appears 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

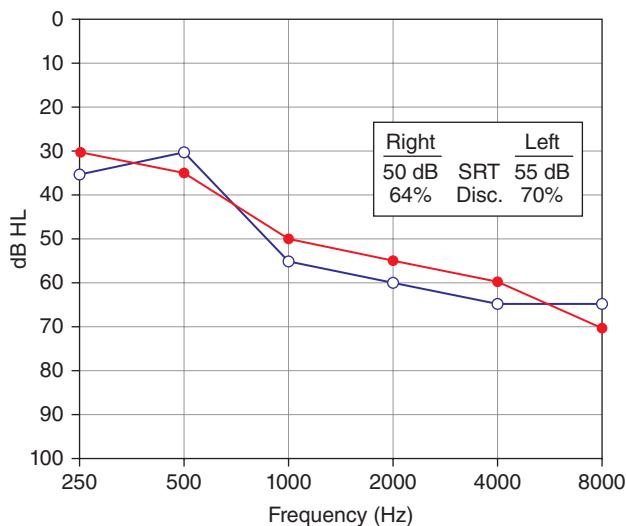


FIGURE 30-3 Presbycusis or age-related hearing loss. The audiogram shows a moderate to severe downsloping sensorineural hearing loss typical of presbycusis. The loss of high-frequency hearing is associated with a decreased speech discrimination score; consequently, patients complain of lack of clarity of hearing, especially in a noisy background. HL, hearing threshold level; SRT, speech reception threshold.

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. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present. An abnormal VEMP test may be helpful in detecting Ménière’s disease in a clinically unaffected contralateral ear. Magnetic resonance imaging (MRI) should be obtained to exclude retrocochlear pathology such as a cerebellopontine angle 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, or degenerative disease or trauma affecting the central auditory pathways. Characteristically, 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) is 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. 86) usually present with asymmetric sensorineural hearing loss with greater deterioration of speech understanding than pure tone hearing. Multiple sclerosis (Chap. 436) 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. 419). HIV (Chap. 197), 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 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.

■ 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 30-1**). 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, 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 are 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.

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. 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 30-2**).

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 30-3**). 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. 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. 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

TABLE 30-1 Hereditary Hearing Impairment Genes

DESIGNATION	GENE	FUNCTION	DESIGNATION	GENE	FUNCTION
Autosomal Dominant					
	CRYM	Thyroid hormone-binding protein	DFNB25	GRXCR1	Reversible S-glutathionylation of proteins
DFNA1	DIAPH1	Cytoskeletal protein	DFNB28	TRIOBP	Cytoskeletal-organizing protein
DFNA2A	KCNQ4	Potassium channel	DFNB29	CLDN14	Tight junctions
DFNA2B	GJB3 (Cx31)	Gap junction	DFNB30	MYO3A	Hybrid motor-signaling myosin
DFNA3A	GJB2 (Cx26)	Gap junction	DFNB31	WHRN	PDZ domain-containing protein
DFNA3B	GJB6 (Cx30)	Gap junction	DFNB35	ESRRB	Estrogen-related receptor beta protein
DFNA4	MYH14	Class II nonmuscle myosin	DFNB36	ESPN	Ca-insensitive actin-bundling protein
	CEACAM16	Cell adhesion molecule	DFNB37	MYO6	Unconventional myosin
DFNA5	DFNA5	Unknown	DFNB39	HFG	Hepatocyte growth factor
DFNA6/14/38	WFS1	Transmembrane protein	DFNB42	ILDR1	Ig-like domain-containing receptor
DFNA8/12	TECTA	Tectorial membrane protein	DFNB44	ADCY1	Adenylate cyclase
DFNA9	COCH	Unknown	DFNB48	CIB2	Calcium and integrin binding protein
DFNA10	EYA4	Developmental gene	DFNB49	BDP1	Subunit of RNA polymerase
DFNA11	MYO7A	Cytoskeletal protein	DFNB49	MARVELD2	Tight junction protein
DFNA13	COL11A2	Cytoskeletal protein	DFNB53	COL11A2	Collagen protein
DFNA15	POU4F3	Transcription factor	DFNB59	PJVK	Zn-binding protein
DFNA17	MYH9	Cytoskeletal protein	DFNB60	SLC22A4	Prestin, motor protein of cochlear outer hair cell
DFNA20/26	ACTG1	Cytoskeletal protein	DFNB61	SLC26A5	Motor protein
DFNA22	MYO6	Unconventional myosin	DFNB63	LRTOMT/COMT2	Putative methyltransferase
DFNA23	SIX1	Developmental gene	DFNB66	DCDC2	Ciliary protein
DFNA25	SLC17A8	Vesicular glutamate transporter	DFNB66/67	LHFPL5	Tetraspan protein
DFNA28	GRHL2	Transcription factor	DFNB68	S1PR2	Tetraspan membrane protein of hair cell stereocilia
DFNA36	TMC1	Transmembrane protein	DFNB70	PNPT1	Mitochondrial-RNA-import protein
DFNA41	P2RX2	Purinergic receptor	DFNB73	BSND	Beta subunit of chloride channel
DFNA44	CCDC50	Effector of epidermal growth factor-mediated signaling	DFNB74	MSRB3	Methionine sulfoxide reductase
DFNA50	MIRN96	MicroRNA	DFNB76	SYNE4	Part of LINC tethering complex
DFNA51	TJP2	Tight junction protein	DFNB77	LOXHD1	Stereociliary protein
DFNA56	TNC	Extracellular matrix protein	DFNB79	TPRN	Unknown
DFNA64	SMAC/DIABLO	Mitochondrial proapoptotic protein	DFNB82	GPSM2	G protein signaling modulator
DFNA65	TBC1D24	ARF6-interacting protein	DFNB84	PTPRQ	Type III receptor-like protein-tyrosine phosphatase family
DFNA66	CD164	Sialomucin	DFNB84	OTOG	Otogelin-like protein
DFNA67	OSBPL2	Intracellular lipid receptor	DFNB86	TBC1D24	GTPase-activating protein
DFNA68	HOMER2	Stereociliary scaffolding protein	DFNB88	ELMOD3	GTPase-activating protein
DFNA69	KITLG	Ligand for KIT receptor	DFNB89	KARS	Lysyl-tRNA synthetase
DFNA70	MCM2	Initiation and elongation during DNA replication	DFNB91	SERPINB6	Protease inhibitor
DFNA71	DMXL2	Regulator of Notch signaling	DFNB93	CABP2	Calcium-binding protein
Autosomal Recessive					
DFNB1A	GJB2 (CX26)	Gap junction	DFNA97	MET	Oncogene/hepatocyte growth factor receptor
DFNB1B	GJB6 (CX30)	Gap junction	DFNB98	TSPEAR	Epilepsy-associated repeats containing protein
DFNB2	MYO7A	Cytoskeletal protein	DFNB99	TMEM132E	Transmembrane protein
DFNB3	MYO15	Cytoskeletal protein	DFNB101	GRXCR2	Maintaining stereocilia bundles
DFNB4	PDS (SLC26A4)	Chloride/iodide transporter	DFNB102	EPS8	Epidermal growth factor receptor
DFNB6	TMIE	Transmembrane protein	DFNB103	CLIC5	Chloride ion transport
DFNB7/B11	TMC1	Transmembrane protein	DFNB105	CDC14A	Protein phosphatase involved in hair cell ciliogenesis
DFNB9	OTOF	Trafficking of membrane vesicles		FAM65B	Membrane-associated protein in stereocilia
DFNB8/10	TMPRSS3	Transmembrane serine protease		EPS8L2	Actin remodeling in response to EGF stimulation
DFNB12	CDH23	Intercellular adherence protein		ROR1	Receptor tyrosine kinase-like orphan receptor
DFNB15/72/95	GIPC3	PDZ domain-containing protein			
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			

TABLE 30-2 Syndromic Hereditary Hearing Impairment Genes

SYNDROME	GENE	FUNCTION
Alport's syndrome	COL4A3-5	Cytoskeletal protein
BOR syndrome	EYA1	Developmental gene
	SIX5 SIX1	Developmental gene Developmental gene
Jervell and Lange-Nielsen syndrome	KCNQ1	Delayed rectifier K ⁺ channel
	KCNE1	Delayed rectifier K ⁺ channel
Norrie's disease	NDP	Cell-cell interactions
Pendred's syndrome	SLC26A4	Chloride/iodide transporter
	FOXI1	Transcriptional activator of SLC26A4
	KCNJ10	Inwardly rectifying K ⁺ channel
Treacher Collins syndrome	TCOF1	Nucleolar-cytoplasmic transport
	POLR1D	Subunit of RNA polymerases I and III
	POLR1C	Subunit of RNA polymerases I and III
Usher's syndrome	MYO7A	Cytoskeletal protein
	USH1C	Unknown
	CDH23	Intercellular adherence protein
	PCDH15	Cell adhesion molecule
	SANS	Harmonin-associated protein
	CIB2	Calcium- and integrin-binding protein
	USH2A	Cell adhesion molecule
	VLGR1	G protein-coupled receptor
	WHRN	PDZ domain-containing protein
	CLRN1	Cellular synapse protein
	HARS	Histidyl-tRNA synthetase
	PDZD7	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
	EDN3	Endothelin B receptor ligand
	SOX10	Transcription factor

Abbreviations: BOR, branchio-oto-renal syndrome; WS, Waardenburg's syndrome.

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 "Audиologic 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.

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 and 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. 30-3). 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 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

TABLE 30-3 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

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 *Electrocotchleography* 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 summatting 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 electrocotchleography. Clinically, the test is useful in the diagnosis of Ménière’s disease, in which an elevation of the ratio of summatting 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 stimuli and evoking a short-latency electromyographic potential; cVEMP (or cervical VEMP) and oVEMP

(or ocular VEMP) have been described. The cVEMP elicits a vestibulo-collic 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, 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 can be placed entirely within the ear canal, thus reducing any 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, efforts are underway to create a new category for “basic” hearing

aids that could be sold over-the-counter, similar to some eyeglasses and contact lenses. By reducing the cost of hearing aids 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 investigated for the treatment of patients with single-sided deafness; early reports 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.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants may be appropriate (Fig. 30-4). Criteria for implantation include severe to profound hearing loss with open-set sentence cognition of $\leq 40\%$ 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 acoustical 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 increasingly being 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.

The first hybrid cochlear implant for the treatment of high-frequency hearing loss has now been approved by the U.S. Food and Drug Administration. 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

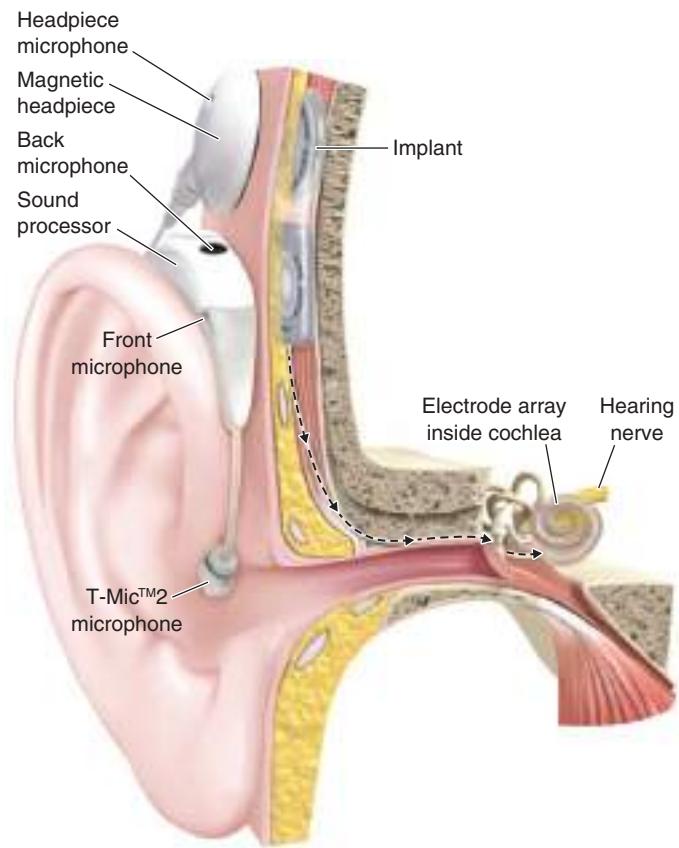


FIGURE 30-4 A cochlear implant is composed of an external microphone and speech processor worn on the ear and a receiver implanted underneath the temporalis muscle. The internal receiver is attached to an electrode that is placed surgically in the cochlea.

"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 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. As for 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. 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.

TABLE 30-4 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 feet	140
12-gauge shotgun blast	165
Loudest sound that can occur	194

Abbreviation: OSHA, Occupational Safety and Health Administration.

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 ear plugs or fluid-filled ear muffs to attenuate intense sound. **Table 30-4** 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 metal working 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 30-5**).

TABLE 30-5 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.

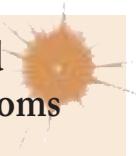
FURTHER READING

- ESPINOSA-SANCHEZ JM, LOPEZ-ESCAmez JA: Menière's disease. *Handb Clin Neurol* 137:257, 2016.
 MOSER T, STARR A: Auditory neuropathy—neural and synaptic mechanisms. *Nat Rev Neurol* 12:135, 2016.
 PATEL M et al: Intradynaptic methylprednisolone versus gentamicin in patients with unilateral Ménière's disease: A randomised, double-blind, comparative effectiveness trial. *Lancet* 388:2753, 2016.
 TIKKA C et al: Interventions to prevent occupational noise-induced hearing loss. *Cochrane Database Syst Rev* 7:CD006396, 2017.
 WILSON BS et al: Global hearing health care: New findings and perspectives. *Lancet* 390:2503, 2017.

31

Sore Throat, Earache, and Upper Respiratory Symptoms

Michael A. Rubin, Larry C. Ford,
 Ralph Gonzales



Infections of the upper respiratory tract (URIs) have a tremendous impact on public health. They are among the most common reasons for visits to primary care providers, and although the illnesses are typically mild, their high incidence and transmission rates place them among the leading causes of time lost from work or school. Even though a minority (~25%) of cases are caused by bacteria, URIs are the leading diagnoses for which antibiotics are prescribed on an outpatient basis in the United States, often inappropriately. Antibiotics are more often misprescribed in adults than in pediatric populations. The enormous consumption of antibiotics for these illnesses has contributed to the rise in antibiotic resistance among common community-acquired pathogens such as *Streptococcus pneumoniae*—a trend that in itself has an enormous influence on public health and on the individual patient.

Although most URIs are caused by viruses, distinguishing patients with primary viral infection from those with primary bacterial infection is difficult. Signs and symptoms of bacterial and viral URIs are typically indistinguishable. Until consistent, inexpensive, and rapid testing becomes available and is used widely, acute infections will be diagnosed largely on clinical grounds. The judicious use and potential for misuse of antibiotics in this setting pose ongoing challenges.

NONSPECIFIC INFECTIONS OF THE UPPER RESPIRATORY TRACT

Nonspecific URIs are a broadly defined group of disorders that collectively constitute the leading cause of ambulatory care visits in the United States. By definition, nonspecific URIs have no prominent localizing features. They are identified by a variety of descriptive names, including *acute infective rhinitis*, *acute rhinopharyngitis/nasopharyngitis*, *acute coryza*, and *acute nasal catarrh*, as well as by the inclusive label *common cold*.

ETIOLOGY

The large assortment of URI classifications reflects the wide variety of causative infectious agents and the varied manifestations of common pathogens. Nearly all nonspecific URIs are caused by viruses spanning multiple virus families and many antigenic types. For instance, there are at least 100 immunotypes of rhinovirus (**Chap. 194**), the most common cause of URI (~30–40% of cases); other causes include influenza virus (three immunotypes; **Chap. 195**) as well as parainfluenza virus (four immunotypes), coronavirus (at least three immunotypes), and adenovirus (47 immunotypes) (**Chap. 194**). Respiratory syncytial virus (RSV), a well-established pathogen in pediatric populations, is also a recognized cause of significant disease in elderly and immunocompromised individuals. A host of additional viruses, including some viruses not typically associated with URIs (e.g., enteroviruses, rubella virus,

and varicella-zoster virus), account for a small percentage of cases in adults each year. Although new diagnostic modalities (e.g., nasopharyngeal swab for polymerase chain reaction) can assign a viral etiology, there are few specific treatment options, and no pathogen is identified in a substantial proportion of cases. A specific diagnostic workup beyond a clinical diagnosis is generally unnecessary in an otherwise healthy adult.

■ CLINICAL MANIFESTATIONS

The signs and symptoms of nonspecific URI are similar to those of other URIs but lack a pronounced localization to one particular anatomic location, such as the sinuses, pharynx, or lower airway. Nonspecific URI commonly presents as an acute, mild, and self-limited catarrhal syndrome with a median duration of ~1 week (range, 2–10 days). Signs and symptoms are diverse and frequently variable across patients, even when caused by the same virus. The principal signs and symptoms of nonspecific URI include rhinorrhea (with or without purulence), nasal congestion, cough, and sore throat. Other manifestations, such as fever, malaise, sneezing, lymphadenopathy, and hoarseness, are more variable, with fever more common among infants and young children. This varying presentation may reflect differences in host response as well as in infecting organisms; myalgias and fatigue, for example, sometimes are seen with influenza and parainfluenza infections, whereas conjunctivitis may suggest infection with adenovirus or enterovirus. Cough secondary to upper respiratory inflammation after such an illness frequently lasts 2–3 weeks and can be misinterpreted as an indication of a process that necessitates antibiotic therapy. Findings on physical examination are frequently nonspecific and unimpressive. Between 0.5 and 2% of colds are complicated by secondary bacterial infections (e.g., rhinosinusitis, otitis media, and pneumonia), particularly in higher-risk populations such as infants, elderly persons, and chronically ill or immunosuppressed individuals. Secondary bacterial infections usually are associated with a prolonged course of illness, increased severity of illness, and localization of signs and symptoms, often as a rebound after initial clinical improvement (the “double-dip” sign). Purulent secretions from the nares or throat often are misinterpreted as an indication of bacterial sinusitis or pharyngitis. These secretions, however, can be seen in nonspecific URI and, in the absence of other clinical features, are poor predictors of bacterial infection.

TREATMENT

Nonspecific Upper Respiratory Infections

Antibiotics have no role in the treatment of uncomplicated nonspecific URI, and their misuse facilitates the emergence of antimicrobial resistance; in healthy volunteers, a single course of a commonly prescribed antibiotic like azithromycin can result in macrolide resistance in oral streptococci many months later. In the absence of clinical evidence of bacterial infection, treatment remains entirely symptom based, with use of decongestants and nonsteroidal anti-inflammatory drugs. Clinical trials of zinc, vitamin C, echinacea, and other alternative remedies have revealed no consistent benefit in the treatment of nonspecific URI.

INFECTIONS OF THE SINUS

Rhinosinusitis refers to an inflammatory condition involving the nasal sinuses. Although most cases of sinusitis involve more than one sinus, the maxillary sinus is most commonly involved; next, in order of frequency, are the ethmoid, frontal, and sphenoid sinuses. Each sinus is lined with a respiratory epithelium that produces mucus, which is transported out by ciliary action through the sinus ostium and into the nasal cavity. Normally, mucus does not accumulate in the sinuses, which remain mostly sterile despite their adjacency to the bacterium-filled nasal passages. When the sinus ostia are obstructed or when ciliary clearance is impaired or absent, the secretions can be retained, producing the typical signs and symptoms of sinusitis. As these secretions accumulate with obstruction, they become more

susceptible to infection with a variety of pathogens, including viruses, bacteria, and, rarely, fungi. Sinusitis affects a tremendous proportion of the population, accounts for millions of visits to primary care physicians each year, and is the fifth leading diagnosis for which antibiotics are prescribed. It typically is classified by duration of illness (acute vs. chronic); by etiology (infectious vs. noninfectious); and, when infectious, by the offending pathogen type (viral, bacterial, or fungal).

■ ACUTE RHINOSINUSITIS

Acute rhinosinusitis—defined as sinusitis of <4 weeks’ duration—constitutes the vast majority of sinusitis cases. Most cases are diagnosed in the ambulatory care setting and occur primarily as a consequence of a preceding viral URI. Differentiating acute bacterial from viral sinusitis on clinical grounds is difficult. Therefore, it is perhaps not surprising that antibiotics are prescribed frequently (in 85–98% of all cases) for this condition.

Etiology The ostial obstruction in rhinosinusitis can arise from both infectious and noninfectious causes. Noninfectious etiologies include allergic rhinitis (with either mucosal edema or polyp obstruction), barotrauma (e.g., from deep-sea diving or air travel), and exposure to chemical irritants. Obstruction can also occur with nasal and sinus tumors (e.g., squamous cell carcinoma) or granulomatous diseases (e.g., granulomatosis with polyangiitis, rhinoscleroma), and conditions leading to altered mucus content (e.g., cystic fibrosis) can cause sinusitis through impaired mucus clearance. In intensive care units, nasotracheal intubation and nasogastric tubes are major risk factors for nosocomial sinusitis.

Viral rhinosinusitis is far more common than bacterial sinusitis, although relatively few studies have sampled sinus aspirates for the presence of different viruses. In the studies that have done so, the viruses most commonly isolated—both alone and with bacteria—have been rhinovirus, parainfluenza virus, and influenza virus. Bacterial causes of sinusitis have been better described. Among community-acquired cases, *S. pneumoniae* and nontypable *Haemophilus influenzae* are the most common pathogens, accounting for 50–60% of cases. *Moraxella catarrhalis* causes disease in a significant percentage (20%) of children but a lesser percentage of adults. Other streptococcal species and *Staphylococcus aureus* cause only a small percentage of cases, although there is increasing concern about methicillin-resistant *S. aureus* (MRSA) as an emerging cause. It is difficult to assess whether a cultured bacterium represents a true infecting organism, an insufficiently deep sample (which would not be expected to be sterile), or—especially in the case of previous sinus surgeries—a colonizing organism. Anaerobes occasionally are found in association with infections of the roots of premolar teeth that spread to the adjacent maxillary sinuses. The role of atypical organisms like *Chlamydia pneumoniae* and *Mycoplasma pneumoniae* in the pathogenesis of acute sinusitis is unclear. Nosocomial cases commonly are associated with bacteria prevalent in the hospital environment, including *S. aureus*, *Pseudomonas aeruginosa*, *Serratia marcescens*, *Klebsiella pneumoniae*, and *Enterobacter* species. Often, these infections are polymicrobial and can involve organisms that are highly resistant to numerous antibiotics. Fungi also are established causes of sinusitis, although most acute cases affect immunocompromised patients and represent invasive, life-threatening infections. The best-known example is rhinocerebral mucormycosis caused by fungi of the order Mucorales, which includes *Rhizopus*, *Rhizomucor*, *Mucor*, *Lichtheimia* (formerly *Mycocladus*, formerly *Absidia*), and *Cunninghamella* (Chap. 213). These infections classically occur in diabetic patients with ketoacidosis but can also develop in transplant recipients, patients with hematologic malignancies, and patients receiving chronic glucocorticoid or deferoxamine therapy. Other hyaline molds, such as *Aspergillus* and *Fusarium* species, also are occasional causes of this disease.

Clinical Manifestations Most cases of acute sinusitis present after or in conjunction with a viral URI, and it can be difficult to discriminate the clinical features of one from the other, with timing becoming important in diagnosis (see below). A large proportion of patients with colds have sinus inflammation, although true bacterial sinusitis complicates only 0.2–2% of these viral infections. Common presenting

symptoms of sinusitis include nasal drainage and congestion, facial pain or pressure, and headache. Thick, purulent or discolored nasal discharge is often thought to indicate bacterial sinusitis but also occurs early in viral infections such as the common cold and is not specific to bacterial infection. Other nonspecific manifestations include cough, sneezing, and fever. Tooth pain, most often involving the upper molars, as well as halitosis are occasionally associated with bacterial sinusitis.

In acute sinusitis, sinus pain or pressure often localizes to the involved sinus (particularly the maxillary sinus) and can be worse when the patient bends over or is supine. Although rare, manifestations of advanced sphenoid or ethmoid sinus infection can be profound, including severe frontal or retroorbital pain radiating to the occiput, thrombosis of the cavernous sinus, and signs of orbital cellulitis. Acute focal sinusitis is uncommon but should be considered with severe symptoms involving the maxillary sinus and fever, regardless of illness duration. This condition is typically associated with red, hot, and swollen sinuses that are extremely tender to palpation; is of staphylococcal etiology; and requires emergent debridement and initial IV administration of antibiotics. Similarly, patients with advanced frontal sinusitis can present with a condition known as *Pott's puffy tumor*, with soft-tissue swelling and pitting edema over the frontal bone from a communicating subperiosteal abscess. Life-threatening complications of sinusitis are rare but include meningitis, epidural abscess, and cerebral abscess.

Patients with acute fungal rhinosinusitis (such as mucormycosis; Chap. 213) often present with symptoms related to pressure effects, particularly when the infection has spread to the orbits and cavernous sinus. Signs such as orbital swelling and cellulitis, proptosis, ptosis, and decreased extraocular movement are common, as is retro- or periorbital pain. Nasopharyngeal ulcerations, epistaxis, and headaches are also common, and involvement of cranial nerves V and VII has been described in more advanced cases. Bony erosion may be evident on examination or endoscopy. Often the patient does not appear to be seriously ill despite the rapidly progressive nature of these infections.

Patients with acute nosocomial sinusitis are often critically ill and thus do not manifest the typical clinical features of sinus disease. This diagnosis should be suspected, however, when hospitalized patients with appropriate risk factors (e.g., nasotracheal intubation) develop fever without another apparent cause.

Diagnosis Distinguishing viral from bacterial rhinosinusitis in the ambulatory setting is usually difficult because of the relatively low sensitivity and specificity of the common clinical features. One clinical feature that has been used to help guide diagnostic and therapeutic decision-making is illness duration. Because acute bacterial sinusitis is uncommon in patients whose symptoms have lasted <10 days, expert panels now recommend reserving this diagnosis for patients with "persistent" symptoms (i.e., symptoms lasting >10 days in adults or >10–14 days in children) accompanied by the three cardinal signs of purulent nasal discharge, nasal obstruction, and facial pain (Table 31-1). The fact that, even among patients who meet these criteria, only 40–50% have true bacterial sinusitis prompts some authorities to favor 14 days of symptoms before considering treatment. The use of CT or sinus radiography is not recommended for acute disease, particularly early in the course of illness (i.e., at <10 days) in light of the high prevalence of similar findings among patients with acute viral rhinosinusitis. In the evaluation of persistent, recurrent, or chronic sinusitis, CT of the sinuses becomes the radiographic study of choice.

The clinical history and/or setting often can identify cases of acute anaerobic bacterial sinusitis, acute fungal sinusitis, or sinusitis from noninfectious causes (e.g., allergic rhinosinusitis). In the case of an immunocompromised patient with acute fungal sinus infection, immediate examination by an otolaryngologist is required. In addition to cultures, biopsy specimens from involved areas should be examined by a pathologist for evidence of fungal hyphal elements and tissue invasion. Cases of suspected acute nosocomial sinusitis should be confirmed by sinus CT. Because therapy should target the offending organism, a sinus aspirate for culture and susceptibility testing should be obtained, whenever possible, before the initiation of antimicrobial therapy. As the ability to isolate the sometimes-myriad components of

TABLE 31-1 Guidelines for the Diagnosis and Treatment of Acute Bacterial Sinusitis in Adults

DIAGNOSTIC CRITERIA	TREATMENT RECOMMENDATIONS*
Moderate symptoms (e.g., nasal purulence/congestion or cough) for >10 d or	<i>Initial therapy:</i> Amoxicillin/clavulanate, 500/125 mg PO tid or 875/125 mg PO bid ^b <i>Penicillin allergy:</i> Doxycycline, 100 mg PO bid; or An antipneumococcal fluoroquinolone (e.g., moxifloxacin, 400 mg/d PO daily) ^c <i>Exposure to antibiotics within 30 d or >30% prevalence of penicillin-resistant <i>Streptococcus pneumoniae</i>:</i> Amoxicillin/clavulanate (extended release), 2000/125 mg PO bid; or Doxycycline, 100 mg PO bid; or An antipneumococcal fluoroquinolone (e.g., moxifloxacin, 400 mg PO daily) ^c
Severe symptoms of any duration, including unilateral/focal facial swelling or tooth pain	<i>Recent treatment failure:</i> Amoxicillin/clavulanate (extended release), 2000 mg PO bid; or An antipneumococcal fluoroquinolone (e.g., moxifloxacin, 400 mg PO daily) ^c

^aThe duration of therapy is 5–7 days if symptoms improve within the first few days of treatment but can be up to 7–10 days, with appropriate follow-up. Severe disease may warrant IV antibiotics and consideration of hospital admission.

^bIn areas where the prevalence of antibiotic resistance is low, amoxicillin can be considered as initial therapy in patients without recent antibiotic exposure.

^cFluoroquinolones carry a risk of tendinitis and neuropathy and should be used only if other options are not reasonable, with consideration of risks and benefits.

the sinus microbiome is augmented by molecular techniques, the hope is for an even more tailored treatment regimen.

TREATMENT

Acute Rhinosinusitis

Most patients with a clinical diagnosis of acute rhinosinusitis improve without antibiotic therapy. The preferred initial approach in patients with mild to moderate symptoms of short duration is therapy aimed at symptom relief and facilitation of sinus drainage, such as with oral and topical decongestants, nasal saline lavage, and—at least in patients with a history of chronic sinusitis or allergies—nasal glucocorticoids. Newer studies have cast doubt on the role of antibiotics and nasal glucocorticoids in acute rhinosinusitis. In one notable double-blind, randomized, placebo-controlled trial, neither antibiotics nor topical glucocorticoids had a significant impact on cure in the study population of patients, the majority of whom had had symptoms for <7 days. Similarly, another high-profile randomized trial comparing antibiotics to placebo in patients with acute rhinosinusitis demonstrated no significant improvement in symptoms by the third day of therapy. Still, antibiotic therapy can be considered for adult patients whose condition does not improve after 10–14 days, and patients with more severe symptoms (regardless of duration) should be treated with antibiotics (Table 31-1). However, watchful waiting remains a viable option in many cases.

Empirical antibiotic therapy for community-acquired sinusitis in adults should consist of the narrowest-spectrum agent active against the most common bacterial pathogens, including *S. pneumoniae* and *H. influenzae*—e.g., amoxicillin/clavulanate (with the decision guided by local rates of β-lactamase-producing *H. influenzae*). No clinical trials support the use of broader-spectrum agents for routine cases of bacterial sinusitis, even in the current era of drug-resistant *S. pneumoniae*. For those patients who do not respond to initial antimicrobial therapy, sinus aspiration and/or lavage by an otolaryngologist should be considered. Antibiotic prophylaxis to prevent episodes of recurrent acute bacterial sinusitis is not recommended.

Surgical intervention and IV antibiotic administration usually are reserved for patients with severe disease or those with intracranial

complications such as abscess and orbital involvement. Immunocompromised patients with acute invasive fungal sinusitis usually require extensive surgical debridement and treatment with IV antifungal agents active against fungal hyphal forms, such as amphotericin B. Specific therapy should be individualized according to the fungal species and its susceptibilities as well as the individual patient's characteristics.

Treatment of nosocomial sinusitis should begin with broad-spectrum antibiotics to cover common and often resistant pathogens such as *S. aureus* and gram-negative bacilli. Therapy then should be tailored to the results of culture and susceptibility testing of sinus aspirates.

■ CHRONIC SINUSITIS

Chronic sinusitis is characterized by symptoms of sinus inflammation lasting >12 weeks. This illness is most commonly associated with either bacteria or fungi, and clinical cure in most cases is very difficult. Many patients have undergone treatment with repeated courses of antibacterial agents and multiple sinus surgeries, increasing their risk of colonization with antibiotic-resistant pathogens and of surgical complications. These patients often have high rates of morbidity, sometimes over many years.

In *chronic bacterial sinusitis*, infection is thought to be due to the impairment of mucociliary clearance from repeated infections rather than to persistent bacterial infection. The pathogenesis of this condition, however, is poorly understood. The role of biofilms in such chronic infections continues to be explored, including the contribution that low-virulence pathogens may play in this complex, interacting milieu. Although certain conditions (e.g., cystic fibrosis) can predispose patients to chronic bacterial sinusitis, most patients with chronic rhinosinusitis do not have obvious underlying conditions that result in the obstruction of sinus drainage, the impairment of ciliary action, or immune dysfunction. Patients experience constant nasal congestion and sinus pressure, with intermittent periods of greater severity, which may persist for years. CT can be helpful in determining the extent of disease, detecting an underlying anatomic defect or obstructing process (e.g., a polyp), and assessing the response to therapy. Management should involve an otolaryngologist to conduct endoscopic examinations and obtain tissue samples for histologic examination and culture. An endoscopy-derived culture not only has a higher yield but also allows direct visualization for abnormal anatomy.

Chronic fungal sinusitis is a disease of immunocompetent hosts and is usually noninvasive, although slowly progressive invasive disease is sometimes seen. Noninvasive disease, which typically is associated with hyaline molds such as *Aspergillus* species and dematiaceous molds such as *Curvularia* or *Bipolaris* species, can present as a number of different scenarios. In mild, indolent disease, which usually occurs in the setting of repeated failures of antibacterial therapy, only nonspecific mucosal changes may be seen on sinus CT. Although there is some controversy on this point, endoscopic surgery is usually curative in these cases, with no need for antifungal therapy. Another form of disease presents as long-standing, often unilateral symptoms and opacification of a single sinus on imaging studies as a result of a mycetoma (fungus ball) within the sinus. Treatment for this condition also is surgical, although systemic antifungal therapy may be warranted in the rare case in which bony erosion occurs. A third form of disease, known as *allergic fungal sinusitis*, is seen in patients with a history of nasal polyposis and asthma, who often have had multiple sinus surgeries. Patients with this condition produce a thick, eosinophil-laden mucus with the consistency of peanut butter that contains sparse fungal hyphae on histologic examination. These patients often present with pansinusitis.

TREATMENT

Chronic Sinusitis

Treatment of chronic bacterial sinusitis can be challenging and consists primarily of repeated culture-guided courses of antibiotics, sometimes for 3–4 weeks or longer at a time; administration of

intranasal glucocorticoids; and mechanical irrigation of the sinus with sterile saline solution. When this management approach fails, sinus surgery may be indicated and sometimes provides significant, albeit short-term, alleviation. Treatment of chronic fungal sinusitis consists of surgical removal of impacted mucus. Recurrence, unfortunately, is common.

INFECTIONS OF THE EAR AND MASTOID

Infections of the ear and associated structures can involve both the middle and the external ear, including the skin, cartilage, periosteum, ear canal, and tympanic and mastoid cavities. Both viruses and bacteria are known causes of these infections, some of which result in significant morbidity if not treated appropriately.

■ INFECTIONS OF EXTERNAL EAR STRUCTURES

Infections involving the structures of the external ear are often difficult to differentiate from noninfectious inflammatory conditions with similar clinical manifestations. Clinicians should consider inflammatory disorders as possible causes of external ear irritation, particularly in the absence of local or regional adenopathy. Aside from the more salient causes of inflammation, such as trauma, insect bite, and overexposure to sunlight or extreme cold, the differential diagnosis should include less common conditions such as autoimmune disorders (e.g., lupus or relapsing polychondritis) and vasculitides (e.g., granulomatosis with polyangiitis).

Auricular Cellulitis Auricular cellulitis is an infection of the skin overlying the external ear and typically follows minor local trauma. It presents as the typical signs and symptoms of cellulitis, with tenderness, erythema, swelling, and warmth of the external ear (particularly the lobule) but without apparent involvement of the ear canal or inner structures. Treatment consists of warm compresses and oral antibiotics such as cephalexin or dicloxacillin that are active against typical skin and soft-tissue pathogens (specifically, *S. aureus* and streptococci). IV antibiotics such as a first-generation cephalosporin (e.g., cefazolin) or a penicillinase-resistant penicillin (e.g., nafcillin) occasionally are needed for more severe cases, with consideration of MRSA if either risk factors or failure of therapy point to this organism.

Perichondritis Perichondritis, an infection of the perichondrium of the auricular cartilage, typically follows local trauma (e.g., piercings, burns, or lacerations). Occasionally, when the infection spreads down to the cartilage of the pinna itself, patients may develop chondritis. The infection may closely resemble auricular cellulitis, with erythema, swelling, and extreme tenderness of the pinna, although the lobule is less often involved in perichondritis. The most common pathogens are *P. aeruginosa* and *S. aureus*, although other gram-negative and gram-positive organisms occasionally are involved. Treatment consists of systemic antibiotics active against both *P. aeruginosa* and *S. aureus*. An antipseudomonal penicillin (e.g., piperacillin) or a combination of a penicillinase-resistant penicillin and an antipseudomonal quinolone (e.g., nafcillin plus ciprofloxacin) is typically used. Incision and drainage may be helpful for culture and for resolution of infection, which often takes weeks. When perichondritis fails to respond to adequate antimicrobial therapy, clinicians should consider a noninfectious inflammatory etiology such as relapsing polychondritis.

Otitis Externa The term *otitis externa* refers to a collection of diseases involving primarily the auditory meatus. Otitis externa usually results from a combination of heat and retained moisture, with desquamation and maceration of the epithelium of the outer ear canal. The disease exists in several forms: localized, diffuse, chronic, and invasive. All forms are predominantly bacterial in origin, with *P. aeruginosa* and *S. aureus* the most common pathogens.

Acute localized otitis externa (furunculosis) can develop in the outer third of the ear canal, where skin overlies cartilage and hair follicles are numerous. As in furunculosis elsewhere on the body, *S. aureus* is the usual pathogen, and treatment typically consists of an oral antistaphylococcal penicillin (e.g., dicloxacillin or cephalexin), with incision and drainage in cases of abscess formation.

Acute diffuse otitis externa is also known as *swimmer's ear*, although it can develop in patients who have not recently been swimming. Heat, humidity, and the loss of protective cerumen lead to excessive moisture and elevation of the pH in the ear canal, which in turn lead to skin maceration and irritation. Infection may then follow; the predominant pathogen is *P. aeruginosa*, although other bacteria—and rarely yeasts—have been recovered from patients with this condition. The illness often starts with itching and progresses to severe pain, which is usually elicited by manipulation of the pinna or tragus. The onset of pain is generally accompanied by the development of an erythematous, swollen ear canal, often with scant white, clumpy discharge. Treatment consists of cleansing the canal to remove debris and enhance the activity of topical therapeutic agents—usually hypertonic saline or mixtures of alcohol and acetic acid. Inflammation can also be decreased by adding glucocorticoids to the treatment regimen or by using Burow's solution (aluminum acetate in water). Antibiotics are most effective when given topically. Otic mixtures provide adequate pathogen coverage; these preparations usually combine neomycin with polymyxin, with or without glucocorticoids. Systemic antimicrobial agents typically are reserved for severe disease or infections in immunocompromised hosts.

Chronic otitis externa is caused primarily by repeated local irritation, most commonly arising from persistent drainage from a chronic middle-ear infection. Other causes of repeated irritation, such as insertion of cotton swabs or other foreign objects into the ear canal, can lead to this condition, as can rare chronic infections such as syphilis, tuberculosis, and leprosy. Chronic otitis externa typically presents as erythematous, scaling dermatitis in which the predominant symptom is pruritus rather than pain; this condition must be differentiated from several others that produce a similar clinical picture, such as atopic dermatitis, seborrheic dermatitis, psoriasis, and dermatomycosis. Therapy consists of identifying and treating or removing the offending process, although successful resolution is frequently difficult.

Invasive otitis externa, also known as *malignant* or *necrotizing otitis externa*, is an aggressive and potentially life-threatening disease that occurs predominantly in elderly diabetic patients and other immunocompromised persons. The disease begins in the external canal as a soft-tissue infection that progresses slowly over weeks to months and often is difficult to distinguish from a severe case of chronic otitis externa because of the presence of purulent otorrhea and an erythematous swollen ear and external canal. Severe, deep-seated otalgia, frequently out of proportion to findings on examination, is often noted and can help differentiate invasive from chronic otitis externa. The characteristic finding on examination is granulation tissue in the posteroinferior wall of the external canal, near the junction of bone and cartilage. If left unchecked, the infection can migrate to the base of the skull (resulting in skull-base osteomyelitis) and onward to the meninges and brain, with a high mortality rate. Cranial nerve involvement is seen occasionally, with the facial nerve usually affected first and most often. Thrombosis of the sigmoid sinus can occur if the infection extends to the area. CT, which can reveal osseous erosion of the temporal bone and skull base, can be used to help determine the extent of disease, as can gallium and technetium-99 scintigraphy studies. *P. aeruginosa* is by far the most common offender, although *S. aureus*, *Staphylococcus epidermidis*, *Aspergillus*, *Actinomycetes*, and some gram-negative bacteria also have been associated with this disease. In all cases, the external ear canal should be cleansed and a biopsy specimen of the granulation tissue within the canal (or of deeper tissues) obtained for culture of the offending organism. IV antibiotic therapy should be given for a prolonged course (6–8 weeks) and directed specifically toward the recovered pathogen. For *P. aeruginosa*, the regimen typically includes an antipseudomonal penicillin or cephalosporin (e.g., piperacillin or cefepime), sometimes with an aminoglycoside or a fluoroquinolone, the latter of which can even be administered orally given its excellent bioavailability. In addition, antibiotic drops containing an agent active against *Pseudomonas* (e.g., ciprofloxacin) are usually prescribed and are combined with glucocorticoids to reduce inflammation. Cases of invasive *Pseudomonas* otitis externa recognized in the early stages can sometimes be treated with oral and otic

fluoroquinolones alone, albeit with close follow-up. Extensive surgical debridement, once an important component of the treatment approach, is now rarely indicated.

In *necrotizing otitis externa*, recurrence is documented up to 20% of the time. Aggressive glycemic control in diabetics is important not only for effective treatment but also for prevention of recurrence. The role of hyperbaric oxygen has not been clearly established.

■ INFECTIONS OF MIDDLE-EAR STRUCTURES

Otitis media is an inflammatory condition of the middle ear that results from dysfunction of the eustachian tube in association with a number of illnesses, including URIs and chronic rhinosinusitis. The inflammatory response in these conditions leads to the development of a sterile transudate within the middle-ear and mastoid cavities. Infection may occur if bacteria or viruses from the nasopharynx contaminate this fluid, producing an acute (or sometimes chronic) illness.

Acute Otitis Media Acute otitis media results when pathogens from the nasopharynx are introduced into the inflammatory fluid collected in the middle ear (e.g., by nose blowing during a URI). Pathogenic proliferation in this space leads to the development of the typical signs and symptoms of acute middle-ear infection. The diagnosis of acute otitis media requires the demonstration of fluid in the middle ear (with tympanic membrane [TM] immobility) and the accompanying signs or symptoms of local or systemic illness (**Table 31-2**).

ETIOLOGY Acute otitis media typically follows a viral URI. The causative viruses (most commonly RSV, influenza virus, rhinovirus, and enterovirus) can themselves cause subsequent acute otitis media; more often, they predispose the patient to bacterial otitis media. Studies using tympanocentesis have consistently found *S. pneumoniae* to be the most important bacterial cause, isolated in up to 35% of cases. *H. influenzae* (nontypable strains) and *M. catarrhalis* also are common bacterial causes of acute otitis media, and concern is increasing with MRSA as an emerging etiologic agent. Viruses, such as those mentioned above, have been recovered either alone or with bacteria in 17–40% of cases.

CLINICAL MANIFESTATIONS Fluid in the middle ear is typically demonstrated or confirmed with pneumatic otoscopy. In the absence of fluid, the TM moves visibly with the application of positive and negative pressure, but this movement is dampened when fluid is present. With bacterial infection, the TM can also be erythematous, bulging, or retracted and occasionally can perforate spontaneously. The signs and symptoms accompanying infection can be local or systemic, including otalgia, otorrhea, diminished hearing, and fever. Erythema of the TM is often evident but is nonspecific as it frequently is seen in association with inflammation of the upper respiratory mucosa. Other signs and symptoms occasionally reported include vertigo, nystagmus, and tinnitus.

TREATMENT

Acute Otitis Media

There has been considerable debate on the usefulness of antibiotics for the treatment of acute otitis media. A higher proportion of treated than untreated patients are free of illness 3–5 days after diagnosis. The difficulty of predicting which patients will benefit from antibiotic therapy has led to different approaches. In the Netherlands, for instance, physicians typically manage acute otitis media with initial observation, administering anti-inflammatory agents for aggressive pain management and reserving antibiotics for high-risk patients, patients with complicated disease, or patients whose condition does not improve after 48–72 h. In contrast, many experts in the United States continue to recommend antibiotic therapy for children <6 months old in light of the higher frequency of secondary complications in this young and functionally immunocompromised population. However, observation without antimicrobial therapy is now the recommended option in the United States for acute otitis media in children >2 years of age and for mild to moderate disease without middle-ear effusion in children 6 months to 2 years of age. Treatment

TABLE 31-2 Guidelines for the Diagnosis and Treatment of Acute Otitis Media

ILLNESS SEVERITY	DIAGNOSTIC CRITERIA	TREATMENT RECOMMENDATIONS
Mild to moderate	>2 yrs or 6 mo to 2 yrs without middle-ear effusion <6 mo; or 6 mo to 2 yrs with middle-ear effusion (fluid in the middle ear, evidenced by decreased TM mobility, air/fluid level behind TM, bulging TM, purulent otorrhea) <i>and</i> acute onset of signs and symptoms of middle-ear inflammation, including fever, otalgia, decreased hearing, tinnitus, vertigo, erythematous TM; or >2 yrs with bilateral disease, TM perforation, high fever, immunocompromise, emesis	<i>Observation alone</i> (deferring antibiotic therapy for 48–72 h and limiting management to symptom relief) <i>Initial therapy^a:</i> Amoxicillin, 80–90 mg/kg qd (up to 2 g) PO in divided doses (bid or tid); or Cefdinir, 14 mg/kg qd PO in 1 dose or divided doses (bid); or Cefuroxime, 30 mg/kg qd PO in divided doses (bid); or Azithromycin, 10 mg/kg qd PO on day 1 followed by 5 mg/kg qd PO for 4 d <i>Exposure to antibiotics within 30 d or recent treatment failure^{a,b}:</i> Amoxicillin, 90 mg/kg qd (up to 2 g) PO in divided doses (bid), plus clavulanate, 6.4 mg/kg qd PO in divided doses (bid); or Ceftriaxone, 50 mg/kg IV/IM qd for 3 d; or Clindamycin, 30–40 mg/kg qd PO in divided doses (tid)
Severe	As above, with temperature ≥39.0°C (≥102°F); or Moderate to severe otalgia	<i>Initial therapy^a:</i> Amoxicillin, 90 mg/kg qd (up to 2 g) PO in divided doses (bid), plus clavulanate, 6.4 mg/kg qd PO in divided doses (bid); or Ceftriaxone, 50 mg/kg IV/IM qd for 3 d <i>Exposure to antibiotics within 30 d or recent treatment failure^{a,b}:</i> Ceftriaxone, 50 mg/kg IV/IM qd for 3 d; or Clindamycin, 30–40 mg/kg qd PO in divided doses (tid); or Consider tympanocentesis with culture

^aDuration (unless otherwise specified): 10 days for patients <6 years old and patients with severe disease; 5–7 days (with consideration of observation only in previously healthy individuals with mild disease) for patients ≥6 years old. ^bFailure to improve and/or clinical worsening after 48–72 h of observation or treatment.

Abbreviation: TM, tympanic membrane.

Source: American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media, 2004.

is typically indicated for patients <6 months old; for children 6 months to 2 years old who have middle-ear effusion and signs/symptoms of middle-ear inflammation; for all patients >2 years old who have bilateral disease, TM perforation, immunocompromise, or emesis; and for any patient who has severe symptoms, including a fever ≥39°C or moderate to severe otalgia (Table 31-2).

Because most studies of the etiologic agents of acute otitis media consistently document similar pathogen profiles, therapy is generally empirical except in those few cases in which tympanocentesis is warranted—e.g., cases refractory to therapy and cases in patients who are severely ill or immunodeficient. Despite resistance to penicillin and amoxicillin in roughly one-quarter of *S. pneumoniae* isolates, one-third of *H. influenzae* isolates, and nearly all *M. catarrhalis* isolates, outcome studies continue to find that amoxicillin is as successful as any other agent, and it remains the drug of first choice in recommendations from multiple sources (Table 31-2). Therapy for uncomplicated acute otitis media typically is administered for 5–7 days to patients aged ≥6 years; longer courses (e.g., 10 days) should be reserved for immunocompromised patients or patients with severe disease, in whom short-course therapy may be inadequate.

A switch in regimen is recommended if there is no clinical improvement by the third day of therapy, given the possibility of infection with a β-lactamase-producing strain of *H. influenzae* or *M. catarrhalis* or with a strain of penicillin-resistant *S. pneumoniae*. Decongestants and antihistamines are frequently used as adjunctive agents to reduce congestion and relieve obstruction of the eustachian tube, but clinical trials have yielded no significant evidence of benefit with either class of agents.

Recurrent Acute Otitis Media Recurrent acute otitis media (more than three episodes within 6 months or four episodes within 12 months) generally is due to relapse or reinfection, although data indicate that the majority of early recurrences are new infections. In general, the same pathogens responsible for acute otitis media cause recurrent disease; even so, the recommended treatment consists of antibiotics active against β-lactamase-producing organisms. Antibiotic prophylaxis (e.g., with amoxicillin) can reduce recurrences in patients with recurrent acute otitis media by an average of one episode per year, which benefit is small compared with the high likelihood of colonization with antibiotic-resistant pathogens. Other approaches, including

placement of tympanostomy tubes, adenoidectomy, and tonsillectomy plus adenoidectomy, are of questionable overall value in light of the relatively small benefit compared with the potential for complications.

Serous Otitis Media In serous otitis media (otitis media with effusion), fluid is present in the middle ear for an extended period in the absence of signs and symptoms of infection. In general, acute effusions are self-limited; most resolve in 2–4 weeks. In some cases, however (in particular after an episode of acute otitis media), effusions can persist for months. These chronic effusions are often associated with significant hearing loss in the affected ear. The great majority of cases of otitis media with effusion resolve spontaneously within 3 months without antibiotic therapy. Antibiotic therapy or myringotomy with insertion of tympanostomy tubes typically is reserved for patients in whom bilateral effusion (1) has persisted for at least 3 months and (2) is associated with significant bilateral hearing loss. With this conservative approach and the application of strict diagnostic criteria for acute otitis media and otitis media with effusion, an estimated 6–8 million courses of antibiotics could be avoided each year in the United States.

Chronic Otitis Media Chronic suppurative otitis media is characterized by persistent or recurrent purulent otorrhea in the setting of TM perforation. Usually, there is also some degree of conductive hearing loss. This condition can be categorized as active or inactive. Inactive disease is characterized by a central perforation of the TM, which allows drainage of purulent fluid from the middle ear. When the perforation is more peripheral, squamous epithelium from the auditory canal may invade the middle ear through the perforation, forming a mass of keratinaceous debris (*cholesteatoma*) at the site of invasion. This mass can enlarge and has the potential to erode bone and promote further infection, which can lead to meningitis, brain abscess, or paralysis of cranial nerve VII. Treatment of chronic active otitis media is surgical; mastoidectomy, myringoplasty, and tympanoplasty can be performed as outpatient surgical procedures, with an overall success rate of ~80%. Chronic inactive otitis media is more difficult to cure, usually requiring repeated courses of topical antibiotic drops during periods of drainage. Systemic antibiotics may offer better cure rates, but their role in the treatment of this condition remains unclear.

Mastoiditis Acute mastoiditis was relatively common among children before the introduction of antibiotics. Because the mastoid air

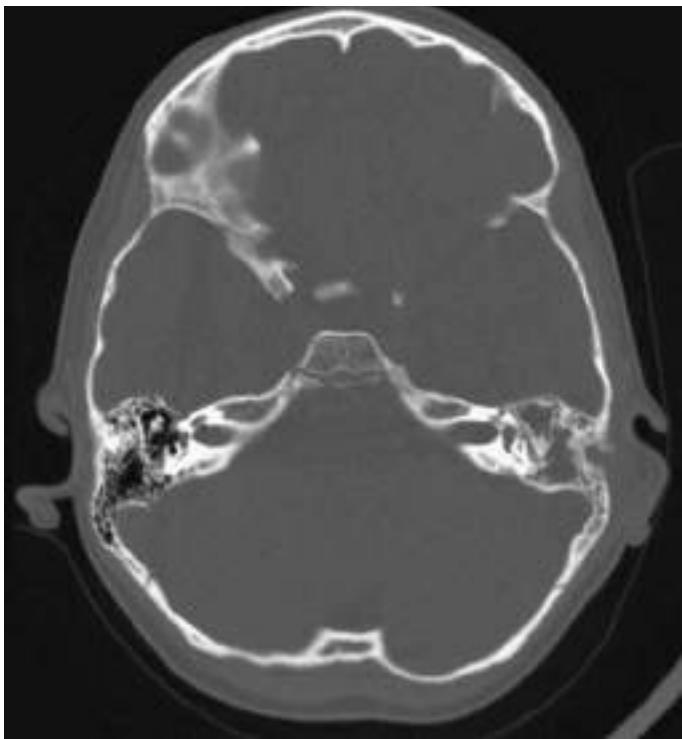


FIGURE 31-1 Acute mastoiditis. Axial CT image shows an acute fluid collection within the mastoid air cells on the left.

cells connect with the middle ear, the process of fluid collection and infection is usually the same in the mastoid as in the middle ear. Early and frequent treatment of acute otitis media is most likely the reason that the incidence of acute mastoiditis has declined to only 1.2–2.0 cases per 100,000 person-years in countries with high prescribing rates for acute otitis media.

In countries such as the Netherlands, where antibiotics are used sparingly for acute otitis media, the incidence rate of acute mastoiditis is roughly twice that in countries like the United States. However, neighboring Denmark has a rate of acute mastoiditis similar to that in the Netherlands but an antibiotic-prescribing rate for acute otitis media more similar to that in the United States.

In typical acute mastoiditis, purulent exudate collects in the mastoid air cells (Fig. 31-1), producing pressure that may result in erosion of the surrounding bone and formation of abscess-like cavities that are usually evident on CT. Patients typically present with pain, erythema, and swelling of the mastoid process along with displacement of the pinna, usually in conjunction with the typical signs and symptoms of acute middle-ear infection. Rarely, patients can develop severe complications if the infection tracks under the periosteum of the temporal bone to cause a subperiosteal abscess, erodes through the mastoid tip to cause a deep neck abscess, or extends posteriorly to cause septic thrombosis of the lateral sinus.

Purulent fluid should be cultured whenever possible to help guide antimicrobial therapy. Initial empirical therapy usually is directed against the typical organisms associated with acute otitis media, such as *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. Patients with more severe or prolonged courses of illness should be treated for infection with *S. aureus* and gram-negative bacilli (including *Pseudomonas*). Broad-spectrum empirical therapy should be narrowed once culture results become available. Most patients can be treated conservatively with IV antibiotics; surgery (cortical mastoidectomy) is reserved for complicated cases and those in which conservative treatment has failed.

INFECTIONS OF THE PHARYNX AND ORAL CAVITY

Oropharyngeal infections range from mild, self-limited viral illnesses to serious, life-threatening bacterial infections. The most common presenting symptom is sore throat—one of the most common reasons

for ambulatory care visits by both adults and children. Although sore throat is a symptom in many noninfectious illnesses as well, the overwhelming majority of patients with a new sore throat have acute pharyngitis of viral or bacterial etiology.

■ ACUTE PHARYNGITIS

Millions of visits to primary care providers each year are for sore throat; the majority of cases of acute pharyngitis are caused by typical respiratory viruses. The most important source of concern is infection with group A β-hemolytic *Streptococcus* (*S. pyogenes*), which is associated with acute glomerulonephritis and acute rheumatic fever. The risk of rheumatic fever can be reduced by timely penicillin therapy.

Etiology A wide variety of organisms cause acute pharyngitis. The relative importance of the different pathogens can only be estimated, since a significant proportion of cases (~30%) have no identified cause. Together, respiratory viruses are the most common identifiable cause of acute pharyngitis, with rhinoviruses and coronaviruses accounting for large proportions of cases (~20% and at least 5%, respectively). Influenza virus, parainfluenza virus, and adenovirus also account for a measurable share of cases, with the former two more seasonal and the latter as part of the more clinically severe syndrome of pharyngoconjunctival fever. Other important but less common viral causes include herpes simplex virus (HSV) types 1 and 2, coxsackievirus A, cytomegalovirus (CMV), and Epstein-Barr virus (EBV). Acute HIV infection can present as acute pharyngitis and should always be considered in at-risk populations.

Acute bacterial pharyngitis is typically caused by *S. pyogenes*, which accounts for ~5–15% of all cases of acute pharyngitis in adults; rates vary with the season and with utilization of the health care system. Group A streptococcal pharyngitis is primarily a disease of children aged 5–15 years; it is uncommon among children <3 years old, as is rheumatic fever. Streptococci of groups C and G account for a minority of cases, although these serogroups are nonrheumatogenic. *Fusobacterium necrophorum* has been increasingly recognized as a cause of pharyngitis in adolescents and young adults and, when sought, is isolated nearly as often as group A streptococci. This organism is important because of the rare but life-threatening *Lemierre disease*, which is generally associated with *F. necrophorum* and is usually preceded by pharyngitis (see “Oral Infections,” below). The remaining bacterial causes of acute pharyngitis are seen infrequently (<1% of cases each) but should be considered in appropriate exposure groups because of the severity of illness if left untreated; these etiologic agents include *Neisseria gonorrhoeae*, *Corynebacterium diphtheriae*, *Corynebacterium ulcerans*, *Yersinia enterocolitica*, and *Treponema pallidum* (in secondary syphilis). Anaerobic bacteria also can cause acute pharyngitis (*Vincent angina*) and can contribute to more serious polymicrobial infections, such as peritonitis or retropharyngeal abscesses (see below). Atypical organisms such as *M. pneumoniae* and *C. pneumoniae* have been recovered from patients with acute pharyngitis; whether these agents are commensals or causes of acute infection is debatable.

Clinical Manifestations Although the signs and symptoms accompanying acute pharyngitis are not reliable predictors of the etiologic agent, the clinical presentation occasionally suggests one etiology over another. Acute pharyngitis due to respiratory viruses such as rhinovirus or coronavirus usually is not severe and typically is associated with a constellation of coryzal symptoms better characterized as non-specific URI. Findings on physical examination are uncommon; fever is rare, and tender cervical adenopathy and pharyngeal exudates are not seen. In contrast, acute pharyngitis from influenza virus can be severe and is much more likely to be associated with fever as well as with myalgias, headache, and cough. The presentation of pharyngoconjunctival fever due to adenovirus infection is similar. Since pharyngeal exudate may be present on examination, this condition can be difficult to differentiate from streptococcal pharyngitis. However, adenoviral pharyngitis is distinguished by the presence of conjunctivitis in one-third to one-half of patients. Acute pharyngitis from primary HSV infection can also mimic streptococcal pharyngitis in some cases, with

pharyngeal inflammation and exudate, but the presence of vesicles and shallow ulcers on the palate can help differentiate the two diseases. This HSV syndrome is distinct from pharyngitis caused by coxsackievirus (*herpangina*), which is associated with small vesicles that develop on the soft palate and uvula and then rupture to form shallow white ulcers. Acute pharyngitis coupled with fever, fatigue, generalized lymphadenopathy, and (on occasion) splenomegaly is characteristic of infectious mononucleosis due to EBV or CMV. Acute primary infection with HIV is frequently associated with fever and acute pharyngitis as well as with myalgias, arthralgias, malaise, and occasionally a nonpruritic maculopapular rash, which may be followed by lymphadenopathy and mucosal ulcerations without exudate.

The clinical features of acute pharyngitis caused by streptococci of groups A, C, and G are similar, ranging from a relatively mild illness without many accompanying symptoms to clinically severe cases with profound pharyngeal pain, fever, chills, and abdominal pain. A hyperemic pharyngeal membrane with tonsillar hypertrophy and exudate is usually seen, along with tender anterior cervical adenopathy. Coryzal manifestations, including cough, are typically absent; when present, they suggest a viral etiology. Strains of *S. pyogenes* that generate erythrogenic toxin can also produce scarlet fever characterized by an erythematous rash and strawberry tongue. The other types of acute bacterial pharyngitis (e.g., gonococcal, diphtherial, and yersinial) often present as exudative pharyngitis with or without other clinical features. Their etiologies are often suggested only by the clinical history.

Diagnosis The primary goal of diagnostic testing is to separate acute streptococcal pharyngitis from pharyngitis of other etiologies (particularly viral) so that antibiotics can be prescribed more efficiently for patients in whom they may be beneficial. The most appropriate standard for the diagnosis of streptococcal pharyngitis, however, has not been established definitively. Throat swab culture is generally regarded as the most appropriate but cannot distinguish between infection and colonization and requires 24–48 h to yield results that vary with technique and culture conditions. Rapid antigen-detection tests offer good specificity (>90%) but lower sensitivity when implemented in routine practice. Sensitivity has also been shown to vary across the clinical spectrum of disease (65–90%). Several clinical prediction systems (Fig. 31-2) can increase the sensitivity of rapid antigen-detection tests to >90% in controlled settings. Since the sensitivities achieved in routine clinical practice are often lower, several medical and professional societies continue to recommend that all negative rapid antigen-detection tests in children be confirmed by a throat culture to limit transmission and complications of illness caused by group A streptococci. The Centers for Disease Control and Prevention, the Infectious Diseases Society of America, and the American Academy of Family Physicians do not recommend backup culture when adults have negative results from a highly sensitive rapid antigen-detection test, however, because of the lower prevalence and smaller benefit in this age group.

Cultures and rapid diagnostic tests for other causes of acute pharyngitis, such as influenza virus, adenovirus, HSV, EBV, CMV, and *M. pneumoniae*, are available in many locations and can be used when these pathogens are suspected. The diagnosis of acute EBV infection depends primarily on the detection of antibodies to the virus with a heterophile agglutination assay (monospot slide test) or enzyme-linked immunosorbent assay. Testing for HIV, ideally through a combination antigen/antibody method, should be performed when acute primary HIV infection is suspected. If other bacterial causes are suspected (particularly *N. gonorrhoeae*, *C. diphtheriae*, or *Y. enterocolitica*), specific cultures should be requested since these organisms may be missed on routine throat swab culture.

TREATMENT

Pharyngitis

Antibiotic treatment of pharyngitis due to *S. pyogenes* confers numerous benefits, including a decrease in the risk of rheumatic fever—the primary focus of treatment. The magnitude of this benefit is fairly

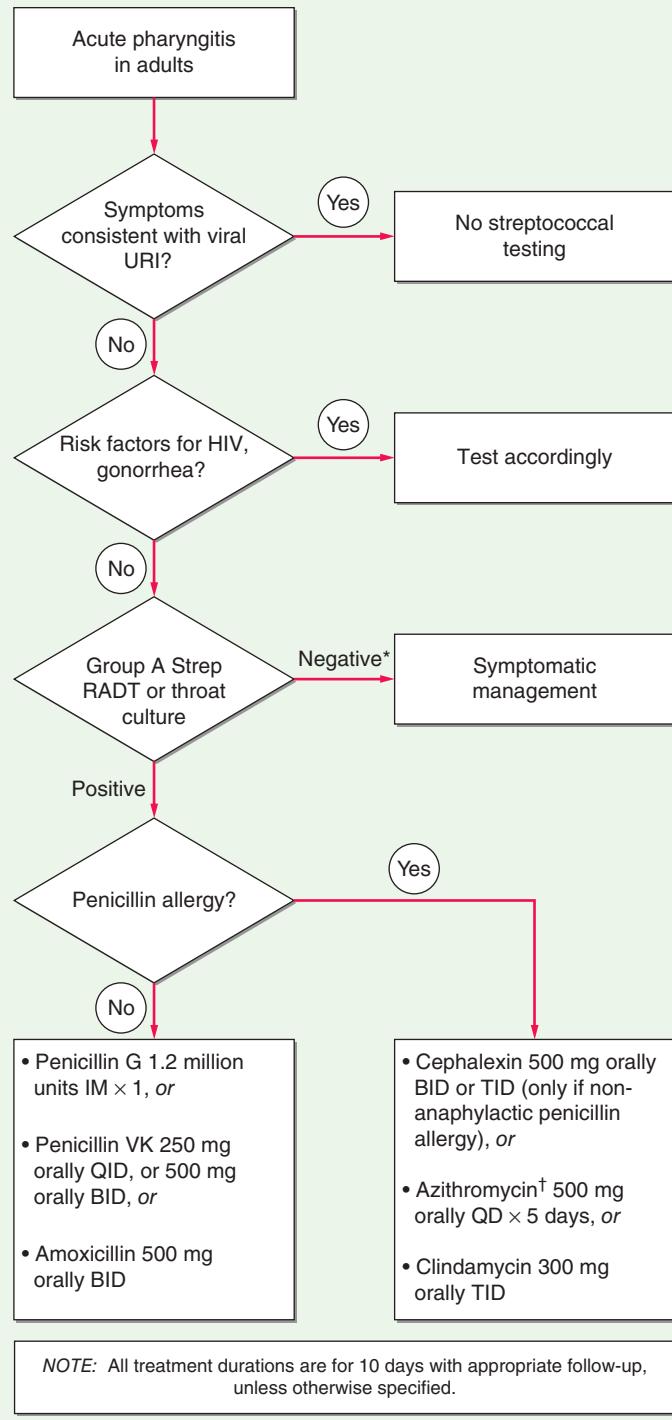
small, since rheumatic fever is now a rare disease, even among untreated patients. Nevertheless, when therapy is started within 48 h of illness onset, symptom duration is decreased modestly. An additional benefit of therapy is the potential to reduce the transmission of streptococcal pharyngitis, particularly in areas of overcrowding or close contact. Antibiotic therapy for acute pharyngitis is therefore recommended in cases in which *S. pyogenes* is confirmed as the etiologic agent by rapid antigen-detection test or throat swab culture. Otherwise, antibiotics should be given in routine cases only when another bacterial cause has been identified. Effective therapy for streptococcal pharyngitis consists of either a single dose of IM benzathine penicillin or a full 10-day course of oral penicillin (Fig. 31-2).

 Azithromycin can be used in place of penicillin, although its potential utility is waning and its use in some parts of the world (particularly Europe) is prohibited as a result of resistance among *S. pyogenes* strains. Broader-spectrum (and often more expensive) antibiotics also are active against streptococci but offer no greater efficacy than the agents mentioned above. Testing for cure is unnecessary and may reveal only chronic colonization. There is no evidence to support antibiotic treatment of group C or G streptococcal pharyngitis or pharyngitis in which mycoplasmas or chlamydiae have been recovered. Cultures can be of benefit because *F. necrophorum*, an increasingly common cause of bacterial pharyngitis in young adults, is not covered by macrolide therapy. Long-term penicillin prophylaxis (benzathine penicillin G, 1.2 million units IM every 3–4 weeks; or penicillin VK, 250 mg PO twice daily) is indicated for patients at risk of recurrent rheumatic fever in order to prevent what could be catastrophic sequelae of recurrent streptococcal pharyngitis.

 Antibiotic shortages, sometimes the result of manufacturing difficulties or delays, natural disasters, and regulatory or other issues, can preclude the use of the optimal antibiotic. These shortages can be regional, national, or international. Communication with pharmacists and the use of antibiotic stewardship teams can help mitigate the effects of shortages, yield recommendations for the use of alternative agents, and prevent delays in treatment that can affect patients' access to antibiotics.

Treatment of viral pharyngitis is entirely symptom-based except in infection with influenza virus or HSV. For influenza, the armamentarium includes the adamantanes amantadine and rimantadine and the neuraminidase inhibitors oseltamivir and zanamivir. Administration of all these agents needs to be started within 48 h of symptom onset to reduce illness duration meaningfully. Among these agents, only oseltamivir and zanamivir are active against both influenza A and influenza B and therefore can be used when local patterns of infection and antiviral resistance are unknown. Oropharyngeal HSV infection sometimes responds to treatment with antiviral agents such as acyclovir, although these drugs are often reserved for immunosuppressed patients.

Complications Although rheumatic fever is the best-known complication of acute streptococcal pharyngitis, the risk of its following acute infection remains quite low. Other complications include acute glomerulonephritis and numerous suppurative conditions, such as peritonsillar abscess (*quinsy*), otitis media, mastoiditis, sinusitis, bacteremia, and pneumonia—all of which occur at low rates. Although antibiotic treatment of acute streptococcal pharyngitis can prevent the development of rheumatic fever, there is no evidence that it can prevent acute glomerulonephritis. Some evidence supports antibiotic use to prevent the suppurative complications of streptococcal pharyngitis, particularly peritonsillar abscess, which can also involve oral anaerobes such as *Fusobacterium*. Abscesses usually are accompanied by severe pharyngeal pain, dysphagia, fever, and dehydration; in addition, medial displacement of the tonsil and lateral displacement of the uvula are often evident on examination. Although early use of IV antibiotics (e.g., clindamycin, penicillin G with metronidazole) may eliminate the need for surgical drainage in some cases, treatment typically involves needle aspiration or incision and drainage.



*Confirmation of a negative rapid antigen-detection test by a throat culture is not required in adults.

†Macrolides do not treat *F. necrophorum*, a cause of pharyngitis in young adults (see text).

Abbreviations: URI, upper respiratory infection; RADT, rapid antigen detection test

FIGURE 31-2 Algorithm for the diagnosis and treatment of acute pharyngitis.

ORAL INFECTIONS

Aside from periodontal diseases such as gingivitis, infections of the oral cavity most commonly involve HSV or *Candida* species. In addition to causing painful cold sores on the lips, HSV can infect the tongue and buccal mucosa, causing the formation of irritating vesicles. Although topical antiviral agents (e.g., acyclovir and penciclovir) can be used externally for cold sores, with possible benefit, oral or IV acyclovir is often needed for primary infections, extensive oral infections, and infections in immunocompromised patients. Oropharyngeal candidiasis (*thrush*) is caused by a variety of *Candida* species, most often

C. albicans. Thrush occurs predominantly in neonates, immunocompromised patients (especially those with AIDS), and recipients of prolonged antibiotic or glucocorticoid therapy. In addition to sore throat, patients often report a burning tongue or abnormal taste, and physical examination reveals friable white or gray plaques on the gingiva, tongue, and oral mucosa, often with underlying erythema. Treatment, which usually consists of a topical antifungal (nystatin or clotrimazole) or oral fluconazole, is typically successful. In the uncommon cases of fluconazole-refractory thrush that are seen in some patients with HIV/AIDS or in patients with resistant organisms that can sometimes complicate the treatment of recurrent oral candidiasis, other therapeutic options include oral voriconazole, an IV echinocandin (caspofungin, micafungin, or anidulafungin), or amphotericin B deoxycholate, if needed. In these cases, therapy based on culture and susceptibility test results is ideal.

Vincent angina, also known as *acute necrotizing ulcerative gingivitis* or *trench mouth*, is a unique and dramatic form of gingivitis characterized by painful, inflamed gingiva with ulcerations of the interdental papillae that bleed easily. Since oral anaerobes are the cause, patients typically have halitosis and frequently present with fever, malaise, and lymphadenopathy. Treatment consists of debridement and oral administration of penicillin plus metronidazole, with clindamycin or doxycycline alone as an alternative.

Ludwig angina is a rapidly progressive, potentially fulminant form of cellulitis that involves the bilateral sublingual and submandibular spaces and that typically originates from an infected or recently extracted tooth, most commonly a lower second or third molar. Improved dental care has reduced the incidence of this disorder substantially. Infection in these areas leads to dysphagia, odynophagia, and “woody” edema in the sublingual region, forcing the tongue up and back with the potential for airway obstruction. Fever, dysarthria, and drooling also may occur, and patients may speak in a “hot potato” voice. Intubation or tracheostomy may be necessary to secure the airway, as asphyxiation is the most common cause of death. Patients should

be admitted to the hospital and closely monitored during treatment with IV antibiotics directed against streptococci and oral anaerobes. Recommended agents include ampicillin/sulbactam, clindamycin, or high-dose penicillin plus metronidazole.

Septic thrombophlebitis of the internal jugular vein (*Lemierre disease*) is a rare anaerobic oropharyngeal infection caused predominantly by *F. necrophorum*. The illness typically starts as a sore throat (most commonly in adolescents and young adults), which may present as exudative tonsillitis or peritonsillar abscess. Infection of the deep pharyngeal tissue allows organisms to drain into the lateral pharyngeal space,

which contains the carotid artery and internal jugular vein. Septic thrombophlebitis of the internal jugular vein can result, with associated pain, dysphagia, and unilateral neck swelling and stiffness. Sepsis usually occurs 3–10 days after the onset of sore throat and is often coupled with metastatic infection to the lung and other distant sites, with pulmonary abscess or empyema. Occasionally, the infection can extend along the carotid sheath and into the posterior mediastinum, resulting in mediastinitis, or it can erode into the carotid artery, with the early sign of repeated small bleeds into the mouth. The mortality rate from these invasive infections can be as high as 50%. Treatment consists of IV antibiotics (clindamycin or ampicillin/sulbactam) and surgical drainage of any purulent collections. The concomitant use of anticoagulants to prevent embolization remains controversial and is not typically advised; both the risks and the benefits of their use must be carefully considered.

INFECTIONS OF THE LARYNX AND EPIGLOTTIS

LARYNGITIS

Laryngitis is defined as any inflammatory process involving the larynx and can be caused by a variety of infectious and noninfectious processes. The vast majority of laryngitis cases seen in clinical practice in developed countries are acute. Acute laryngitis is a common syndrome caused predominantly by the same viruses responsible for many other URIs. In fact, most cases of acute laryngitis occur in the setting of a viral URI.

Etiology Nearly all major respiratory viruses have been implicated in acute viral laryngitis, including rhinovirus, influenza virus, parainfluenza virus, adenovirus, coxsackievirus, coronavirus, and RSV. Acute laryngitis can also be associated with acute bacterial respiratory infections such as those caused by group A *Streptococcus* or *C. diphtheriae* (although diphtheria has been virtually eliminated in the United States). Another bacterial pathogen thought to play a role (albeit unclear) in the pathogenesis of acute laryngitis is *M. catarrhalis*, which has been recovered from nasopharyngeal cultures in a significant percentage of cases.



Chronic laryngitis of infectious etiology is much less common in developed than in developing countries. Laryngitis due to

Mycobacterium tuberculosis is often difficult to distinguish from laryngeal cancer, in part because of the frequent absence of signs, symptoms, and radiographic findings typical of pulmonary disease. *Histoplasma* and *Blastomyces* may cause laryngitis, often as a complication of systemic infection. *Candida* species can cause laryngitis as well, often in association with thrush or esophagitis and particularly in immunosuppressed patients. Rare cases of chronic laryngitis are due to *Coccidioides* and *Cryptococcus*.

Clinical Manifestations Laryngitis is characterized by hoarseness and also can be associated with reduced vocal pitch or aphonia. As acute laryngitis is caused primarily by respiratory viruses, these symptoms usually occur in association with other symptoms and signs of URI, including rhinorrhea, nasal congestion, cough, and sore throat. Direct laryngoscopy often reveals diffuse laryngeal erythema and edema, along with vascular engorgement of the vocal folds. In addition, chronic disease (e.g., tuberculous laryngitis) often includes mucosal nodules and ulcerations visible on laryngoscopy; these lesions are sometimes mistaken for laryngeal cancer.

TREATMENT

Laryngitis

Acute laryngitis is usually treated with humidification and voice rest alone. Antibiotics are not recommended except when group A *Streptococcus* is cultured, in which case penicillin is the drug of choice. The choice of therapy for chronic laryngitis depends on the pathogen, whose identification usually requires biopsy with culture.

Patients with laryngeal tuberculosis are highly contagious because of the large number of organisms that are easily aerosolized. These patients should be managed in the same way as patients with active pulmonary disease.

CRUP

The term *croup* actually denotes a group of diseases collectively referred to as “croup syndrome,” all of which are acute and predominantly viral respiratory illnesses characterized by marked swelling of the subglottic region of the larynx. Croup primarily affects children <6 years old. For a detailed discussion of this entity, the reader should consult a textbook of pediatric medicine.

EPIGLOTTITIS

Acute epiglottitis (supraglottitis) is an acute, rapidly progressive form of cellulitis of the epiglottis and adjacent structures that can result in complete—and potentially fatal—airway obstruction in both children and adults. Before the widespread use of *H. influenzae* type b (Hib) vaccine, this entity was much more common among children, with a peak incidence at ~3.5 years of age. In some countries, mass vaccination against Hib has reduced the annual incidence of acute epiglottitis in children by >90%; in contrast, the annual incidence in adults has changed little since the introduction of Hib vaccine. Because of the danger of airway obstruction, acute epiglottitis constitutes a medical emergency, particularly in children, and prompt diagnosis and airway protection are of the utmost importance.

Etiology After the introduction of the Hib vaccine in the mid-1980s, disease incidence among children in the United States declined dramatically. Nevertheless, lack of vaccination or vaccine failure has meant that many pediatric cases seen today are still due to Hib. In adults and (more recently) in children, a variety of other bacterial pathogens have been associated with epiglottitis, the most common being group A *Streptococcus*. Other pathogens—seen less frequently—include *S. pneumoniae*, *Haemophilus parainfluenzae*, and *S. aureus* (including MRSA). Viruses have not been established as causes of acute epiglottitis.

Clinical Manifestations and Diagnosis Epiglottitis typically presents more acutely in young children than in adolescents or adults. On presentation, most children have had symptoms for <24 h, including high fever, severe sore throat, tachycardia, systemic toxicity, and (in many cases) drooling while sitting forward. Symptoms and signs of respiratory obstruction also may be present and may progress rapidly. The somewhat milder illness in adolescents and adults often follows 1–2 days of severe sore throat and is commonly accompanied by dyspnea, drooling, and stridor. Physical examination of patients with acute epiglottitis may reveal moderate or severe respiratory distress, with inspiratory stridor and retractions of the chest wall. These findings diminish as the disease progresses and the patient tires. Conversely, oropharyngeal examination reveals infection that is much less severe than would be predicted from the symptoms—a finding that should alert the clinician to a cause of symptoms and obstruction that lies beyond the tonsils. The diagnosis often is made on clinical grounds, although direct fiberoptic laryngoscopy is frequently performed in a controlled environment (e.g., an operating room) to visualize and culture the typical edematous “cherry-red” epiglottis and facilitate placement of an endotracheal tube. Direct visualization in an examination room (i.e., with a tongue blade and indirect laryngoscopy) is not recommended because of the risk of immediate laryngospasm and complete airway obstruction. Lateral neck radiographs and laboratory tests can assist in the diagnosis but may delay the critical securing of the airway and cause the patient to be moved or repositioned more than is necessary, thereby increasing the risk of further airway compromise. Neck radiographs typically reveal an enlarged edematous epiglottis (the “thumbprint sign,” Fig. 31-3), usually with a dilated hypopharynx and normal subglottic structures. Laboratory tests characteristically document mild to moderate leukocytosis with a predominance of neutrophils. Blood cultures are positive in a significant proportion of cases.



FIGURE 31-3 Acute epiglottitis. In this lateral soft-tissue radiograph of the neck, the arrow indicates the enlarged edematous epiglottis (the “thumbprint sign”).

TREATMENT

Epiglottitis

Security of the airway is always of primary concern in acute epiglottitis, even if the diagnosis is only suspected. Mere observation for signs of impending airway obstruction is not routinely recommended, particularly in children. Many adults have been managed with observation only since the illness is perceived to be milder in this age group, but some data suggest that this approach may be risky and probably should be reserved only for adult patients who have yet to develop dyspnea or stridor. Once the airway has been secured and specimens of blood and epiglottis tissue have been obtained for culture, treatment with IV antibiotics should be given to cover the most likely organisms, particularly *H. influenzae*. Because rates of ampicillin resistance in this organism have risen significantly in recent years, therapy with a β -lactam/ β -lactamase inhibitor combination or a third-generation cephalosporin is recommended. Typically, ampicillin/sulbactam, cefotaxime, or ceftriaxone is given, with clindamycin and trimethoprim-sulfamethoxazole reserved for patients allergic to β -lactams. Antibiotic therapy should be continued for 7–10 days and should be tailored to the organism recovered in culture. If the household contacts of a patient with *H. influenzae* epiglottitis include an unvaccinated child aged <4 years, all members of the household (including the patient) should receive prophylactic rifampin for 4 days to eradicate carriage of *H. influenzae*.

INFECTIONS OF DEEP NECK STRUCTURES

Deep neck infections are usually extensions of infection from other primary sites, most often within the pharynx or oral cavity. Many of these infections are life-threatening but are difficult to detect at early stages, when they may be more easily managed. Three of the most clinically relevant spaces in the neck are the submandibular (and sublingual) space, the lateral pharyngeal (or parapharyngeal) space, and the retropharyngeal space. These spaces communicate with one another and with other important structures in the head, neck, and thorax, providing pathogens with easy access to areas that include the mediastinum,

carotid sheath, skull base, and meninges. Once infection reaches these sensitive areas, mortality rates can be as high as 20–50%.

Infection of the submandibular and/or sublingual space typically originates from an infected or recently extracted lower tooth. The result is the severe, life-threatening infection referred to as Ludwig angina (see “Oral Infections,” above). Infection of the lateral pharyngeal (or parapharyngeal) space is most often a complication of common infections of the oral cavity and upper respiratory tract, including tonsillitis, peritonsillar abscess, pharyngitis, mastoiditis, and periodontal infection. This space, situated deep in the lateral wall of the pharynx, contains a number of sensitive structures, including the carotid artery, internal jugular vein, cervical sympathetic chain, and portions of cranial nerves IX through XII; at its distal end, it opens into the posterior mediastinum. Involvement of this space with infection can therefore be rapidly fatal. Examination may reveal some tonsillar displacement, trismus, and neck rigidity, but swelling of the lateral pharyngeal wall can easily be missed. The diagnosis can be confirmed by CT. Treatment consists of airway management, operative drainage of fluid collections, and at least 10 days of IV therapy with an antibiotic active against streptococci and oral anaerobes (e.g., ampicillin/sulbactam). A particularly severe form of this infection involving the components of the carotid sheath (poststernal septicemia, Lemierre disease) is described above (see “Oral Infections”). Infection of the retropharyngeal space also can be extremely dangerous, as this space runs posterior to the pharynx from the skull base to the superior mediastinum. Infections in this space are more common among children <5 years old because of the presence of several small retropharyngeal lymph nodes that typically atrophy by age 4 years. Infection is usually a consequence of extension from another site of infection—most commonly, acute pharyngitis. Other sources include otitis media, tonsillitis, dental infections, Ludwig angina, and anterior extension of vertebral osteomyelitis. Retropharyngeal space infection also can follow penetrating trauma to the posterior pharynx (e.g., from an endoscopic procedure). Infections are commonly polymicrobial, involving a mixture of aerobes and anaerobes; group A β -hemolytic streptococci and *S. aureus* are the most common pathogens. *M. tuberculosis* was a common cause in the past but now is rarely involved in the United States.

Patients with retropharyngeal abscess typically present with sore throat, fever, dysphagia, and neck pain and are often drooling because of difficulty and pain with swallowing. Examination may reveal tender cervical adenopathy, neck swelling, and diffuse erythema and edema of the posterior pharynx as well as a bulge in the posterior pharyngeal wall that may not be obvious on routine inspection. A soft-tissue mass is usually demonstrable by lateral neck radiography or CT. Because of the risk of airway obstruction, treatment begins with securing of the airway, which is followed by a combination of surgical drainage and IV antibiotic administration. Initial empirical therapy should cover streptococci, oral anaerobes, and *S. aureus*; ampicillin/sulbactam, clindamycin plus ceftriaxone, or meropenem is usually effective. Complications result primarily from extension to other areas (e.g., rupture into the posterior pharynx may lead to aspiration pneumonia and empyema). Extension may also occur to the lateral pharyngeal space and mediastinum, resulting in mediastinitis and pericarditis, or into nearby major blood vessels. All these events are associated with a high mortality rate.

FURTHER READING

- BROOK I: Microbiology of chronic rhinosinusitis. *Eur J Clin Microbiol Infect Dis* 35:1059, 2016.
- FLETCHER-LARTEY S et al: Why do general practitioners prescribe antibiotics for upper respiratory tract infections to meet patient expectations: A mixed methods study. *BMJ Open* 6:e012244, 2016.
- JENSEN A et al: *Fusobacterium necrophorum* tonsillitis: An important cause of tonsillitis in adolescents and young adults. *Clin Microbiol Infect* 21:266.e1, 2015.
- LEE GC et al: Outpatient antibiotic prescribing in the United States: 2000 to 2010. *BMC Med* 12:96, 2014.

32

Oral Manifestations of Disease

Samuel C. Durso



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. A2). 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 asymptotically 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. 172). Elderly patients, patients with diabetes mellitus, and patients taking glucocorticoids may experience little or no pain or fever as these complications develop.

Periodontal Disease 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 atherosclerosis, 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*. Stress and poor oral hygiene are risk factors. 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. 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-Lefèvre 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.

Pregnancy may produce gingivitis and localized *pyogenic granulomas*. Severe periodontal disease occurs in uncontrolled diabetes mellitus. *Gingival hyperplasia* may be caused by phenytoin, calcium channel blockers (e.g., nifedipine), and cyclosporine, though 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, though 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 32-1).

Pigmented Lesions See Table 32-2.

Dermatologic Diseases See Tables 32-1, 32-2, and 32-3 and Chaps. 52–57.

Diseases of the Tongue See Table 32-4.

HIV Disease and AIDS See Tables 32-1, 32-2, 32-3, and 32-5; Chap. 197; and Fig. 189-3.

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 32-1). 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 *Behcet's syndrome* (Chap. 357). 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. 353). Aphthous-like ulcers are seen in *Crohn's disease* (Chap. 319), 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 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.

TABLE 32-1 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 virus)	Oral mucosa	Fatigue, sore throat, malaise, fever, and cervical lymphadenopathy; numerous small ulcers usually appear several days before lymphadenopathy; gingival bleeding and multiple petechiae at junction of hard and soft palates	Oral lesions disappear during convalescence; no treatment is given, though glucocorticoids 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	Gummatus 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	Gummatus 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 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

(Continued)

TABLE 32-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa (Continued)

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Bacterial or Fungal Diseases (Continued)			
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 mouth, soft palate, lateral and ventral tongue)	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 5–15 mm (major)	Lesions heal in 1–2 weeks but may recur 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

^aSee Table 32-3.

Abbreviations: CNS, central nervous system; EM, erythema multiforme; HSV, herpes simplex virus; VZV, varicella-zoster virus.

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 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 B₁₂ deficiency, iron deficiency, diabetes

TABLE 32-2 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
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

TABLE 32-3 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	<p>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</p> <p>Erythematous type: flat, red, sometimes sore areas in same groups of patients</p> <p>Candidal leukoplakia: nonremovable white thickening of epithelium due to <i>Candida</i></p> <p>Angular cheilitis: sore fissures at corner of mouth</p>	<p>Responds favorably to antifungal therapy and correction of predisposing causes where possible</p> <p>Course same as for pseudomembranous type</p> <p>Responds to prolonged antifungal therapy</p> <p>Responds to topical antifungal therapy</p>
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 <i>Candida</i>
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

TABLE 32-4 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

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.

TABLE 32-5 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 <i>Mycobacterium tuberculosis</i> or <i>Mycobacterium avium-intracellulare</i> Fungal infection (histoplasmosis, cryptococcosis, candidiasis, geotrichosis, aspergillosis) Bacterial infection (<i>Escherichia coli</i> , <i>Enterobacter cloacae</i> , <i>Klebsiella pneumoniae</i> , <i>Pseudomonas aeruginosa</i>) 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

^aStrongly associated with HIV infection.

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 is 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 comprises 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 as post-resection for certain histologic types with a high risk of recurrence. Malignant salivary gland tumors have a 5-year survival rate of ~68%.

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 anti-coagulation 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 in increasingly common use. Simple extractions of 1–3 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. 123**) 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 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.
- ELAD S et al: Novel anticoagulants: General overview and practical considerations for dental practitioners. *Oral Dis* 22:23, 2016.
- SOLLECTO TP et al: The use of prophylactic antibiotics prior to dental procedures in patients with prosthetic joints: Evidence-based guidelines for dental practitioners—a report of the American Dental Association Council on Scientific Affairs. *J Am Dent Assoc* 146:11, 2015.

Section 5 Alterations in Circulatory and Respiratory Functions

33

Dyspnea

Rebecca M. Baron



DYSPNEA

Definition: The American Thoracic Society consensus statement defines *dyspnea* as a “subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity. The experience derives from interactions among multiple physiologic, psychological, social, and environmental factors and may induce secondary physiological and behavioral responses.” Dyspnea, a symptom, can be perceived only by the person experiencing it and, therefore, must be self-reported. In contrast, signs of increased work of breathing, such as tachypnea, accessory muscle use, and intercostal retraction, can be measured and reported by clinicians.

Epidemiology: Dyspnea is a common, and it has been reported that up to one half of inpatients and one quarter of ambulatory patients experience dyspnea, with a prevalence of 9–13% in the community that increases to as high as 37% for adults aged ≥70 years. Dyspnea is a frequent cause for emergency room visits, accounting for as many as 3–4 million visits per year. Furthermore, it is increasingly appreciated that the degree of dyspnea may better predict outcomes in chronic obstructive pulmonary disease (COPD) than does the forced expiratory volume in 1 s (FEV1), and formal measures of dyspnea have been incorporated into the Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2017 COPD severity assessment guidelines. Dyspnea may also predict outcomes in other chronic heart and lung diseases as well. Dyspnea can arise from a diverse array of pulmonary, cardiac, and neurologic underlying causes, and elucidation of particular symptoms may point toward a specific etiology and/or mechanism driving dyspnea (although additional diagnostic testing is often required as will be further discussed below).

MECHANISMS UNDERLYING DYSPNEA

The mechanisms underlying dyspnea are complex, as it can arise from different contributory respiratory sensations. While a large body of research has increased our understanding of mechanisms underlying particular respiratory sensations such as “chest tightness” or “air hunger” it is likely that a given disease state might produce the sensation of dyspnea via more than one underlying mechanism. Dyspnea can arise from a variety of pathways, including generation of *afferent* signals from the respiratory system to the central nervous system (CNS), *efferent* signals from the CNS to the respiratory muscles, and particularly when there is a mismatch in the integrative signaling between these two pathways, termed “efferent-reafferent mismatch” (Fig. 33-1).

Afferent signals trigger the CNS (brainstem and/or cortex) and include primarily: (a) peripheral chemoreceptors in the carotid body and aortic arch and central chemoreceptors in the medulla that are activated by hypoxemia, hypercapnia, or academia, and might produce a sense of “air hunger”; and (b) mechanoreceptors in the upper airways, lungs (including stretch receptors, irritant receptors, and J receptors), and chest wall (including muscle spindles as stretch receptors and tendon organs that monitor force generation) that are activated in the setting of an increased work load from a disease state producing an increase in airway resistance that may be associated with symptoms of chest tightness (e.g., asthma or COPD) or decreased lung or chest wall compliance (e.g., pulmonary fibrosis). Other afferent signals that trigger dyspnea within the respiratory system can arise from pulmonary vascular receptor responses to changes in pulmonary artery pressure and skeletal muscle (termed metaboreceptors) that are believed to sense changes in the biochemical environment.

Efferent signals are sent from the CNS (motor cortex and brainstem) to the respiratory muscles, and are also transmitted by corollary discharge to the sensory cortex that are believed to underlie sensations of respiratory effort (or “work of breathing”) and perhaps contribute to sensations of “air hunger,” especially in response to an increased ventilatory load in a disease state such as COPD. In addition, fear or anxiety may heighten the sense of dyspnea through exacerbating the underlying physiologic disturbance in response to an increased respiratory rate or disordered breathing pattern.

ASSESSING DYSPNEA

While it is well appreciated that dyspnea is a difficult quality to reliably measure due to multiple relevant possible domains that can be measured (e.g., sensory-perceptual experience, affective distress, and symptom impact or burden), and there exist no uniformly agreed upon tools for dyspnea assessment, consensus opinion is that dyspnea should be formally assessed in a context most relevant and beneficial for patient management; furthermore, that the specific domains being measured are adequately described. There are a number of emerging tools that have been developed for formal dyspnea assessment. As an example, the GOLD 2017 criteria advocate use of a dyspnea assessment tool such as the Modified Medical Research Council Dyspnea Scale (MMRC, Table 33-1) to assess symptom/impact burden in COPD.

DIFFERENTIAL DIAGNOSIS

This chapter focuses largely on chronic dyspnea, which is defined as symptoms lasting longer than 1 month and can arise from a broad array of different underlying conditions, most commonly attributable to pulmonary or cardiac conditions that account for as many as 85% of the underlying causes of dyspnea. However, as many as one-third of patients may have multifactorial reasons underlying dyspnea. Examples of a wide array of conditions that underlie dyspnea with possible mechanisms underlying the presenting symptoms are described in Table 33-2.

Respiratory system causes include diseases of the airways (e.g., asthma and COPD), diseases of the parenchyma (more commonly interstitial lung diseases are seen in the setting of chronic dyspnea, but alveolar filling processes, such as hypersensitivity pneumonitis or bronchiolitis obliterans organizing pneumonia [BOOP], can also present with similar symptoms), diseases affecting the chest wall (e.g., bony abnormalities such as kyphoscoliosis, or neuromuscular weakness conditions such as amyotrophic lateral sclerosis), and diseases affecting the pulmonary vasculature (e.g., pulmonary hypertension that can arise from a variety of underlying causes, or chronic thromboembolic disease). Diseases affecting the cardiovascular system that can present with dyspnea include processes affecting left heart function, such as coronary artery disease and cardiomyopathy, as well as disease processes affecting the pericardium, including restrictive pericarditis and cardiac tamponade. Other conditions underlying dyspnea that might not directly emanate from the pulmonary or cardiovascular systems include anemia (thereby potentially affecting oxygen-carrying capacity), deconditioning, and psychological processes such as anxiety. Distinguishing between the myriad of underlying processes that might present with dyspnea can be challenging. A graded approach that begins with a history and physical examination, followed by selected laboratory testing that might then advance to additional diagnostics and potentially subspecialty referral may help elucidate the underlying cause of dyspnea. However, a substantial proportion of patients may have persistent dyspnea despite treatment for an underlying process, or may not have a specific underlying process identified that is driving the dyspnea.

APPROACH TO THE PATIENT

Dyspnea (See Fig. 33-2)

OVERALL

For patients with a known prior pulmonary, cardiac, or neuromuscular condition and worsening dyspnea, the initial focus of the evaluation will usually address determining whether the known condition has progressed or whether a new process has developed

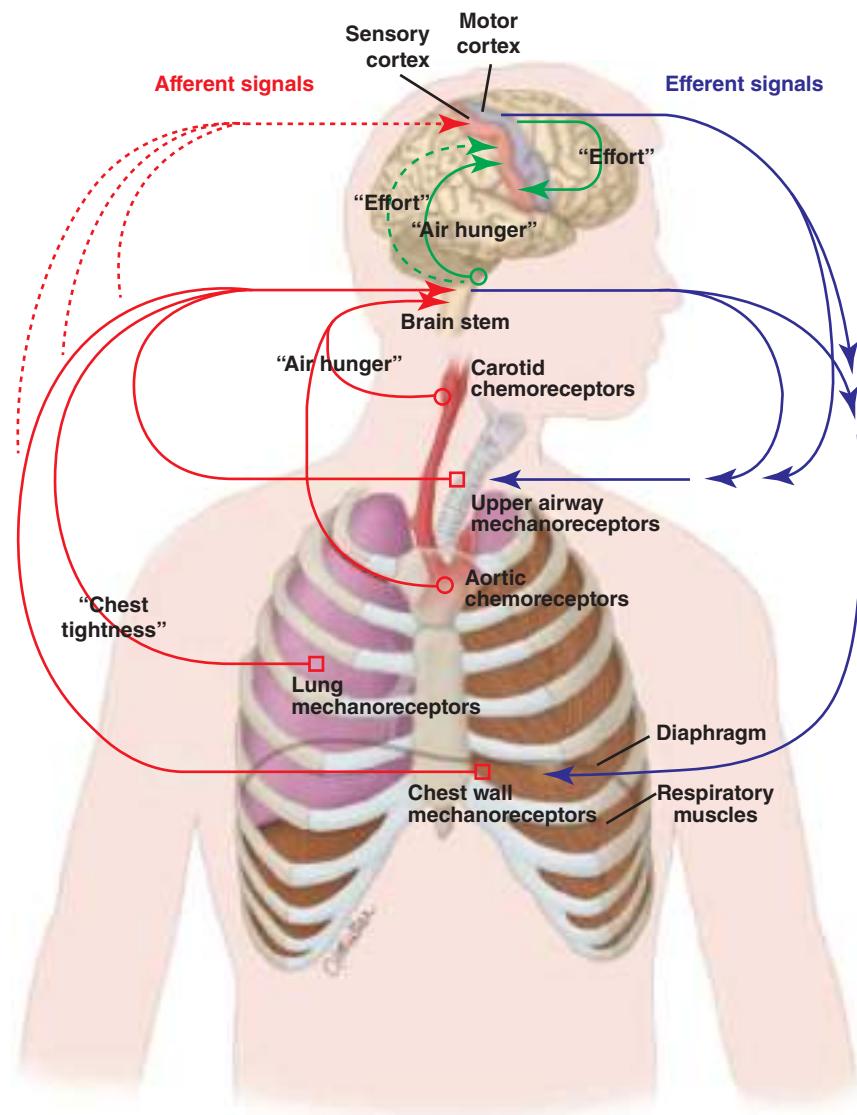


FIGURE 33-1 Signalling pathways underlying dyspnea. Dyspnea arises from a range of sensory inputs, many of which lead to distinct descriptive phrases used by patients (shown in quotes in the figure). The sensation of respiratory effort likely arises from signals transmitted from the motor cortex to the sensory cortex (green arrow) when outgoing motor commands are sent to the ventilatory muscles (efferent signals, blue arrow). Motor output from the brain stem (blue arrow) may also be accompanied by signals transmitted to the sensory cortex and contribute to the sensation of effort (dotted green arrow). The sensation of air hunger probably derives from a combination of stimuli that increase the drive to breathe such as hypoxemia or hypercapnia (mediated by signals from chemoreceptors in the carotid body and aortic arch, indicated by afferent signals in red), acute hypercapnia or acidemia (mediated by signals from the peripheral and central chemoreceptors, indicated by afferent signals in red), airway and interstitial inflammation (mediated by pulmonary afferents, indicated by afferent signals in red), and pulmonary vascular receptors. Dyspnea arises in part from a perceived mismatch between the outgoing efferent messages to the ventilatory muscles and incoming afferent signals from the lungs and chest wall. Chest tightness, often associated with bronchospasm, is largely mediated by stimulation of vagal-irritant receptors. Afferent signals (red arrows) from airway, lung, and chest wall mechanoreceptors most likely pass through the brain stem before being transmitted to sensory cortex, although it is also possible that some afferent information bypasses the brain stem and goes directly to sensory cortex (dotted arrow).

Red arrows and text: afferent signals; Blue arrows and text: efferent signals; Green arrows: signals within the central nervous system; Dotted lines: hypothetical pathways; Hollow Red Circles: chemoreceptors; Hollow Red Squares: mechanoreceptors. (Adapted from UpToDate 2017.)

that is causing dyspnea. For patients without a prior known potential cause of dyspnea, the initial evaluation will focus on determining an underlying etiology. Determining the underlying cause, if possible, is extremely important, as the treatment may vary dramatically based upon the predisposing condition. An initial history and physical examination remain fundamental to the evaluation followed by initial diagnostic testing as indicated that might prompt subspecialty referral (e.g., pulmonary, cardiology, neurology, sleep, and/or specialized dyspnea clinic) if the cause of dyspnea remains elusive (Fig. 33-2). As many as two-thirds of patients will require diagnostic testing beyond the initial clinical presentation.

HISTORY

The patient should be asked to describe in his/her own words what the discomfort feels like as well as the effect of position, infections,

and environmental stimuli on the dyspnea, as descriptors may be helpful in pointing toward an etiology. For example, symptoms of chest tightness might suggest the possibility of bronchoconstriction, and the sensation of inability to take a deep breath may correlate with dynamic hyperinflation from COPD. Orthopnea is a common indicator of congestive heart failure (CHF), mechanical impairment of the diaphragm associated with obesity, or asthma triggered by esophageal reflux. Nocturnal dyspnea suggests CHF or asthma. Acute, intermittent episodes of dyspnea are more likely to reflect episodes of myocardial ischemia, bronchospasm, or pulmonary embolism, while chronic persistent dyspnea is more typical of COPD, interstitial lung disease, and chronic thromboembolic disease. Information on risk factors for drug-induced or occupational lung disease and for coronary artery disease should be elicited. Left atrial myxoma or hepatopulmonary syndrome should be considered

TABLE 33-1 An Example of a Clinical Method for Rating Dyspnea: The Modified Medical Research Council Dyspnea Scale^a

GRADE OF DYSPNEA	DESCRIPTION
0	Not troubled by breathlessness, except with strenuous exercise
1	Shortness of breath walking on level ground or with walking up a slight hill
2	Walks slower than people of similar age on level ground due to breathlessness, or has to stop to rest when walking at own pace on level ground
3	Stops to rest after walking 100 m or after walking a few minutes on level ground
4	Too breathless to leave the house, or breathless with activities of daily living (e.g., dressing/undressing)

^aWhich has been incorporated into the GOLD 2017 guidelines as a possible tool for rating dyspnea in COPD.

Source: Modified from DA Mahler, CK Wells: Evaluation of clinical methods for rating dyspnea. *Chest* 93:580, 1988.

when the patient complains of *platypnea*—i.e., dyspnea in the upright position with relief in the supine position.

PHYSICAL EXAMINATION

Initial vital signs might be helpful in pointing toward an underlying etiology in the context of the remainder of the evaluation. For

example, the presence of fever might point toward an underlying infectious or inflammatory process; the presence of hypertension in the setting of a heart failure might point toward diastolic dysfunction; the presence of tachycardia might be associated with many different underlying processes including fever, cardiac dysfunction, and deconditioning; and the presence of resting hypoxemia suggests processes involving hypercapnia, ventilation-perfusion mismatch, shunt, or impairment in diffusion capacity might be involved. An exertional oxygen saturation should also be obtained as described below. The physical examination should begin during the interview of the patient. Inability of the patient to speak in full sentences before stopping to get a deep breath suggests a condition that leads to stimulation of the controller or impairment of the ventilatory pump with reduced vital capacity. Evidence of increased work of breathing (suprACLAVICULAR retractions; use of accessory muscles of ventilation; and the tripod position, characterized by sitting with the hands braced on the knees) is indicative of increased airway resistance or stiffness of the lungs and the chest wall. When measuring the vital signs, the physician should accurately assess the respiratory rate and measure the pulsus paradoxus (**Chap. 265**); if the systolic pressure decreases by >10 mmHg, the presence of COPD, acute asthma, or pericardial disease should be considered. During the general examination, signs of anemia (pale conjunctivae), cyanosis, and cirrhosis (spider angiomas, gynecomastia) should be sought. Examination of the chest should focus on symmetry of movement; percussion (dullness is indicative of pleural effusion; hyperresonance is a sign of

TABLE 33-2 Differential Diagnosis of Disease Processes Underlying Dyspnea

SYSTEM	TYPE OF PROCESS	EXAMPLE OF DISEASE PROCESS	POSSIBLE PRESENTING DYSPNEA SYMPTOMS	POSSIBLE PHYSICAL FINDINGS	POSSIBLE MECHANISMS UNDERLYING DYSPNEA	INITIAL DIAGNOSTIC STUDIES (AND POSSIBLE FINDINGS)
Pulmonary	Airways disease	Asthma, COPD	Chest tightness, tachypnea, increased WOB, air hunger, inability to get a deep breath	Wheezing, accessory muscle use, exertional hypoxemia (especially with COPD)	Increased WOB, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Peak flow (reduced); Spirometry (OVD); CXR (hyper-inflation; loss of lung parenchyma in COPD)
	Parenchymal disease	Interstitial lung disease ^a	Air hunger, inability to get a deep breath	Dry end-inspiratory crackles, clubbing, exertional hypoxemia	Increased WOB, increased respiratory drive, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Spirometry and lung volumes (RVD); CXR and chest CT (interstitial lung disease)
	Chest wall disease	Kyphoscoliosis, Neuromuscular (NM) weakness	Increased WOB, inability to get a deep breath	Decreased diaphragm excursion; atelectasis	Increased WOB; stimulation of pulmonary receptors (if atelectasis is present)	Spirometry and lung volumes (RVD); MIP and MEPs (reduced in NM weakness)
Pulmonary and cardiac	Pulmonary vasculature	Pulmonary Hypertension	Tachypnea	Elevated R heart pressures, exertional hypoxemia	Increased respiratory drive, hypoxemia, stimulation of vascular receptors	Diffusion capacity (reduced); ECG; ECHO (to evaluate PA pressures) ^b
Cardiac	Left heart failure	Coronary artery disease, cardio-myopathy ^c	Chest tightness, air hunger	Elevated L heart pressures; wet crackles on lung examination; pulsus paradoxus (pericardial disease)	Increased WOB and drive, hypoxemia, stimulation of vascular and pulmonary receptors ^d	Consider BNP testing in the acute setting; ECG, ECHO, may need stress testing and/or LHC
	Pericardial disease	Restrictive pericarditis; Cardiac tamponade				
Other	Variable	Anemia Deconditioning Psychological	Exertional breathlessness Poor fitness Anxiety	Variable	Metabo-receptors (anemia, poor fitness); chemoreceptors (anaerobic metabolism from poor fitness); some subjects may have increased sensitivity to hypercapnia	Hematocrit for anemia; exclude other causes

^aDifferential diagnosis of interstitial lung disease includes idiopathic pulmonary fibrosis, collagen vascular disease, drug or occupation-induced pneumonitis, lymphangitic spread of malignancy; processes that are more alveolar rather than interstitial in nature can also less commonly contribute to parenchymal lung disease underlying chronic dyspnea and include entities such as hypersensitivity pneumonitis, bronchiolitis obliterans organizing pneumonia, etc. ^bWould additionally consider these patients for CT angiography to evaluate for presence of thromboemboli, ventilation/perfusion scanning to evaluate for the presence of chronic thromboembolic disease, and right heart catheterization (RHC) to further evaluate pulmonary hypertension. ^cDiastolic dysfunction in the setting of a stiff left ventricle is often seen and contributes significantly to insidious dyspnea that can be difficult to treat. ^dMay stimulate metaboreceptors if cardiac output is sufficiently reduced to a result in a lactic acidosis.

Abbreviations: BNP brain natriuretic peptide; COPD, chronic obstructive pulmonary disease; CT, computed tomography; CT angio, CT angiography; CXR, chest x-ray; ECHO, echocardiogram; ECG, electrocardiogram; LHC, left heart catheterization; MIP/MEP, maximal inspiratory and maximal expiratory pressures (obtained in the PFT laboratory); OVD, obstructive ventilatory defect; RVD, restrictive ventilatory defect; WOB, work of breathing.

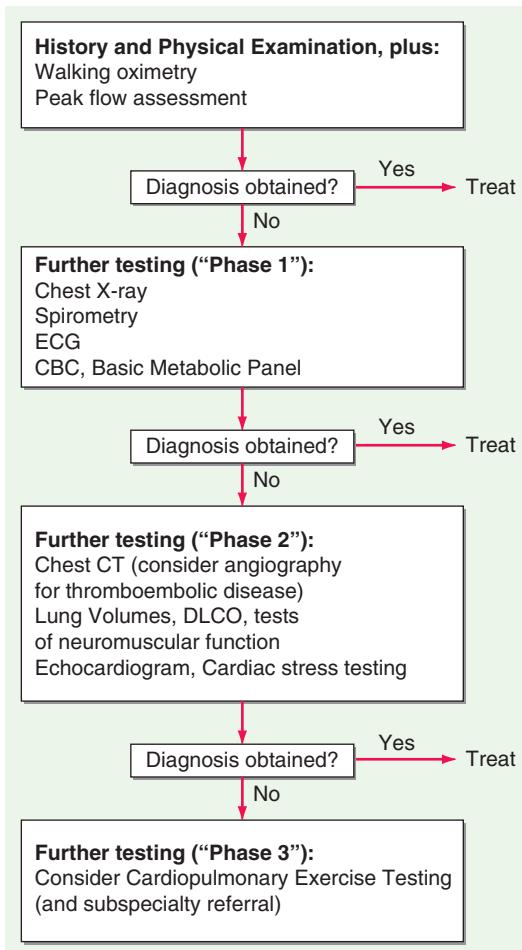


FIGURE 33-2 Possible algorithm for the evaluation of the patient with dyspnea. As described in the text, the approach should begin with a detailed history and physical examination, followed by progressive testing and ultimately more invasive testing and subspecialty referral as is indicated to determine the underlying cause of dyspnea. (Adapted from NG Karnani et al: *Am Fam Physician* 71:1529, 2005.)

emphysema); and auscultation (wheezes, rhonchi, prolonged expiratory phase, and diminished breath sounds are clues to disorders of the airways; rales suggest interstitial edema or fibrosis). The cardiac examination should focus on signs of elevated right heart pressures (jugular venous distention, edema, accentuated pulmonic component to the second heart sound); left ventricular dysfunction (S3 and S4 gallops); and valvular disease (murmurs). When examining the abdomen with the patient in the supine position, the physician should note whether there is paradoxical movement of the abdomen as well as the presence of increased respiratory distress in the supine position: inward motion during inspiration is a sign of diaphragmatic weakness, and rounding of the abdomen during exhalation is suggestive of pulmonary edema. Clubbing of the digits may be an indication of interstitial pulmonary fibrosis or bronchiectasis, and joint swelling or deformation as well as changes consistent with Raynaud's disease may be indicative of a collagen-vascular process that can be associated with pulmonary disease.

Patients should be asked to walk under observation with oximetry in order to reproduce the symptoms. The patient should be examined during and at the end of exercise for new findings that were not present at rest (e.g., presence of wheezing), and for changes in oxygen saturation.

CHEST IMAGING

After the history elicitation and the physical examination, a chest radiograph should be obtained if the diagnosis remains elusive. The

lung volumes should be assessed: hyperinflation is consistent with obstructive lung disease, whereas low lung volumes suggest interstitial edema or fibrosis, diaphragmatic dysfunction, or impaired chest wall motion. The pulmonary parenchyma should be examined for evidence of interstitial disease, infiltrates, and emphysema. Prominent pulmonary vasculature in the upper zones indicates pulmonary venous hypertension, while enlarged central pulmonary arteries may suggest pulmonary arterial hypertension. An enlarged cardiac silhouette can point toward dilated cardiomyopathy or valvular disease. Bilateral pleural effusions are typical of CHF and some forms of collagen-vascular disease. Unilateral effusions raise the specter of carcinoma and pulmonary embolism but may also occur in heart failure or in the case of a parapneumonic effusion. CT of the chest is generally reserved for further evaluation of the lung parenchyma (interstitial lung disease) and possible pulmonary embolism if there remains diagnostic uncertainty.

LABORATORY STUDIES

Initial laboratory testing should include a hematocrit to exclude occult anemia as an underlying cause of reduced oxygen-carrying capacity contributing to dyspnea, and a basic metabolic panel may be helpful to exclude a significant underlying metabolic acidosis (and conversely, an elevated bicarbonate might point toward the possibility of carbon dioxide retention that might be seen in chronic respiratory failure—in such a setting, an arterial blood gas may provide useful additional information). Additional laboratory studies should include electrocardiography to seek evidence of ventricular hypertrophy and prior myocardial infarction and spirometry that can be diagnostic of the presence of an obstructive ventilatory defect, and suggest the possibility of a restrictive ventilatory defect (that then might prompt additional pulmonary function laboratory testing, including lung volumes, diffusion capacity, and possible tests of neuromuscular function). Echocardiography is indicated when systolic dysfunction, pulmonary hypertension, or valvular heart disease is suspected. Bronchoprovocation testing and/or home peak-flow monitoring may be useful in patients with intermittent symptoms suggestive of asthma who have a normal physical examination and spirometry; up to one-third of patients with the clinical diagnosis of asthma do not have reactive airways disease when formally tested. Measurement of brain natriuretic peptide levels in serum is increasingly used to assess for CHF in patients presenting with acute dyspnea but may be elevated in the presence of right ventricular strain as well.

DISTINGUISHING CARDIOVASCULAR FROM RESPIRATORY SYSTEM DYSPNEA

If a patient has evidence of both pulmonary and cardiac disease that is either not responsive to treatment, or it remains unclear what factors are primarily driving dyspnea, a cardiopulmonary exercise test (CPET) can be carried out to determine which system is responsible for the exercise limitation. CPET includes incremental symptom-limited exercise (cycling or treadmill) with measurements of ventilation and pulmonary gas exchange, and in some cases includes non-invasive and invasive measures of pulmonary vascular pressures and cardiac output. If, at peak exercise, the patient achieves predicted maximal ventilation, demonstrates an increase in dead space or hypoxemia, or develops bronchospasm, the respiratory system may be the cause of the problem. Alternatively, if the heart rate is >85% of the predicted maximum, if the anaerobic threshold occurs early, if the blood pressure becomes excessively high or decreases during exercise, if the O₂ pulse (O₂ consumption/heart rate, an indicator of stroke volume) falls, or if there are ischemic changes on the electrocardiogram, an abnormality of the cardiovascular system is likely the explanation for the breathing discomfort. Additionally, a CPET may also help point toward a peripheral extraction deficit, or metabolic/neuromuscular disease as potential underlying processes driving dyspnea.

TREATMENT

Dyspnea

The first goal is to correct the underlying condition(s) driving dyspnea and address potentially reversible causes with appropriate treatment for the particular condition. Multiple different interventions may be necessary, given that dyspnea often arises from multifactorial causes. If relief of dyspnea with treatment of the underlying condition(s) is not fully possible, an effort is made to lessen the intensity of the symptom and its effect on the patient's quality of life. Despite an increased understanding of the mechanisms underlying dyspnea, there has been limited progress in treatment strategies for dyspnea. Supplemental O₂ should be administered if the resting O₂ saturation is ≤88% or if the patient's saturation drops to these levels with activity or sleep. In particular, for patients with COPD, supplemental oxygen for those with hypoxemia has been shown to improve mortality, and pulmonary rehabilitation programs have demonstrated positive effects on dyspnea, exercise capacity, and rates of hospitalization. Opioids have been shown to reduce symptoms of dyspnea, largely through reducing air hunger, thus, likely suppressing respiratory drive and influencing cortical activity. However, opioids should be considered for each patient individually based upon the risk-benefit profile as regards the effects of respiratory depression. Studies of anxiolytics for dyspnea have not demonstrated consistent benefit. Additional approaches are under study for dyspnea, including inhaled furosemide that might alter afferent sensory information.

ACKNOWLEDGMENT

With prior contributions from Richard M. Schwartzstein.

FURTHER READING

- BANZETT RB et al: Multidimensional dyspnea profile: An instrument for clinical and laboratory research. *Eur Respir J* 45:1681, 2015.
- LAVIOLETTE L, LAVENEZIANA P ON BEHALF OF THE ERS RESEARCH SEMINAR FACULTY: Dyspnoea: A multidimensional and multidisciplinary approach. *Eur Respir J* 43:1750, 2014.
- PARSHALL MB et al: An Official American Thoracic Society Statement: Update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 185:435, 2012.
- WAHLS SA: Causes and evaluation of chronic dyspnea. *Am Fam Physician* 86:173, 2012.

34

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

Spontaneous cough is triggered by stimulation of sensory nerve endings that are thought to be primarily rapidly adapting receptors and

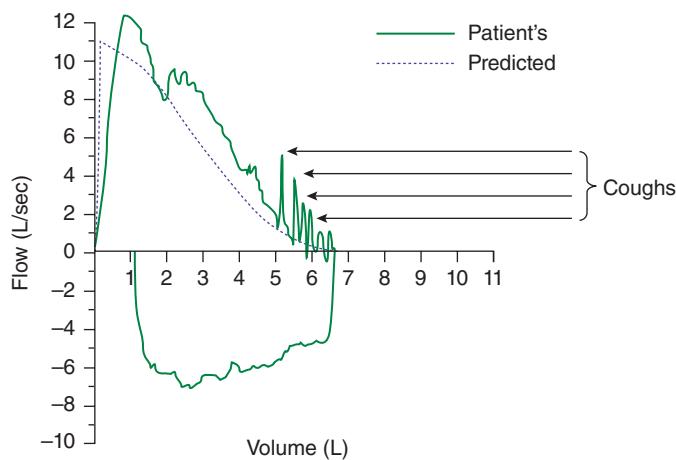


FIGURE 34-1 Flow-volume curve shows spikes of high expiratory flow achieved with cough.

C fibers. Both chemical (e.g., capsaicin) and mechanical (e.g., particulates in air pollution) stimuli may initiate the cough reflex. A cationic ion channel—the transient receptor potential vanilloid 1 (TRPV1)—found on rapidly adapting receptors and C fibers is the receptor for capsaicin, and its expression is increased in patients with chronic cough. Afferent nerve endings richly innervate the pharynx, larynx, and airways to the level of the terminal bronchioles and extend into the lung parenchyma. They may also be located in the external auditory meatus (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 vaguely identified as the “cough center.” The cough reflex involves a highly orchestrated series of involuntary muscular actions, with the potential for input from cortical pathways as well. 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. 34-1). 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.

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 34-1). 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 strength, running the gamut from

TABLE 34-1 Causes of Impaired Cough

- Decreased respiratory muscle strength
- Chest wall or abdominal pain
- Chest wall deformity (e.g., severe kyphoscoliosis)
- Impaired glottic closure or tracheostomy
- Tracheobronchomalacia
- Abnormal airway secretions
- Central respiratory depression (e.g., anesthesia, sedation, or coma)

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 despite a preserved ability to generate normal expiratory velocities; such failure may be due to either abnormal airway secretions (e.g., bronchiectasis due to cystic fibrosis) or structural abnormalities of the airways (e.g., tracheomalacia with excessive expiratory collapse of the trachea during cough).

■ 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, 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 (“post-bronchitic cough”) is a very common cause of chronic cough, especially in the winter months.

■ 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 does the cough produce 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 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 critically important. 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.

■ CHRONIC COUGH WITH A NORMAL CHEST RADIOPHGRAPH

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. Clues suggesting this etiology include 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. 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, non-acid 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, 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.

Chronic eosinophilic bronchitis causes chronic cough with a normal chest radiograph. This condition is characterized by sputum eosinophilia in excess of 3% without airflow obstruction or bronchial hyperresponsiveness and is successfully treated with inhaled glucocorticoids.

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 (H₂) 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 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 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.

■ 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 or hydrocodone, 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. 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 limitations of currently available agents are greatly needed. Approaches that are being explored include the development of neurokinin receptor antagonists, TRPV1 ion channel antagonists, 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.
- GIBSON PG, VERTIGAN AE: Management of chronic refractory cough. *BMJ* 351:h5590, 2015.
- KAHRILAS PJ et al: Chronic cough due to gastroesophageal reflux in adults: CHEST Guideline and Expert Panel Report. *Chest* 150:1341, 2016.
- 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.
- RYAN NM et al: Gabapentin for refractory chronic cough: a randomized, double-blind, placebo-controlled trial. *Lancet* 380:1583, 2012.
- SMITH JA, WOODCOCK A: Chronic cough. *N Engl J Med* 375:1544, 2016.

35

Hemoptysis

Anna K. Brady, Patricia A. Kritek



Hemoptysis is the expectoration of blood from the respiratory tract. The first step in evaluation is to ascertain whether the bleeding is coming from the respiratory tree or instead originating from the nasal cavities (i.e., epistaxis) or the gastrointestinal tract (i.e., hematemesis) as the therapies for these etiologies will be significantly different. Once established as hemoptysis, the exact nature of the expectoration is important as the term can be applied to blood-tinged phlegm, the pink frothy sputum of pulmonary edema, or frank blood. Next steps include identifying the source and etiology of bleeding.

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 often necessary.

A unique feature of the lung that predisposes to hemoptysis of varied severity is its dual blood supply—the pulmonary and bronchial circulations. The former is a low-pressure system that is essential to gas exchange at the alveolar level; in contrast, the bronchial arteries originate from the aorta and are under systemic pressure. The bronchial arteries supply the airways and have the ability to neovascularize tumors, dilate airways of bronchiectasis, and cavitary lesions. Most hemoptysis is due to vessels in the bronchial circulation and is, therefore, under systemic pressure, making it more challenging to arrest the bleeding.

ETIOLOGY

Hemoptysis commonly results from infection, malignancy, or vascular disease; however, the differential for bleeding from the respiratory tree is varied and broad.

Infections Most blood-tinged sputum and small-volume hemoptysis is 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 with exacerbations of disease. 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 non-tuberculous 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, 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 commonly results from pulmonary edema due to elevated left ventricular end-diastolic pressure. While the classic description of the sputum expectorated in pulmonary edema is "pink and frothy," a spectrum of hemoptysis including frank blood can be seen.

A pulmonary embolism with parenchymal infarction can present with hemoptysis, although most pulmonary emboli do not cause hemoptysis and will present with other signs and symptoms. An ectatic vessel in an airway or a pulmonary arteriovenous malformation can be a source of bleeding. While rare, rupture of an aortobronchial fistula can result in massive bleeding and sudden death; these fistulae arise in the setting of aortic pathology such as aneurysm or pseudoaneurysm and can cause small bleeding episodes that herald massive hemoptysis.

Diffuse alveolar hemorrhage (DAH), despite causing significant bleeding into the lung parenchyma, uncommonly results in hemoptysis. 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). DAH more commonly presents with diffuse ground glass opacities on imaging and anemia, so the absence of hemoptysis should not exclude the diagnosis.

Malignancy Bronchogenic carcinoma of any histology is a common cause of hemoptysis (both massive and non-massive) in modern published series. Hemoptysis often indicates 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 first step in evaluating hemoptysis is to determine the amount or severity of bleeding. 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. 35-1.

It is crucial to determine whether the amount of blood expectorated is *massive*; while there is no agreed-upon volume, blood loss of 400 mL in 24 hours or 100–150 mL expectorated at one time are considered *massive hemoptysis*. These numbers derive from the volume of the tracheobronchial tree (generally 100–200 mL). This determination is clinically important as patients rarely die of exsanguination and, instead, 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 mL, so using referents like cups (one U.S. cup is 236 mL) can be helpful. Fortunately, massive hemoptysis only accounts for 5–15% of cases of hemoptysis.

Careful 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. A thorough medical history with careful attention to chronic pulmonary disease should

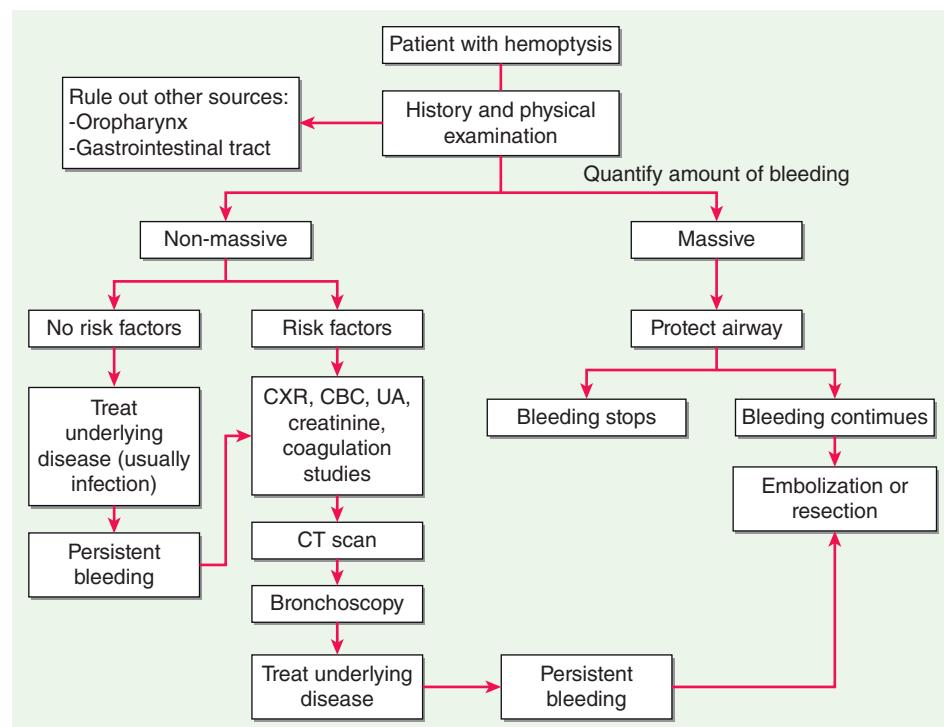


FIGURE 35-1 Approach to the management of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.

be obtained, and the clinician should determine risk factors for malignancy and bronchiectatic lung disease (e.g., cystic fibrosis, sarcoidosis).

Physical Examination Reviewing the vital signs is an important first step. The presence of hypoxemia, tachypnea, and tachycardia should raise concern. Clinicians should examine the nasal and oral cavities; observe the patient's breathing pattern, with careful attention to any respiratory distress; and auscultate the lungs. Clubbing can suggest underlying lung disease such as lung cancer or cystic fibrosis. Signs of bleeding diathesis (e.g., skin or mucosal ecchymoses and petechiae) or teleangiectasias may suggest other predispositions to 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, as well as urinalysis to exclude pulmonary-renal disease.

In patients with small, non-massive hemoptysis, outpatient evaluation can be pursued. All patients with hemoptysis need chest imaging. A chest radiograph is usually obtained first, though it frequently does not localize bleeding and can appear normal. In patients without risk factors for malignancy and with a normal chest radiograph, treating for bronchitis and ensuring close follow-up is a reasonable strategy, with further diagnostic workup if bleeding persists.

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) should be obtained to better identify masses, bronchiectasis, and parenchymal lesions. Following 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.

Interventions When the amount of hemoptysis is massive, there are three simultaneous goals: first, protect the non-bleeding lung; second, locate the site of bleeding; and third, control the bleeding.

Protecting the airway and non-bleeding lung is paramount in the management of massive hemoptysis, since 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 non-bleeding 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 non-bleeding lung either by selective intubation of one lung (i.e., the non-bleeding lung) or insertion of a double-lumen endotracheal tube.

Locating the bleeding site is sometimes obvious, but frequently it can be difficult to determine the source of hemoptysis. A chest radiograph, if it shows new opacities, can be helpful in localizing the side or site of bleeding, though 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, though 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 massive 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 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 cauterization. Because most massive hemoptysis arises from the bronchial circulation, bronchial artery embolization is the procedure of choice for control of massive hemoptysis. It is not

without risk—embolization of the anterior spinal artery is a known complication—but is generally successful in the short term, with >80% success rate at controlling bleeding immediately, though 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.

FURTHER READING

- ADELMAN M et al: Cryptogenic hemoptysis: Clinical features, bronchoscopic findings, and natural history in 67 patients. Ann Int Med 102:829, 1985.
- FLUME PA et al: CF pulmonary guidelines. Pulmonary complications: Hemoptysis and pneumothorax. AJRCCM 182:298, 2010.
- HIRSHBERG B et al: Hemoptysis: Etiology, evaluation, and outcome in a tertiary care hospital. Chest 112:440, 1997.
- JOHNSON JL: Manifestations of hemoptysis: How to manage minor, moderate, and massive bleeding. Postgrad Med 112:4:101, 2002.
- LORDAN JL et al: The pulmonary physician in critical care: Illustrative case 7. Assessment and management of massive hemoptysis. Thorax 58:814, 2003.
- SOPKO DR, SMITH TP: Bronchial artery embolization for massive hemoptysis. Semin Intervent Radiol 28:48, 2011.

36

Hypoxia and Cyanosis

Joseph Loscalzo



HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O₂ and nutrients to cells and to remove CO₂ 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 O₂.

RESPONSES TO HYPOXIA

Decreased O₂ availability to cells 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 Ca²⁺ influx and activation of Ca²⁺-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 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 Ca²⁺ channels raising the cytosolic [Ca²⁺] 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; 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. 33), 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 vasoconstriction 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

Respiratory Hypoxia When hypoxia occurs from respiratory failure, Pao_2 declines, and when respiratory failure is persistent, the hemoglobin-oxygen (Hb-O_2) dissociation curve (see Fig. 94-2) 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 disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air (Fio_2). In this latter situation, Paco_2 falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in SaO_2 at any level of Pao_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 Paco_2 (Chap. 279). 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 (*intraluminal right-to-left shunting*) by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through pulmonary arteriovenous connections. The low Pao_2 in this situation is only partially corrected by an Fio_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 (Fio_2) leads to a decrease in alveolar Po_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, and Eisenmenger's syndrome (Chap. 264). As in pulmonary right-to-left shunting, the Pao_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 Pao_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 Po_2 and saturation in the venous blood decline to a greater extent than normal.

Carbon Monoxide (CO) Intoxication (See also Chap. S11)

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. 94-2) so that O_2 is unloaded only at lower tensions, further contributing to tissue hypoxia.

Circulatory Hypoxia As in anemic hypoxia, the Pao_2 is usually normal, but venous and tissue Po_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 ($a-v\text{-O}_2$ difference), or gradient. Generalized circulatory hypoxia occurs in heart failure (Chap. 252) and in most forms of shock (Chap. 296).

Specific Organ Hypoxia 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. 275). 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. 94-2), 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. 450) and several other similarly acting poisons cause cellular hypoxia. 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. 51).

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 secondary to erythropoietin production (Chap. 99). 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 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. 99) must be distinguished from the true cyanosis discussed here. A cherry-colored flush, rather than cyanosis, is caused by COHb (Chap. 450).

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 Sao_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 Sao_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 the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of Sao_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. 94), is present in blood.

Cyanosis may be subdivided into central and peripheral types. In *central* cyanosis, the Sao_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 or those beneath the tongue 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 (Table 36-1) Decreased Sao_2 results from a marked reduction in the Pao_2 . This reduction may be brought about by a decline in the Fio_2 without sufficient compensatory alveolar hyperventilation to maintain alveolar Po_2 . Cyanosis usually becomes manifest in an ascent to an altitude of 4000 m (13,000 ft).

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. 279). 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 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. 264).

Pulmonary arteriovenous fistulae may be congenital or acquired, solitary or multiple, 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. 94); 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

TABLE 36-1 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

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. 275), 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, 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, asbestos, 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.

37

Edema

Eugene Braunwald, Joseph Loscalzo



PLASMA AND INTERSTITIAL FLUID EXCHANGE

About two-thirds of total body water is intracellular and one-third is extracellular. Approximately 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.

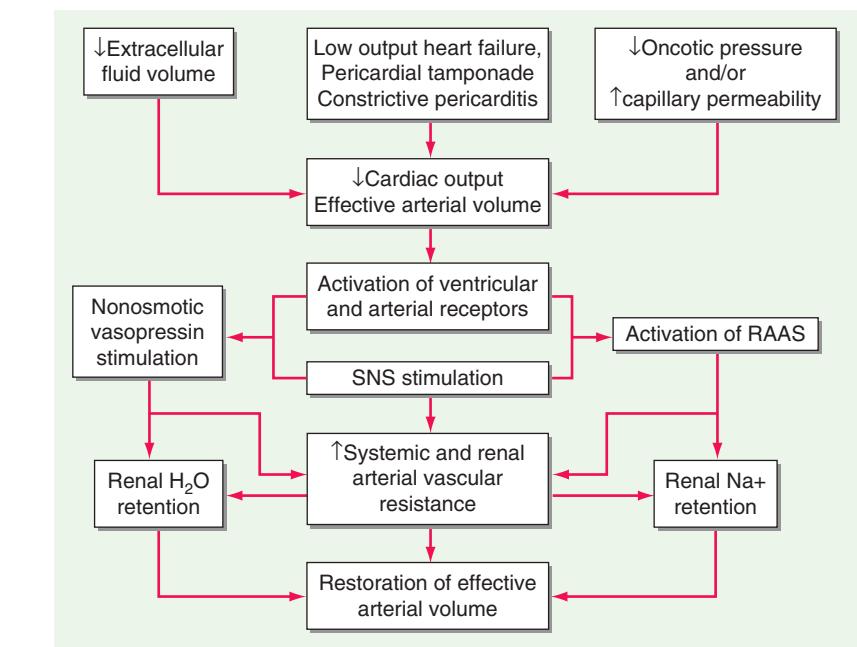
There is constant interchange of fluid between the two compartments of the extracellular fluid. The hydrostatic pressure within the capillaries and the colloid oncotic pressure in the interstitial fluid promote the movement of water and diffusible solutes from plasma to the interstitium. This movement 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

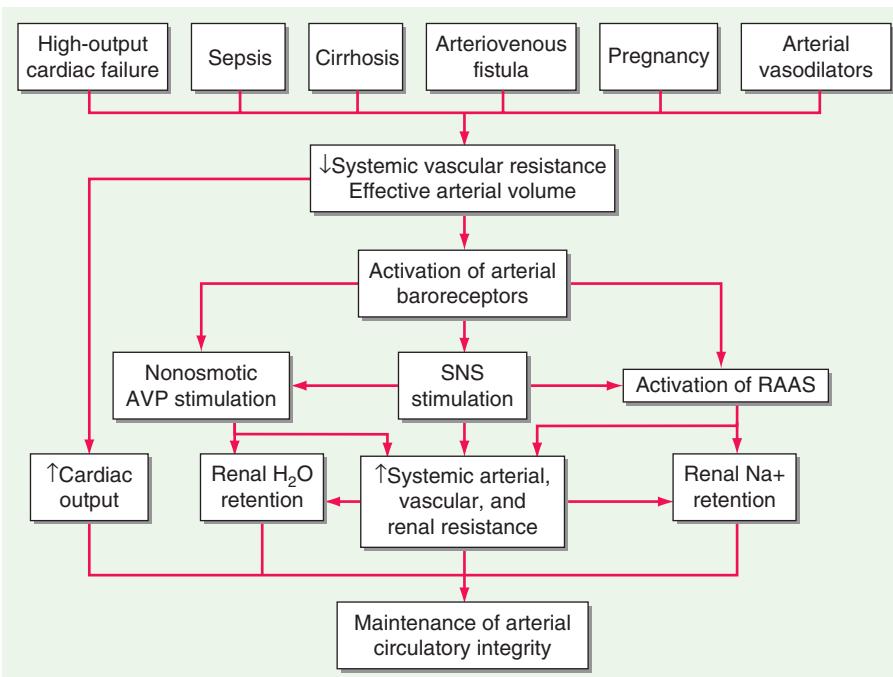
In many forms of edema, the effective arterial blood volume, a parameter that represents the filling of the arterial tree and that effectively perfuses the tissues, is reduced. Underfilling of the arterial tree may be caused by a reduction of cardiac output and/or systemic vascular resistance, by the pooling of blood in the splanchnic veins (as in cirrhosis), and by hypoalbuminemia (Fig. 37-1A). 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.

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 oncotic 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.



A



B

FIGURE 37-1 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. RAAS, renin-angiotensin aldosterone system; SNS, sympathetic nervous system. (Modified from RW Schrier: Ann Intern Med 113:155, 1990.)

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. 374) The secretion of arginine vasopressin (AVP) by the posterior pituitary gland occurs in response to increased intracellular osmolar concentration; by stimulating V₂ 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 distension 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) 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. 46) 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.

■ GENERALIZED EDEMA

The differences among the major causes of generalized edema are shown in **Table 37-1**. 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.

Heart Failure (See also Chap. 252) 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 resulting in edema (Fig. 37-1).

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 such as electrocardiography, echocardiography, and measurements of BNP (or NTproBNP) 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. 308) 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. 337) 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 of albumin, 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. 37-1B). 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. 46). 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 37-2). Mechanisms include renal vasoconstriction (NSAIDs and cyclosporine), arteriolar dilation (vasodilators), augmented renal sodium reabsorption (steroid hormones), and capillary damage.

Edema of Nutritional Origin 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

TABLE 37-1 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 angioma, 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: Modified from GM Chertow: Approach to the patient with edema, in Primary Cardiology, 2nd ed, E Braunwald, L Goldman (eds). Philadelphia, Saunders, 2003, pp 117–128.

TABLE 37-2 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

Progesterins

Cyclosporine

Growth hormone

Immunotherapies

Interleukin 2

OKT3 monoclonal antibody

Source: Modified from GM Chertow: Approach to the patient with edema, in *Primary Cardiology*, 2nd ed, E Braunwald, L Goldman (eds). Philadelphia, Saunders, 2003, pp 117-128.

reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation (*Chap. 326*) (Fig. 37-1B). 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 in 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, and hyperthyroidism (pretibial myxedema secondary to Graves' disease), in which edema is typically nonpitting and, in Graves' disease, exogenous hyperadrenocortism; 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. 48, 304, and 305.**

■ FURTHER READING

CLARK AL, CLELAND JG: Causes and treatment of oedema in patients with heart failure. *Nature Rev Cardiol* 10:156, 2013.

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.

FERRELL RE et al: G/C2 missense mutations cause human lymphedema. *Am J Hum Genet* 86:943, 2010.

FRISON S et al: Omitting edema measurement: How much acute malnutrition are we missing? *Am J Clin Nutr* 102:1176, 2015.

LEVICK JR, MICHEL CC: Microvascular fluid exchange and the revised Starling principle. *Cardiovascular Res* 87:198, 2010.

MORTIMER PS, ROCKSON SG: New developments in clinical aspects of lymphatic disease. *J Clin Invest* 124:915, 2014.

38

Approach to the Patient with a Heart Murmur

Patrick T. O'Gara, Joseph Loscalzo



The differential diagnosis of a heart murmur begins with a careful assessment of its major attributes and response to bedside maneuvers. The history, clinical context, and associated physical examination findings provide additional clues to help establish the significance of a heart murmur. Accurate bedside identification of a heart murmur can inform decisions regarding the indications for noninvasive testing and the need for referral to a cardiovascular specialist. Preliminary discussions can be held with the patient regarding antibiotic or rheumatic fever prophylaxis, the need to restrict various forms of physical activity, and the potential role for family screening.

Heart murmurs are caused by audible vibrations that are due to increased turbulence from accelerated blood flow through normal or abnormal orifices; flow through a narrowed or irregular orifice into a dilated vessel or chamber; or backward flow through an incompetent valve, ventricular septal defect, or patent ductus arteriosus. They traditionally are defined by their timing within the cardiac cycle (*Fig. 38-1*). *Systolic murmurs* begin with or after the first heart sound (S_1) and terminate at or before the component (A_2 or P_2) of the second heart

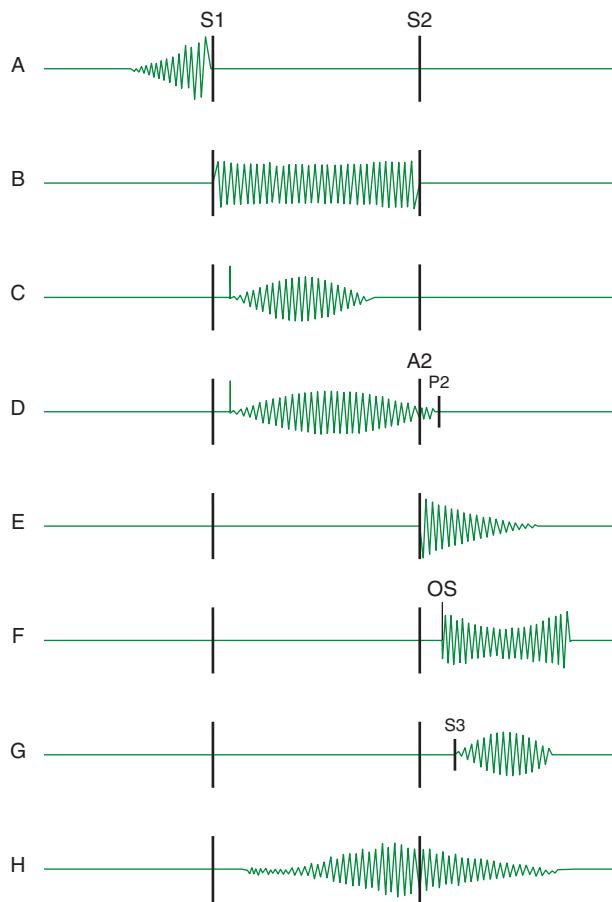


FIGURE 38-1 Diagram depicting principal heart murmurs. **A.** Presystolic murmur of mitral or tricuspid stenosis. **B.** Holosystolic (pansystolic) murmur of mitral or tricuspid regurgitation or of ventricular septal defect. **C.** Aortic ejection murmur beginning with an ejection click and fading before the second heart sound. **D.** Systolic murmur in pulmonic stenosis spilling through the aortic second sound, pulmonic valve closure being delayed. **E.** Aortic or pulmonary diastolic murmur. **F.** Long diastolic murmur of mitral stenosis after the opening snap (OS). **G.** Short mid-diastolic inflow murmur after a third heart sound. **H.** Continuous murmur of patent ductus arteriosus. (Adapted from P Wood: Diseases of the Heart and Circulation, London, Eyre & Spottiswoode, 1968. Permission granted courtesy of Antony and Julie Wood.)

sound (S_2) that corresponds to their site of origin (left or right, respectively). *Diastolic murmurs* begin with or after the associated component of S_2 and end at or before the subsequent S_1 . *Continuous murmurs* are not confined to either phase of the cardiac cycle but instead begin in early systole and proceed through S_2 into all or part of diastole. The accurate timing of heart murmurs is the first step in their identification. The distinction between S_1 and S_2 and therefore systole and diastole is usually a straightforward process but can be difficult in the setting of a tachyarrhythmia, in which case the heart sounds can be distinguished by simultaneous palpation of the carotid upstroke, which should closely follow S_1 .

Duration and Character The duration of a heart murmur depends on the length of time over which a pressure difference exists between two cardiac chambers, the left ventricle and the aorta, the right ventricle and the pulmonary artery, or the great vessels. The magnitude and variability of this pressure difference, coupled with the geometry and compliance of the involved chambers or vessels, dictate the velocity of flow; the degree of turbulence; and the resulting frequency, configuration, and intensity of the murmur. The diastolic murmur of chronic aortic regurgitation (AR) is a blowing, high-frequency event, whereas the murmur of mitral stenosis (MS), indicative of the left atrial-left ventricular diastolic pressure gradient, is a low-frequency event, heard as a rumbling sound with the bell of the stethoscope. The frequency components of a heart murmur may vary at different sites of auscultation. The coarse systolic murmur of aortic stenosis (AS)

may sound higher pitched and more acoustically pure at the apex, a phenomenon eponymously referred to as the *Gallavardin effect*. Some murmurs may have a distinct or unusual quality, such as the “honking” sound appreciated in some patients with mitral regurgitation (MR) due to mitral valve prolapse (MVP).

The configuration of a heart murmur may be described as crescendo, decrescendo, crescendo-decrescendo, or plateau. The decrescendo configuration of the murmur of chronic AR (Fig. 38-1E) can be understood in terms of the progressive decline in the diastolic pressure gradient between the aorta and the left ventricle. The crescendo-decrescendo configuration of the murmur of AS reflects the changes in the systolic pressure gradient between the left ventricle and the aorta as ejection occurs, whereas the plateau configuration of the murmur of chronic MR (Fig. 38-1B) is consistent with the large and nearly constant pressure difference between the left ventricle and the left atrium.

Intensity The intensity of a heart murmur is graded on a scale of 1–6 (or I–VI). A grade 1 murmur is very soft and is heard only with great effort. A grade 2 murmur is easily heard but not particularly loud. A grade 3 murmur is loud but is not accompanied by a palpable thrill over the site of maximal intensity. A grade 4 murmur is very loud and accompanied by a thrill. A grade 5 murmur is loud enough to be heard with only the edge of the stethoscope touching the chest, whereas a grade 6 murmur is loud enough to be heard with the stethoscope slightly off the chest. Murmurs of grade 3 or greater intensity usually signify important structural heart disease and indicate high blood flow velocity at the site of murmur production. Small ventricular septal defects (VSDs), for example, are accompanied by loud, usually grade 4 or greater, systolic murmurs as blood is ejected at high velocity from the left ventricle to the right ventricle. Low-velocity events, such as left-to-right shunting across an atrial septal defect (ASD), are usually silent. The intensity of a heart murmur may be diminished by any process that increases the distance between the intracardiac source and the stethoscope on the chest wall, such as obesity, obstructive lung disease, or a large pericardial effusion. The intensity of a murmur also may be misleadingly soft when cardiac output is reduced significantly or when the pressure gradient between the involved cardiac structures is low.

Location and Radiation Recognition of the location and radiation of the murmur help facilitate its accurate identification (Fig. 38-2). Adventitious sounds, such as a systolic click or diastolic snap, or abnormalities of S_1 or S_2 may provide additional clues. Careful attention to the characteristics of the murmur and other heart sounds during the respiratory cycle and the performance of simple bedside maneuvers complete the auscultatory examination. These features, along with recommendations for further testing, are discussed below in the context of specific systolic, diastolic, and continuous heart murmurs (Table 38-1).

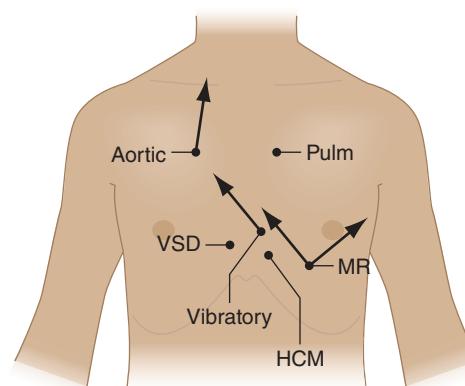


FIGURE 38-2 Maximal intensity and radiation of six isolated systolic murmurs. Aortic, aortic stenosis; HCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; Pulm, pulmonary stenosis; VSD, ventricular septal defect. (From JB Barlow: Perspectives on the Mitral Valve. Philadelphia, FA Davis, 1987, p 140.)

TABLE 38-1 Principal Causes of Heart Murmurs

Systolic Murmurs	
Early systolic	
Mitral	
Acute MR	
VSD	
Muscular	
Nonrestrictive with pulmonary hypertension	
Tricuspid	
TR with normal pulmonary artery pressure	
Midsystolic	
Aortic	
Obstructive	
Supravalvular-supravalvular AS, coarctation of the aorta	
Valvular-AS and aortic sclerosis	
Subvalvular-discrete, tunnel or HOCM	
Increased flow, hyperkinetic states, AR, complete heart block	
Dilation of ascending aorta, atherosoma, aortitis	
Pulmonary	
Obstructive	
Supravalvular-pulmonary artery stenosis	
Valvular-pulmonic valve stenosis	
Subvalvular-infundibular stenosis (dynamic)	
Increased flow, hyperkinetic states, left-to-right shunt (e.g., ASD)	
Dilation of pulmonary artery	
Late systolic	
Mitral	
MVP, acute myocardial ischemia	
Tricuspid	
TVP	
Holosystolic	
Atrioventricular valve regurgitation (MR, TR)	
Left-to-right shunt at ventricular level (VSD)	
Early Diastolic Murmurs	
AR	
Valvular: congenital (bicuspid valve), rheumatic deformity, endocarditis, prolapse, trauma, post-valvulotomy	
Dilation of valve ring: aorta dissection, annuloaortic ectasia, cystic medial degeneration, hypertension, ankylosing spondylitis	
Widening of commissures: syphilis	
Pulmonic regurgitation	
Valvular: post-valvulotomy, endocarditis, rheumatic fever, carcinoid	
Dilation of valve ring: pulmonary hypertension; Marfan syndrome	
Congenital: isolated or associated with tetralogy of Fallot, VSD, pulmonic stenosis	
Mid-Diastolic Murmurs	
Mitral	
MS	
Carey-Coombs murmur (mid-diastolic apical murmur in acute rheumatic fever)	
Increased flow across nonstenotic mitral valve (e.g., MR, VSD, PDA, high-output states, and complete heart block)	
Tricuspid	
Tricuspid stenosis	
Increased flow across nonstenotic tricuspid valve (e.g., TR, ASD, and anomalous pulmonary venous return)	
Left and right atrial tumors (myxoma)	
Severe AR (Austin Flint murmur)	
Continuous Murmurs	
Patent ductus arteriosus	Proximal coronary artery stenosis
Coronary AV fistula	Mammary souffle of pregnancy
Ruptured sinus of Valsalva aneurysm	Pulmonary artery branch stenosis
Aortic septal defect	Bronchial collateral circulation
Cervical venous hum	Small (restrictive) ASD with MS
Anomalous left coronary artery	Intercostal AV fistula

Abbreviations: AR, aortic regurgitation; AS, aortic stenosis; ASD, atrial septal defect; AV, arteriovenous; HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; MS, mitral stenosis; MVP, mitral valve prolapse; PDA, patent ductus arteriosus; TR, tricuspid regurgitation; TVP, tricuspid valve prolapse; VSD, ventricular septal defect.

Source: E Braunwald, JK Perloff, in D Zipes et al (eds): *Braunwald's Heart Disease*, 7th ed. Philadelphia, Elsevier, 2005; PJ Norton, RA O'Rourke, in E Braunwald, L Goldman (eds): *Primary Cardiology*, 2nd ed. Philadelphia, Elsevier, 2003.

■ SYSTOLIC HEART MURMURS

Early Systolic Murmurs Early systolic murmurs begin with S_1 and extend for a variable period, ending well before S_2 . Their causes are relatively few in number. *Acute, severe MR* into a normal-sized, relatively noncompliant left atrium results in an early, decrescendo systolic murmur best heard at or just medial to the apical impulse. These characteristics reflect the progressive attenuation of the pressure gradient between the left ventricle and the left atrium during systole owing to the rapid rise in left atrial pressure caused by the sudden volume load into an unprepared, noncompliant chamber and contrast sharply with the auscultatory features of chronic MR. Clinical settings in which acute, severe MR occur include (1) papillary muscle rupture complicating acute myocardial infarction (MI) (Chap. 269), (2) rupture of chordae tendineae in the setting of myxomatous mitral valve disease (MVP, Chap. 260), (3) infective endocarditis (Chap. 123), and (4) blunt chest wall trauma.

Acute, severe MR from papillary muscle rupture usually accompanies an inferior, posterior, or lateral MI, and occurs 2–7 days after presentation. It often is signaled by chest pain, hypotension, and pulmonary edema, but a murmur may be absent in up to 50% of cases. The posteromedial papillary muscle is involved 6 to 10 times more frequently than the anterolateral papillary muscle. The murmur is to be distinguished from that associated with post-MI ventricular septal rupture, which is accompanied by a systolic thrill at the left sternal border in nearly all patients and is holosystolic in duration. A new heart murmur after an MI is an indication for transthoracic echocardiography (TTE) (Chap. 236), which allows bedside delineation of its etiology and pathophysiologic significance. The distinction between acute MR and ventricular septal rupture also can be achieved with right-sided heart catheterization, sequential determination of oxygen saturations, and analysis of the pressure waveforms (tall v wave in the pulmonary artery wedge pressure in MR). Post-MI mechanical complications of this nature mandate aggressive medical stabilization and prompt referral for surgical repair.

Spontaneous chordal rupture can complicate the course of myxomatous mitral valve disease (MVP) and result in new-onset or “acute on chronic” severe MR. MVP may occur as an isolated phenomenon, or the lesion may be part of a more generalized connective tissue disorder as seen, for example, in patients with Marfan syndrome. Acute, severe MR as a consequence of infective endocarditis results from destruction of leaflet tissue, chordal rupture, or both. Blunt chest wall trauma is usually self-evident but may be disarmingly trivial; it can result in papillary muscle contusion and rupture, chordal detachment, or leaflet avulsion. TTE is indicated in all cases of suspected acute, severe MR to define its mechanism and severity, delineate left ventricular size and systolic function, and provide an assessment of suitability for primary valve repair.

A congenital, small muscular VSD (Chap. 264) may be associated with an early systolic murmur. The defect closes progressively during septal contraction, and thus the murmur is confined to early systole. It is localized to the left sternal border (Fig. 38-2) and is usually of grade 4 or 5 intensity. Signs of pulmonary hypertension or left ventricular volume overload are absent. Anatomically large and uncorrected VSDs, which usually involve the membranous portion of the septum, may lead to pulmonary hypertension. The murmur associated with the left-to-right shunt, which earlier may have been holosystolic, becomes limited to the first portion of systole as the elevated pulmonary vascular resistance leads to an abrupt rise in right ventricular pressure and an attenuation of the interventricular pressure gradient during the remainder of the cardiac cycle. In such instances, signs of pulmonary hypertension (right ventricular lift, loud and single or closely split S_2) may predominate. The murmur is best heard along the left sternal border but is softer. Suspicion of a VSD is an indication for TTE.

Tricuspid regurgitation (TR) with normal pulmonary artery pressures, as may occur with infective endocarditis, may produce an early systolic murmur. The murmur is soft (grade 1 or 2), is best heard at the lower left sternal border and may increase in intensity with inspiration

(Carvallo's sign). Regurgitant *c-v* waves may be visible in the jugular venous pulse. TR in this setting is not associated with signs of right heart failure.

Midsystolic Murmurs Midsystolic murmurs begin at a short interval after S_1 , end before S_2 (Fig. 38-1C) and are usually crescendo-decrescendo in configuration. AS is the most common cause of a midsystolic murmur in an adult. The murmur of AS is usually loudest to the right of the sternum in the second intercostal space (aortic area, Fig. 38-2) and radiates into the carotids. Transmission of the midsystolic murmur to the apex, where it becomes higher-pitched, is common (Gallavardin effect; see above).

Differentiation of this apical systolic murmur from MR can be difficult. The murmur of AS will increase in intensity or become louder, in the beat after a premature beat, whereas the murmur of MR will have constant intensity from beat to beat. The intensity of the AS murmur also varies directly with the cardiac output. With a normal cardiac output, a systolic thrill and a grade 4 or higher murmur suggest severe AS. The murmur is softer in the setting of heart failure and low cardiac output. Other auscultatory findings of severe AS include a soft or absent A_2 , paradoxical splitting of S_2 , an apical S_4 , and a late-peaking systolic murmur. In children, adolescents, and young adults with congenital valvular AS, an early ejection sound (click) is usually audible, more often along the left sternal border than at the base. Its presence signifies a flexible, noncalcified bicuspid valve (or one of its variants) and localizes the left ventricular outflow obstruction to the valvular (rather than sub- or supravalvular) level.

Assessment of the volume and rate of rise of the carotid pulse can provide additional information. A small and delayed upstroke (*parvus et tardus*) is consistent with severe AS. The carotid pulse examination is less discriminatory, however, in older patients with stiffened arteries. The electrocardiogram (ECG) shows signs of left ventricular hypertrophy (LVH) as the severity of the stenosis increases. TTE is indicated to assess the anatomic features of the aortic valve, the severity of the stenosis, left ventricular size, wall thickness and function, and the size and contour of the aortic root and proximal ascending aorta.

The obstructive form of hypertrophic cardiomyopathy (HOCM) is associated with a midsystolic murmur that is usually loudest along the left sternal border or between the left lower sternal border and the apex (Chap. 254, Fig. 38-2). The murmur is produced by both dynamic left ventricular outflow tract obstruction and MR, and thus, its configuration is a hybrid between ejection and regurgitant phenomena. The intensity of the murmur may vary from beat to beat and after provocative maneuvers but usually does not exceed grade 3. The murmur classically will increase in intensity with maneuvers that result in increasing degrees of outflow tract obstruction, such as a reduction in preload or afterload (Valsalva, standing, vasodilators), or with an augmentation of contractility (inotropic stimulation). Maneuvers that increase preload (squatting, passive leg raising, volume administration) or afterload (squatting, vasopressors) or that reduce contractility (β -adrenoreceptor blockers) decrease the intensity of the murmur. In rare patients, there may be reversed splitting of S_2 . A sustained left ventricular apical impulse and an S_4 may be appreciated. In contrast to AS, the carotid upstroke is rapid and of normal volume. Rarely, it is bisferiens or bifid in contour (see Fig. 234-2D) due to midsystolic closure of the aortic valve. LVH is present on the ECG, and the diagnosis is confirmed by TTE. Although the systolic murmur associated with MVP behaves similarly to that due to HOCM in response to the Valsalva maneuver and to standing/squatting (Fig. 38-3), these two lesions can be distinguished on the basis of their associated findings, such as the presence of LVH in HOCM or a nonejection click in MVP.

The midsystolic, crescendo-decrescendo murmur of congenital pulmonic stenosis (PS, Chap. 264) is best appreciated in the second and third left intercostal spaces (pulmonic area) (Figs. 38-2 and 38-4). The duration of the murmur lengthens and the intensity of P_2 diminishes with increasing degrees of valvular stenosis (Fig. 38-1D). An early ejection sound, the intensity of which decreases with inspiration, is heard in younger patients. A parasternal lift and ECG evidence of right ventricular hypertrophy indicate severe pressure overload. If obtained,

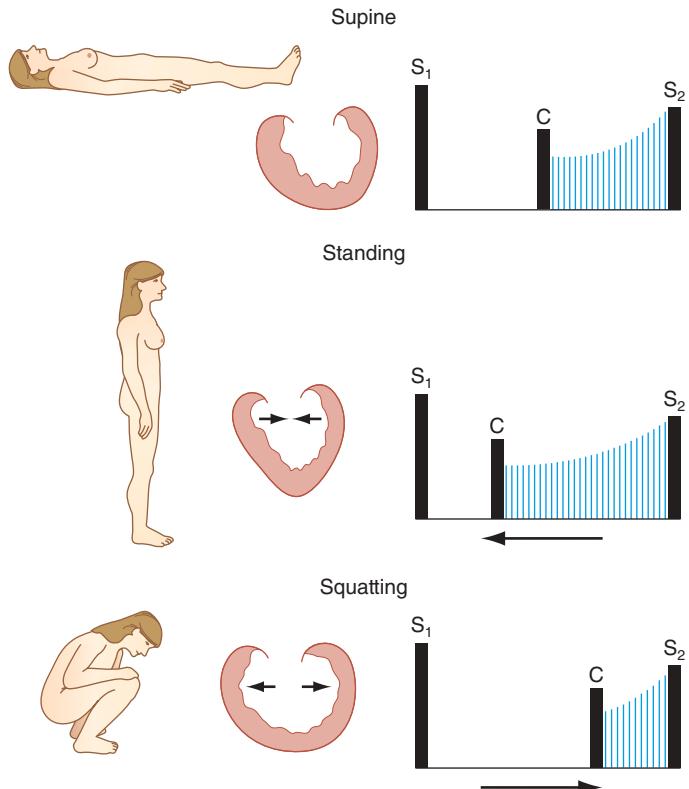
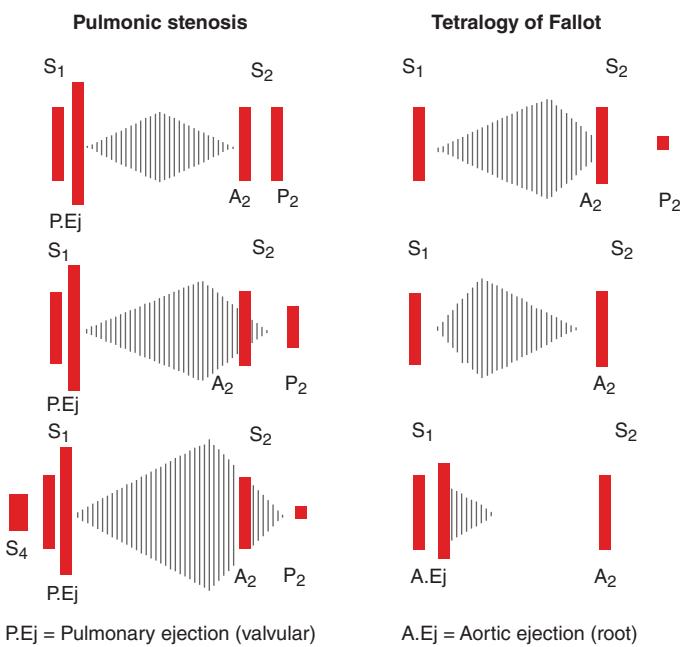


FIGURE 38-3 A midsystolic nonejection sound (C) occurs in mitral valve prolapse and is followed by a late systolic murmur that crescendos to the second heart sound (S_2). Standing decreases venous return; the heart becomes smaller; C moves closer to the first heart sound (S_1), and the mitral regurgitant murmur has an earlier onset. With prompt squatting, venous return and afterload increase; the heart becomes larger; C moves toward S_2 ; and the duration of the murmur shortens. The systolic murmur of hypertrophic obstructive cardiomyopathy behaves similarly. (From JA Shaver, JJ Leonard, DF Leon: Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 13. Copyright, American Heart Association.)

the chest x-ray may show poststenotic dilation of the main pulmonary artery. TTE is recommended for complete characterization.

Significant left-to-right intracardiac shunting due to an ASD (Chap. 264) leads to an increase in pulmonary blood flow and a grades 2–3 midsystolic murmur at the middle to upper left sternal border attributed to increased flow rates across the pulmonic valve with fixed splitting of S_2 . Ostium secundum ASDs are the most common cause of these shunts in adults. Features suggestive of a primum ASD include the coexistence of MR due to a cleft anterior mitral valve leaflet and left axis deviation of the QRS complex on the ECG. With sinus venosus ASDs, the left-to-right shunt is usually not large enough to result in a systolic murmur, although the ECG may show abnormalities of sinus node function. A grade 2 or 3 midsystolic murmur may also be heard best at the upper left sternal border in patients with idiopathic dilation of the pulmonary artery; a pulmonary ejection sound is also present in these patients. TTE is indicated to evaluate a grade 2 or 3 midsystolic murmur when there are other signs of cardiac disease.

An isolated grade 1 or 2 midsystolic murmur, heard in the absence of symptoms or signs of heart disease, is most often a benign finding for which no further evaluation, including TTE, is necessary. The most common example of a murmur of this type in an older adult patient is the crescendo-decrescendo murmur of aortic valve sclerosis, heard at the second right interspace (Fig. 38-2). Aortic sclerosis is defined as focal thickening and calcification of the aortic valve to a degree that does not interfere with leaflet opening. The carotid upstrokes are normal, and electrocardiographic LVH is not present. A grade 1 or 2 midsystolic murmur often can be heard at the left sternal border with pregnancy, hyperthyroidism, or anemia, physiologic states that are associated with accelerated blood flow. Still's murmur refers to a benign



P.Ej = Pulmonary ejection (valvular)

A.Ej = Aortic ejection (root)

FIGURE 38-4 **Left.** In valvar pulmonic stenosis with intact ventricular septum, right ventricular systolic ejection becomes progressively longer, with increasing obstruction to flow. As a result, the murmur becomes longer and louder, enveloping the aortic component of the second heart sound (A₂). The pulmonic component (P₂) occurs later, and splitting becomes wider but more difficult to hear because A₂ is lost in the murmur and P₂ becomes progressively fainter and lower pitched. As the pulmonic gradient increases, the isometric contraction phase shortens until the pulmonic valve ejection sound fuses with the first heart sound (S₁). In severe pulmonic stenosis with concentric hypertrophy and decreasing right ventricular compliance, a fourth heart sound appears. **Right.** In tetralogy of Fallot with increasing obstruction at the pulmonic infundibular area, an increasing amount of right ventricular blood is shunted across the silent ventricular septal defect and flow across the obstructed outflow tract decreases. Therefore, with increasing obstruction the murmur becomes shorter, earlier, and fainter. P₂ is absent in severe tetralogy of Fallot. A large aortic root receives almost all cardiac output from both ventricular chambers, and the aorta dilates and is accompanied by a root ejection sound that does not vary with respiration. (From JA Shaver, JJ Leonard, DF Leon: Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 45. Copyright, American Heart Association.)

grade 2, vibratory or musical midsystolic murmur at the mid or lower left sternal border in normal children and adolescents, best heard in the supine position (Fig. 38-2).

Late Systolic Murmurs A late systolic murmur that is best heard at the left ventricular apex is usually due to MVP (Chap. 260). Often, this murmur is introduced by one or more nonejection clicks. The radiation of the murmur can help identify the specific mitral leaflet involved in the process of prolapse or flail. The term *flail* refers to the movement made by an unsupported portion of the leaflet (usually the tip) after loss of its chordal attachment(s). With posterior leaflet prolapse or flail, the resultant jet of MR is directed anteriorly and medially, as a result of which the murmur radiates to the base of the heart and masquerades as AS. Anterior leaflet prolapse or flail results in a posteriorly directed MR jet that radiates to the axilla or left infrascapular region. Leaflet flail is associated with a murmur of grade 3 or 4 intensity that can be heard throughout the precordium in thin-chested patients. The presence of an S₃ or a short, rumbling mid-diastolic murmur due to enhanced flow signifies severe MR.

Bedside maneuvers that decrease left ventricular preload, such as standing, will cause the click and murmur of MVP to move closer to the first heart sound, as leaflet prolapse occurs earlier in systole. Standing also causes the murmur to become louder and longer. With squatting, left ventricular preload and afterload are increased abruptly, leading to an increase in left ventricular volume, and the click and murmur move away from the first heart sound as leaflet prolapse is delayed; the murmur becomes softer and shorter in duration (Fig. 38-3). As noted above,

these responses to standing and squatting are directionally similar to those observed in patients with HOCM.

A late, apical systolic murmur indicative of MR may be heard transiently in the setting of acute myocardial ischemia; it is due to apical tethering and malcoaptation of the leaflets in response to structural and functional changes of the ventricle and mitral annulus. The intensity of the murmur varies as a function of left ventricular afterload and will increase in the setting of hypertension. TTE is recommended for assessment of late systolic murmurs.

Holosystolic Murmurs (Figs. 38-1B and 38-5) Holosystolic murmurs begin with S₁ and continue through systole to S₂. They are usually indicative of chronic mitral or tricuspid valve regurgitation or a VSD and warrant TTE for further characterization. The holosystolic murmur of chronic MR is best heard at the left ventricular apex and radiates to the axilla (Fig. 38-2); it is usually high-pitched and plateau in configuration because of the wide difference between left ventricular and left atrial pressure throughout systole. In contrast to acute MR, left atrial compliance is normal or even increased in chronic MR. As a result, there is only a small increase in left atrial pressure for any increase in regurgitant volume.

Several conditions are associated with chronic MR and an apical holosystolic murmur, including rheumatic scarring of the leaflets, mitral annular calcification, postinfarction left ventricular remodeling, and severe left ventricular chamber enlargement. The circumference of the mitral annulus increases as the left ventricle enlarges and leads to failure of leaflet coaptation with central MR in patients with dilated cardiomyopathy (Chap. 254). The severity of the MR is worsened by any contribution from apical displacement of the papillary muscles and leaflet tethering (remodeling). Because the mitral annulus is contiguous with the left atrial endocardium, gradual enlargement of the left atrium from chronic MR will result in further stretching of the annulus and more MR; thus, "MR begets MR." Chronic severe MR results in enlargement and leftward displacement of the left ventricular apex beat and, in some patients, a diastolic filling complex, as described previously (Fig. 38-1G).

The holosystolic murmur of chronic TR is generally softer than that of MR, is loudest at the left lower sternal border, and usually increases in intensity with inspiration (Carvallo's sign). Associated signs include c-v waves in the jugular venous pulse, an enlarged and pulsatile liver, ascites, and peripheral edema. The abnormal jugular venous waveforms are the predominant finding and seen very often in the absence of an audible murmur despite Doppler echocardiographic verification of TR. Causes of primary TR include myxomatous disease (prolapse), endocarditis, rheumatic disease, radiation, carcinoid, Ebstein's anomaly, and chordal detachment as a complication of right ventricular endomyocardial biopsy. TR is much more commonly a passive process that results secondarily from annular enlargement due to right ventricular dilation in the face of volume or pressure overload or adverse right ventricular remodeling.

The holosystolic murmur of a VSD is loudest at the mid- to lower-left sternal border (Fig. 38-2) and radiates widely. A thrill is present at the site of maximal intensity in the majority of patients. There is no change in the intensity of the murmur with inspiration. The intensity of the murmur varies as a function of the anatomic size of the defect. Small, restrictive VSDs, as exemplified by the *maladie de Roger*, create a very loud murmur due to the significant and sustained systolic pressure gradient between the left and right ventricles. With large defects, the ventricular pressures tend to equalize, shunt flow is balanced, and a murmur is not appreciated. The distinction between post-MI ventricular septal rupture and MR has been reviewed previously.

■ DIASTOLIC HEART MURMURS

Early Diastolic Murmurs (Fig. 38-1E) Chronic AR results in a high-pitched, blowing, decrescendo, early- to mid-diastolic murmur that begins after the aortic component of S₂ (A₂), and is best heard at the second right interspace. The murmur may be soft and difficult to hear unless auscultation is performed with the patient leaning forward at end expiration. This maneuver brings the aortic root closer to the

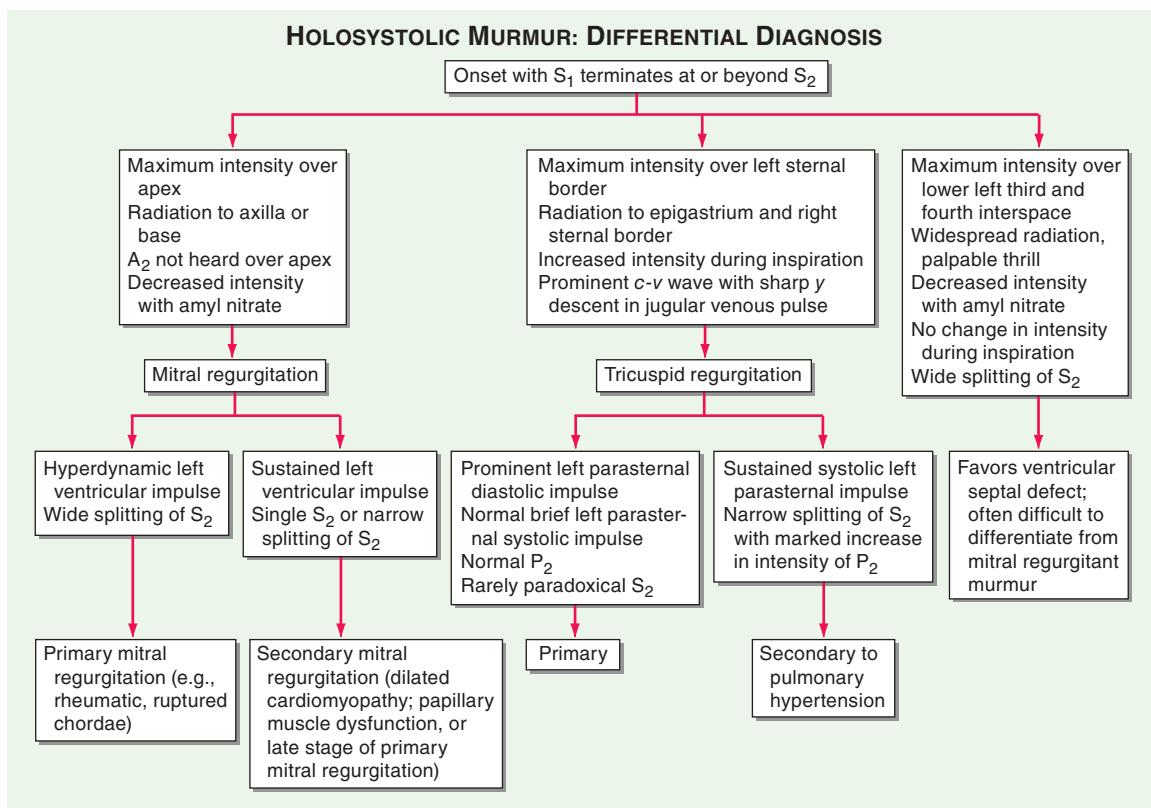


FIGURE 38-5 Differential diagnosis of a holosystolic murmur.

anterior chest wall. Radiation of the murmur may provide a clue to the cause of the AR. With primary valve disease, such as that due to congenital bicuspid disease, prolapse, or endocarditis, the diastolic murmur tends to radiate along the left sternal border, where it is often louder than appreciated in the second right interspace. When AR is caused by aortic root disease, the diastolic murmur may radiate along the right sternal border. Diseases of the aortic root cause dilation or distortion of the aortic annulus and failure of leaflet coaptation. Causes include Marfan syndrome with aneurysm formation, annuloaortic ectasia, ankylosing spondylitis, and aortic dissection.

Chronic, severe AR also may produce a lower-pitched mid to late, grade 1 or 2 diastolic murmur at the apex (Austin Flint murmur), which is thought to reflect turbulence at the mitral inflow area from the admixture of regurgitant (aortic) and forward (mitral) blood flow. This lower-pitched, apical diastolic murmur can be distinguished from that due to MS by the absence of an opening snap and the response of the murmur to a vasodilator challenge. Lowering afterload with an agent such as amyl nitrite will decrease the duration and magnitude of the aortic-left ventricular diastolic pressure gradient, and thus the Austin Flint murmur of severe AR will become shorter and softer. The intensity of the diastolic murmur of MS (Fig. 38-6) may either remain constant or increase with afterload reduction because of the reflex increase in cardiac output and mitral valve flow.

Although AS and AR may coexist, a grade 2 or 3 crescendo-decrescendo midsystolic murmur frequently is heard at the base of the heart in patients with isolated, severe AR and is due to an increased volume and rate of systolic flow. Accurate bedside identification of coexistent AS can be difficult unless the carotid pulse examination is abnormal or the midsystolic murmur is of grade 4 or greater intensity. In the absence of heart failure, chronic severe AR is accompanied by several peripheral signs of significant diastolic runoff, including a wide pulse pressure, a “water-hammer” carotid upstroke (Corrigan’s pulse), and Quincke’s pulsations of the nail beds. The diastolic murmur of acute, severe AR is notably shorter in duration and lower pitched than the murmur of chronic AR. It can be very difficult to appreciate in the presence of a rapid heart rate. These attributes reflect the abrupt rate of rise of diastolic pressure within the unprepared and noncompliant left

ventricle and the correspondingly rapid decline in the aortic-left ventricular diastolic pressure gradient. Left ventricular diastolic pressure may increase sufficiently to result in premature closure of the mitral valve and a soft first heart sound. Peripheral signs of significant diastolic runoff are not present.

Pulmonic regurgitation (PR) results in a decrescendo, early to mid-diastolic murmur (*Graham Steell murmur*) that begins after the pulmonic component of S₂ (P₂), is best heard at the second left interspace, and radiates along the left sternal border. The intensity of the murmur may increase with inspiration. PR is most commonly due to dilation of the valve annulus from chronic elevation of the pulmonary

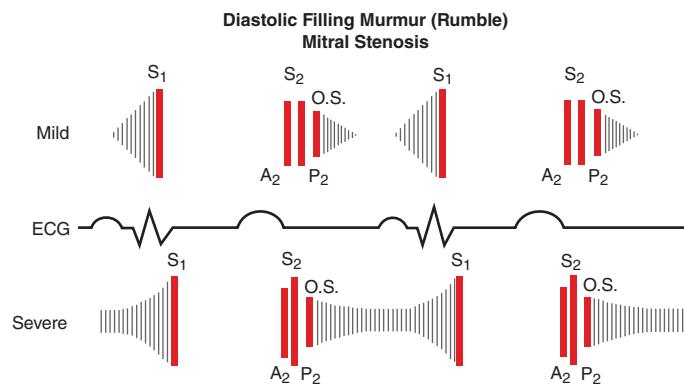


FIGURE 38-6 Diastolic filling murmur (rumble) in mitral stenosis. In mild mitral stenosis, the diastolic gradient across the valve is limited to the phases of rapid ventricular filling in early diastole and presystole. The rumble may occur during either or both periods. As the stenotic process becomes severe, a large pressure gradient exists across the valve during the entire diastolic filling period, and the rumble persists throughout diastole. As the left atrial pressure becomes greater, the interval between A₂ (or P₂) and the opening snap (O.S.) shortens. In severe mitral stenosis, secondary pulmonary hypertension develops and results in a loud P₂ and the splitting interval usually narrows. ECG, electrocardiogram. (From JA Shaver, JJ Leonard, DF Leon: Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 55. Copyright, American Heart Association.)

artery pressure. Signs of pulmonary hypertension, including a right ventricular lift and a loud, single or narrowly split S_2 , are present. These features also help distinguish PR from AR as the cause of a decrescendo diastolic murmur heard along the left sternal border. PR in the absence of pulmonary hypertension can occur with endocarditis or a congenitally deformed valve. It is usually present after repair of tetralogy of Fallot in childhood. When pulmonary hypertension is not present, the diastolic murmur is softer and lower pitched than the classic Graham Steell murmur, and the severity of the PR can be difficult to appreciate.

TTE is indicated for the further evaluation of a patient with an early to mid-diastolic murmur. Longitudinal assessment of lesion severity, ventricular size, and of systolic function helps guide a potential decision for surgical management. TTE also can provide anatomic information regarding the root and proximal ascending aorta, although computed tomographic or magnetic resonance angiography may be indicated for more precise characterization (Chap. 236).

Mid-Diastolic Murmurs (Figs. 38-1F and 38-1G) Mid-diastolic murmurs result from obstruction and/or augmented flow at the level of the mitral or tricuspid valve. Rheumatic fever is the most common cause of MS (Fig. 38-6). In younger patients with pliable valves, S_1 is loud and the murmur begins after an opening snap, which is a high-pitched sound that occurs shortly after S_2 . The interval between the pulmonic component of the second heart sound (P_2) and the opening snap is inversely related to the magnitude of the left atrial-left ventricular pressure gradient. The murmur of MS is low-pitched and thus is best heard with the bell of the stethoscope. It is loudest at the left ventricular apex and often is appreciated only when the patient is turned in the left lateral decubitus position. It is usually of grade 1 or 2 intensity but may be absent when the cardiac output is severely reduced despite significant obstruction. The intensity of the murmur increases during maneuvers that increase cardiac output and mitral valve flow, such as exercise. The duration of the murmur reflects the length of time over which left atrial pressure exceeds left ventricular diastolic pressure. An increase in the intensity of the murmur just before S_1 , a phenomenon known as *presystolic accentuation* (Figs. 38-1A and 38-6), occurs in patients in sinus rhythm and is due to a late increase in trans-mitral flow with atrial contraction. Presystolic accentuation does not occur in patients with atrial fibrillation.

The mid-diastolic murmur associated with tricuspid stenosis is best heard at the lower left sternal border and increases in intensity with inspiration. A prolonged y descent may be visible in the jugular venous waveform. This murmur is very difficult to hear and often is obscured by left-sided acoustical events.

There are several other causes of mid-diastolic murmurs. Large left atrial myxomas may prolapse across the mitral valve and cause variable degrees of obstruction to left ventricular inflow (Chap. 266). The murmur associated with an atrial myxoma may change in duration and intensity with changes in body position. An opening snap is not present, and there is no presystolic accentuation. Augmented mitral diastolic flow can occur with isolated severe MR or with a large left-to-right shunt at the ventricular or great vessel level and produce a soft, rapid filling sound (S_1) followed by a short, low-pitched mid-diastolic apical murmur (Fig. 38-1G). The Austin Flint murmur of severe, chronic AR has already been described.

A short, mid-diastolic murmur is rarely heard during an episode of acute rheumatic fever (Carey-Coombs murmur) and probably is due to flow through an edematous mitral valve. An opening snap is not present in the acute phase, and the murmur dissipates with resolution of the acute attack. Complete heart block with dysynchronous atrial and ventricular activation may be associated with intermittent mid- to late diastolic murmurs if atrial contraction occurs when the mitral valve is partially closed. Mid-diastolic murmurs indicative of increased tricuspid valve flow can occur with severe, isolated TR and with large ASDs and significant left-to-right shunting. Other signs of an ASD are present (Chap. 264), including fixed splitting of S_2 and a midsystolic murmur at the mid- to upper left sternal border. TTE is indicated for evaluation of a patient with a mid- to late diastolic murmur. Findings specific to the diseases discussed above will help guide management.

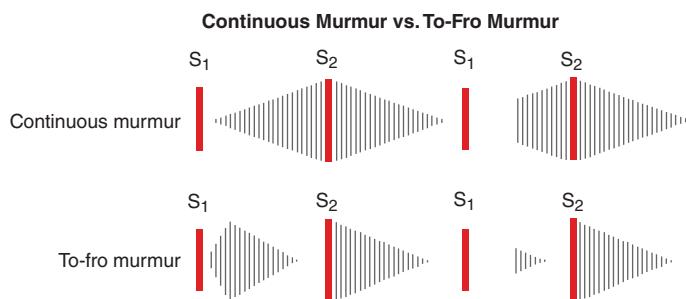


FIGURE 38-7 Comparison of the continuous murmur and the to-fro murmur. During abnormal communication between high-pressure and low-pressure systems, a large pressure gradient exists throughout the cardiac cycle, producing a continuous murmur. A classic example is patent ductus arteriosus. At times, this type of murmur can be confused with a to-fro murmur, which is a combination of systolic ejection murmur and a murmur of semilunar valve incompetence. A classic example of a to-fro murmur is aortic stenosis and regurgitation. A continuous murmur crescendos to near the second heart sound (S_2), whereas a to-fro murmur has two components. The midsystolic ejection component decrescends and disappears as it approaches S_2 . (From JA Shaver, JJ Leonard, DF Leon: *Examination of the Heart, Part IV, Auscultation of the Heart*. Dallas, American Heart Association, 1990, p 55. Copyright, American Heart Association.)

■ CONTINUOUS MURMURS

(Figs. 38-1H and 38-7) Continuous murmurs begin in systole, peak near the second heart sound, and continue into all or part of diastole. Their presence throughout the cardiac cycle implies a pressure gradient between two chambers or vessels during both systole and diastole. The continuous murmur associated with a patent ductus arteriosus is best heard at the upper left sternal border. Large, uncorrected shunts may lead to pulmonary hypertension, attenuation or obliteration of the diastolic component of the murmur, reversal of shunt flow, and differential cyanosis of the lower extremities. A ruptured sinus of Valsalva aneurysm creates a continuous murmur of abrupt onset at the upper right sternal border. Rupture typically occurs into a right heart chamber, and the murmur is indicative of a continuous pressure difference between the aorta and either the right ventricle or the right atrium. A continuous murmur also may be audible along the left sternal border with a coronary arteriovenous fistula and at the site of an arteriovenous fistula used for hemodialysis access. Enhanced flow through enlarged intercostal collateral arteries in patients with aortic coarctation may produce a continuous murmur along the course of one or more ribs. A cervical bruit with both systolic and diastolic components (a to-fro murmur, Fig. 38-7) usually indicates a high-grade carotid artery stenosis.

Not all continuous murmurs are pathologic. A continuous venous hum can be heard in healthy children and young adults, especially during pregnancy; it is best appreciated in the right supraclavicular fossa and can be obliterated by pressure over the right internal jugular vein or by having the patient turn his or her head toward the examiner. The continuous mammary souffle of pregnancy is created by enhanced arterial flow through engorged breasts and usually appears during the late third trimester or early puerperium. The murmur is louder in systole. Firm pressure with the diaphragm of the stethoscope can eliminate the diastolic portion of the murmur.

■ DYNAMIC AUSCULTATION

(Table 38-2; see Table 234-1) Careful attention to the behavior of heart murmurs during simple maneuvers that alter cardiac hemodynamics can provide important clues to their cause and significance.

Respiration Auscultation should be performed during quiet respiration or with a modest increase in inspiratory effort, as more forceful movement of the chest tends to obscure the heart sounds. Left-sided murmurs may be best heard at end expiration, when lung volumes are minimized and the heart and great vessels are brought closer to the chest wall. This phenomenon is characteristic of the murmur of AR. Murmurs of right-sided origin, such as tricuspid or pulmonic regurgitation, increase in intensity during inspiration. The intensity of left-sided murmurs either remains constant or decreases with inspiration.

TABLE 38-2 Dynamic Auscultation: Bedside Maneuvers That Can Be Used to Change the Intensity of Cardiac Murmurs (See Text)

1. Respiration
2. Isometric exercise (handgrip)
3. Transient arterial occlusion
4. Pharmacologic manipulation of preload and/or afterload
5. Valsalva maneuver
6. Rapid standing/squatting
7. Passive leg raising
8. Post-premature beat

Bedside assessment also should evaluate the behavior of S_2 with respiration and the dynamic relationship between the aortic and pulmonic components (Fig. 38-8). Reversed splitting can be a feature of severe AS, HOCM, left bundle branch block, right ventricular pacing, or acute myocardial ischemia. Fixed splitting of S_2 in the presence of a grade 2 or 3 midsystolic murmur at the mid- or upper left sternal border indicates an ASD. Physiologic but wide splitting during the respiratory cycle implies either premature aortic valve closure, as can occur with severe MR, or delayed pulmonic valve closure due to PS or right bundle branch block.

Alterations of Systemic Vascular Resistance Murmurs can change characteristics after maneuvers that alter systemic vascular resistance and left ventricular afterload. The systolic murmurs of MR and VSD become louder during sustained handgrip, simultaneous inflation of blood pressure cuffs on both upper extremities to pressures 20–40 mmHg above systolic pressure for 20 s, or infusion of a vasopressor agent. The murmurs associated with AS or HOCM will

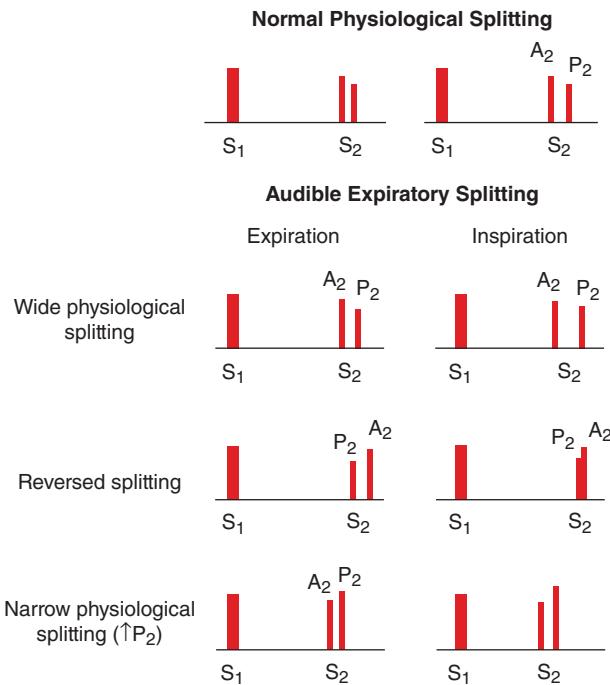


FIGURE 38-8 **Top.** Normal physiologic splitting. During expiration, the aortic (A_2) and pulmonic (P_2) components of the second heart sound are separated by <30 ms and are appreciated as a single sound. During inspiration, the splitting interval widens, and A_2 and P_2 are clearly separated into two distinct sounds. **Bottom.** Audible expiratory splitting. Wide physiologic splitting is caused by a delay in P_2 (as, for example, with right bundle branch block) or by early closure of the aortic valve (A_2 , as for example with severe mitral regurgitation). Reversed splitting is caused by a delay in A_2 , resulting in paradoxical movement; i.e., with inspiration P_2 moves toward A_2 , and the splitting interval narrows. Narrow physiologic splitting occurs in pulmonary hypertension, and both A_2 and P_2 are heard during expiration at a narrow splitting interval because of the increased intensity and high-frequency composition of P_2 . (From JA Shaver, JJ Leonard, DF Leon: *Examination of the Heart, Part IV, Auscultation of the Heart*. Dallas, American Heart Association, 1990, p 17. Copyright, American Heart Association.)

become softer or remain unchanged with these maneuvers. The diastolic murmur of AR becomes louder in response to interventions that raise systemic vascular resistance.

Opposite changes in systolic and diastolic murmurs may occur with the use of pharmacologic agents that lower systemic vascular resistance. Inhaled amyl nitrite is now rarely used for this purpose but can help distinguish the murmur of AS or HOCM from that of either MR or VSD, if necessary. The former two murmurs increase in intensity, whereas the latter two become softer after exposure to amyl nitrite. As noted previously, the Austin Flint murmur of severe AR becomes softer, but the mid-diastolic rumble of MS becomes louder, in response to the abrupt lowering of systemic vascular resistance with amyl nitrite.

Changes in Venous Return The Valsalva maneuver results in an increase in intrathoracic pressure, followed by a decrease in venous return, ventricular filling, and cardiac output. The majority of murmurs decrease in intensity during the strain phase of the maneuver. Two notable exceptions are the murmurs associated with MVP and obstructive HOCM, both of which become louder during the Valsalva maneuver. The murmur of MVP may also become longer as leaflet prolapse occurs earlier in systole at smaller ventricular volumes. These murmurs behave in a similar and parallel fashion with standing. Both the click and the murmur of MVP move closer in timing to S_1 on rapid standing from a squatting position (Fig. 38-3). The increase in the intensity of the murmur of HOCM is predicated on the augmentation of the dynamic left ventricular outflow tract gradient that occurs with reduced ventricular filling. Squatting results in abrupt increases in both venous return (preload) and left ventricular afterload that increase ventricular volume, changes that predictably cause a decrease in the intensity and duration of the murmurs associated with MVP and HOCM; the click and murmur of MVP move away from S_1 with squatting. Passive leg raising can be used to increase venous return in patients who are unable to squat and stand. This maneuver may lead to a decrease in the intensity of the murmur associated with HOCM but has less effect in patients with MVP.

Post-premature Ventricular Contraction A change in the intensity of a systolic murmur in the first beat after a premature beat, or in the beat after a long cycle length in patients with atrial fibrillation, can help distinguish AS from MR, particularly in an older patient in whom the murmur of AS is well transmitted to the apex. Systolic murmurs due to left ventricular outflow obstruction, including that due to AS, increase in intensity in the beat after a premature beat because of the combined effects of enhanced left ventricular filling and post-extrasystolic potentiation of contractile function. Forward flow accelerates, causing an increase in the gradient and a louder murmur. The intensity of the murmur of MR does not change in the post-premature beat as there is relatively little further increase in mitral valve flow or change in the left ventricular-left atrial gradient.

THE CLINICAL CONTEXT

Additional clues to the etiology and importance of a heart murmur can be gleaned from the history and other physical examination findings. Symptoms suggestive of cardiovascular, neurologic, or pulmonary disease help focus the differential diagnosis, as do findings relevant to the jugular venous pressure and waveforms, the arterial pulses, other heart sounds, the lungs, the abdomen, the skin, and the extremities. In many instances, laboratory studies, an ECG, and/or a chest x-ray may have been obtained earlier and may contain valuable information. A patient with suspected infective endocarditis, for example, may have a murmur in the setting of fever, chills, anorexia, fatigue, dyspnea, splenomegaly, petechiae, and positive blood cultures. A new systolic murmur in a patient with a marked fall in blood pressure after a recent MI suggests myocardial rupture. By contrast, an isolated grade 1 or 2 midsystolic murmur at the left sternal border in a healthy, active, and asymptomatic young adult is most likely a benign finding for which no further evaluation is indicated. The context in which the murmur is appreciated often dictates the need for further testing and the pace of the evaluation.

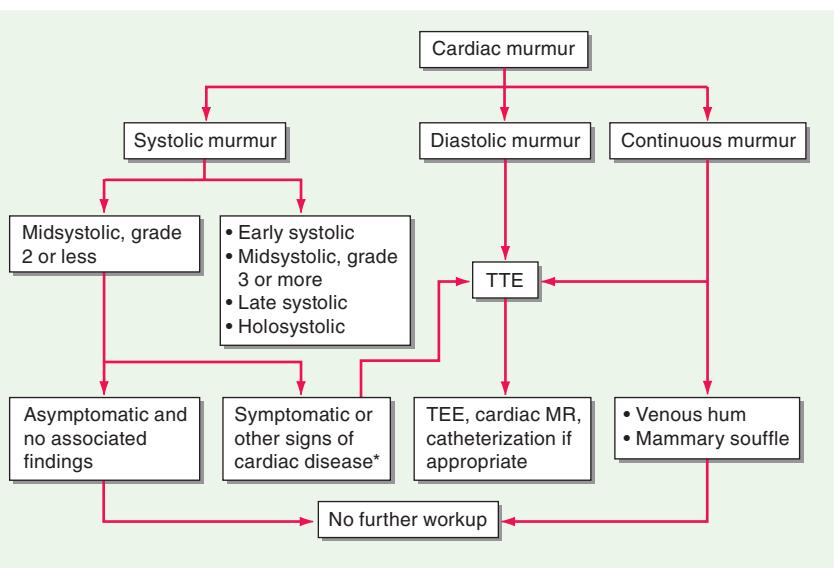


FIGURE 38-9 Strategy for evaluating heart murmurs. *If an electrocardiogram or chest x-ray has been obtained and is abnormal, echocardiography is indicated. TTE, transthoracic echocardiography; TEE, transesophageal echocardiography; MR, magnetic resonance. (Adapted from RO Bonow et al: *J Am Coll Cardiol* 32:1486, 1998.)

ECHOCARDIOGRAPHY

(Fig. 38-9; Chaps. 234 and 236) Echocardiography with color flow and spectral Doppler is a valuable tool for the assessment of cardiac murmurs. Information regarding valve structure and function, chamber size, wall thickness, ventricular function, estimated pulmonary artery pressures, intracardiac shunt flow, pulmonary and hepatic vein flow, and aortic flow can be ascertained readily. It is important to note that Doppler signals of trace or mild valvular regurgitation of no clinical consequence can be detected with structurally normal tricuspid, pulmonic, and mitral valves. Such signals are not likely to generate enough turbulence to create an audible murmur.

Echocardiography is indicated for the evaluation of patients with early, late, or holosystolic murmurs and patients with grade 3 or louder midsystolic murmurs. Patients with grade 1 or 2 midsystolic murmurs but other symptoms or signs of cardiovascular disease, including those from ECG or chest x-ray, should also undergo echocardiography. Echocardiography is also indicated for the evaluation of any patient with a diastolic murmur and for patients with continuous murmurs not due to a venous hum or mammary souffle. Echocardiography should be considered when there is a clinical need to verify normal cardiac structure and function in a patient whose symptoms and signs are probably noncardiac in origin. The performance of serial echocardiography to follow the course of asymptomatic individuals with valvular heart disease is a central feature of their longitudinal assessment, and it provides valuable information that may have an impact on decisions regarding the timing of surgery. Routine echocardiography is *not* recommended for asymptomatic patients with a grade 1 or 2 midsystolic murmur without other signs of heart disease. For this category of patients, referral to a cardiovascular specialist should be considered if there is doubt about the significance of the murmur after the initial examination.

The selective use of echocardiography outlined above has not been subjected to rigorous analysis of its cost-effectiveness. For some clinicians, handheld or miniaturized cardiac ultrasound devices have replaced the stethoscope. Although several reports attest to the improved sensitivity of such devices for the detection of valvular heart disease (e.g., rheumatic heart disease in susceptible populations), accuracy is highly operator-dependent, and incremental cost considerations and outcomes have not been addressed adequately for most patient scenarios. The use of electronic or digital stethoscopes with spectral display capabilities has also been proposed as a method to improve the characterization of heart murmurs and the mentored teaching of cardiac auscultation.

OTHER CARDIAC TESTING

(Chap. 236, Fig. 38-9) In relatively few patients, clinical assessment and TTE do not adequately characterize the origin and significance of a heart murmur. Transesophageal echocardiography (TEE) can be considered for further evaluation, especially when the TTE windows are limited by body size, chest configuration, or intrathoracic pathology. TEE offers enhanced sensitivity for the detection of a wide range of structural cardiac disorders. Electrocardiographically gated cardiac magnetic resonance (CMR) imaging, although limited in its ability to display valvular morphology, can provide quantitative information regarding valvular function, stenosis severity, regurgitant fraction, regurgitant volume, shunt flow, chamber and great vessel size, ventricular function, and myocardial perfusion. CMR has largely supplanted the need for cardiac catheterization and invasive hemodynamic assessment when there is a discrepancy between the clinical and echocardiographic findings. Invasive coronary angiography is performed routinely in most adult patients before valve surgery, especially when there is a suspicion of coronary artery disease predicated on symptoms, risk factors, and/or age. The use of

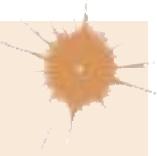
computed tomography coronary angiography (CTCA) to exclude coronary artery disease in selected patients with a low pretest probability of disease before valve surgery has gained wider acceptance.

INTEGRATED APPROACH

The accurate identification of a heart murmur begins with a systematic approach to cardiac auscultation. Characterization of its major attributes, as reviewed above, allows the examiner to construct a preliminary differential diagnosis, which is then refined by integration of information available from the history, associated cardiac findings, the general physical examination, and the clinical context. The need for and urgency of further testing follow sequentially. Correlation of the findings on auscultation with the noninvasive data provides an educational feedback loop and an opportunity for improving physical examination skills. Cost constraints mandate that noninvasive imaging be justified on the basis of its incremental contribution to diagnosis, treatment, and outcome. Cardiac auscultation using a stethoscope remains a time-honored tradition in medicine, the benefits of which extend beyond accurate recognition of heart sounds. Selective augmentation with, rather than wholesale replacement by, handheld ultrasound and newer technologies may improve diagnostic accuracy and better guide therapeutic decisions.

FURTHER READING

- EDELMAN ER, WEBER BN: Tenuous tether. *N Engl J Med* 373:2199, 2015.
- FANG LC, O'GARA PT: The history and physical examination. An evidence-based approach, in *Braunwald's Heart Disease. A Textbook of Cardiovascular Medicine*, 10th ed, DL Mann et al (eds). Philadelphia, Elsevier/Saunders, 2015, pp 95-113.
- FUSTER V: The stethoscope's prognosis. Very much alive and very necessary. *J Am Coll Cardiol* 67:1118, 2016.
- KIMURA BJ et al: Cardiac limited ultrasound examination techniques to augment the bedside cardiac physical examination. *J Ultrasound Med* 34:1683, 2015.
- LAI LS et al: Computerized automatic diagnosis of innocent and pathologic murmurs in pediatrics: A pilot study. *Congen Heart Dis* 11:386, 2016.
- NISHIMURA R et al: 2014 AHA/ACC guideline for the management of patients with valvular heart disease. *J Am Coll Cardiol* 63:2438, 2014.
- SHRESTHA NR et al: Prevalence of subclinical rheumatic heart disease in Eastern Nepal: A school-based cross-sectional study. *JAMA Cardiol* 1:89, 2016.
- STOKKE TM et al: Brief group training of medical students in focused cardiac ultrasound may improve diagnostic accuracy of physical examination. *J Am Soc Echocardiogr* 27:1238, 2014.

39**Palpitations****Joseph Loscalzo**

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/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. 241**) 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 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.

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.
 JAMSHED N, DUBIN J, ELDAGAH Z: Emergency management of palpitations in the elderly: Epidemiology, diagnostic approaches, and therapeutic options. Clin Geriatr Med 29:205, 2013.
 SEDAGHAT-YAZDI F, KOENIG PR: The teenager with palpitations. Pediatr Clin North Am 61:63, 2014.
 WEBER BE, KAPOOR WN: Evaluation and outcomes of patients with palpitations. Am J Med 100:138, 1996.

Section 6 Alterations in Gastrointestinal Function**40****Dysphagia****Ikuko Hirano, Peter J. Kahrilas**

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 and 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.

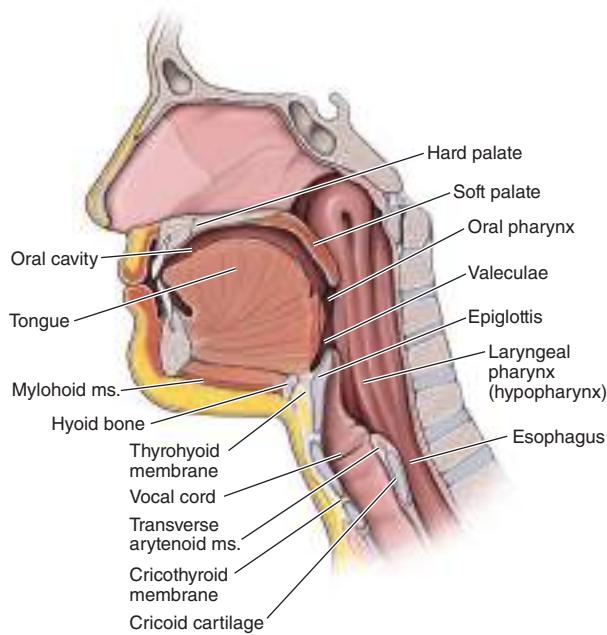
The musculature of the oral cavity, pharynx, UES, and cervical esophagus is striated and directly innervated by lower motor neurons carried in cranial nerves (Fig. 40-1). Oral cavity muscles are innervated by the fifth (trigeminal) and seventh (facial) cranial nerves; the tongue,

by the twelfth (hypoglossal) cranial nerve. Pharyngeal muscles are innervated by the ninth (glossopharyngeal) and tenth (vagus) cranial nerves.

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 suprathyroid 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.

Sagittal view of the pharynx



Musculature of the pharynx

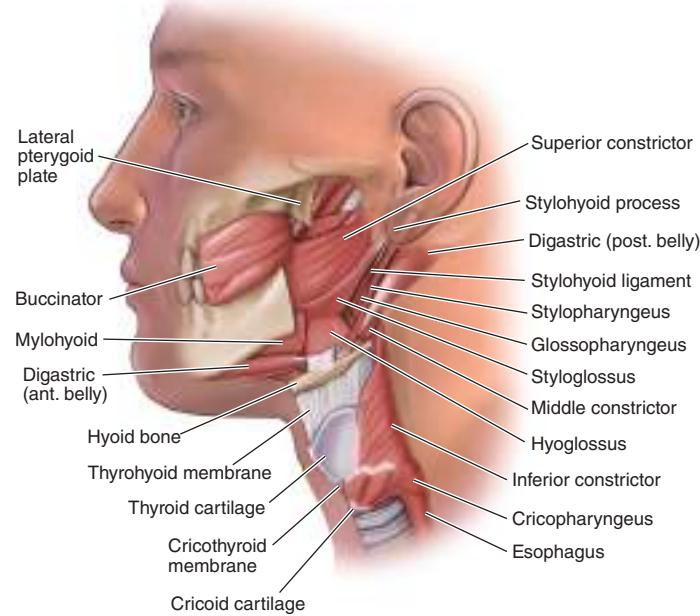


FIGURE 40-1 Sagittal and diagrammatic views of the musculature involved in enacting oropharyngeal swallowing. Note the dominance of the tongue in the sagittal view and the intimate relationship between the entrance to the larynx (airway) and the esophagus. In the resting configuration illustrated, the 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.)

■ 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 degluttitive 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.

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 which 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 degluttitive 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 non-peristaltic, disordered contractions. Absent peristalsis and failure of degluttitive 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 nonspecific pattern commonly found in patients with scleroderma.

APPROACH TO THE PATIENT

Dysphagia

Figure 40-2 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 restricting 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 a crucial detail. Intermittent dysphagia that occurs only with solid food implies

APPROACH TO THE PATIENT WITH DYSPHAGIA

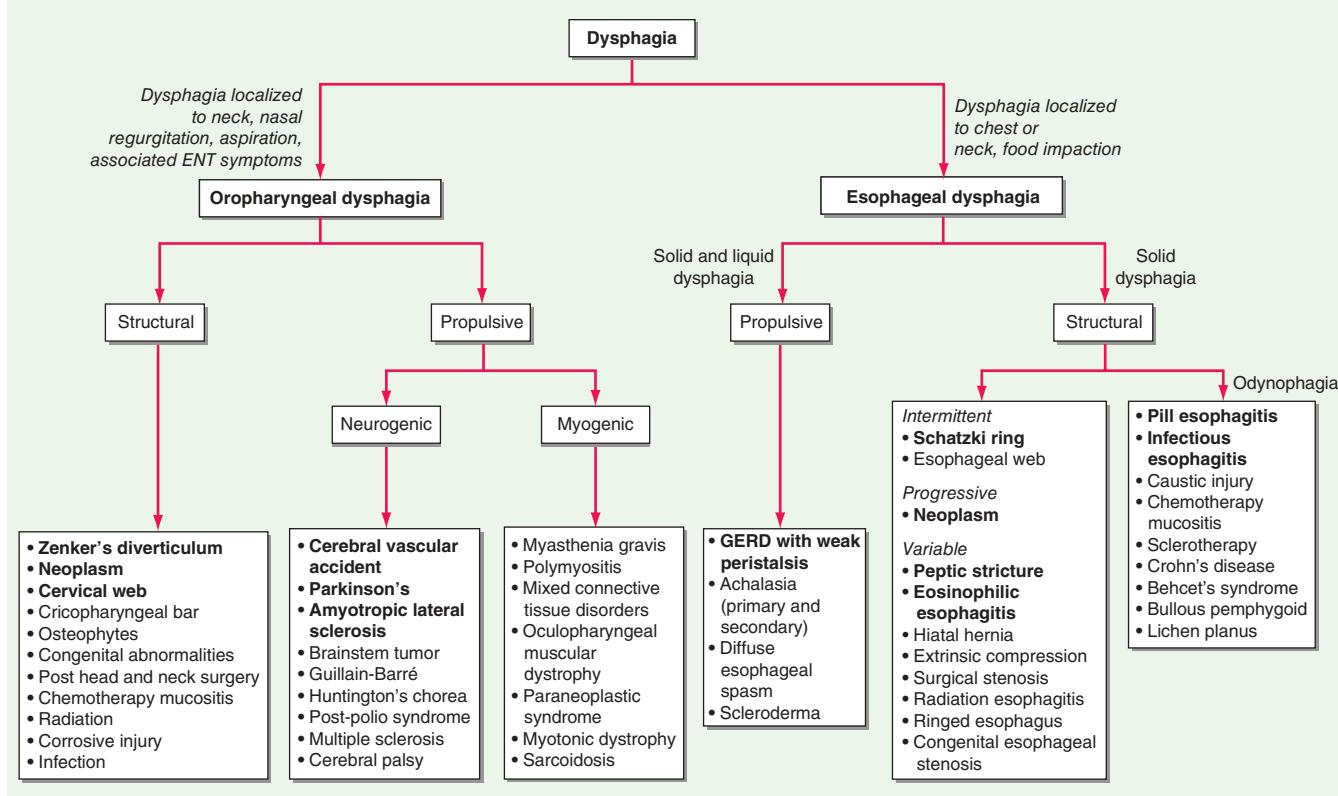


FIGURE 40-2 Approach to the patient with dysphagia. Etiologies in bold print are the most common. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.

structural dysphagia, whereas constant dysphagia with both liquids and solids strongly suggests a 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, somewhat paradoxically, 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 frequently accompanies 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 strong history of atopy increases concerns for eosinophilic esophagitis, especially in younger Caucasian male patients.

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, tongue atrophy, and hyperactive jaw jerk, in addition to evidence

of generalized neuromuscular disease, should be elicited. The neck should be examined for thyromegaly. A careful inspection of the mouth and pharynx should disclose lesions that may interfere with passage of food. 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 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 due to intraluminal obstruction (esophageal or proximal gastric cancer, metastatic deposits), 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. 40-2). 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 one to obtain 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

41

Nausea, Vomiting, and Indigestion

William L. Hasler



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 esophageal strictures, prior esophageal surgery, esophageal diverticula, or paraesophageal herniation. In specific cases, computed tomography (CT) examination 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 trained 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 gastrostomy and enteral feeding. Patients with myasthenia gravis (Chap. 440) and polymyositis (Chap. 358) 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 the idiopathic 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. 316. The majority of causes of esophageal dysphagia are effectively managed by means of esophageal dilatation 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 has emerged as an important cause of dysphagia that is amenable to treatment by elimination of dietary allergens or administration of swallowed, topically acting glucocorticoids.

FURTHER READING

- Cook IJ: Oropharyngeal dysphagia. *Gastroenterol Clin North Am* 38:411, 2009.
- Hirano I: Esophagus: Anatomy and structural anomalies, in *Yamada Atlas of Gastroenterology*, 6th ed. Wiley-Blackwell Publishing Co. 2016, pp 42–59.
- Kahrilas PJ et al: The Chicago Classification of esophageal motility disorders, v3.0. *Neurogastroenterol Motil* 27:160, 2015.
- 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.

Nausea is the subjective feeling of a need to vomit. **Vomiting** (emesis) is the oral expulsion of gastrointestinal contents due to 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** is a term encompassing 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 (an inability to complete a meal due to premature fullness), bloating, eructation (belching), and anorexia. Others report predominantly epigastric burning or pain.

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 conscious perception. This is supported by functional brain imaging studies showing activation of cerebral cortical regions 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 NK₁, serotonin 5-HT₃, and vasopressin pathways.

Somatic and visceral muscles respond stereotypically during emesis. Inspiratory thoracic and abdominal wall muscles contract, producing high intrathoracic and intraabdominal pressures that evacuate the stomach. The gastric cardia herniates above the diaphragm, and the larynx moves upward to propel the vomitus. Distally migrating gut contractions are normally regulated 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, the slow wave is abolished and replaced by orally propagating spikes that evoke retrograde contractions that assist in expulsion of gut contents.

Activators of Emesis Emetic stimuli act at several sites. Emesis evoked by unpleasant thoughts or smells originates in the brain, whereas cranial nerves mediate vomiting after gag reflex activation. 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 M₁ and histaminergic H₁ receptors. Vagal afferent stimuli activate 5-HT₃ receptors. The area postrema is served by nerves acting on 5-HT₃, M₁, H₁, and dopamine D₂ subtypes. Central NK₁ receptors mediate both nausea and vomiting. Cannabinoid CB₁ pathways may participate in the cerebral cortex and brainstem. Optimal pharmacologic therapy of vomiting requires understanding these pathways.

DIFFERENTIAL DIAGNOSIS

Nausea and vomiting are caused by conditions within and outside the gut, by drugs, and by circulating toxins (Table 41-1). Unexplained causes of chronic nausea and vomiting are relatively rare, being reported by 2–3% of the population.

TABLE 41-1 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	Antibiotics
Colonic obstruction	Labyrinthine disease	Cardiac antiarrhythmics
Superior mesenteric artery syndrome	Motion sickness	Digoxin
Enteric infections	Labyrinthitis	Oral hypoglycemics
Viral	Malignancy	Oral contraceptives
Bacterial	Intracerebral disorders	Antidepressants
Inflammatory diseases	Malignancy	Restless legs/Parkinson's therapies
Cholecystitis	Hemorrhage	Smoking cessation agents
Pancreatitis	Abscess	Endocrine/metabolic disease
Appendicitis	Hydrocephalus	Pregnancy
Hepatitis	Psychiatric illness	Uremia
Altered sensorimotor function	Anorexia and bulimia nervosa	Ketoacidosis
Gastroparesis	Depression	Thyroid and parathyroid disease
Intestinal pseudoobstruction	Postoperative vomiting	Adrenal insufficiency
Gastroesophageal reflux		Toxins
Chronic nausea vomiting syndrome		Liver failure
Cyclic vomiting syndrome		Ethanol
Cannabinoid hyperemesis syndrome		
Rumination syndrome		
Biliary colic		
Abdominal irradiation		

Intrapерitoneal Disorders Visceral obstruction and inflammation of hollow and solid viscera may elicit vomiting. Gastric obstruction results from ulcers and malignancy. Small-bowel and colon blockage occur because of adhesions, benign or malignant tumors, volvulus, intussusception, or inflammatory diseases like Crohn's disease. 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. Abdominal irradiation impairs intestinal motor function and induces strictures. Biliary colic causes nausea by acting on local afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis result from visceral irritation and induction of ileus. Enteric infections with viruses like norovirus or rotavirus or bacteria like *Staphylococcus aureus* and *Bacillus cereus* cause vomiting, especially in children. Opportunistic infections like cytomegalovirus or herpes simplex virus induce emesis in immunocompromised individuals.

Gut sensorimotor dysfunction often causes nausea and vomiting. *Gastroparesis* presents with symptoms of gastric retention 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 common etiology. It occurs in the absence of systemic illness and follows a viral illness in ~15–20% of cases, suggesting an infectious trigger. *Intestinal pseudoobstruction* is characterized by disrupted intestinal and colonic motor activity 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 as a familial visceral myopathy or neuropathy, 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 report nausea and vomiting, as do some with irritable bowel syndrome (IBS) or chronic constipation.

Other functional gastroduodenal disorders without organic abnormalities have been characterized. *Chronic nausea vomiting syndrome* is defined as bothersome nausea at least one day and/or one or more vomiting episodes weekly in the absence of an eating disorder or psychiatric disease. *Cyclic vomiting syndrome* causes 3–14% of cases of unexplained nausea and vomiting and presents with periodic discrete

episodes of relentless vomiting in children and adults and shows an association with migraine headaches, suggesting that some cases may be migraine variants. Some adult cases have been associated with rapid gastric emptying. A related condition, *cannabinoid hyperemesis syndrome*, presents with cyclical vomiting with intervening well periods in individuals (mostly men) who use large quantities of cannabis over many years and resolves with its discontinuation. Pathologic behaviors such as taking prolonged hot baths or showers are associated with the syndrome. *Rumination syndrome*, characterized by repetitive regurgitation of recently ingested food, 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, most commonly 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 psychiatric illnesses including anorexia nervosa, bulimia nervosa, anxiety, and depression often report significant nausea that may be associated with delayed gastric emptying.

Medications and Metabolic Disorders Drugs evoke vomiting by action on the stomach (analgesics, erythromycin) or area postrema (opiates, 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 vomiting that is acute (within hours of administration), delayed (after 1 or more days), or anticipatory. Acute emesis from highly emetogenic agents (e.g., cisplatin) is mediated by 5-HT₃ pathways. Delayed emesis is less dependent on 5-HT₃ pathways with greater mediation by NK₁ mechanisms. Anticipatory nausea may respond to anxiolytic therapy rather than antiemetics.

Metabolic disorders elicit nausea and vomiting. Pregnancy is the most prevalent endocrinologic cause, and nausea affects 70% of women in the first trimester. Hyperemesis gravidarum is a severe form of nausea of pregnancy that produces significant dehydration and electrolyte disturbances. 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, the vomitus may contain food residue ingested days before. Hematemesis raises suspicion of an 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. Profound weight loss raises concern about malignancy or obstruction. Fevers suggest inflammation. An intracranial source is considered if there are headaches or visual field 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 may be absent with ileus. High-pitched rushes suggest bowel obstruction, whereas a succussion splash upon abrupt lateral movement of the patient is found with gastroparesis or pyloric obstruction. Tenderness or involuntary guarding raises suspicion of inflammation. Fecal blood suggests mucosal injury from ulcer, ischemia, or tumor. Neurologic disease presents with papilledema, visual field loss, or focal neural abnormalities. Neoplasm is suggested by palpable masses or adenopathy.

DIAGNOSTIC TESTING

For intractable symptoms or an elusive diagnosis, selected screening tests can direct clinical care. Electrolyte replacement is indicated for hypokalemia or metabolic alkalosis. Iron-deficiency anemia mandates a search for mucosal injury. Pancreaticobiliary disease is indicated by abnormal pancreatic or liver biochemistries. Endocrinologic, rheumatologic, or paraneoplastic etiologies are suggested by hormone or serologic abnormalities. If obstruction is suspected, supine and upright abdominal radiographs may show intestinal air-fluid levels with reduced colonic air. Ileus is characterized by diffusely dilated air-filled bowel loops.

Anatomic studies may be indicated if initial testing is nondiagnostic. Upper endoscopy detects ulcers, malignancy, and retained food residue 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 provide define inflammation in Crohn's disease. Brain CT or MRI can delineate

intracranial disease. Mesenteric angiography, CT, or MRI is useful for suspected ischemia.

Gastrointestinal motility testing may detect an underlying motor disorder. Gastroparesis commonly is diagnosed by gastric scintigraphy, by which measures emptying of a radiolabeled meal. A non-radioactive ¹³C-labelled gastric emptying breath test was FDA-approved in 2015 and may be a cost-effective alternative to scintigraphy. Intestinal pseudoobstruction is suggested by abnormal barium transit and luminal dilation on small-bowel contrast radiography. Wireless motility capsule methods measure transit in the stomach, small bowel, and colon by detecting pH changes between regions and also can diagnose gastroparesis and small bowel dysmotility. Small-intestinal manometry can confirm the diagnosis of pseudoobstruction and characterize the motor abnormality as neuropathic or myopathic 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.

TREATMENT

Nausea and Vomiting

GENERAL PRINCIPLES

Therapy of vomiting is tailored to correcting remediable abnormalities if possible. Hospitalization is considered for severe dehydration, especially if oral fluid replenishment cannot be sustained. Once oral intake is tolerated, nutrients are restarted with low-fat liquids, because lipids delay gastric emptying. A low residue, small particle diet has shown efficacy in gastroparesis in a controlled study. Controlling blood glucose in poorly controlled diabetics can reduce hospitalizations in gastroparesis and may improve nausea and vomiting.

ANTIEMETIC MEDICATIONS

The most commonly used antiemetic agents act on central nervous system sites (Table 41-2). Antihistamines like dimenhydrinate and meclizine and anticholinergics like scopolamine act on labyrinthine pathways to treat motion sickness and labyrinthine disorders. D₂ antagonists treat emesis evoked by area postrema stimuli and are used for medication, toxic, and metabolic etiologies. Dopamine antagonists cross the blood-brain barrier and cause anxiety,

TABLE 41-2 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	Medication-, toxin-, or metabolic-induced emesis
	5-HT ₃ antagonist	Ondansetron, granisetron	Chemotherapy- and radiation-induced emesis, postoperative emesis
	NK ₁ antagonist	Aprepitant	Chemotherapy-induced nausea and vomiting
	Tricyclic antidepressant	Amitriptyline, nortriptyline	Chronic nausea vomiting syndrome, cyclic vomiting syndrome, ?gastroparesis
	Other antidepressant	Mirtazapine, olanzapine	?Chronic nausea vomiting syndrome, ?gastroparesis
Prokinetic agents	5-HT ₄ agonist and antidopaminergic	Metoclopramide	Gastroparesis
	Motilin agonist	Erythromycin	Gastroparesis, ?intestinal pseudoobstruction
	Peripheral antidopaminergic	Domperidone	Gastroparesis
	Somatostatin analogue	Octreotide	Intestinal pseudoobstruction
	Acetylcholinesterase inhibitor	Pyridostigmine	?Small-intestinal dysmotility/pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy
	Glucocorticoids	Methylprednisolone, dexamethasone	Chemotherapy-induced emesis
	Cannabinoids	Tetrahydrocannabinol	Chemotherapy-induced emesis

Note: ?, indication is uncertain.

movement disorders, and hyperprolactinemic effects (galactorrhea, sexual dysfunction).

Other classes exhibit antiemetic properties. 5-HT₃ antagonists like ondansetron and granisetron prevent postoperative vomiting, radiation therapy-induced symptoms, and cancer chemotherapy-induced emesis, but also are used for other causes of emesis. NK₁ antagonists like aprepitant are approved for chemotherapy-induced vomiting and also reduce 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 may exhibit antiemetic effects.

GASTROINTESTINAL MOTOR STIMULANTS

Drugs that stimulate gastric emptying are used for gastroparesis (Table 41-2). Metoclopramide, a combined 5-HT₄ agonist and D₂ 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 fasting motor stimulant. Intravenous erythromycin is useful for inpatients with refractory gastroparesis. Utility of oral forms is limited by development of tolerance. Domperidone, a D₂ antagonist not available in the United States, exhibits prokinetic and antiemetic effects but does not cross into most brain regions; thus, dystonic reactions are rare. Domperidone can induce hyperprolactinemic side effects via effects on pituitary regions served by a porous blood-brain barrier. Prucalopride, a 5-HT₄ agonist available in Canada and Europe, has shown efficacy in a preliminary gastroparesis trial.

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 may benefit some patients with small-bowel dysmotility. Pyloric botulinum toxin injections are reported in uncontrolled studies to reduce gastroparesis symptoms, but small controlled trials observe benefits no greater than sham treatments. Surgical pyloroplasty and peroral endoscopic myotomy (POEM) of the pylorus has improved symptoms in case series. Placing a feeding jejunostomy reduces hospitalizations and improves overall health in some patients with refractory gastroparesis. Postvagotomy gastroparesis may improve with near-total gastric resection; similar operations are being tried for other gastroparesis etiologies. Implanted gastric electrical stimulators may reduce symptoms, enhance nutrition, improve quality of life, and decrease health care expenditures in medication-refractory gastroparesis, but small controlled trials do not report convincing benefits.

SAFETY CONSIDERATIONS

Safety concerns have been raised about selected antiemetics. Centrally acting antidopaminergics, especially 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-HT₃ antagonists can induce dangerous cardiac arrhythmias, especially in those with QTc interval prolongation on electrocardiography (ECG). Surveillance ECG testing has been advocated for some of these agents.

SELECTED CLINICAL SETTINGS

Some cancer chemotherapies are intensely emetogenic (Chap. 69). Combining a 5-HT₃ antagonist, an NK₁ antagonist, and a glucocorticoid can control both acute and delayed vomiting after highly emetogenic chemotherapy. Unlike other drugs in the same class, the 5-HT₃ antagonist palonosetron can prevent delayed chemotherapy-induced vomiting. Benzodiazepines like lorazepam reduce anticipatory nausea and vomiting. Miscellaneous therapies with benefit in chemotherapy-induced emesis include cannabinoids,

olanzapine, and alternative therapies like ginger. Most antiemetic regimens produce greater reductions in vomiting than nausea.

Clinicians should exercise caution in managing pregnant patients with nausea. Studies of the teratogenic effects of antiemetic agents provide conflicting results. Few controlled trials have been performed in nausea of pregnancy. Antihistamines like meclizine and doxylamine, antidopaminergics like prochlorperazine, and antiserotonergics like ondansetron demonstrate limited efficacy. Some obstetricians offer alternative therapies including pyridoxine, acupressure, or ginger.

Managing cyclic vomiting syndrome is challenging. Prophylaxis with tricyclic antidepressants, cyproheptadine, or β-adrenoceptor antagonists can reduce the severity and frequency of attacks. Intravenous 5-HT₃ antagonists combined with the sedating effects of a benzodiazepine like lorazepam are a mainstay for treating acute flares. Small studies report benefits with antimigraine agents, including the 5-HT₁ agonist sumatriptan, and selected anticonvulsants like topiramate, zonisamide, and levetiracetam.

INDIGESTION

MECHANISMS

The most common causes of indigestion are gastroesophageal reflux and functional dyspepsia. Other cases are a consequence of organic illness.

Gastroesophageal Reflux Gastroesophageal reflux results from many physiologic defects. Reduced lower esophageal sphincter (LES) tone contributes to reflux in scleroderma and pregnancy and may be a factor in some patients without systemic illness. Others exhibit frequent transient LES relaxations (TLESRs) that permit bathing of the esophagus by acid or nonacidic fluid. Reductions in esophageal body motility or salivary secretion prolong fluid exposure. Increased intragastric pressure promotes gastroesophageal reflux in obese patients. The role of hiatal hernias is controversial—most reflux patients have hiatal hernias, but most with hiatal hernias do not report excess heartburn.

Gastric Motor Dysfunction Disturbed gastric motility may contribute to gastroesophageal reflux in up to one-third of cases. Delayed gastric emptying is also found in ~30% of functional dyspeptics, while rapid gastric emptying affects 5%. The relation of these defects to symptom induction is uncertain; studies show poor correlation between symptom severity and degrees of motor dysfunction. 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. Others with dyspepsia exhibit hypersensitivity to chemical stimulation with capsaicin or with acid or lipid perfusion of the duodenum. Some individuals with functional heartburn without increased acid or nonacid reflux may have heightened perception of normal esophageal acidity.

Other Factors *Helicobacter pylori* has a clear etiologic role in peptic ulcer disease, but ulcers cause a minority of dyspepsia cases. *H. pylori* 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 contributions from central nervous system pathways. Inflammatory factors like duodenal eosinophilia (and possibly increased duodenal mast cells) may contribute to early satiety and pain in functional dyspepsia. Up to 20% of functional dyspepsia patients report symptom onset after a viral illness, suggestive of an infectious cause. Analgesics cause dyspepsia, whereas nitrates, calcium channel blockers, theophylline, and progesterone promote

gastroesophageal reflux. Other stimuli that induce reflux include ethanol, tobacco, and caffeine via LES relaxation. Genetic factors predispose to development of reflux and dyspepsia.

■ DIFFERENTIAL DIAGNOSIS

Gastroesophageal Reflux Disease Gastroesophageal reflux disease (GERD) is prevalent. Heartburn or regurgitation are reported weekly by 18–28%. Most cases of heartburn result from excess acid reflux, but reflux of nonacidic fluid produces similar symptoms. Alkaline reflux esophagitis produces GERD-like symptoms most often in patients who have had surgery for peptic ulcer disease. Ten percent of patients with heartburn exhibit no increase in acid or nonacidic esophageal reflux (functional heartburn).

Functional Dyspepsia Nearly 25% 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% of dyspeptic patients, is defined as bothersome postprandial fullness, early satiety, or epigastric pain or burning with symptom onset at least 6 months before diagnosis in the absence of organic cause. Functional dyspepsia is subdivided into postprandial distress syndrome, characterized by meal-induced fullness and early satiety, and epigastric pain syndrome, which presents with epigastric pain or burning which may or may not be meal-related. Most cases follow a benign course, but some with *H. pylori* infection or on nonsteroidal anti-inflammatory drugs (NSAIDs) develop ulcers.

Ulcer Disease In most GERD patients, there is no injury to the esophagus. However, 5% develop esophageal ulcers, and some form strictures. 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. 319) and Zollinger-Ellison syndrome (Chap. 317), resulting from gastrin overproduction by an endocrine tumor.

Malignancy Dyspeptic patients often 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. 76). 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 inflammation include eosinophilic esophagitis and pill esophagitis. Biliary colic is in the differential diagnosis of unexplained upper abdominal pain, but most patients with biliary colic report discrete acute episodes of right upper quadrant or epigastric pain rather than the chronic burning or fullness of dyspepsia. Twenty percent of gastroparesis patients report a predominance of pain rather than nausea and vomiting. Intestinal lactase deficiency as a cause of gas, bloating, and discomfort occurs in 15–25% of whites of northern European descent but is more common in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms. Small-intestinal bacterial overgrowth may cause dyspepsia, often associated with bowel dysfunction, distention, and malabsorption. Celiac disease, pancreatic disease (chronic pancreatitis, malignancy), hepatocellular carcinoma, Ménétrier's disease, infiltrative diseases (sarcoidosis, eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Gluten sensitivity in the absence of celiac disease can elicit unexplained upper abdominal symptoms. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis.

APPROACH TO THE PATIENT

Indigestion

HISTORY AND PHYSICAL EXAMINATION

Management of 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 salivary secretions 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 do not report heartburn, but note abdominal pain or other symptoms.

Dyspeptic patients typically report symptoms referable to the upper abdomen that may be meal-related, as with postprandial distress syndrome, or possibly independent of food ingestion in epigastric pain syndrome. Functional dyspepsia overlaps with other disorders including GERD, IBS, and idiopathic gastroparesis.

The physical examination with GERD and functional dyspepsia usually is normal. In atypical GERD, pharyngeal erythema and wheezing may be noted. Recurrent acid regurgitation may cause poor dentition. Dyspeptics may exhibit epigastric tenderness or distention.

Discriminating functional from organic causes of indigestion mandates excluding certain historic and examination features. Odynophagia suggests esophageal infection. Dysphagia is concerning for a benign or malignant esophageal blockage. Other alarm features include unexplained weight loss, recurrent vomiting, occult or gross bleeding, jaundice, palpable mass or adenopathy, and a family history of gastrointestinal neoplasm.

DIAGNOSTIC TESTING

Because indigestion is prevalent and most cases result from GERD or functional dyspepsia, a general principle is to perform only limited and directed diagnostic testing in selected individuals.

Once alarm factors are excluded (Table 41-3), patients with typical GERD do not need further evaluation and are treated empirically. Upper endoscopy is indicated to exclude mucosal injury in cases with atypical symptoms or 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 exhibit a therapeutic response 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 >55 years old or who have

TABLE 41-3 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

alarm factors because of the purported elevated risks of malignancy and ulcer in these groups. However, findings of endoscopy performed for uninvestigated dyspepsia include erosive esophagitis in 13%, peptic ulcer in 8%, and gastric or esophageal malignancy in only 0.3%. Management of patients <55 years old without alarm factors depends on the local prevalence of *H. pylori* infection. In regions with low *H. pylori* prevalence (<10%), a 4-week trial of an acid-suppressing medication such as a proton pump inhibitor (PPI) is recommended. If this fails, a “test and treat” approach is most commonly applied. *H. pylori* status is determined with urea breath testing or stool antigen measurement. Those who are *H. pylori* positive are given therapy to eradicate the infection. If symptoms resolve on either regimen, no further intervention is required. For patients in areas with high *H. pylori* prevalence (>10%), an initial test and treat approach is advocated, with a subsequent trial of an acid-suppressing regimen offered for those in whom *H. pylori* treatment fails or for those who are negative for the infection. In each of these patient subsets, upper endoscopy is reserved for those whose symptoms fail to respond to therapy.

Further testing is indicated in some settings. If bleeding is noted, a blood count can exclude anemia. Thyroid chemistries or calcium levels screen for metabolic disease, whereas specific serologies may suggest celiac disease. Pancreatic and liver chemistries are obtained for possible pancreaticobiliary causes which are further investigated with ultrasound, CT, or MRI. Gastric emptying testing is considered to exclude gastroparesis for dyspeptic symptoms that resemble postprandial distress when drug therapy fails and in some GERD patients, especially if surgical intervention is an option. Breath testing after carbohydrate ingestion detects lactase deficiency, intolerance to other carbohydrates, or small-intestinal bacterial overgrowth.

TREATMENT

Indigestion

GENERAL PRINCIPLES

For mild indigestion, reassurance that a careful evaluation revealed no serious organic disease may be the only intervention needed. Drugs that cause gastroesophageal reflux or dyspepsia should be stopped, if possible. Patients with GERD should limit ethanol, caffeine, chocolate, and tobacco use due to their effects on the LES. Other measures in GERD include ingesting a low-fat diet, avoiding snacks before bedtime, and elevating the head of the bed. Patients with functional dyspepsia also may be advised to reduce intake of fat, spicy foods, caffeine, and alcohol.

Specific therapies for organic disease should be offered when possible. Surgery is appropriate for biliary colic. Diet changes are indicated for lactase deficiency or celiac disease. Peptic ulcers may be cured by specific medical regimens. However, because most indigestion is caused by GERD or functional dyspepsia, medications that reduce gastric acid, modulate motility, or blunt gastric sensitivity are used.

ACID-SUPPRESSING OR NEUTRALIZING MEDICATIONS

Drugs that reduce or neutralize gastric acid are often prescribed for GERD. Histamine H₂ 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 H₂ 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 typically responds better to PPI therapy than regurgitation or atypical GERD symptoms. Some individuals respond to doubling of the PPI dose or adding an H₂ antagonist at bedtime. Infrequent complications of long-term

PPI therapy include diarrhea (from *Clostridium difficile* infection or microscopic colitis), small-intestinal bacterial overgrowth, nutrient deficiency (vitamin B₁₂, iron, calcium), hypomagnesemia, bone demineralization, interstitial nephritis, and impaired medication absorption (e.g., clopidogrel). Many patients started on a PPI can be stepped down to an H₂ antagonist or switched to an on-demand schedule.

Acid-suppressing drugs are also effective in selected patients with functional dyspepsia. A meta-analysis of 10 controlled trials calculated a risk ratio of 0.87, with a 95% confidence interval of 0.80–0.96, favoring PPI therapy over placebo. H₂ antagonists also reportedly improve symptoms in functional dyspepsia; however, findings of trials of this drug class likely are influenced by inclusion of large numbers of GERD patients.

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). Algiric 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 H₂ antagonists.

HELICOBACTER PYLORI ERADICATION

H. pylori eradication is definitively indicated only 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. 158 and 317**) include 10–14 days of a PPI or bismuth subsalicylate with two antibiotics. *H. pylori* infection is associated with reduced prevalence of GERD, especially in the elderly. However, eradication of the 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

Prokinetics like metoclopramide, erythromycin, and domperidone have limited utility in GERD. The γ-aminobutyric acid B (GABA-B) agonist baclofen reduces esophageal exposure to acid and nonacidic fluids by reducing TLESRs by 40%; this drug is proposed as adjunctive therapy for refractory acid and nonacid reflux. Several studies have promoted the efficacy of motor-stimulating drugs 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 prokinetic drugs. The 5-HT_{1A} agonists buspirone and tandospirone may improve some functional dyspepsia symptoms by enhancing meal-induced gastric accommodation. Acotiamide promotes 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.

ANTIDEPRESSANTS

Some patients with refractory functional heartburn may respond to antidepressants in 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 recent controlled trial in functional dyspepsia, the tricyclic drug amitriptyline produced symptom reductions while the SSRI escitalopram had no benefit in a 3-way comparison with placebo. In another controlled trial in functional dyspepsia, the antidepressant mirtazapine produced superior symptom reductions versus placebo.

OTHER OPTIONS

Antireflux surgery (fundoplication) to increase LES pressure may be offered to GERD patients who are young and require lifelong therapy, have typical heartburn and regurgitation, 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 or 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. Studies assessing the utility and safety of gastroesophageal junction endoscopic therapies (radiofrequency therapy, transoral fundoplication, endoscopic stapling, antireflux mucosectomy) and laparoscopic magnetic sphincter augmentation to enhance gas-troesophageal barrier function in GERD are ongoing.

Gas and bloating can be troubling symptoms in some patients with indigestion that are difficult to treat. Dietary exclusion of gas-producing foods such as legumes and use of simethicone or activated charcoal provide benefits in some cases. Low FODMAP (fermentable oligosaccharide, disaccharide, monosaccharide, and polyol) diets and therapies to modify gut flora (nonabsorbable antibiotics, probiotics) reduce gaseous symptoms in some IBS patients. The utility of low-FODMAP diets, antibiotics, and probiotics in functional dyspepsia is unproven. Herbal remedies such as STW 5 (Iberogast, a mixture of nine herbal agents) are useful in some dyspeptic patients. Psychological treatments (e.g., behavioral therapy, psychotherapy, hypnotherapy) may be offered for refractory functional dyspepsia, but no convincing data confirm their efficacy.

FURTHER READING

- HASLER WL: Newest drugs for unexplained nausea and vomiting. *Curr Treat Options Gastroenterol* 14:371, 2016.
- PATTI MG: An evidence-based approach to the treatment of gastroesophageal reflux disease. *JAMA Surg* 151:73, 2016.
- SCARPELLINI E et al: Management of refractory typical GERD symptoms. *Nat Rev Gastroenterol Hepatol* 13:281, 2016.
- STANGHELLINI V et al: Gastroduodenal disorders. *Gastroenterology* 150:1380, 2016.
- TALLEY NJ, FORD AC: Functional dyspepsia. *N Engl J Med* 373:1853, 2015.

and constipation are among the most common patient complaints presenting to internists and primary care physicians, and they 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 42-1**. 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.

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

42

Diarrhea and Constipation

Michael Camilleri, Joseph A. Murray

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). The annual economic burden to society may exceed \$20 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. 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. Diarrhea

TABLE 42-1 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.

arterial vessels. Sympathetic input to the gut is generally excitatory to sphincters and inhibitory to non-sphincteric 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.

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-5, creating a glucose concentration and osmotic gradient between the lumen and the intercellular space, drawing water and electrolytes passively from the lumen. A variety of neural and nonneuronal 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 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 increases 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. 42-1**). 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.

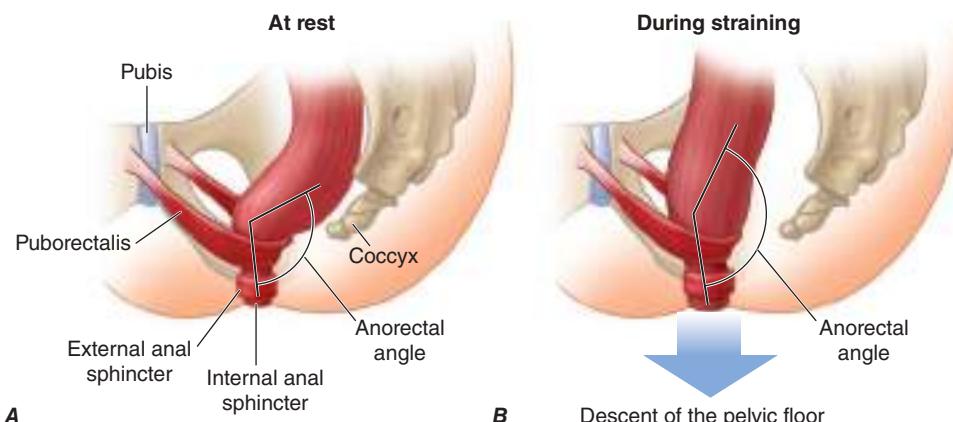


FIGURE 42-1 Sagittal view of the anorectum (A) at rest and (B) during straining to defecate. Continence is maintained by normal rectal sensation and tonic contraction of the internal anal sphincter and the puborectalis muscle, which wraps around the anorectum, maintaining an anorectal angle between 80° 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. (Reproduced with permission from A Lembo, M Camilleri: *N Engl J Med* 349:1360, 2003.)

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. 129). 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.

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 or soft cheeses; and *Vibrio* species, *Salmonella*, or acute hepatitis A from seafood, especially if raw. State departments of public health issue communications regarding food-related illnesses, which may have originated domestically or been imported, but ultimately cause epidemics in the United States (e.g., the *Cyclospora* epidemic of 2013 in midwestern states that resulted from bagged salads).

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*, Microsporida, and *Blastocystis hominis*) may also play a role (Chap. 197). 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.

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 may be acquired in the community.

The pathophysiology underlying acute diarrhea by infectious agents produces specific clinical features that may also be helpful in diagnosis (Table 42-2). 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.

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-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

TABLE 42-2 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					
<i>Bacillus cereus</i> , <i>Staphylococcus aureus</i> , <i>Clostridium perfringens</i>	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
<i>Vibrio cholerae</i> , enterotoxigenic <i>Escherichia coli</i> , <i>Klebsiella pneumoniae</i> , <i>Aeromonas</i> species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent <i>E. coli</i> , <i>Giardia</i> organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
<i>Clostridium difficile</i>	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic <i>E. coli</i>	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					
<i>Salmonella</i> , <i>Campylobacter</i> , and <i>Aeromonas</i> species, <i>Vibrio parahaemolyticus</i> , <i>Yersinia</i>	12 h–11 d	0–3+	2–4+	3–4+	1–4+, watery or bloody
Severe inflammation					
<i>Shigella</i> species, enteroinvasive <i>E. coli</i> , <i>Entamoeba histolytica</i>	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.

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. 42-2). 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 $\geq 38.5^{\circ}\text{C}$ ($\geq 101^{\circ}\text{F}$), 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

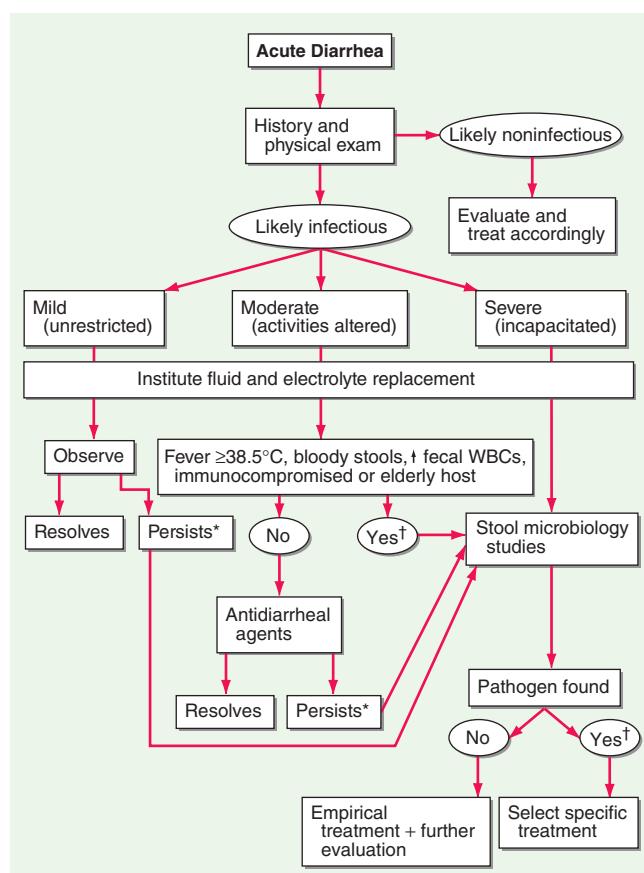


FIGURE 42-2 Algorithm for the management of acute diarrhea. Consider empirical treatment before evaluation with (*) metronidazole and with (†) quinolone. WBCs, white blood cells.

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. 218), 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.

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 exacerbated or prolonged by them. 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. 42-2). 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. 128, 156, and 160–166). 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 42-3).

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

TABLE 42-3 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 (\downarrow 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 neuromopathies
Hyperthyroidism
Drugs (prokinetic agents)
Postvagotomy
Factitious Causes
Munchausen
Eating disorders
Iatrogenic Causes
Cholecystectomy
Ileal resection
Bariatric surgery
Vagotomy, fundoplication

Abbreviation: FODMAP fermentable oligosaccharides, disaccharides, monosaccharides, and polyols.

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 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 $\text{Cl}^-/\text{HCO}_3^-$ 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 non-celiac 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), 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 lymphopenia. 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 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 to 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*, *Behcet'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.

Factitious Causes Factitious 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. 42-3). When this strategy is unrevealing, simple triage tests are often warranted to direct the choice of more complex investigations (Fig. 42-3). The history, physical examination (Table 42-4), 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 inflammatory bowel disease (IBD) or sprue 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

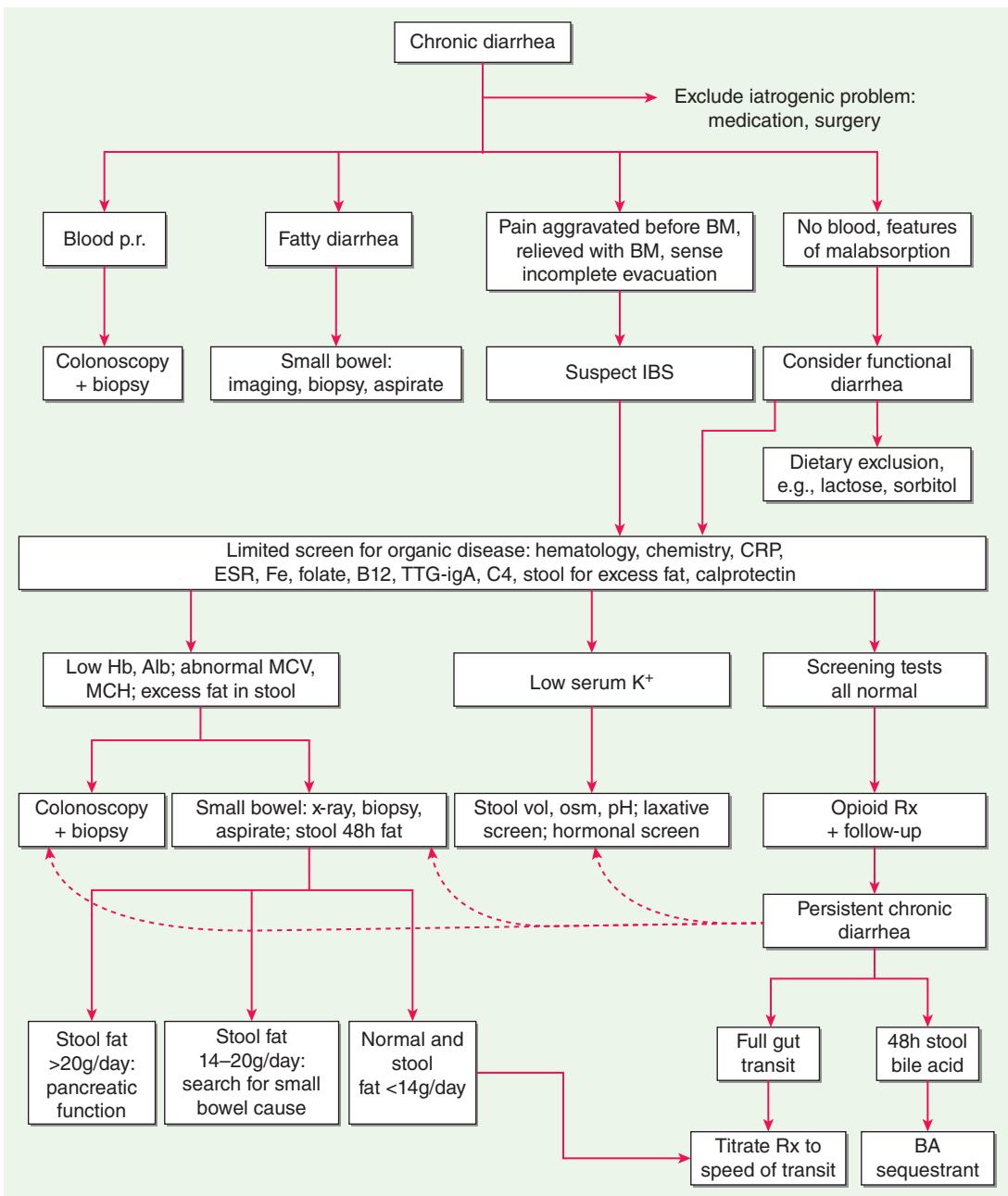


FIGURE 42-3 Algorithm for management of chronic diarrhea. Patients undergo an initial evaluation based on different symptom presentations, leading to selection of patients for imaging, biopsy analysis, and limited screens for organic diseases. Alb, albumin; BA, bile acid; BM, bowel movement; C4, 7 α-hydroxy-4-cholesten-3-one; 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. (Reprinted from M Camilleri, JH Sellin, KE Barrett: Pathophysiology, evaluation, and management of chronic watery diarrhea. *Gastroenterology* 152:515, 2017.)

TABLE 42-4 Physical Examination in Patients with Chronic Diarrhea

- Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, dermatitis herpetiformis, edema, or clubbing?
- Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
- Is there an abdominal mass or tenderness?
- Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?
- 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?

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 or colestevam).

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 or colestevam 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. 42-3). 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, and 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 *Giardia* and quantitative cultures) 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.

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 contrast studies.

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-HT₃ 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, pellet stools occur with slow transit, whereas loose, watery stools are associated with rapid transit. Both small pellet 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 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

TABLE 42-5 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

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

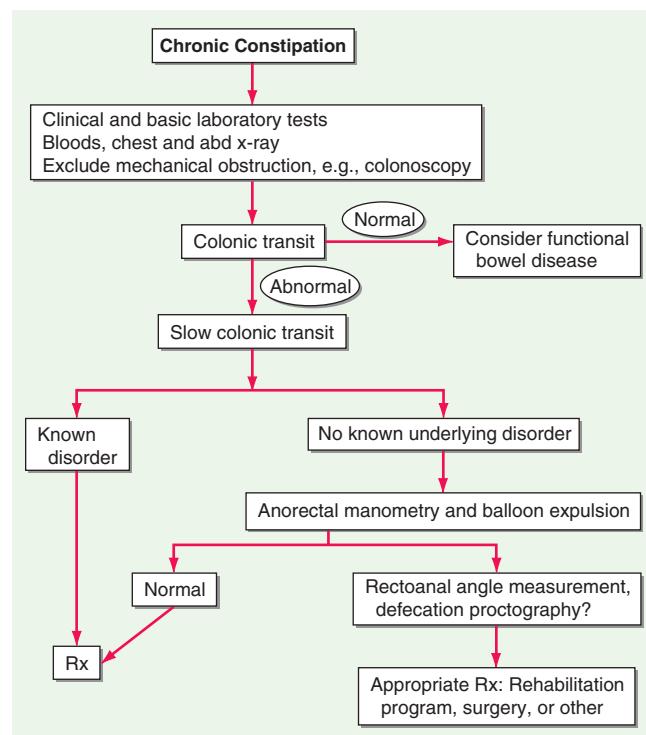
Pathophysiologically, chronic constipation generally results from inadequate fiber or fluid intake or from disordered colonic transit or anorectal function. These result from neurogastroenterologic disturbance, certain drugs, advancing age, or in association with a large number of systemic diseases that affect the GI tract (Table 42-5). 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.

APPROACH TO THE PATIENT

Constipation

A careful history should explore the patient's symptoms and confirm whether he or she 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. 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, a rectal examination 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 during straining to simulate stool evacuation).

The presence of weight loss, rectal bleeding, or anemia with constipation mandates either flexible sigmoidoscopy plus barium

**FIGURE 42-4 Algorithm for the management of constipation.** abd, abdominal.

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. 42-4). 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.

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.

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 radionuclabeled 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.

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 enterocoele.

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. 42-4). The remaining ~60% of patients with constipation has 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, osmotic, prokinetic, secretory, and stimulant laxatives including fiber, psyllium, milk of magnesia, lactulose, polyethylene glycol (colonic lavage solution), lubiprostone, linaclootide, and bisacodyl, or, in some countries, prucalopride, a 5-HT₄ agonist. If a 3- to 6-month trial of medical therapy fails, unassociated with obstructed defecation, the patients should be considered for laparoscopic colectomy with ileorectalostomy; however, this should not be undertaken 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 in 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 ileorectalostomy 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.
- BHARUCHA AE, RAO SS: An update on anorectal disorders for gastroenterologists. *Gastroenterology* 146:37, 2014.
- BHARUCHA AE, PEMBERTON JH, LOCKE GR 3rd: American Gastroenterological Association technical review on constipation. *Gastroenterology* 144:218, 2013.
- BOECKXSTAENS G et al: Fundamentals of neurogastroenterology: Physiology/motility—sensation. *Gastroenterology* pii: S0016-5085(16)00221-3, 2016. doi: 10.1053/j.gastro.2016.02.030. [Epub ahead of print]
- CAMILLERI M, SELLIN JH, BARRETT KE: Pathophysiology, evaluation, and management of chronic watery diarrhea. *Gastroenterology* 152:515, 2017.
- LEMBO A, CAMILLERI M: Chronic constipation. *N Engl J Med* 349:1360, 2003.
- RIDDLE MS, DUPONT HL, CONNOR BA: ACG Clinical Guideline: Diagnosis, treatment, and prevention of acute diarrheal infections in adults. *Am J Gastroenterol* 111:602, 2016.
- RUBIO-TAPIA A et al: American College of Gastroenterology. ACG clinical guidelines: Diagnosis and management of celiac disease. *Am J Gastroenterol* 108:656, 2013.
- SCHILLER LR, PARDI DS, SELLIN JH: Chronic diarrhea: Diagnosis and management. *Clin Gastroenterol Hepatol* 15:182, 2017.
- UZZAN M et al: Gastrointestinal disorders associated with common variable immune deficiency (CVID) and chronic granulomatous disease (CGD). *Curr Gastroenterol Rep* 18:17, 2016.

PHYSIOLOGY OF WEIGHT REGULATION WITH AGING

(See also Chaps. 463 and 394) 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 with aging 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 and/or poverty, 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 43-1). Not infrequently, more than one of these causes can be responsible for UWL. In most series, 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.

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 causes 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.

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

43

Unintentional Weight Loss

J. Larry Jameson

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.

TABLE 43-1 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
Malabsorption	Caries
Peptic ulcer	Dysgeusia
Inflammatory bowel disease	
Pancreatitis	
Obstruction/constipation	
Pernicious anemia	
Endocrine and metabolic	Age-related factors
Hyperthyroidism	Physiologic changes
Diabetes mellitus	Visual impairment
Pheochromocytoma	Decreased taste and smell
Adrenal insufficiency	Functional disabilities
Cardiac disorders	Neurologic
Chronic ischemia	Stroke
Chronic congestive heart failure	Parkinson's disease
Respiratory disorders	Neuromuscular disorders
Emphysema	Dementia
Chronic obstructive pulmonary disease	
Renal insufficiency	Social
Rheumatologic disease	Isolation
Infections	Economic hardship
HIV	
Tuberculosis	
Parasitic infection	
Subacute bacterial endocarditis	
Idiopathic	Psychiatric and behavioral
	Depression
	Anxiety
	Paranoia
	Bereavement
	Alcoholism
	Eating disorders
	Increased activity or exercise

features and may present as “apathetic hyperthyroidism” or T_3 toxicosis (**Chap. 375**).

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. 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 43-2**). Age, sex, and risk factor-specific cancer screening tests, such as mammography and colonoscopy, should be performed (**Chap. 66**). 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. 464**). The Mini Nutritional Assessment (www.mna-elderly.com) and the Nutrition Screening Initiative (<http://www.ncbi.nlm.nih.gov/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 since the prognosis of weight loss of undetermined cause is generally favorable.

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.

TABLE 43-2 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	

^aMay be more specific to assess weight loss in the elderly.

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.
- COMPSTON JE et al: Increase in fracture risk following unintentional weight loss in postmenopausal women: The global longitudinal study of osteoporosis in women. *J Bone Miner Res* 31:1466, 2016.
- GADDEY HL, HOLDER K: Unintentional weight loss in older adults. *Am Fam Physician* 89:718, 2014.
- McMINN J et al: Investigation and management of unintentional weight loss in older adults. *BMJ* 342:d1732, 2011.
- MILLER SL, WOLFE RR: The danger of weight loss in the elderly. *J Nutr Health and Aging* 12:487, 2008.
- VANDERSCHUEREN S et al: The diagnostic spectrum of unintentional weight loss. *Eur J Intern Med* 16:160, 2005.

one-third of such patients have further bleeding that requires urgent surgery if they are treated conservatively. These patients benefit from endoscopic therapy with 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.

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. Recent meta-analysis of randomized trials documents that high-dose intermittent PPIs are non-inferior to constant-infusion PPI therapy and thus may be substituted in this population. 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 will 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 failure to restart aspirin was associated with no significant difference in rebleeding (5% vs 10%) at 30 days but a significant increase in mortality (9% vs 1%) compared with immediate reinstitution of aspirin. 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. 317.**

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 gastoesophageal 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. 316.**

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. Urgent endoscopy within 12 h is recommended in cirrhotics with UGIB, and if esophageal varices are present, endoscopic ligation is performed and an IV vasoactive medication (octreotide, somatostatin, vaptoreotide, terlipressin) is given 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 should also be considered in the first 1–2 days of hospitalization for acute variceal bleeding in patients with advanced liver disease (e.g., Child-Pugh class C with Child-Pugh score 10–13), because randomized trials show significant decreases in rebleeding and mortality compared with standard endoscopic and medical therapy.

44

Gastrointestinal Bleeding

Loren Laine

Gastrointestinal bleeding (GIB) is the most common gastrointestinal condition leading to hospitalization in the United States, accounting for over 507,000 admissions and \$4.85 billion in direct costs annually. Upper GIB (UGIB) incidence has decreased in recent decades, primarily due to decreases in GIB from ulcers. The ratio of UGIB to lower GIB (LGIB) among GIB admissions from U.S. emergency rooms is ~1.3. The case fatality of patients hospitalized with GIB has also decreased and is <3% in the United States. Patients generally die from decompensation of other underlying illnesses rather than exsanguination.

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; or with iron-deficiency anemia or a positive fecal occult blood test on routine testing. GIB is also categorized by the site of bleeding as UGIB (esophagus, stomach, duodenum), LGIB (colonic), small intestinal, or obscure GIB (if the source is unclear).

SOURCES OF GASTROINTESTINAL BLEEDING

Upper Gastrointestinal Sources of Bleeding

PEPTIC ULCERS Peptic ulcers are the most common cause of UGIB, accounting for ~50% of UGIB hospitalizations. Features of an ulcer at endoscopy provide important prognostic information that guides subsequent management decisions as outlined in **Figs. 315-3 and 315-4**. Approximately 20% of patients with bleeding ulcers have the highest risk findings of active bleeding or a nonbleeding visible vessel:

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 due to cirrhosis 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 which are confined to the mucosa and do not cause major bleeding due to the absence of arteries and veins 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) ~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 who have experienced serious trauma, major surgery, burns covering more than one-third of the body surface area, major intracranial disease, or severe medical illness (i.e., 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. Pharmacologic prophylaxis for bleeding may be considered in the high-risk patients mentioned above. Meta-analyses of randomized trials indicate that PPIs are more effective than H₂-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 up to ~5–10% of GIB cases. The most common causes in adults >40 years are vascular ectasias, neoplasm (e.g., GI 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 causes in patients <40 years include Crohn's disease, polyposis syndromes, or neoplasm. Less common causes of small-intestinal GIB include 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 recent 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 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, followed by 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, trauma, varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, 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. Colonic diverticula stop bleeding spontaneously in ~80–90% of patients and, on long-term follow-up, rebleed in ~15–40% of patients. 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 is hemodynamically significant. Endoscopic hemostatic therapy may be used in the treatment of vascular ectasias, as well as discrete bleeding ulcers and post-polypectomy 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 the 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 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).

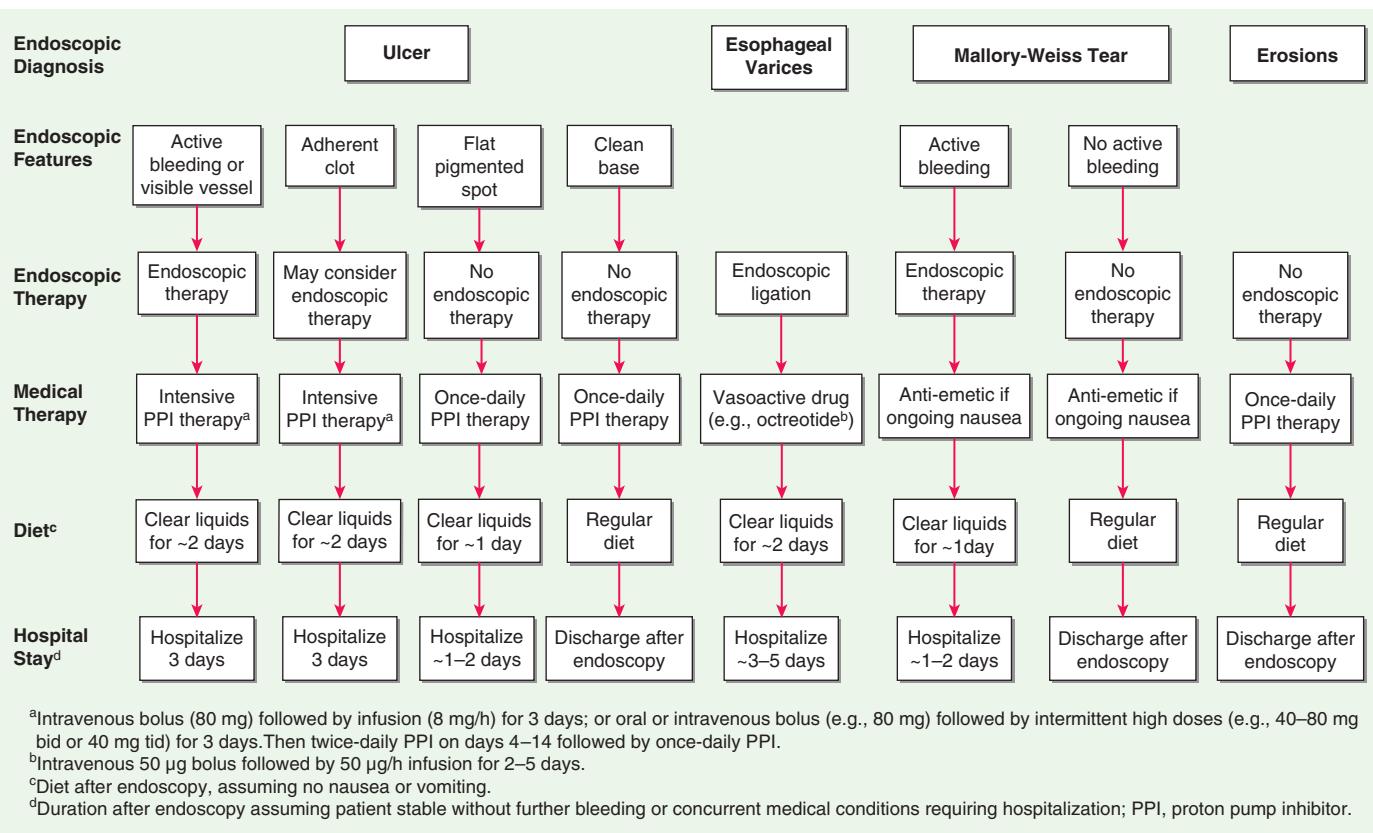


FIGURE 44-1 Suggested algorithm for patients with acute upper gastrointestinal bleeding based on endoscopic findings.

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. 44-1)

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 44-1) of 0–1 or 0–2 among patients <70 years because when hospitalized <1% of such patients require intervention and <0.5% die.

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 min before endoscopy, also may be considered to improve visualization at endoscopy: it provides a small but significant increase in diagnostic yield and decrease in red cell transfusions. Cirrhotic patients presenting with UGIB should be given an antibiotic (quinolone or ceftriaxone) and IV vasoactive medication upon presentation, even before endoscopy. Antibiotics decrease bacterial infections, rebleeding, and mortality, and vasoactive medications may improve control of bleeding in the first 12 h after presentation.

Upper endoscopy should be performed within 24 h in most patients with UGIB. Patients at higher risk (e.g., hemodynamic instability, cirrhosis) may benefit from more urgent endoscopy within 12 h. Early endoscopy is also beneficial in low-risk patients for management decisions (e.g., discharge). Patients with major bleeding and high-risk endoscopic findings (e.g., varices, ulcers with active bleeding or a visible vessel) benefit from endoscopic hemostatic therapy, whereas patients with low-risk lesions (e.g., clean-based

ulcers, erosions, nonbleeding Mallory-Weiss tears) who have stable vital signs and hemoglobin and no other medical problems may be discharged home.

EVALUATION AND MANAGEMENT OF LGIB (FIG. 44-2)

Patients with hematochezia and hemodynamic instability should have upper endoscopy to rule out an upper GI source before evaluation of the lower GI tract.

TABLE 44-1 Glasgow-Blatchford Score

ADMISSION MARKER	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
Other markers	
Melena	1
Syncope	2
Hepatic disease	2
Cardiac failure	2

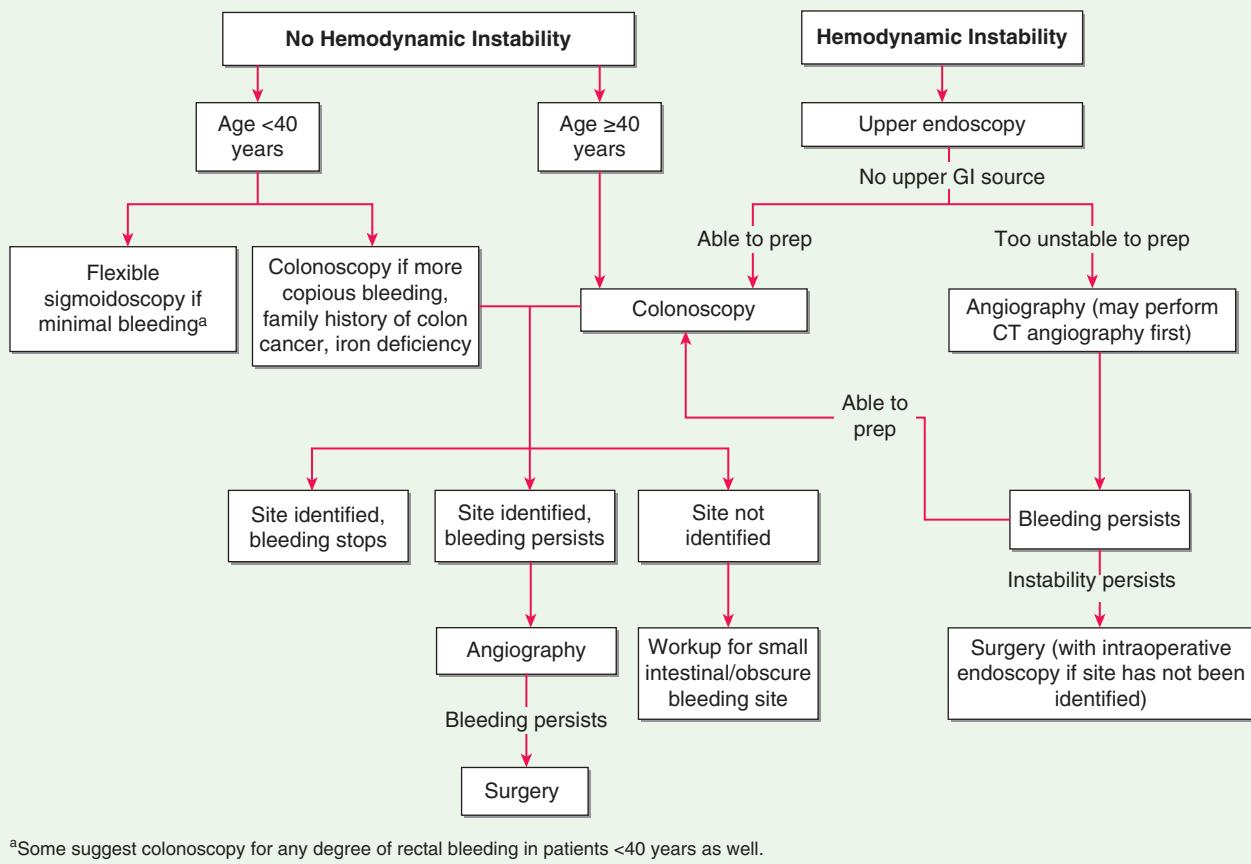


FIGURE 44-2 Suggested algorithm for patients with acute lower gastrointestinal bleeding.

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, radionuclide scans should be interpreted with caution because results, especially from later images, are highly variable. Multidetector CT angiography is likely superior to nuclear scintigraphy and increasingly used in its place. In active LGIB, angiography can detect the site of bleeding (extravasation of contrast into the gut) and permits treatment with 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" 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, current recommendations suggest patients may either be observed or, if their clinical course mandates (e.g., need for transfusions), undergo further testing. "Deep" enteroscopy (double-balloon, single-balloon, or spiral enteroscopy) is commonly the next test undertaken 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 50 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.

■ FURTHER READING

- DE FRANCHIS R: Expanding consensus in portal hypertension. *J Hepatol* 63:743, 2015.
- 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.

- GERSON LB et al: ACG clinical guideline: Diagnosis and management of small bowel bleeding. *Am J Gastroenterol* 110:1265, 2015.
- GRALNEK IM et al: Diagnosis and management of upper gastrointestinal hemorrhage: European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy* 47:1, 2015.
- LAINÉ L: Upper gastrointestinal bleeding due to a peptic ulcer. *N Engl J Med* 374:2367, 2016.
- LAINÉ L, JENSEN DM: ACG Practice Guidelines: Management of patients with ulcer bleeding. *Am J Gastroenterol* 107:345, 2012.
- STRATE LL, GRALNEK KM: ACG clinical guideline: Management of patients with acute lower gastrointestinal bleeding. *Am J Gastroenterol* 111:459, 2016.
- SUNG JJ et al: Continuation of low-dose aspirin therapy in peptic ulcer bleeding: A randomized trial. *Ann Intern Med* 152:1, 2010.
- VILLANEUVA C et al: Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med* 368:11, 2013.

45

Jaundice

Savio John, Daniel S. Pratt



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 the drug quinacrine, and excessive exposure to phenols. Carotenoderma is the yellow color imparted to the skin of healthy individuals who ingest excessive amounts of vegetables and fruits that contain carotene, such as carrots, leafy vegetables, squash, peaches, and oranges. 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.

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. 331) 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. 331, Fig. 331-1).

After entering the hepatocyte, unconjugated bilirubin is bound in the cytosol to a number of 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 dipyrromethene 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 $\mu\text{mol/L}$ (1 and 1.5 mg/dL). Total serum bilirubin concentrations are between 3.4 and 15.4 $\mu\text{mol/L}$ (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 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 all of the conditions that can cause 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. 45-1).

Simply stated, 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? All of these questions can 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. 45-1). 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 45-1). 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 $\mu\text{mol/L}$ (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 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 $\mu\text{mol/L}$ [>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 $\mu\text{mol/L}$ (6–25 mg/dL). In these patients, mutations in the bilirubin

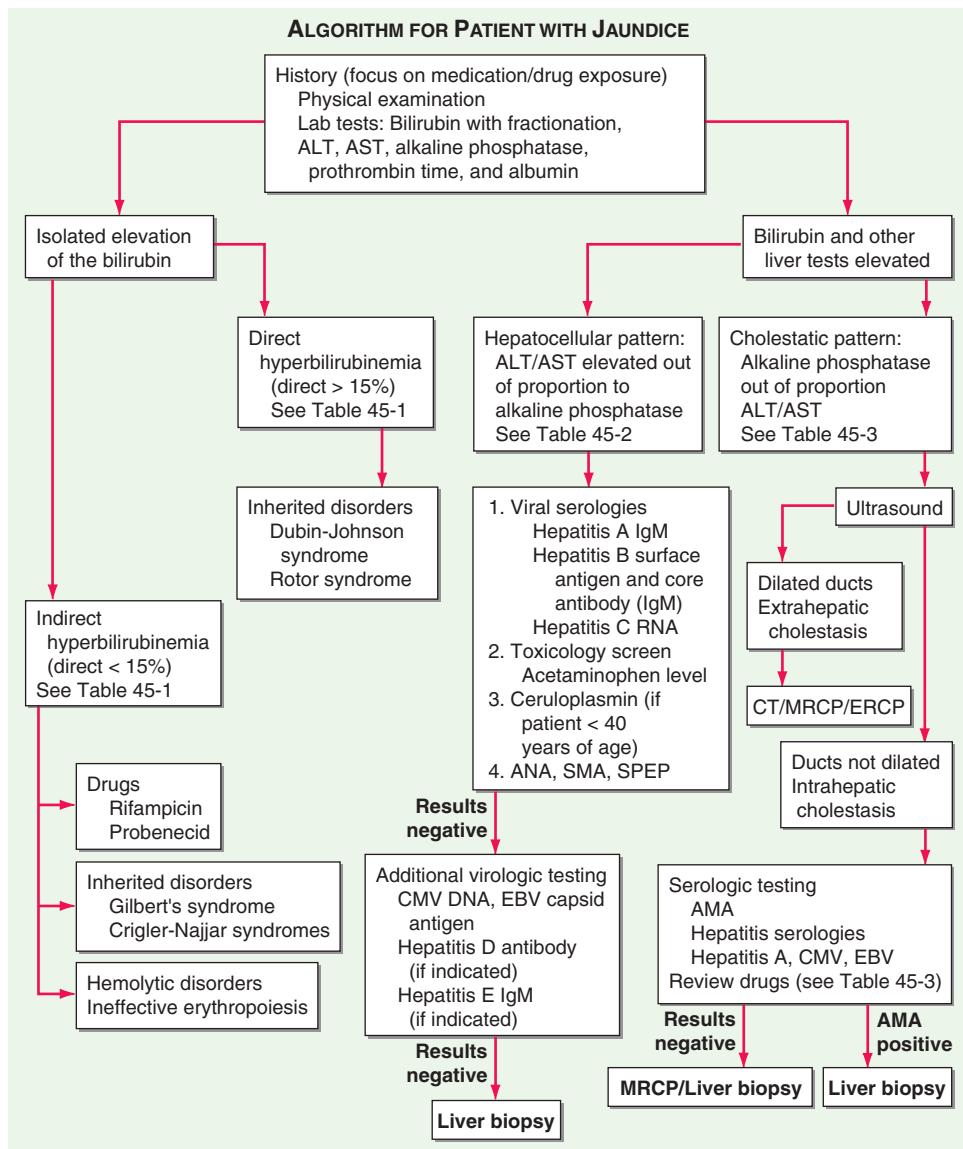


FIGURE 45-1 Evaluation of the patient with jaundice. ALT, alanine aminotransferase; AMA, antimitochondrial antibody; ANA, antinuclear antibody; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; LKM, liver-kidney microsomal antibody; MRCP, magnetic resonance cholangiopancreatography; SMA, smooth-muscle antibody; SPEP, serum protein electrophoresis.

TABLE 45-1 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. Ribavirin
 - 4. Protease inhibitors (Atazanavir, Indinavir)
 - 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

UDPGT gene cause the reduction—typically $\leq 10\%$ —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 $<103 \mu\text{mol/L}$ (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 45-1). 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. 45-1).

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 particular 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 alcoholic (Laennec's) 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. 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 330-1; Fig. 45-1)—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 45-2). 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. Because it can take

TABLE 45-2 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— <i>Amanita phalloides</i> , <i>A. verna</i>
Wilson's disease
Autoimmune hepatitis

many weeks for hepatitis C antibody to become detectable, its assay is an unreliable test if acute hepatitis C is suspected. Studies for hepatitis D and E viruses, 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 an antinuclear antibody assay 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 next step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 45-1). 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 CT, magnetic resonance cholangiopancreatography (MRCP), endoscopic retrograde cholangiopancreatography (ERCP), percutaneous transhepatic cholangiography (PTC), and endoscopic ultrasound (EUS). CT scanning 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 cases where the need for intervention is thought to be small. EUS displays sensitivity and specificity comparable to that of MRCP in the detection of bile duct obstruction. EUS also allows biopsy of suspected malignant lesions, but is invasive and requires sedation.

In patients with apparent *intrahepatic cholestasis*, the diagnosis is often made by serologic testing in combination with percutaneous liver biopsy. The list of possible causes of intrahepatic cholestasis is long and varied (Table 45-3). 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

TABLE 45-3 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
 - 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. Extrahepatic
 - 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)

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 middle-aged 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 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. The histologic picture is similar to that in primary biliary cholangitis. This picture is 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, 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 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 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 45-3). 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.

■ 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, Ebola virus, late phases of yellow fever, dengue hemorrhagic fever, schistosomiasis, fascioliasis, clonorchiasis, opisthorchiasis, ascariasis, echinococcosis, hepatosplenic candidiasis, disseminated histoplasmosis, cryptococcosis, coccidioidomycosis, 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 together with Daniel Pratt.

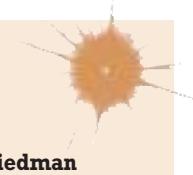
■ FURTHER READING

- ERLINGER S, ARIAS IM, DHUMEAX D: Inherited disorders of bilirubin transport and conjugation: New insights into molecular mechanisms and consequences. *Gastroenterology* 146:1625, 2014.
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.

46

Abdominal Swelling and Ascites

Kathleen E. Corey, 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 F*s: 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 an increase in intraabdominal volume loads. 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.

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.

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 chronic 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 alcoholic 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. 46-1).

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.

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. 337). In some cases, a liver biopsy may be necessary to confirm cirrhosis.



FIGURE 46-1 CT of a patient with a cirrhotic, nodular liver (white arrow), splenomegaly (yellow arrow), and ascites (arrowheads).

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 (contrary 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. 46-2). 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

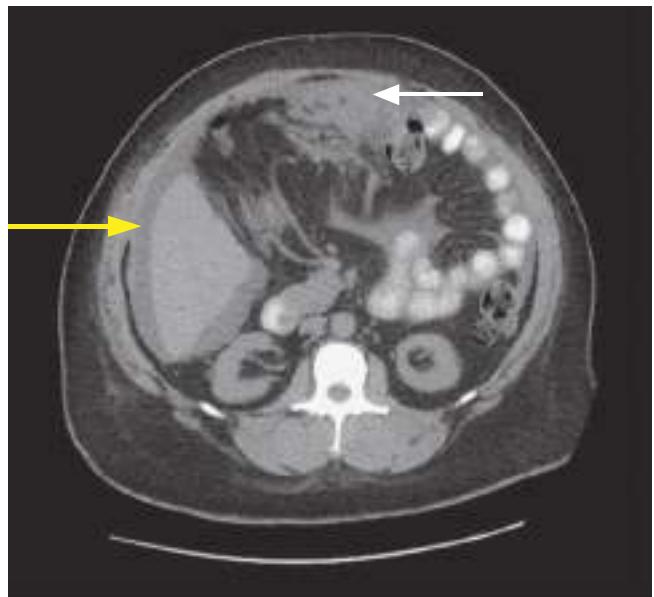


FIGURE 46-2 CT of a patient with peritoneal carcinomatosis (white arrow) and ascites (yellow arrow).

on the peritoneum exude a proteinaceous fluid. Pancreatic ascites results from leakage of pancreatic enzymes into the peritoneum.

■ 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 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. 46-3). 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

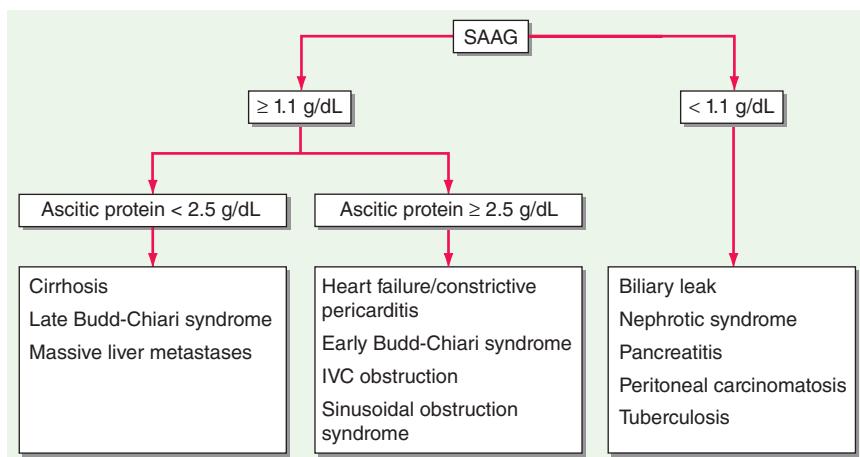


FIGURE 46-3 Algorithm for the diagnosis of ascites according to the serum-ascites albumin gradient (SAAG). IVC, inferior vena cava.

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.

For high-SAAG (≥ 1.1) ascites, the ascitic protein level can provide further clues to the etiology (Fig. 46-3). 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 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 to 3%; a culture increases the sensitivity to 35–50%. In patients without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of $> 90\%$ 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. 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 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 infusion of albumin accompanying LVP decreases the risk of "post-paracentesis 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.

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).

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. 127) 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 of or 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 $\geq 250/\mu\text{L}$ in the ascitic fluid. Cultures of ascitic fluid 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. 127). 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 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 $<1 \text{ g/dL}$, or active gastrointestinal bleeding should receive prophylactic antibiotics to prevent SBP; oral daily norfloxacin is commonly used. 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.

FURTHER READING

- BECKER G et al: Malignant ascites: Systematic review and guideline for treatment. *Eur J Cancer* 42:589, 2006.
- BERNARDI M et al: Albumin infusion in patients undergoing large-volume paracentesis: A meta-analysis of randomized trials. *Hepatology* 55:1172, 2012.
- 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.
- FERNANDEZ J et al: Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: A prospective study. *Hepatology* 55:1551, 2012.
- GE PS, RUNYON BA: Role of plasma BNP in patients with ascites: Advantages and pitfalls. *Hepatology* 59: 751, 2014.
- ORMAN ES et al: Paracentesis is associated with reduced mortality in patients hospitalized with cirrhosis and ascites. *Clin Gastroenterol Hepatol* 12:496, 2014.
- 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.
- 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.
- 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.

Section 7 Alterations in Renal and Urinary Tract

Function

47

Dysuria, Bladder Pain, and the Interstitial Cystitis/Bladder Pain Syndrome

John W. Warren

Dysuria and bladder pain are two symptoms that commonly call attention to the lower urinary tract.

DYSURIA

Dysuria, or pain that occurs during urination, is commonly perceived as burning or stinging in the urethra and is a symptom of several syndromes. The presence or absence of *other* symptoms is often helpful in distinguishing among these conditions. Some of these syndromes differ between men and women.

Women Approximately 50% of women experience dysuria at some time in their lives; ~20% report having had dysuria within the past year. Most dysuria syndromes in women can be categorized into two broad groups: bacterial cystitis and lower genital tract infections.

Bacterial cystitis is usually caused by *Escherichia coli*; a few other gram-negative rods and *Staphylococcus saprophyticus* also can be responsible. Bacterial cystitis is acute in onset and manifests not only as dysuria but also as urinary frequency, urinary urgency, suprapubic pain, and/or hematuria.

The lower genital tract infections include vaginitis, urethritis, and ulcerative lesions; many of these infections are caused by sexually transmitted organisms and should be considered particularly in young women who have new or multiple sexual partners or whose partners do not use condoms. The onset of dysuria associated with these syndromes is more gradual than in bacterial cystitis and is thought (but not proven) to result from the flow of urine over damaged epithelium. Frequency, urgency, suprapubic pain, and hematuria are reported less frequently than in bacterial cystitis. Vaginitis, caused by *Candida albicans* or *Trichomonas vaginalis*, presents as vaginal discharge or irritation. Urethritis is a consequence of infection by *Chlamydia trachomatis* or *Neisseria gonorrhoeae*. Ulcerative genital lesions may be caused by herpes simplex virus and several other specific organisms.

Among women presenting with dysuria, the probability of bacterial cystitis is ~50%. This figure rises to >90% if four criteria are met: dysuria and frequency without vaginal discharge or irritation. Present standards suggest that women meeting these four criteria, if they are otherwise healthy, are not pregnant, and have an apparently normal urinary tract, can be diagnosed with uncomplicated bacterial cystitis and treated empirically with appropriate antibiotics. Other women with dysuria should be further evaluated by urine dipstick, urine culture, and a pelvic examination.

Men Dysuria is less common among men. The syndromes presenting as dysuria are similar to those in women but with some important distinctions.

In the majority of men with dysuria, frequency, urgency, and/or suprapubic, penile, and/or perineal pain, the prostate is involved, either as the source of infection or as an obstruction to urine flow. Bacterial prostatitis is usually caused by *E. coli* or another gram-negative rod, with one of two presentations. *Acute bacterial prostatitis* presents with fever and chills; prostate examination should be gentle or not performed at all, as massage may result in a wave of bacteremia. *Chronic bacterial prostatitis* presents as recurrent episodes of bacterial cystitis; prostate examination with massage demonstrates prostatic bacteria

and leukocytes. *Benign prostatic hyperplasia* (BPH) can obstruct urine flow, with consequent symptoms of weak stream, hesitancy, and dribbling. If a bacterial infection develops behind the obstructing prostate, dysuria and other symptoms of cystitis will occur. Men whose symptoms are consistent with bacterial cystitis should be evaluated with urinalysis and urine culture.

Several sexually transmitted infections can manifest as dysuria. Urethritis (usually without urinary frequency) presents as a urethral discharge and can be caused by *C. trachomatis*, *N. gonorrhoeae*, *Mycoplasma genitalium*, *Ureaplasma urealyticum*, or *T. vaginalis*. Herpes simplex, chancroid, and other ulcerous lesions may present as dysuria, again without urinary frequency.

For further discussion, see Chaps. 130 and 131.

Either Women or Men Other causes of dysuria may be found in patients of either sex. Some cases are acute and include lower urinary tract stones, trauma, and urethral exposure to topical chemicals. Others may be relatively chronic and attributable to lower urinary tract cancers, certain medications, Behcet's syndrome, reactive arthritis, a poorly understood entity known as *chronic urethral syndrome*, or interstitial cystitis/bladder pain syndrome (see below).

■ BLADDER PAIN

Studies indicate that patients perceive pain as coming from the urinary bladder if it is suprapubic in location, alters with bladder filling or emptying, and/or is associated with urinary symptoms such as urgency and frequency. Bladder pain occurring acutely (i.e., over hours or a day or two) is helpful in distinguishing bacterial cystitis from urethritis, vaginitis, and other genital infections. Chronic or recurrent bladder pain may accompany lower urinary tract stones; bladder, uterine, cervical, vaginal, urethral, or prostate cancer; urethral diverticulum; cystitis induced by radiation or certain medications; tuberculous cystitis; bladder neck obstruction; neurogenic bladder; urogenital prolapse; or BPH. In the absence of these conditions, the diagnosis of interstitial cystitis/bladder pain syndrome (IC/BPS) should be considered.

■ INTERSTITIAL CYSTITIS/BLADDER PAIN SYNDROME

Most clinicians with outpatient practices see undiagnosed cases of IC/BPS. This chronic condition is characterized by pain perceived to be from the urinary bladder, urinary urgency and frequency, and nocturia. As currently diagnosed, the majority of cases occur in women. Symptoms wax and wane for months or years or possibly even for the rest of the patient's life. The spectrum of symptom intensity is broad. The pain can be excruciating, urgency can be distressing, frequency can be up to 60 times per 24 h, and nocturia can cause sleep deprivation. These symptoms can disrupt daily activities, work schedules, and personal relationships; patients with IC/BPS report less life satisfaction than do those with end-stage renal disease.

IC/BPS is not a new disease, having first been described in the late nineteenth century in a patient with the symptoms described above and a single ulcer visible on cystoscopy (now called a *Hunner lesion* after the urologist who first reported it). Over the ensuing decades, it became clear that many patients with similar symptoms had no ulcer. It is now appreciated that $\leq 10\%$ of patients with IC/BPS have a Hunner lesion. The definition of IC/BPS, its diagnostic features, and even its name continue to evolve. The American Urological Association has defined IC/BPS as "an unpleasant sensation (pain, pressure, discomfort) perceived to be related to the urinary bladder, associated with lower urinary tract symptoms of more than 6 weeks' duration, in the absence of infection or other identifiable causes."

Many patients with IC/BPS also have other syndromes, such as fibromyalgia, chronic fatigue syndrome, and irritable bowel syndrome. These syndromes collectively are known as *functional somatic syndromes* (FSSs): chronic conditions in which pain and fatigue are prominent features but laboratory tests and histologic findings are normal. Like IC/BPS, the FSSs often are associated with depression and anxiety. The majority of FSSs affect more women than men, and more than one FSS can affect a single patient. Because of its similar features and comorbidity, IC/BPS sometimes is considered an FSS.

Epidemiology Contemporary population studies of IC/BPS in the United States indicate a prevalence of 3–6% among women and 2–4% among men. For decades, it was thought that IC/BPS occurred mostly in women. These prevalence findings, however, have generated research aimed at determining the proportion of men who have symptoms usually diagnosed as chronic prostatitis (now known as *chronic prostatitis/chronic pelvic pain syndrome*) but who actually have IC/BPS.

Among women, the average age at onset of IC/BPS symptoms is the early forties, but the range is from childhood through the early sixties. Risk factors (antecedent features that distinguish cases from controls) primarily have been FSSs. Indeed, the odds of IC/BPS increase with the number of such syndromes present. Surgery was long thought to be a risk factor for IC/BPS, but analyses adjusting for FSSs refuted that association. About one-third of patients appear to have bacterial cystitis at the onset of IC/BPS.

The natural history of IC/BPS is not known. Although studies from urology and urogynecology practices have been interpreted as showing that IC/BPS lasts for the lifetime of the patient, population studies suggest that some individuals with IC/BPS do not consult specialists and may not seek medical care at all, and most prevalence studies do not show an upward trend with age—a pattern that would be expected with incident cases throughout adulthood followed by lifetime persistence of a nonfatal disease. It may be reasonable to conclude that patients in a urology practice represent those with the most severe and recalcitrant IC/BPS.

Pathology For the $\leq 10\%$ of IC/BPS patients who have a Hunner lesion, the term *interstitial cystitis* may indeed describe the histopathologic picture. Most of these patients have substantive inflammation, mast cells, and granulation tissue. However, in the 90% of patients without such lesions, the bladder mucosa and interstitium are relatively normal, with scant inflammation.

Etiology Numerous hypotheses about the pathogenesis of IC/BPS have been put forward. It is not surprising that most early theories focused on the bladder. For instance, IC/BPS has been investigated as a chronic bladder infection. Sophisticated technologies have not identified a causative organism in urine or in bladder tissue; however, the patients studied by these methods had IC/BPS of long duration, and the results do not preclude the possibility that infection may trigger the syndrome or may be a feature of early IC/BPS. Other inflammatory factors, including a role for mast cells, have been postulated, but (as noted above) the 90% of patients who do not have a Hunner ulcer have little bladder inflammation and do not have a prominence of mast cells in urine or in bladder tissue. Autoimmunity has been considered, but autoantibodies are low in titer, nonspecific, and thought to be a result rather than a cause of IC/BPS. Increased permeability of the bladder mucosa due to defective epithelium or glycosaminoglycan (the bladder's mucous coating) has been studied frequently, but the findings have been inconclusive.

Investigations of causes outside the bladder have been prompted by the presence of comorbid FSSs. Many patients with FSSs have abnormal pain sensitivity as evidenced by (1) low pain thresholds in body areas unrelated to the diagnosed syndrome, (2) dysfunctional descending neurologic control of tactile signals, and (3) enhanced brain responses to touch in functional neuroimaging studies. Moreover, in patients with IC/BPS, body surfaces remote from the bladder are more sensitive to pain than is the case in individuals without IC/BPS. All these findings are consistent with upregulation of sensory processing in the brain. Indeed, a prevailing theory is that these concomitantly occurring syndromes have in common an abnormality of brain processing of sensory input. However, antecedence is a critical criterion for causality, and no study has demonstrated that abnormal pain sensitivity precedes either IC/BPS or the FSSs.

Clinical Presentation In some patients, IC/BPS has a gradual onset and/or the cardinal symptoms of pain, urgency, frequency, and nocturia appear sequentially in no consistent order. Other patients can identify the exact date of onset of IC/BPS symptoms. More than half of the latter patients describe dysuria beginning on that date.

Only a minority of IC/BPS patients who obtain medical care soon after symptom onset have uropathogenic bacteria or leukocytes in the urine. These patients—and many others with new-onset IC/BPS—are treated with antibiotics for presumptive bacterial cystitis or, if male, chronic bacterial prostatitis. Persistent or recurring symptoms without bacteriuria eventually prompt a differential diagnosis, and IC/BPS is considered. Traditionally, the diagnosis of IC/BPS has been delayed for years, but recent interest in the disease appears to have shortened this interval.

Two-thirds of women with IC/BPS report two or more sites of pain. The most common site (involved in 80% of women) and generally the one with the most severe pain is the suprapubic area. About 35% of female patients have pain in the urethra, 25% in other parts of the vulva, and 30% in non-urogenital areas, mostly the low back and also the anterior or posterior thighs or the buttocks. The pain of IC/BPS is most commonly described as aching, pressing, throbbing, tender, and/or piercing. What may distinguish IC/BPS from other pelvic pain is that, in 95% of patients, bladder filling exacerbates the pain and/or bladder emptying relieves it. Almost as many patients report a puzzling pattern in which certain dietary substances worsen the pain of IC/BPS. Smaller majorities report that their IC/BPS pain is worsened by menstruation, stress, tight clothing, exercise, and riding in a car as well as during or after vaginal intercourse.

The urethral and vulvar pains of IC/BPS merit special mention. In addition to the descriptive adjectives for IC/BPS mentioned above, these pains commonly are described as burning, stinging, and sharp and as being worsened by touch, tampons, and vaginal intercourse. Patients report that urethral pain increases during urination and generally lessens afterward. These characteristics have commonly resulted in the diagnosis of the urethral pain of IC/BPS as chronic urethral syndrome and of the vulvar pain as vulvodynia.

In many patients with IC/BPS, there is a link between pain and urinary urgency; that is, two-thirds of patients describe the urge to urinate as a desire to relieve their bladder pain. Only 20% report that the urge stems from a desire to prevent incontinence; indeed, very few patients with IC/BPS are incontinent. As mentioned above, urinary frequency can be severe, with ~85% of patients voiding more than 10 times per 24 h and some as often as 60 times. Voiding continues through the night, and nocturia is common, frequent, and often associated with sleep deprivation.

Beyond these common symptoms of IC/BPS, additional urinary and other symptoms may be present. Among the urinary symptoms are difficulty in starting urine flow, perceptions of difficulty in emptying the bladder, and bladder spasms. Among the non-urinary symptoms are the manifestations of comorbid FSSs as well as symptoms that do not constitute recognized syndromes, such as numbness, muscle spasms, dizziness, ringing in the ears, and blurred vision.

The pain, urgency, and frequency of IC/BPS can be debilitating. Proximity to a bathroom is a continual focus, and patients report difficulties in the workplace, leisure activities, travel, and simply leaving home. Familial and sexual relationships can be strained.

Diagnosis Traditionally, IC/BPS has been considered a rare condition that is diagnosed by urologists at cystoscopy. However, this disorder is much more common than once was thought; it is now being considered earlier in its course and is being diagnosed and managed more often by primary care clinicians. Results of physical examination, urinalysis, and urologic procedures are insensitive and/or nonspecific. Thus, diagnosis is based on the presence of appropriate symptoms and the exclusion of diseases with a similar presentation.

Three categories of disorders can be considered in the differential diagnosis of IC/BPS. The first comprises diseases that manifest as bladder pain or urinary symptoms. Among the latter diseases is *overactive bladder*, a chronic condition of women and men that presents as urgency and frequency and can be distinguished from IC/BPS by the patient's history: pain is not a feature of overactive bladder, and its urgency arises from the need to avoid incontinence. Endometriosis is a special case: it can be asymptomatic or can cause pelvic pain, dysmenorrhea, and dyspareunia—i.e., types of pain that mimic IC/BPS. Endometrial

implants on the bladder (although uncommon) can cause urinary symptoms, and the resulting syndrome can mimic IC/BPS. Even if endometriosis is identified, it is difficult in the absence of bladder implants to determine whether it is causative of or incidental to the symptoms of IC/BPS in a specific woman.

The second category of disorders encompasses the FSSs that can accompany IC/BPS. IC/BPS can be misdiagnosed as gynecologic chronic pelvic pain, irritable bowel syndrome, or fibromyalgia. The correct diagnosis may be entertained only when either changes of pain with altered bladder volume or urinary symptoms become more prominent.

The third category involves syndromes that IC/BPS mimics by way of its referred pain, such as vulvodynia and chronic urethral syndrome. Therefore, IC/BPS should be considered in the differential diagnosis of persistent or recurrent "urinary tract infection" (UTI) with sterile urine cultures; "overactive bladder" with pain; chronic pelvic pain, endometriosis, vulvodynia, or FSSs with urinary symptoms; and "chronic prostatitis." Important clues to the diagnosis of IC/BPS are changes of pain with bladder volume or with certain foods or drinks.

Cystoscopy under anesthesia formerly was thought to be necessary for the diagnosis of IC/BPS because of its capacity to reveal a Hunner lesion or—in the 90% of patients without an ulcer—petechial hemorrhages after bladder distention. However, because Hunner lesions are uncommon in IC/BPS and petechiae are nonspecific, cystoscopy is no longer necessary for diagnosis. Accordingly, the indications for urologic referral have evolved toward the need to rule out other diseases or to administer more advanced treatment.

A typical patient presents to the primary clinician after days, weeks, or months of pain, urgency, frequency, and/or nocturia. The presence of urinary nitrites, leukocytes, or uropathogenic bacteria should prompt treatment for UTI in women and for chronic bacterial prostatitis in men. Persistence or recurrence of symptoms in the absence of bacteriuria should prompt a pelvic examination for women, an assay for serum prostate-specific antigen for men, and urine cytology and inclusion of IC/BPS in the differential diagnosis for both sexes.

In the diagnosis of IC/BPS, inquiries about pain, pressure, and discomfort are useful; IC/BPS should be considered if any of these sensations are noted in one or more anterior or posterior sites between the umbilicus and the upper thighs. Nondirective questions about the effect of bladder volume changes include "As your next urination approaches, does this pain get better, get worse, or stay the same?" and "After you urinate, does this pain get better, get worse, or stay the same?" Establishing that the pain is exacerbated by the consumption of certain foods and drinks not only supports the diagnosis of IC/BPS but also serves as the basis for one of the first steps in managing this syndrome. A nondirective way to ask about urgency is to describe it to the patient as a compelling urge to urinate that is difficult to postpone; follow-up questions can determine whether this urge is intended to relieve pain or prevent incontinence. To assess severity and provide quantitative baseline measures, pain and urgency should be estimated by the patient on a scale of 0–10, with 0 being none and 10 the worst imaginable. Frequency per 24-h period should be determined and nocturia assessed as the number of times per night the patient is awakened by the need to urinate.

About half of patients with IC/BPS have intermittent or persistent microscopic hematuria; this manifestation and the need to exclude bladder stones or cancer require urologic or urogynecologic referral. Initiation of therapy for IC/BPS does not hamper subsequent urologic evaluation.

TREATMENT

Interstitial Cystitis/Bladder Pain Syndrome

The goal of therapy is to relieve the symptoms of IC/BPS; the challenge lies in the fact that no treatment is uniformly successful. However, most patients eventually obtain relief, generally with a multifaceted approach. The American Urological Association's guidelines for management of IC/BPS are an excellent resource.

The correct strategy is to begin with conservative therapies and proceed to riskier measures only if necessary and under the supervision of a urologist or urogynecologist. Conservative tactics include education, stress reduction, dietary changes, medications, pelvic-floor physical therapy, and treatment of associated FSSs.

Months or even years may have passed since the onset of symptoms, and the patient's life may have been disrupted continually, with repeated medical visits provoking frustration and dismay in both the patient and the physician. In this circumstance, simply giving a name to the syndrome is beneficial. The physician should discuss the disease, the diagnostic and therapeutic strategies, and the prognosis with the patient and with the spouse and/or other pertinent family members, who may need to be made aware that although IC/BPS has no visible manifestations, the patient is undergoing substantial pain and suffering. This information is particularly important for sexual partners, as exacerbation of pain during and after intercourse is a common feature of IC/BPS. Because stress can worsen IC/BPS symptoms, stress reduction and active measures such as yoga or meditation exercises may be suggested. The Interstitial Cystitis Association (www.ic-help.com) and the Interstitial Cystitis Network (www.ic-network.com) can be useful in this educational process.

Over time, many patients identify particular foods and drinks that exacerbate their symptoms. Common among these are chilies, chocolate, citrus fruits, tomatoes, alcohol, caffeinated drinks, and carbonated beverages; full lists of common trigger foods are available at the websites cited above. In constructing a benign diet, some patients find it useful to exclude all possible offenders and add items back into the diet one at a time to identify those that worsen their symptoms. Patients also should experiment with fluid volumes; some find relief with less fluid, others with more.

The pelvic floor is often tender in IC/BPS patients. Two randomized controlled trials showed that weekly physical therapy directed at relaxation of the pelvic muscles yielded significantly more relief than a similar schedule of general body massage. This intervention can be initiated under the direction of a knowledgeable physical therapist who recognizes that the objective is to relax the pelvic floor, not to strengthen it.

Among oral medications, nonsteroidal anti-inflammatory drugs are commonly used but are controversial and often unsuccessful. Two randomized controlled trials showed that amitriptyline can diminish IC/BPS symptoms if an adequate dose (≥ 50 mg per night) can be given. This drug is used not for its antidepressant activity but because of its proven effects on neuropathic pain; however, it is not approved by the U.S. Food and Drug Administration for treatment of IC/BPS. An initial dose of 10 mg at bedtime is increased weekly up to 75 mg (or less if a lower dose adequately relieves symptoms). Side effects can be expected and include dry mouth, weight gain, sedation, and constipation. If this regimen does not control symptoms adequately, pentosan polysulfate, a semisynthetic polysaccharide, can be added at a dose of 100 mg three times a day. Its theoretical effect is to replenish a possibly defective glycosaminoglycan layer over the bladder mucosa; randomized controlled trials suggest only a modest benefit over placebo. Adverse reactions are uncommon and include gastrointestinal symptoms, headache, and alopecia. Pentosan polysulfate has weak anticoagulant effects and probably should be avoided by patients with coagulation abnormalities.

Anecdotal reports suggest that successful therapy for one FSS is accompanied by diminished symptoms of other FSSs. As has been noted here, IC/BPS often is associated with one or several FSSs. Thus, it seems reasonable to hope that, to the extent that accompanying FSSs are treated successfully, the symptoms of IC/BPS will be relieved as well.

If several months of these therapies in combination do not relieve symptoms adequately, the patient should be referred to a urologist or urogynecologist who has access to additional modalities. Cystoscopy under anesthesia allows distention of the bladder with water, a procedure that provides ~40% of patients with several months of relief and can be repeated. For those few patients with

a Hunner lesion, fulguration may offer relief. Solutions containing lidocaine, hyaluronic acid, or dimethyl sulfoxide can be instilled into the bladder, or botulinum toxin can be injected into the bladder wall. Physicians experienced in the care of IC/BPS patients have used anticonvulsants, narcotics, and cyclosporine as components of therapy. Pain specialists can be of assistance. Sacral neuromodulation can be tested with a temporary percutaneous electrode and, if effective, can be administered with an implanted device. In a very small number of patients with recalcitrant symptoms, surgeries, including cystoplasty, partial or total cystectomy, and urinary diversion, may provide relief.

FURTHER READING

- FITZGERALD MP et al: Randomized multicenter clinical trial of myofascial physical therapy in women with interstitial cystitis/painful bladder syndrome and pelvic floor tenderness. *J Urol* 187:2113, 2012.
 HANNO PM et al: AUA guideline for the diagnosis and treatment of interstitial cystitis/bladder pain syndrome. *J Urol* 185:2162, 2011.
 HANNO PM et al: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome: AUA guideline amendment. *J Urol* 193:1545, 2015.
 SHORTER B et al: Effect of comestibles on symptoms of interstitial cystitis. *J Urol* 178:145, 2007.

48

Azotemia and Urinary Abnormalities

David B. Mount



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 48-1) 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 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 48-1) 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.

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

TABLE 48-1 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	304, 308, 310, 313
	Oliguria	Proteinuria, pyuria	
	Documented recent decline in GFR	Casts, edema	
Acute nephritis	Hematuria, RBC casts	Proteinuria	308
	Azotemia, reduced GFR, oliguria	Pyuria	
	Edema, hypertension	Circulatory congestion	
Chronic renal failure	Azotemia for >3 months	Proteinuria, casts	305
	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	
	Broad casts in urinary sediment	Hyperkalemia, metabolic acidosis	
Nephrotic syndrome	Proteinuria, with >3.5 g/24 h per 1.73 m ²	Casts	308
	Hypoalbuminemia	Lipiduria	
	Edema	Hypercoagulable state	
	Hyperlipidemia		
Asymptomatic urinary abnormalities	Hematuria		308
	Proteinuria (below nephrotic range)		
	Sterile pyuria, casts		
Urinary tract infection/pyelonephritis	Bacteriuria, with >10 ⁵ cfu/mL	Hematuria	130
	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	309, 310
	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	271, 311
		Casts	
		Azotemia	
Nephrolithiasis	Previous history of stone passage or removal	Hematuria	312
	Previous history of stone seen by x-ray	Pyuria	
	Renal colic	Frequency, urgency	
Urinary tract obstruction	Azotemia, oliguria, anuria	Hematuria	313
	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.

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. 48-1). 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

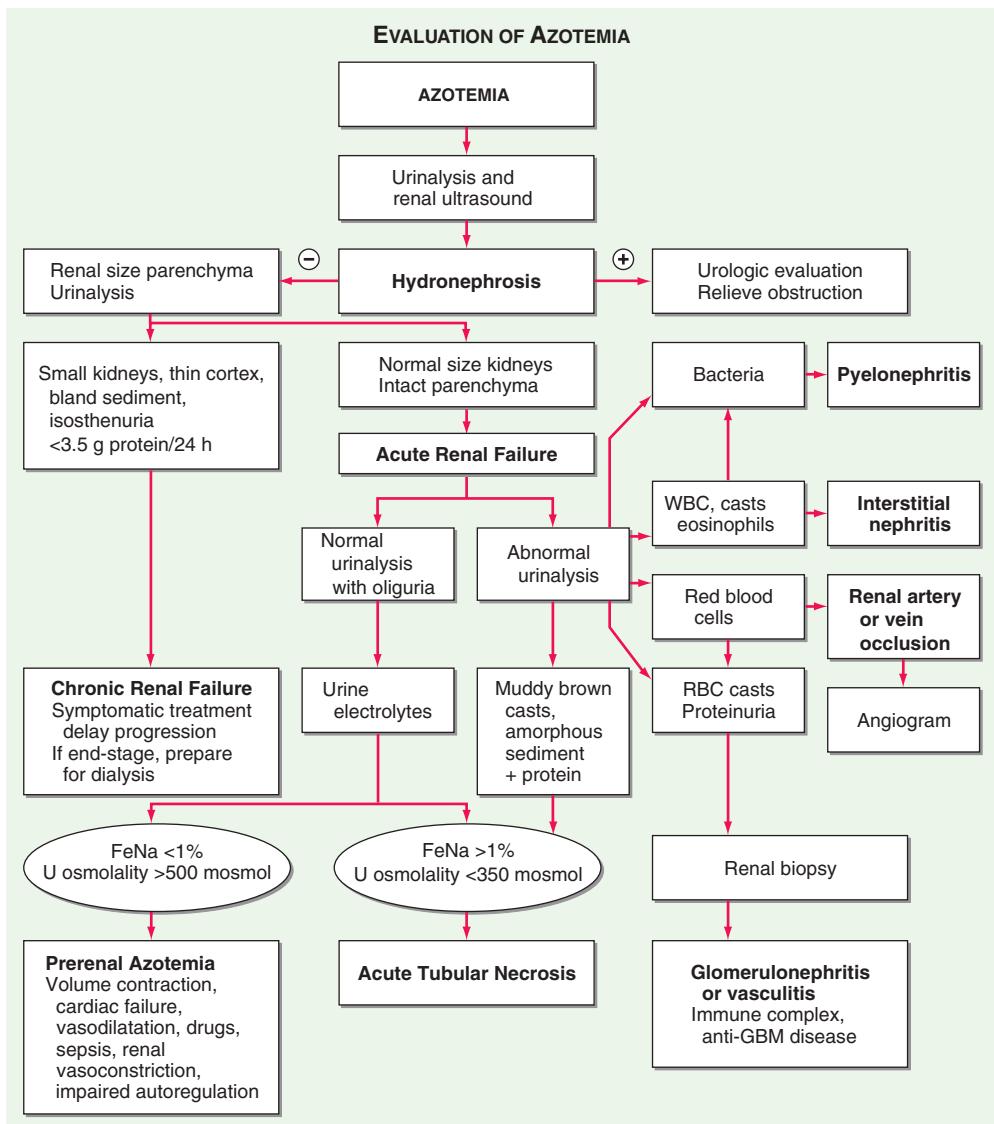


FIGURE 48-1 Approach to the patient with azotemia. FeNa, fractional excretion of sodium; GBM, glomerular basement membrane; RBC, red blood cell; WBC, white blood cell.

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.

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: $\text{CrCl} = (\text{U}_{\text{vol}} \times \text{U}_{\text{Cr}}) / (\text{P}_{\text{Cr}} \times \text{T}_{\text{min}})$. 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), 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).

$$\text{Cockcroft-Gault: } \text{CrCl} (\text{mL/min}) = (140 - \text{age (years)}) \times \text{weight (kg)} \\ \times [0.85 \text{ if female}] / (72 \times \text{P}_{\text{Cr}} (\text{mg/dL}))$$

$$\text{MDRD: eGFR (mL/min per } 1.73 \text{ m}^2) = 186.3 \times \text{P}_{\text{Cr}} (\text{e}^{-1.154}) \times \text{age (e}^{-0.203}) \\ \times (0.742 \text{ if female}) \times (1.21 \text{ if black}).$$

Numerous websites are available to assist with these calculations (www.kidney.org/professionals/kdoqi/gfr_calculator.cfm). A newer CKD-EPI 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:

$$\text{CKD-EPI: eGFR} = 141 \times \min(\text{P}_{\text{Cr}}/\text{k}, 1)^{\text{a}} \times \max(\text{P}_{\text{Cr}}/\text{k}, 1)^{-1.209} \times 0.993^{\text{Age}} \\ \times 1.018 \text{ [if female]} \times 1.159 \text{ [if black]},$$

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 $\text{P}_{\text{Cr}}/\text{k}$ or 1, and \max indicates the maximum of $\text{P}_{\text{Cr}}/\text{k}$ or 1 (<http://www.qxmd.com/renal/Calculate-CKD-EPI-GFR.php>).

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} . 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} ; however, like serum creatinine, cystatin C is influenced by the patient's age, race, and sex and also is associated with diabetes, smoking, and markers of inflammation.

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 situation, 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. 305) 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. 48-1. 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. 305. Acute renal failure (Chap. 304) can result from processes that affect and 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. 313).

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 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

TABLE 48-2 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

$$^aFE_{Na} = \frac{U_{Na} \times P_{Cr} \times 100}{P_{Na} \times U_{Cr}}$$

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.

prerenal azotemia from ATN (Table 48-2). 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, <1%), and $U_{Cr}/P_{Cr} > 40$ (Table 48-2). The FE_{Na} is typically >1% in ATN, but may be <1% in patients with milder, nonoliguric ATN (e.g., from rhabdomyolysis) and in pts 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. 48-1). 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. 313.**

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. 48-1, 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 48-2 and Fig. 48-1). 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. 304.**

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. 310**. 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. 48-1). 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. 311.**

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. 48-1), 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

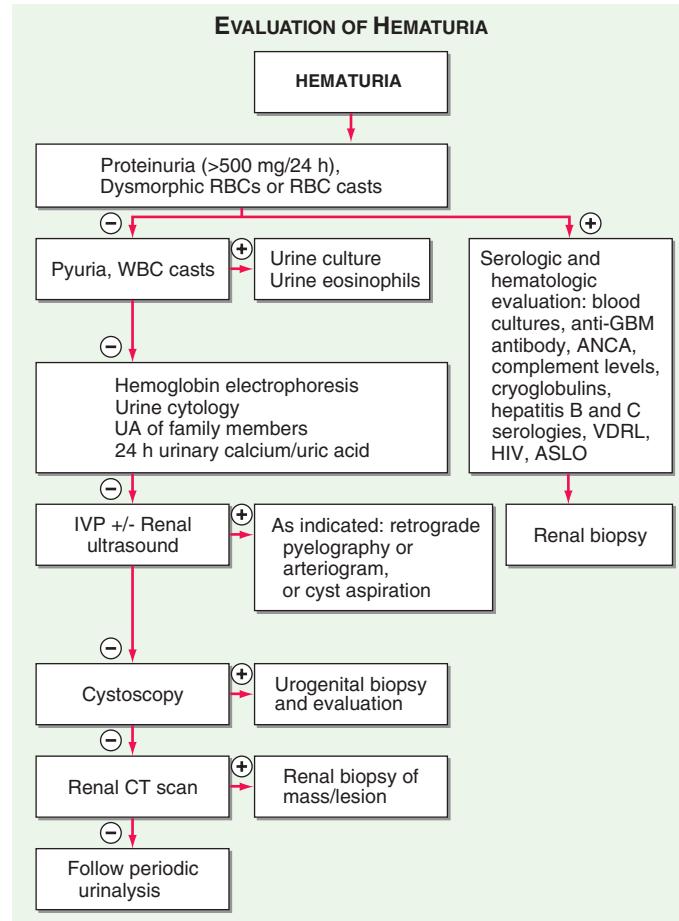


FIGURE 48-2 Approach to the patient with hematuria. ANCA, antineutrophil cytoplasmic antibody; ASLO, antistreptolysin O; CT, computed tomography; GBM, glomerular basement membrane; IVP, intravenous pyelography; RBC, red blood cell; UA, urinalysis; VDRL, Venereal Disease Research Laboratory; WBC, white blood cell.

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. 48-2**. A detailed discussion of glomerulonephritis and diseases of the microvasculature is found in **Chap. 310**.

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 nonsteroidal anti-inflammatory drugs, 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.

ABNORMALITIES OF THE URINE

PROTEINURIA

The evaluation of proteinuria is shown schematically in **Fig. 48-3** and typically is initiated after detection of proteinuria by dipstick examination. The dipstick measurement detects only albumin and gives

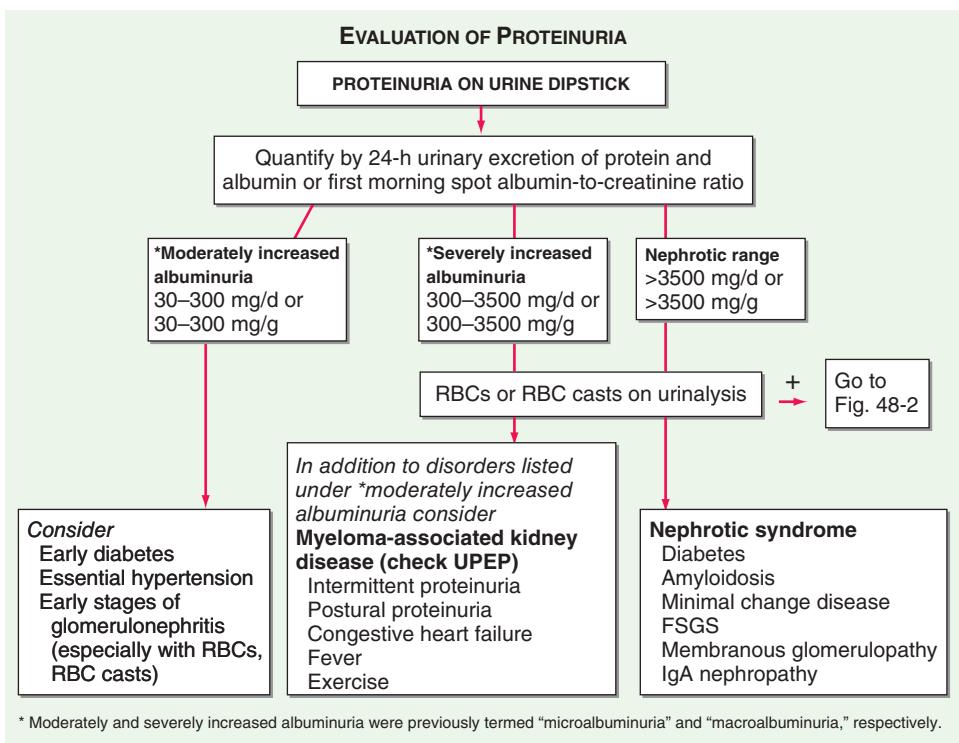


FIGURE 48-3 Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional dipsticks detect predominantly albumin and provide a semiquantitative assessment (trace, 1+, 2+, or 3+), which is influenced by urinary concentration 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.

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) \approx 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. 48-3). 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).

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. 48-3). 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. 48-3) 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 "nonspecific" proteinuria (Fig. 48-3).

When the total daily urinary excretion of protein is >3.5 g, hypoalbuminemia, hyperlipidemia, and edema (nephrotic syndrome; Fig. 48-3) 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. 48-3). 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. 310). 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. 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. 48-3 and Chap. 308). 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. 48-3) and drugs can cause the nephrotic syndrome; a complete list is found in Chap. 308.

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. 48-2. 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. Hypercalcemia 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. 308](#). 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. 48-2.

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. 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. 48-4). 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

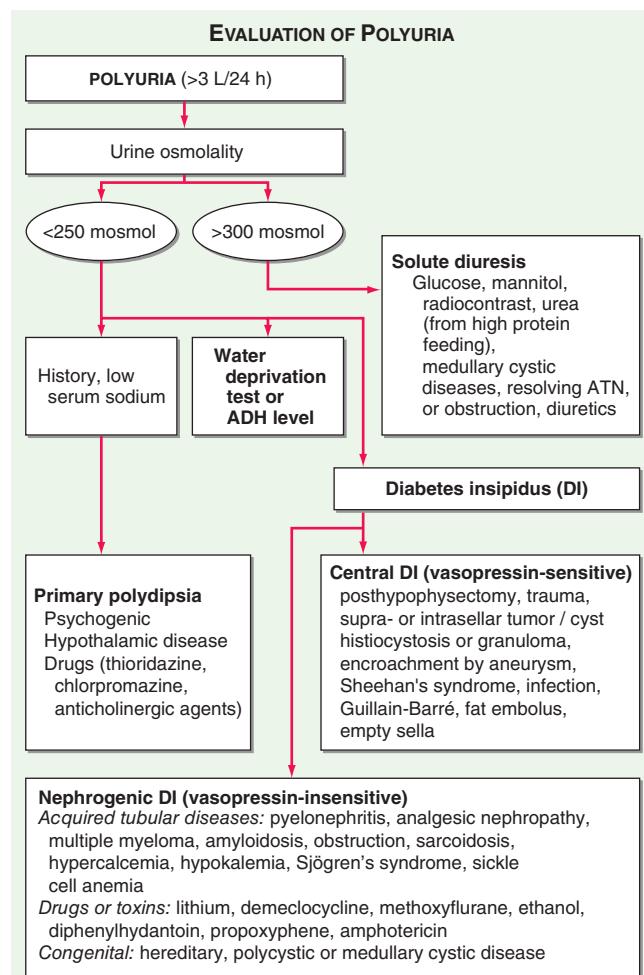


FIGURE 48-4 Approach to the patient with polyuria. AVP, antidiuretic hormone; ATN, acute tubular necrosis.

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 vasopressin (*central diabetes insipidus*), or failure of renal tubules to respond to vasopressin (*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.

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 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, neurological 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 vasopressin-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. 48-4.

A plasma AVP level is recommended as the best method for distinguishing between central and nephrogenic diabetes insipidus. Alternatively, a water deprivation test plus exogenous vasopressin may distinguish primary polydipsia from central and nephrogenic diabetes insipidus. **For a detailed discussion, see Chap. 374.**

ACKNOWLEDGMENT

This chapter was adapted and updated from the prior version written by Julie Lin and Bradley Denker.

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.
- KÖHLER H et al: Acanthocyturia—A characteristic marker for glomerular bleeding. *Kidney Int* 40:115, 1991.
- LEVEY AS et al: Glomerular filtration rate and albuminuria for detection and staging of acute and chronic kidney disease in adults: A systematic review. *JAMA* 313:837, 2015.
- PERAZELLA MA: The urine sediment as a biomarker of kidney disease. *Am J Kidney Dis* 66:748, 2015.
- SHARFUDDIN AA et al: Acute kidney injury, in *Brenner and Rector's The Kidney*, 10th ed, K Skorecki et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp 958–1011.

anions Cl^- and HCO_3^- , whereas K^+ and organic phosphate esters (ATP, creatine phosphate, and phospholipids) are the predominant ICF osmoles. Solutes that are restricted to the ECF or the ICF determine the “tonicity” or effective osmolality of that compartment. Certain solutes, particularly urea, do not contribute to water shifts across most membranes and are thus known as *ineffective osmoles*.

Water Balance Vasopressin secretion, water ingestion, and renal water transport collaborate to maintain human body fluid osmolality between 280 and 295 mOsm/kg. Vasopressin (AVP) is synthesized in magnocellular neurons within the hypothalamus; the distal axons of these neurons project to the posterior pituitary or neurohypophysis, from which AVP is released into the circulation. A network of central “osmoreceptor” neurons, which includes the AVP-expressing magnocellular neurons themselves, sense circulating osmolality via nonselective, stretch-activated cation channels. These osmoreceptor neurons are activated or inhibited by modest increases and decreases in circulating osmolality, respectively; activation leads to AVP release and thirst.

AVP secretion is stimulated as systemic osmolality increases above a threshold level of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP (Fig. 49-1). Thirst and thus water ingestion are also activated at ~285 mOsm/kg, beyond which there is an equivalent linear increase in the perceived intensity of thirst as a function of circulating osmolality. Changes in blood volume and blood pressure are also direct stimuli for AVP release and thirst, albeit with a less sensitive response profile. Of perhaps greater clinical relevance to the pathophysiology of water homeostasis, ECF volume strongly modulates the relationship between circulating osmolality and AVP release, such that hypovolemia reduces the osmotic threshold and increases the slope of the response curve to osmolality; *hypervolemia* has an opposite effect, increasing the osmotic threshold and reducing the slope of the response curve (Fig. 49-1). Notably, AVP has a half-life in the circulation of only 10–20 min; thus, changes in ECF volume and/or circulating osmolality can rapidly affect water homeostasis. In addition to volume status, a number of other “nonosmotic” stimuli have potent activating effects on osmosensitive neurons and AVP release, including nausea, intracerebral angiotensin II, serotonin, and multiple drugs.

The excretion or retention of electrolyte-free water by the kidney is modulated by circulating AVP. AVP acts on renal, V_2 -type receptors in the thick ascending limb of Henle and principal cells of the collecting duct (CD), increasing intracellular levels of cyclic AMP and activating protein kinase A (PKA)-dependent phosphorylation of multiple transport proteins. The AVP- and PKA-dependent activation of Na^+/Cl^- and K^+ transport by the thick ascending limb of the loop of Henle (TALH) is a key participant in the countercurrent mechanism (Fig. 49-2). The countercurrent mechanism ultimately increases the interstitial osmolality in the inner medulla of the kidney, driving water absorption

49

Fluid and Electrolyte Disturbances

David B. Mount

SODIUM AND WATER

COMPOSITION OF BODY FLUIDS

Water is the most abundant constituent in the body, comprising ~50% of body weight in women and 60% in men. Total-body water is distributed in two major compartments: 55–75% is intracellular (intracellular fluid [ICF]), and 25–45% is extracellular (extracellular fluid [ECF]). The ECF is further subdivided into intravascular (plasma water) and extravascular (interstitial) spaces in a ratio of 1:3. Fluid movement between the intravascular and interstitial spaces occurs across the capillary wall and is determined by Starling forces, i.e., capillary hydraulic pressure and colloid osmotic pressure. The transcapillary hydraulic pressure gradient exceeds the corresponding oncotic pressure gradient, thereby favoring the movement of plasma ultrafiltrate into the extravascular space. The return of fluid into the intravascular compartment occurs via lymphatic flow.

The solute or particle concentration of a fluid is known as its osmolality, expressed as milliosmoles per kilogram of water (mOsm/kg). Water easily diffuses across most cell membranes to achieve osmotic equilibrium (ECF osmolality = ICF osmolality). Notably, the extracellular and intracellular solute compositions differ considerably owing to the activity of various transporters, channels, and ATP-driven membrane pumps. The major ECF particles are Na^+ and its accompanying

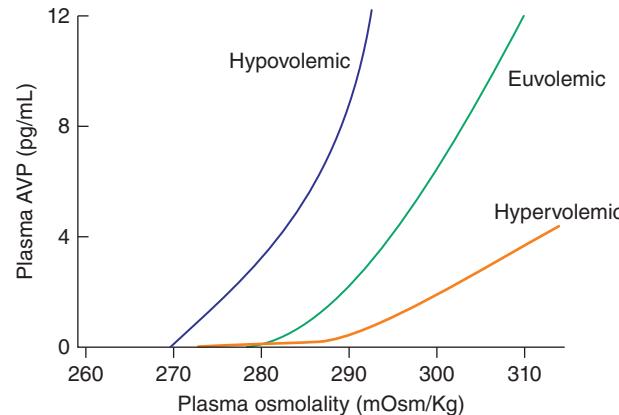


FIGURE 49-1 Circulating levels of vasopressin (AVP) in response to changes in osmolality. Plasma AVP becomes detectable in euvolemic, healthy individuals at a threshold of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP. The vasopressin response to osmolality is modulated strongly by volume status. The osmotic threshold is thus slightly lower in hypovolemia, with a steeper response curve; hypervolemia reduces the sensitivity of circulating AVP levels to osmolality.

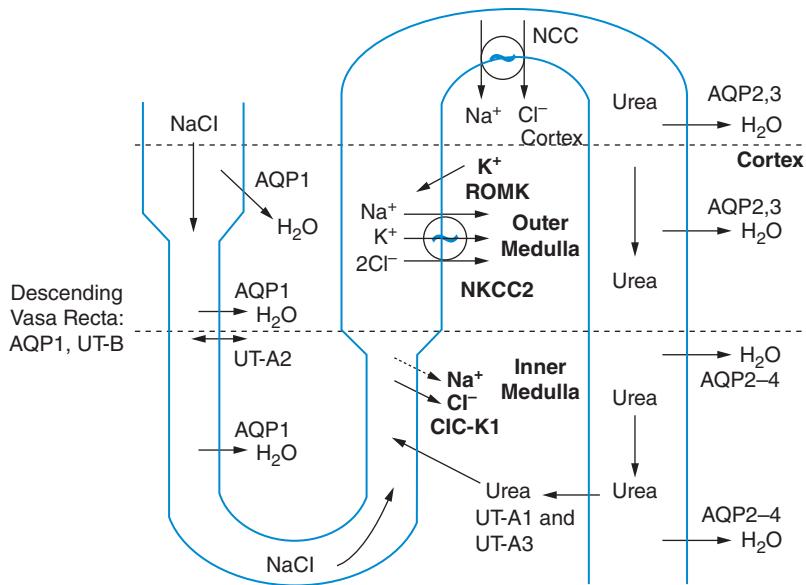


FIGURE 49-2 The renal concentrating mechanism. Water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (see text for details). Diagram showing the location of the major transport proteins involved; a loop of Henle is depicted on the left, collecting duct on the right. AQP, aquaporin; CLC-K1, chloride channel; NKCC2, Na-K- Cl cotransporter; ROMK, renal outer medullary K^+ channel; UT, urea transporter. (Used with permission from JM Sands: Molecular approaches to urea transporters. *J Am Soc Nephrol* 13:2795, 2002.)

across the renal CD. However, water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (Fig. 49-2). Water transport across apical and basolateral aquaporin-1 water channels in the descending thin limb of the loop of Henle is thus involved, as is passive absorption of Na^+ - Cl^- by the thin ascending limb, via apical and basolateral CLC-K1 chloride channels and paracellular Na^+ transport. Renal urea transport in turn plays important roles in the generation of the medullary osmotic gradient and the ability to excrete solute-free water under conditions of both high and low protein intake (Fig. 49-2).

AVP-induced, PKA-dependent phosphorylation of the aquaporin-2 water channel in principal cells stimulates the insertion of active water channels into the lumen of the CD, resulting in transepithelial water absorption down the medullary osmotic gradient (Fig. 49-3). Under "antidiuretic" conditions, with increased circulating AVP, the kidney reabsorbs water filtered by the glomerulus, equilibrating the osmolality across the CD epithelium to excrete a hypertonic, "concentrated" urine (osmolality of up to 1200 mOsm/kg). In the absence of circulating AVP, insertion of aquaporin-2 channels and water absorption across the CD is essentially abolished, resulting in secretion of a hypotonic, dilute urine (osmolality as low as 30–50 mOsm/kg). Abnormalities in this "final common pathway" are involved in most disorders of water homeostasis, e.g., a reduced or absent insertion of active aquaporin-2 water channels into the membrane of principal cells in diabetes insipidus (DI).

Maintenance of Arterial Circulatory Integrity Sodium is actively pumped out of cells by the Na^+/K^+ -ATPase membrane pump. In consequence, 85–90% of body Na^+ is extracellular, and the ECF volume (ECFV) is a function of total-body Na^+ content. Arterial perfusion and circulatory integrity are, in turn, determined by renal Na^+ retention or excretion, in addition to the modulation of systemic arterial resistance. Within the kidney, Na^+ is filtered by the glomeruli and then sequentially reabsorbed by the renal tubules. The Na^+ cation is typically reabsorbed with the chloride anion (Cl^-), and, thus, chloride homeostasis also affects the ECFV. On a quantitative level, at a glomerular filtration rate (GFR) of 180 L/d and serum Na^+ of ~140 mM, the kidney filters some 25,200 mmol/d of Na^+ . This is equivalent to ~1.5 kg of salt, which would occupy roughly 10 times the extracellular space; 99.6% of filtered Na^+ - Cl^- must be reabsorbed to excrete 100 mM per day.

Minute changes in renal Na^+ - Cl^- excretion will thus have significant effects on the ECFV, leading to edema syndromes or hypovolemia.

Approximately two-thirds of filtered Na^+ - Cl^- is reabsorbed by the renal proximal tubule, via both paracellular and transcellular mechanisms. The TALH subsequently reabsorbs another 25–30% of filtered Na^+ - Cl^- via the apical, furosemide-sensitive Na^+ - K^+ - 2Cl^- cotransporter. The adjacent aldosterone-sensitive distal nephron, comprising the distal convoluted tubule (DCT), connecting tubule (CNT), and CD, accomplishes the "fine-tuning" of renal Na^+ - Cl^- excretion. The thiazide-sensitive apical Na^+ - Cl^- cotransporter (NCC) reabsorbs 5–10% of filtered Na^+ - Cl^- in the DCT. Principal cells in the CNT and CD reabsorb Na^+ via electronegenic, amiloride-sensitive epithelial Na^+ channels (ENaC); Cl^- ions are primarily reabsorbed by adjacent intercalated cells, via apical Cl^- exchange (Cl^- -OH⁻ and Cl^- - HCO_3^- exchange, mediated by the SLC26A4 anion exchanger) (Fig. 49-4).

Renal tubular reabsorption of filtered Na^+ - Cl^- is regulated by multiple circulating and paracrine hormones, in addition to the activity of renal nerves. Angiotensin II activates proximal Na^+ - Cl^- reabsorption, as do adrenergic receptors under the influence of renal sympathetic innervation; locally generated dopamine, in contrast, has a *natriuretic* effect. Aldosterone primarily activates Na^+ - Cl^- reabsorption within the aldosterone-sensitive distal nephron. In particular, aldosterone activates the ENaC channel in principal cells, inducing Na^+ absorption and promoting K^+ excretion (Fig. 49-4).

Circulatory integrity is critical for the perfusion and function of vital organs. "Underfilling" of the arterial circulation is sensed by ventricular and vascular pressure receptors, resulting in a neurohumoral activation (increased sympathetic tone, activation of the renin-angiotensin-aldosterone axis, and increased circulating AVP) that synergistically increases renal Na^+ - Cl^- reabsorption, vascular resistance, and renal water reabsorption. This occurs in the context of decreased cardiac

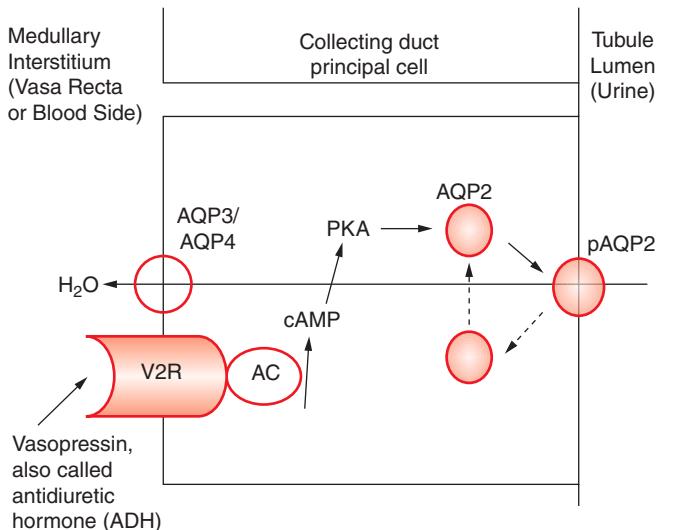


FIGURE 49-3 Vasopressin and the regulation of water permeability in the renal collecting duct. Vasopressin binds to the type 2 vasopressin receptor (V2R) on the basolateral membrane of principal cells, activates adenylyl cyclase (AC), increases intracellular cyclic adenosine monophosphate (cAMP), and stimulates protein kinase A (PKA) activity. Cytoplasmic vesicles carrying aquaporin-2 (AQP) water channel proteins are inserted into the luminal membrane in response to vasopressin, thereby increasing the water permeability of this membrane. When vasopressin stimulation ends, water channels are retrieved by an endocytic process and water permeability returns to its low basal rate. The AQP3 and AQP4 water channels are expressed on the basolateral membrane and complete the transcellular pathway for water reabsorption. pAQP2, phosphorylated aquaporin-2. (From JM Sands, DG Bichet: Nephrogenic diabetes insipidus. *Ann Intern Med* 144:186, 2006, with permission.)

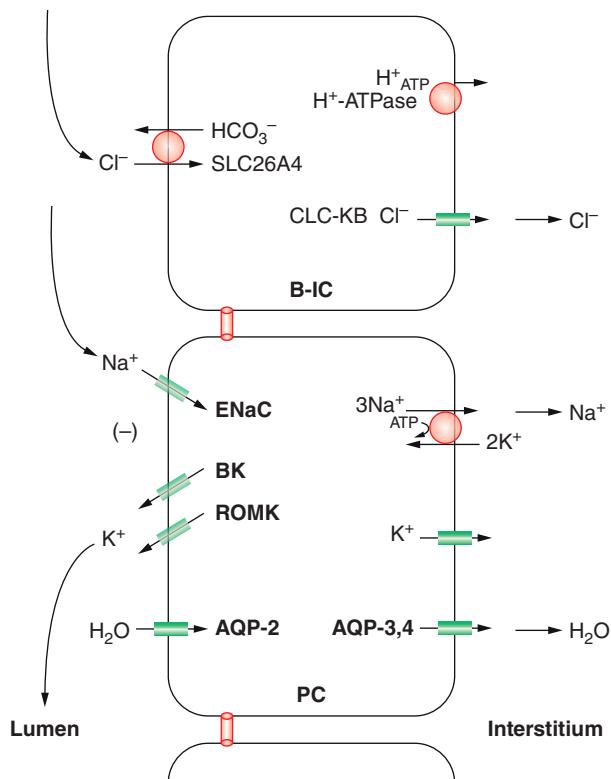


FIGURE 49-4 Sodium, water, and potassium transport in principal cells (PC) and adjacent β -intercalated cells (B-IC). The absorption of Na⁺ via the amiloride-sensitive epithelial sodium channel (ENaC) generates a lumen-negative potential difference, which drives K⁺ excretion through the apical secretory K⁺ channel ROMK (renal outer medullary K⁺ channel) and/or the flow-dependent BK channel. Transepithelial Cl⁻ transport occurs in adjacent β -intercalated cells, via apical Cl⁻-HCO₃⁻ and Cl⁻-OH⁻ exchange (SLC26A4 anion exchanger, also known as pendrin) basolateral CLC chloride channels. Water is absorbed down the osmotic gradient by principal cells, through the apical aquaporin-2 (AQP-2) and basolateral aquaporin-3 and aquaporin-4 (Fig. 49-3).

output, as occurs in hypovolemic states, low-output cardiac failure, decreased oncotic pressure, and/or increased capillary permeability. Alternatively, excessive arterial vasodilation results in *relative* arterial underfilling, leading to neurohumoral activation in the defense of tissue perfusion. These physiologic responses play important roles in many of the disorders discussed in this chapter. In particular, it is important to appreciate that AVP functions in the defense of circulatory integrity, inducing vasoconstriction, increasing sympathetic nervous system tone, increasing renal retention of both water and Na⁺-Cl⁻, and modulating the arterial baroreceptor reflex. Most of these responses involve activation of systemic V_{1A} AVP receptors, but concomitant activation of V₂ receptors in the kidney can result in renal water retention and hyponatremia.

HYPVOLEMIA

Etiology True volume depletion, or hypovolemia, generally refers to a state of combined salt and water loss, leading to contraction of the ECFV. The loss of salt and water may be renal or nonrenal in origin.

RENAL CAUSES Excessive urinary Na⁺-Cl⁻ and water loss is a feature of several conditions. A high filtered load of endogenous solutes, such as glucose and urea, can impair tubular reabsorption of Na⁺-Cl⁻ and water, leading to an osmotic diuresis. Exogenous mannitol, often used to decrease intracerebral pressure, is filtered by glomeruli but not reabsorbed by the proximal tubule, thus causing an osmotic diuresis. Pharmacologic diuretics selectively impair Na⁺-Cl⁻ reabsorption at specific sites along the nephron, leading to increased urinary Na⁺-Cl⁻ excretion. Other drugs can induce natriuresis as a side effect. For example, acetazolamide can inhibit proximal tubular Na⁺-Cl⁻ absorption via its inhibition of carbonic anhydrase; other drugs, such as the antibiotics trimethoprim (TMP) and pentamidine, inhibit distal tubular Na⁺

reabsorption through the amiloride-sensitive ENaC channel, leading to urinary Na⁺-Cl⁻ loss. Hereditary defects in renal transport proteins are also associated with reduced reabsorption of filtered Na⁺-Cl⁻ and/or water. Alternatively, mineralocorticoid deficiency, mineralocorticoid resistance, or inhibition of the mineralocorticoid receptor (MLR) can reduce Na⁺-Cl⁻ reabsorption by the aldosterone-sensitive distal nephron. Finally, tubulointerstitial injury, as occurs in interstitial nephritis, acute tubular injury, or obstructive uropathy, can reduce distal tubular Na⁺-Cl⁻ and/or water absorption.

Excessive excretion of free water, i.e., water without electrolytes, can also lead to hypovolemia. However, the effect on ECFV is usually less marked, given that two-thirds of the water volume is lost from the ICF. Excessive renal water excretion occurs in the setting of decreased circulating AVP or renal resistance to AVP (central and nephrogenic DI, respectively).

EXTRARENAL CAUSES Nonrenal causes of hypovolemia include fluid loss from the gastrointestinal tract, skin, and respiratory system. Accumulations of fluid within specific tissue compartments, typically the interstitium, peritoneum, or gastrointestinal tract, can also cause hypovolemia.

Approximately 9 L of fluid enter the gastrointestinal tract daily, 2 L by ingestion and 7 L by secretion; almost 98% of this volume is absorbed, such that daily fecal fluid loss is only 100–200 mL. Impaired gastrointestinal reabsorption or enhanced secretion of fluid can cause hypovolemia. Because gastric secretions have a low pH (high H⁺ concentration), whereas biliary, pancreatic, and intestinal secretions are alkaline (high HCO₃⁻ concentration), vomiting and diarrhea are often accompanied by metabolic alkalosis and acidosis, respectively.

Evaporation of water from the skin and respiratory tract (so-called “insensible losses”) constitutes the major route for loss of solute-free water, which is typically 500–650 mL/d in healthy adults. This evaporative loss can increase during febrile illness or prolonged heat exposure. Hyperventilation can also increase insensible losses via the respiratory tract, particularly in ventilated patients; the humidity of inspired air is another determining factor. In addition, increased exertion and/or ambient temperature will increase insensible losses via sweat, which is hypotonic to plasma. Profuse sweating without adequate repletion of water and Na⁺-Cl⁻ can thus lead to both hypovolemia and hypertonicity. Alternatively, replacement of these insensible losses with a surfeit of free water, without adequate replacement of electrolytes, may lead to hypovolemic hyponatremia.

Excessive fluid accumulation in interstitial and/or peritoneal spaces can also cause intravascular hypovolemia. Increases in vascular permeability and/or a reduction in oncotic pressure (hypoalbuminemia) alter Starling forces, resulting in excessive “third spacing” of the ECFV. This occurs in sepsis syndrome, burns, pancreatitis, nutritional hypoalbuminemia, and peritonitis. Alternatively, distributive hypovolemia can occur due to accumulation of fluid within specific compartments, for example within the bowel lumen in gastrointestinal obstruction or ileus. Hypovolemia can also occur after extracorporeal hemorrhage or after significant hemorrhage into an expandable space, for example, the retroperitoneum.

Diagnostic Evaluation A careful history will usually determine the etiologic cause of hypovolemia. Symptoms of hypovolemia are nonspecific and include fatigue, weakness, thirst, and postural dizziness; more severe symptoms and signs include oliguria, cyanosis, abdominal and chest pain, and confusion or obtundation. Associated electrolyte disorders may cause additional symptoms, for example, muscle weakness in patients with hypokalemia. On examination, diminished skin turgor and dry oral mucous membranes are less than ideal markers of a decreased ECFV in adult patients; more reliable signs of hypovolemia include a decreased jugular venous pressure (JVP), orthostatic tachycardia (an increase of >15–20 beats/min upon standing), and orthostatic hypotension (a >10–20 mmHg drop in blood pressure on standing). More severe fluid loss leads to hypovolemic shock, with hypotension, tachycardia, peripheral vasoconstriction, and peripheral hypoperfusion; these patients may exhibit peripheral cyanosis, cold extremities, oliguria, and altered mental status.

Routine chemistries may reveal an increase in blood urea nitrogen (BUN) and creatinine, reflective of a decrease in GFR. Creatinine is the more dependable measure of GFR, because BUN levels may be influenced by an increase in tubular reabsorption ("prerenal azotemia"), an increase in urea generation in catabolic states, hyperalimentation, or gastrointestinal bleeding, and/or a decreased urea generation in decreased protein intake. In hypovolemic shock, liver function tests and cardiac biomarkers may show evidence of hepatic and cardiac ischemia, respectively. Routine chemistries and/or blood gases may reveal evidence of acid-base disorders. For example, bicarbonate loss due to diarrheal illness is a very common cause of metabolic acidosis; alternatively, patients with severe hypovolemic shock may develop lactic acidosis with an elevated anion gap.

The neurohumoral response to hypovolemia stimulates an increase in renal tubular Na^+ and water reabsorption. Therefore, the urine Na^+ concentration is typically $<20 \text{ mM}$ in nonrenal causes of hypovolemia, with a urine osmolality of $>450 \text{ mOsm/kg}$. The reduction in both GFR and distal tubular Na^+ delivery may cause a defect in renal potassium excretion, with an increase in plasma K^+ concentration. Of note, patients with hypovolemia and a hypochloremic alkalosis due to vomiting, diarrhea, or diuretics will typically have a urine Na^+ concentration $>20 \text{ mM}$ and urine pH of >7.0 , due to the increase in filtered HCO_3^- ; the urine Cl^- concentration in this setting is a more accurate indicator of volume status, with a level $<25 \text{ mM}$ suggestive of hypovolemia. The urine Na^+ concentration is often $>20 \text{ mM}$ in patients with *renal* causes of hypovolemia, such as acute tubular necrosis; similarly, patients with DI will have an inappropriately dilute urine.

TREATMENT

Hypovolemia

The therapeutic goals in hypovolemia are to restore normovolemia and replace ongoing fluid losses. Mild hypovolemia can usually be treated with oral hydration and resumption of a normal maintenance diet. More severe hypovolemia requires intravenous hydration, tailoring the choice of solution to the underlying pathophysiology. Isotonic, "normal" saline ($0.9\% \text{ NaCl}$, 154 mM Na^+) is the most appropriate resuscitation fluid for normonatremic or hyponatremic patients with severe hypovolemia; colloid solutions such as intravenous albumin are not demonstrably superior for this purpose. Hypernatremic patients should receive a hypotonic solution, 5% dextrose if there has only been water loss (as in DI), or hypotonic

saline ($1/2$ or $1/4$ normal saline) if there has been water and $\text{Na}^+ \text{-Cl}^-$ loss; changes in free water administration should be made if necessary, based on frequent measuring of serum chemistries. Patients with bicarbonate loss and metabolic acidosis, as occur frequently in diarrhea, should receive intravenous bicarbonate, either an isotonic solution ($150 \text{ meq of } \text{Na}^+\text{-HCO}_3^-$ in 5% dextrose) or a more hypotonic bicarbonate solution in dextrose or dilute saline. Patients with severe hemorrhage or anemia should receive red cell transfusions, without increasing the hematocrit beyond 35%.

SODIUM DISORDERS

Disorders of serum Na^+ concentration are caused by abnormalities in water homeostasis, leading to changes in the relative ratio of Na^+ to body water. Water intake and circulating AVP constitute the two key effectors in the defense of serum osmolality; defects in one or both of these two defense mechanisms cause most cases of hyponatremia and hypernatremia. In contrast, abnormalities in sodium homeostasis *per se* lead to a deficit or surplus of whole-body $\text{Na}^+ \text{-Cl}^-$ content, a key determinant of the ECFV and circulatory integrity. Notably, volume status also modulates the release of AVP by the posterior pituitary, such that hypovolemia is associated with higher circulating levels of the hormone at each level of serum osmolality. Similarly, in "hypervolemic" causes of arterial underfilling, e.g., heart failure and cirrhosis, the associated neurohumoral activation encompasses an increase in circulating AVP, leading to water retention and hyponatremia. Therefore, a key concept in sodium disorders is that the absolute plasma Na^+ concentration tells one nothing about the volume status of a given patient, which furthermore must be taken into account in the diagnostic and therapeutic approach.

HYPONATREMIA

Hyponatremia, which is defined as a plasma Na^+ concentration $<135 \text{ mM}$, is a very common disorder, occurring in up to 22% of hospitalized patients. This disorder is almost always the result of an increase in circulating AVP and/or increased renal sensitivity to AVP, combined with an intake of free water; a notable exception is hyponatremia due to low solute intake (see below). The underlying pathophysiology for the exaggerated or "inappropriate" AVP response differs in patients with hyponatremia as a function of their ECFV. Hyponatremia is thus subdivided diagnostically into three groups, depending on clinical history and volume status, i.e., "hypovolemic," "euvolemic," and "hypervolemic" (Fig. 49-5).

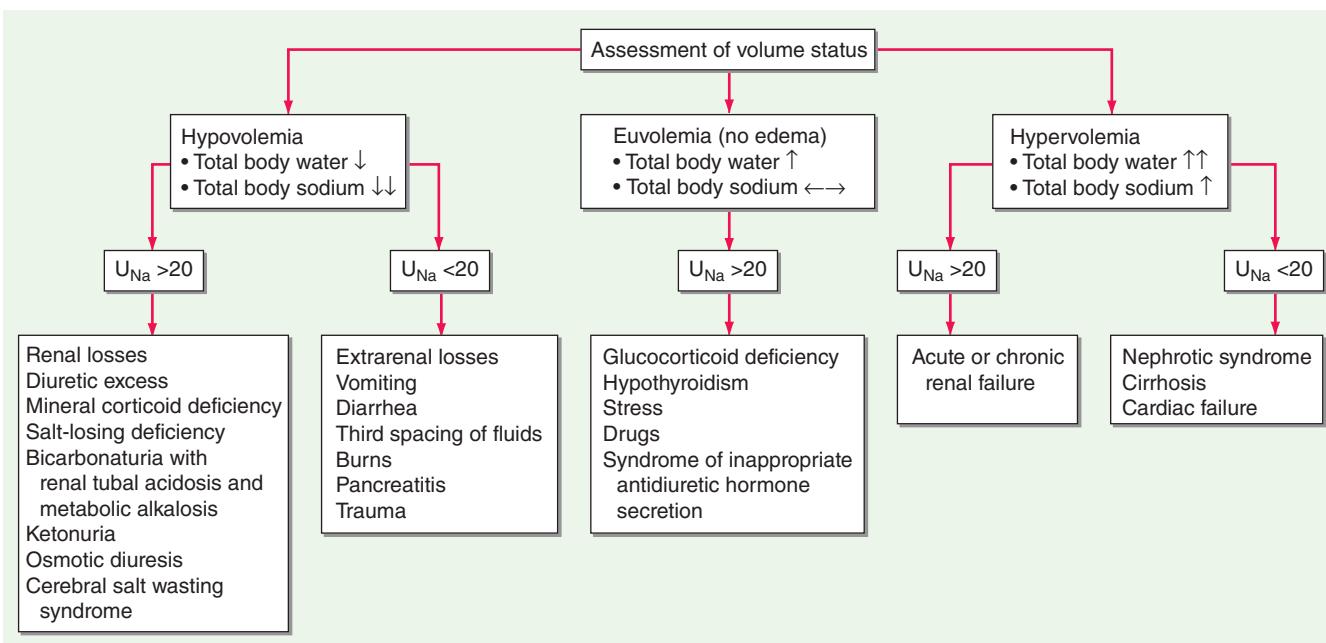


FIGURE 49-5 The diagnostic approach to hyponatremia. (From S Kumar, T Berl: Diseases of water metabolism, in Atlas of Diseases of the Kidney, RW Schrier [ed]. Philadelphia, Current Medicine, Inc, 1999; with permission.)

Hypovolemic Hyponatremia Hypovolemia causes a marked neurohumoral activation, increasing circulating levels of AVP. The increase in circulating AVP helps preserve blood pressure via vascular and baroreceptor V_{1A} receptors and increases water reabsorption via renal V₂ receptors; activation of V₂ receptors can lead to hyponatremia in the setting of increased free water intake. Nonrenal causes of hypovolemic hyponatremia include GI loss (e.g., vomiting, diarrhea, tube drainage) and insensible loss (sweating, burns) of Na⁺-Cl⁻ and water, in the absence of adequate oral replacement; urine Na⁺ concentration is typically <20 mM. Notably, these patients may be clinically classified as euvolemic, with only the reduced urinary Na⁺ concentration to indicate the cause of their hyponatremia. Indeed, a urine Na⁺ concentration <20 mM, in the absence of a cause of hypovolemic hyponatremia, predicts a rapid increase in plasma Na⁺ concentration in response to intravenous normal saline; saline therapy thus induces a water diuresis in this setting, as circulating AVP levels plummet.

The renal causes of hypovolemic hyponatremia share an inappropriate loss of Na⁺-Cl⁻ in the urine, leading to volume depletion and an increase in circulating AVP; urine Na⁺ concentration is typically >20 mM (Fig. 49-5). A deficiency in circulating aldosterone and/or its renal effects can lead to hyponatremia in primary adrenal insufficiency and other causes of hypoaldosteronism; hyperkalemia and hyponatremia in a hypotensive and/or hypovolemic patient with high urine Na⁺ concentration (much greater than 20 mM) should strongly suggest this diagnosis. Salt-losing nephropathies may lead to hyponatremia when sodium intake is reduced, due to impaired renal tubular function; typical causes include reflux nephropathy, interstitial nephropathies, postobstructive uropathy, medullary cystic disease, and the recovery phase of acute tubular necrosis. Thiazide diuretics cause hyponatremia via a number of mechanisms, including polydipsia and diuretic-induced volume depletion. Notably, thiazides do not inhibit the renal concentrating mechanism, such that circulating AVP retains a full effect on renal water retention. In contrast, loop diuretics, which are less frequently associated with hyponatremia, inhibit Na⁺-Cl⁻ and K⁺ absorption by the TALH, blunting the countercurrent mechanism and reducing the ability to concentrate the urine. Increased excretion of an osmotically active nonreabsorbable or poorly reabsorbable solute can also lead to volume depletion and hyponatremia; important causes include glycosuria, ketonuria (e.g., in starvation or in diabetic or alcoholic ketoacidosis), and bicarbonaturia (e.g., in renal tubular acidosis or metabolic alkalosis, where the associated bicarbonaturia leads to loss of Na⁺).

Finally, the syndrome of “cerebral salt wasting” is a rare cause of hypovolemic hyponatremia, encompassing hyponatremia with clinical hypovolemia and inappropriate natriuresis in association with intracranial disease; associated disorders include subarachnoid hemorrhage, traumatic brain injury, craniotomy, encephalitis, and meningitis. Distinction from the more common syndrome of inappropriate antidiuresis (SIAD) is critical because cerebral salt wasting will typically respond to aggressive Na⁺-Cl⁻ repletion.

Hypervolemic Hyponatremia Patients with hypervolemic hyponatremia develop an increase in total-body Na⁺-Cl⁻ that is accompanied by a proportionately *greater* increase in total-body water, leading to a reduced plasma Na⁺ concentration. As in hypovolemic hyponatremia, the causative disorders can be separated by the effect on urine Na⁺ concentration, with acute or chronic renal failure uniquely associated with an increase in urine Na⁺ concentration (Fig. 49-5). The pathophysiology of hyponatremia in the sodium-avid edematous disorders (congestive heart failure [CHF], cirrhosis, and nephrotic syndrome) is similar to that in hypovolemic hyponatremia, except that arterial filling and circulatory integrity is decreased due to the specific etiologic factors (e.g., cardiac dysfunction in CHF, peripheral vasodilation in cirrhosis). Urine Na⁺ concentration is typically very low, i.e., <10 mM, even after hydration with normal saline; this Na⁺-avid state may be obscured by diuretic therapy. The degree of hyponatremia provides an indirect index of the associated neurohumoral activation and is an important prognostic indicator in hypervolemic hyponatremia.

Euvolemic Hyponatremia Euvolemic hyponatremia can occur in moderate to severe hypothyroidism, with correction after achieving a euthyroid state. Severe hyponatremia can also be a consequence of secondary adrenal insufficiency due to pituitary disease; whereas the deficit in circulating aldosterone in primary adrenal insufficiency causes *hypovolemic* hyponatremia, the predominant glucocorticoid deficiency in secondary adrenal failure is associated with *euvolemic* hyponatremia. Glucocorticoids exert a negative feedback on AVP release by the posterior pituitary such that hydrocortisone replacement in these patients can rapidly normalize the AVP response to osmolality, reducing circulating AVP.

The SIAD is the most frequent cause of euvolemic hyponatremia (Table 49-1). The generation of hyponatremia in SIAD requires an intake of free water, with persistent intake at serum osmolalities that are lower than the usual threshold for thirst; as one would expect,

TABLE 49-1 Causes of the Syndrome of Inappropriate Antidiuresis (SIAD)

MALIGNANT DISEASES	PULMONARY DISORDERS	DISORDERS OF THE CENTRAL NERVOUS SYSTEM	DRUGS	OTHER CAUSES
Carcinoma	Infections	Infection	Drugs that stimulate release of AVP or enhance its action	Hereditary (gain-of-function mutations in the vasopressin V ₂ receptor)
Lung	Bacterial pneumonia	Encephalitis	Chlorpropamide	Idiopathic
Small cell	Viral pneumonia	Meningitis	SSRIs	Transient
Mesothelioma	Pulmonary abscess	Brain abscess	Tricyclic antidepressants	Endurance exercise
Oropharynx	Tuberculosis	Rocky Mountain spotted fever	Clofibrate	General anesthesia
Gastrointestinal tract	Aspergillosis	AIDS	Carbamazepine	Nausea
Stomach	Asthma	Bleeding and masses	Vincristine	Pain
Duodenum	Cystic fibrosis	Subdural hematoma	Nicotine	Stress
Pancreas	Respiratory failure associated with positive-pressure breathing	Subarachnoid hemorrhage	Narcotics	
Genitourinary tract		Cerebrovascular accident	Antipsychotic drugs	
Ureter		Brain tumors	Ifosfamide	
Bladder		Head trauma	Cyclophosphamide	
Prostate		Hydrocephalus	Nonsteroidal anti-inflammatory drugs	
Endometrium		Cavernous sinus thrombosis	MDMA (“ecstasy”)	
Endocrine thymoma		Other	AVP analogues	
Lymphomas		Multiple sclerosis	Desmopressin	
Sarcomas		Guillain-Barré syndrome	Oxytocin	
Ewing's sarcoma		Shy-Drager syndrome	Vasopressin	
		Delirium tremens		
		Acute intermittent porphyria		

Abbreviations: AVP vasopressin; MDMA; 3,4-methylenedioxymethamphetamine; SSRI, selective serotonin reuptake inhibitor.

Source: From DH Ellison, T Berl: Syndrome of inappropriate antidiuresis. N Engl J Med 356:2064, 2007.

the osmotic threshold and osmotic response curves for the sensation of thirst are shifted downward in patients with SIAD. Four distinct patterns of AVP secretion have been recognized in patients with SIAD, independent for the most part of the underlying cause. Unregulated, erratic AVP secretion is seen in about a third of patients, with no obvious correlation between serum osmolality and circulating AVP levels. Other patients fail to suppress AVP secretion at lower serum osmolalities, with a normal response curve to hyperosmolar conditions; others have a "reset osmostat," with a lower threshold osmolality and a left-shifted osmotic response curve. Finally, the fourth subset of patients have essentially no detectable circulating AVP, suggesting either a gain in function in renal water reabsorption or a circulating antidiuretic substance that is distinct from AVP. Gain-in-function mutations of a single specific residue in the V₂ AVP receptor have been described in some of these patients, leading to constitutive activation of the receptor in the absence of AVP and "nephrogenic" SIAD.

Strictly speaking, patients with SIAD are not euvolemic but are subclinically volume-expanded, due to AVP-induced water and Na⁺-Cl⁻ retention; "AVP escape" mechanisms invoked by sustained increases in AVP serve to limit distal renal tubular transport, preserving a modestly hypervolemic steady state. Serum uric acid is often low (<4 mg/dL) in patients with SIAD, consistent with suppressed proximal tubular transport in the setting of increased distal tubular Na⁺-Cl⁻ and water transport; in contrast, patients with hypovolemic hyponatremia will often be hyperuricemic, due to a shared activation of proximal tubular Na⁺-Cl⁻ and urate transport.

Common causes of SIAD include pulmonary disease (e.g., pneumonia, tuberculosis, pleural effusion) and central nervous system (CNS) diseases (e.g., tumor, subarachnoid hemorrhage, meningitis). SIAD also occurs with malignancies, most commonly with small-cell lung carcinoma (75% of malignancy-associated SIAD); ~10% of patients with this tumor will have a plasma Na⁺ concentration of <130 mM at presentation. SIAD is also a frequent complication of certain drugs, most commonly the selective serotonin reuptake inhibitors (SSRIs). Other drugs can potentiate the renal effect of AVP, without exerting direct effects on circulating AVP levels (Table 49-1).

Low Solute Intake and Hyponatremia Hyponatremia can occasionally occur in patients with a very low intake of dietary solutes. Classically, this occurs in alcoholics whose sole nutrient is beer, hence the diagnostic label of *beer potomania*; beer is very low in protein and salt content, containing only 1–2 mM of Na⁺. The syndrome has also been described in nonalcoholic patients with highly restricted solute intake due to nutrient-restricted diets, e.g., extreme vegetarian diets. Patients with hyponatremia due to low solute intake typically present with a very low urine osmolality (<100–200 mOsm/kg) with a urine Na⁺ concentration that is <10–20 mM. The fundamental abnormality is the inadequate dietary intake of solutes; the reduced urinary solute excretion limits water excretion such that hyponatremia ensues after relatively modest polydipsia. AVP levels have not been reported in patients with beer potomania but are expected to be suppressed or rapidly suppressible with saline hydration; this fits with the overly rapid correction in plasma Na⁺ concentration that can be seen with saline hydration. Resumption of a normal diet and/or saline hydration will also correct the causative deficit in urinary solute excretion, such that patients with beer potomania typically correct their plasma Na⁺ concentration promptly after admission to the hospital.

Clinical Features of Hyponatremia Hyponatremia induces generalized cellular swelling, a consequence of water movement down the osmotic gradient from the hypotonic ECF to the ICF. The symptoms of hyponatremia are primarily neurologic, reflecting the development of cerebral edema within a rigid skull. The initial CNS response to acute hyponatremia is an increase in interstitial pressure, leading to shunting of ECF and solutes from the interstitial space into the cerebrospinal fluid and then on into the systemic circulation. This is accompanied by an efflux of the major intracellular ions, Na⁺, K⁺, and Cl⁻, from brain cells. Acute hyponatremic encephalopathy ensues when these volume regulatory mechanisms are overwhelmed by a rapid decrease in tonicity, resulting in acute cerebral edema. Early symptoms can include

TABLE 49-2 Causes of Acute Hyponatremia

Iatrogenic
Postoperative: premenopausal women
Hypotonic fluids with cause of ↑ vasopressin
Glycine irrigation: TURP; uterine surgery
Colonoscopy preparation
Recent institution of thiazides
Polydipsia
MDMA ("ecstasy," "Molly") ingestion
Exercise induced
Multifactorial, e.g., thiazide and polydipsia

Abbreviations: MDMA, 3,4-methylenedioxymethamphetamine; TURP, transurethral resection of the prostate.

nausea, headache, and vomiting. However, severe complications can rapidly evolve, including seizure activity, brainstem herniation, coma, and death. A key complication of acute hyponatremia is normocapneic or hypercapneic respiratory failure; the associated hypoxia may amplify the neurologic injury. Normocapneic respiratory failure in this setting is typically due to noncardiogenic, "neurogenic" pulmonary edema, with a normal pulmonary capillary wedge pressure.

Acute symptomatic hyponatremia is a medical emergency, occurring in a number of specific settings (Table 49-2). Women, particularly before menopause, are much more likely than men to develop encephalopathy and severe neurologic sequelae. Acute hyponatremia often has an iatrogenic component, e.g., when hypotonic intravenous fluids are given to postoperative patients with an increase in circulating AVP. Exercise-associated hyponatremia, an important clinical issue at marathons and other endurance events, has similarly been linked to both a "nonosmotic" increase in circulating AVP and excessive free water intake. The recreational drugs Molly and ecstasy, which share an active ingredient (MDMA, 3,4-methylenedioxymethamphetamine), cause a rapid and potent induction of both thirst and AVP, leading to severe acute hyponatremia.

Persistent, chronic hyponatremia results in an efflux of organic osmolytes (creatinine, betaine, glutamate, myoinositol, and taurine) from brain cells; this response reduces intracellular osmolality and the osmotic gradient favoring water entry. This reduction in intracellular osmolytes is largely complete within 48 h, the time period that clinically defines chronic hyponatremia; this temporal definition has considerable relevance for the treatment of hyponatremia (see below). The cellular response to chronic hyponatremia does not fully protect patients from symptoms, which can include vomiting, nausea, confusion, and seizures, usually at plasma Na⁺ concentration <125 mM. Even patients who are judged "asymptomatic" can manifest subtle gait and cognitive defects that reverse with correction of hyponatremia; notably, chronic "asymptomatic" hyponatremia increases the risk of falls. Chronic hyponatremia also increases the risk of bony fractures owing to the associated neurologic dysfunction and to a hyponatremia-associated reduction in bone density. Therefore, every attempt should be made to safely correct the plasma Na⁺ concentration in patients with chronic hyponatremia, even in the absence of overt symptoms (see the section on treatment of hyponatremia below).

The management of chronic hyponatremia is complicated significantly by the asymmetry of the cellular response to correction of plasma Na⁺ concentration. Specifically, the *reaccumulation* of organic osmolytes by brain cells is attenuated and delayed as osmolality increases after correction of hyponatremia, sometimes resulting in degenerative loss of oligodendrocytes and an osmotic demyelination syndrome (ODS). Overly rapid correction of hyponatremia (>8–10 mM in 24 h or 18 mM in 48 h) is also associated with a disruption in integrity of the blood-brain barrier, allowing the entry of immune mediators that may contribute to demyelination. The lesions of ODS classically affect the pons, a neuroanatomic structure wherein the delay in the reaccumulation of osmotic osmolytes is particularly pronounced; clinically, patients with central pontine myelinolysis can present 1 or more days after overcorrection of hyponatremia with paraparesis or quadripareisis,

dysphagia, dysarthria, diplopia, a “locked-in syndrome,” and/or loss of consciousness. Other regions of the brain can also be involved in ODS, most commonly in association with lesions of the pons but occasionally in isolation; in order of frequency, the lesions of extrapontine myelinolysis can occur in the cerebellum, lateral geniculate body, thalamus, putamen, and cerebral cortex or subcortex. Clinical presentation of ODS can, therefore, vary as a function of the extent and localization of extrapontine myelinolysis, with the reported development of ataxia, mutism, parkinsonism, dystonia, and catatonia. Relowering of plasma Na^+ concentration after overly rapid correction can prevent or attenuate ODS (see the section on treatment of hyponatremia below). However, even appropriately slow correction can be associated with ODS, particularly in patients with additional risk factors; these include alcoholism, malnutrition, hypokalemia, and liver transplantation.

Diagnostic Evaluation of Hyponatremia Clinical assessment of hyponatremic patients should focus on the underlying cause; a detailed drug history is particularly crucial (Table 49-1). A careful clinical assessment of volume status is obligatory for the classical diagnostic approach to hyponatremia (Fig. 49-5). Hyponatremia is frequently multifactorial, particularly when severe; clinical evaluation should consider all the possible causes for excessive circulating AVP, including volume status, drugs, and the presence of nausea and/or pain. Radiologic imaging may also be appropriate to assess whether patients have a pulmonary or CNS cause for hyponatremia. A screening chest x-ray may fail to detect a small-cell carcinoma of the lung; computed tomography (CT) scanning of the thorax should be considered in patients at high risk for this tumor (e.g., patients with a smoking history).

Laboratory investigation should include a measurement of serum osmolality to exclude pseudohyponatremia, which is defined as the coexistence of hyponatremia with a normal or increased plasma tonicity. Most clinical laboratories measure plasma Na^+ concentration by testing diluted samples with automated ion-sensitive electrodes, correcting for this dilution by assuming that plasma is 93% water. This correction factor can be inaccurate in patients with pseudohyponatremia due to extreme hyperlipidemia and/or hyperproteinemia, in whom serum lipid or protein makes up a greater percentage of plasma volume. The measured osmolality should also be converted to the effective osmolality (tonicity) by subtracting the measured concentration of urea (divided by 2.8, if in mg/dL); patients with hyponatremia have an effective osmolality of $<275 \text{ mOsm/kg}$.

Elevated BUN and creatinine in routine chemistries can also indicate renal dysfunction as a potential cause of hyponatremia, whereas hyperkalemia may suggest adrenal insufficiency or hypoaldosteronism. Serum glucose should also be measured; plasma Na^+ concentration falls by $\sim 1.6\text{--}2.4 \text{ mM}$ for every 100-mg/dL increase in glucose, due to glucose-induced water efflux from cells; this “true” hyponatremia resolves after correction of hyperglycemia. Measurement of serum uric acid should also be performed; whereas patients with SIAD-type physiology will typically be hypouricemic (serum uric acid $<4 \text{ mg/dL}$), volume-depleted patients will often be hyperuricemic. In the appropriate clinical setting, thyroid, adrenal, and pituitary function should also be tested; hypothyroidism and secondary adrenal failure due to pituitary insufficiency are important causes of euvolemic hyponatremia, whereas primary adrenal failure causes hypovolemic hyponatremia. A cosyntropin stimulation test is necessary to assess for primary adrenal insufficiency.

Urine electrolytes and osmolality are crucial tests in the initial evaluation of hyponatremia. A urine Na^+ concentration $<20\text{--}30 \text{ mM}$ is consistent with hypovolemic hyponatremia, in the clinical absence of a hypervolemic, Na^+ -avid syndrome such as CHF (Fig. 49-5). In contrast, patients with SIAD will typically excrete urine with an Na^+ concentration that is $>30 \text{ mM}$. However, there can be substantial overlap in urine Na^+ concentration values in patients with SIAD and hypovolemic hyponatremia, particularly in the elderly; the ultimate “gold standard” for the diagnosis of hypovolemic hyponatremia is the demonstration that plasma Na^+ concentration corrects after hydration with normal saline. Patients with thiazide-associated hyponatremia may also present with higher than expected urine Na^+ concentration

and other findings suggestive of SIAD; one should defer making a diagnosis of SIAD in these patients until 1–2 weeks after discontinuing the thiazide. A urine osmolality $<100 \text{ mOsm/kg}$ is suggestive of polydipsia; urine osmolality $>400 \text{ mOsm/kg}$ indicates that AVP excess is playing a more dominant role, whereas intermediate values are more consistent with multifactorial pathophysiology (e.g., AVP excess with a significant component of polydipsia). Patients with hyponatremia due to decreased solute intake (beer potomania) typically have urine Na^+ concentration $<20 \text{ mM}$ and urine osmolality in the range of <100 to the low 200s. Finally, the measurement of urine K^+ concentration is required to calculate the urine-to-plasma electrolyte ratio, which is useful to predict the response to fluid restriction (see the section on treatment of hyponatremia below).

TREATMENT

Hyponatremia

Three major considerations guide the therapy of hyponatremia. First, the presence and/or severity of symptoms determine the urgency and goals of therapy. Patients with acute hyponatremia (Table 49-2) present with symptoms that can range from headache, nausea, and/or vomiting, to seizures, obtundation, and central herniation; patients with chronic hyponatremia, present for $>48 \text{ h}$, are less likely to have severe symptoms. Second, patients with chronic hyponatremia are at risk for ODS if plasma Na^+ concentration is corrected by $>8\text{--}10 \text{ mM}$ within the first 24 h and/or by $>18 \text{ mM}$ within the first 48 h. Third, the response to interventions such as hypertonic saline, isotonic saline, or AVP antagonists can be highly unpredictable, such that frequent monitoring of plasma Na^+ concentration during corrective therapy is imperative.

Once the urgency in correcting the plasma Na^+ concentration has been established and appropriate therapy instituted, the focus should be on treatment or withdrawal of the underlying cause. Patients with euvolemic hyponatremia due to SIAD, hypothyroidism, or secondary adrenal failure will respond to successful treatment of the underlying cause, with an increase in plasma Na^+ concentration. However, not all causes of SIAD are immediately reversible, necessitating pharmacologic therapy to increase the plasma Na^+ concentration (see below). Hypovolemic hyponatremia will respond to intravenous hydration with isotonic normal saline, with a rapid reduction in circulating AVP and a brisk water diuresis; it may be necessary to reduce the rate of correction if the history suggests that hyponatremia has been chronic, i.e., present for more than 48 h (see below). Hypervolemic hyponatremia due to CHF will often respond to improved therapy of the underlying cardiomyopathy, e.g., following the institution or intensification of angiotensin-converting enzyme (ACE) inhibition. Finally, patients with hyponatremia due to beer potomania and low solute intake will respond very rapidly to intravenous saline and the resumption of a normal diet. Notably, patients with beer potomania have a very high risk of developing ODS, due to the associated hypokalemia, alcoholism, malnutrition, and high risk of overcorrecting the plasma Na^+ concentration.

Water deprivation has long been a cornerstone of the therapy of chronic hyponatremia. However, patients who are excreting minimal electrolyte-free water will require aggressive fluid restriction; this can be very difficult for patients with SIAD to tolerate, given that their thirst is also inappropriately stimulated. The urine-to-plasma electrolyte ratio ($\text{urinary } [\text{Na}^+] + [\text{K}^+]/\text{plasma } [\text{Na}^+]$) can be exploited as a quick indicator of electrolyte-free water excretion (Table 49-3); patients with a ratio of >1 should be more aggressively restricted ($<500 \text{ mL/d}$), those with a ratio of ~ 1 should be restricted to $500\text{--}700 \text{ mL/d}$, and those with a ratio <1 should be restricted to $<1 \text{ L/d}$. In hypokalemic patients, potassium replacement will serve to increase plasma Na^+ concentration, given that the plasma Na^+ concentration is a functional of both exchangeable Na^+ and exchangeable K^+ divided by total-body water; a corollary is that aggressive repletion of K^+ has the potential to overcorrect the plasma

TABLE 49-3 Management of Hypernatremia**Water Deficit**

- Estimate total-body water (TBW): 50% of body weight in women and 60% in men
- Calculate free-water deficit: $[(\text{Na}^+ - 140)/140] \times \text{TBW}$
- Administer deficit over 48–72 h, without decrease in plasma Na^+ concentration by $>10 \text{ mM}/24 \text{ h}$

Ongoing Water Losses

- Calculate free-water clearance, $C_e \text{H}_2\text{O}$:

$$C_e \text{H}_2\text{O} = V \times \left(1 - \frac{U_{\text{Na}} + U_k}{P_{\text{Na}}} \right)$$

where V is urinary volume, U_{Na} is urinary $[\text{Na}^+]$, U_k is urinary $[\text{K}^+]$, and P_{Na} is plasma $[\text{Na}^+]$

Insensible Losses

- $\sim 10 \text{ mL/kg}$ per day: less if ventilated, more if febrile

Total

- Add components to determine water deficit and ongoing water loss; correct the water deficit over 48–72 h and replace daily water loss. Avoid correction of plasma $[\text{Na}^+]$ by $>10 \text{ mM/d}$.

Na^+ concentration even in the absence of hypertonic saline. Plasma Na^+ concentration will also tend to respond to an increase in dietary solute intake, which increases the ability to excrete free water; this can be accomplished with oral salt tablets and with newly available, palatable preparations of oral urea.

Patients in whom therapy with fluid restriction, potassium replacement, and/or increased solute intake fails may merit pharmacologic therapy to increase their plasma Na^+ concentration. Many patients with SIAD respond to combined therapy with oral furosemide, 20 mg twice a day (higher doses may be necessary in renal insufficiency), and oral salt tablets; furosemide serves to inhibit the renal countercurrent mechanism and blunt urinary concentrating ability, whereas the salt tablets counteract diuretic-associated natriuresis. Demeclocycline is a potent inhibitor of principal cells and can be used in patients whose Na levels do not increase in response to furosemide and salt tablets. However, this agent can be associated with a reduction in GFR, due to excessive natriuresis and/or direct renal toxicity; it should be avoided in cirrhotic patients in particular, who are at higher risk of nephrotoxicity due to drug accumulation. If available, palatable preparations of oral urea can also be used to manage SIAD; the increase in solute excretion with oral urea ingestion increases free water excretion, thus reducing the plasma Na^+ .

AVP antagonists (vaptans) are highly effective in SIAD and in hypervolemic hyponatremia due to heart failure or cirrhosis, reliably increasing plasma Na^+ concentration due to their “aquaretic” effects (augmentation of free water clearance). Most of these agents specifically antagonize the V_2 AVP receptor; tolvaptan is currently the only oral V_2 antagonist to be approved by the U.S. Food and Drug Administration. Conivaptan, the only available intravenous vaptan, is a mixed V_{1A}/V_2 antagonist, with a modest risk of hypotension due to V_{1A} receptor inhibition. Therapy with vaptans must be initiated in a hospital setting, with a liberalization of fluid restriction ($>2 \text{ L/d}$) and close monitoring of plasma Na^+ concentration. Although approved for the management of all but hypovolemic hyponatremia and acute hyponatremia, the clinical indications are limited. Oral tolvaptan is perhaps most appropriate for the management of significant and persistent SIAD (e.g., in small-cell lung carcinoma) that has not responded to water restriction and/or oral furosemide and salt tablets. Abnormalities in liver function tests have been reported with chronic tolvaptan therapy; hence, the use of this agent should be restricted to <1 –2 months.

Treatment of acute symptomatic hyponatremia should include hypertonic 3% saline (513 mM) to acutely increase plasma Na^+ concentration by 1 – 2 mM/h to a total of 4 – 6 mM ; this modest increase is typically sufficient to alleviate severe acute symptoms, after which

corrective guidelines for chronic hyponatremia are appropriate (see below). A number of equations have been developed to estimate the required rate of hypertonic saline, which has an Na^+-Cl^- concentration of 513 mM. The traditional approach is to calculate an Na^+ deficit, where the Na^+ deficit = $0.6 \times \text{body weight} \times (\text{target plasma } \text{Na}^+ \text{ concentration} - \text{starting plasma } \text{Na}^+ \text{ concentration})$, followed by a calculation of the required rate. Regardless of the method used to determine the rate of administration, the increase in plasma Na^+ concentration can be highly unpredictable during treatment with hypertonic saline, due to rapid changes in the underlying physiology; plasma Na^+ concentration should be monitored every 2–4 h during treatment, with appropriate changes in therapy based on the observed rate of change. The administration of supplemental oxygen and ventilatory support is also critical in acute hyponatremia, in the event that patients develop acute pulmonary edema or hypercapneic respiratory failure. Intravenous loop diuretics will help treat acute pulmonary edema and will also increase free water excretion, by interfering with the renal countercurrent multiplication system. AVP antagonists do *not* have an approved role in the management of acute hyponatremia.

The rate of correction should be comparatively slow in *chronic* hyponatremia (< 8 – 10 mM in the first 24 h and < 18 mM in the first 48 h), so as to avoid ODS; lower target rates are appropriate in patients at particular risk for ODS, such as alcoholics or hypokalemic patients. Overcorrection of the plasma Na^+ concentration can occur when AVP levels rapidly normalize, for example following the treatment of patients with chronic hypovolemic hyponatremia with intravenous saline or following glucocorticoid replacement of patients with hypopituitarism and secondary adrenal failure. Approximately 10% of patients treated with vaptans will overcorrect; the risk is increased if water intake is not liberalized. In the event that the plasma Na^+ concentration overcorrects following therapy, be it with hypertonic saline, isotonic saline, or a vaptan, hyponatremia can be safely reinduced or stabilized by the administration of the AVP agonist desmopressin acetate (DDAVP) and/or the administration of free water, typically intravenous D₅W; the goal is to prevent or reverse the development of ODS. Alternatively, the treatment of patients with marked hyponatremia can be initiated with the twice-daily administration of DDAVP to maintain constant AVP bioactivity, combined with the administration of hypertonic saline to slowly correct the serum sodium in a more controlled fashion, thus reducing upfront the risk of overcorrection.

HYPERNATREMIA

Etiology Hyponatremia is defined as an increase in the plasma Na^+ concentration to $>145 \text{ mM}$. Considerably less common than hyponatremia, hyponatremia is nonetheless associated with mortality rates of as high as 40–60%, mostly due to the severity of the associated underlying disease processes. Hyponatremia is usually the result of a combined water and electrolyte deficit, with losses of H_2O in excess of Na^+ . Less frequently, the ingestion or iatrogenic administration of excess Na^+ can be causative, for example after IV administration of excessive hypertonic Na^+-Cl^- or $\text{Na}^+-\text{HCO}_3^-$ (Fig. 49-6).

Elderly individuals with reduced thirst and/or diminished access to fluids are at the highest risk of developing hyponatremia. Patients with hyponatremia may rarely have a central defect in hypothalamic osmoreceptor function, with a mixture of both decreased thirst and reduced AVP secretion. Causes of this adipsic DI include primary or metastatic tumor, occlusion or ligation of the anterior communicating artery, trauma, hydrocephalus, and inflammation.

Hyponatremia can develop following the loss of water via both renal and nonrenal routes. Insensible losses of water may increase in the setting of fever, exercise, heat exposure, severe burns, or mechanical ventilation. Diarrhea is, in turn, the most common gastrointestinal cause of hyponatremia. Notably, osmotic diarrhea and viral gastroenteritis typically generate stools with Na^+ and $\text{K}^+ < 100 \text{ mM}$, thus leading to water loss and hyponatremia; in contrast, secretory diarrhea

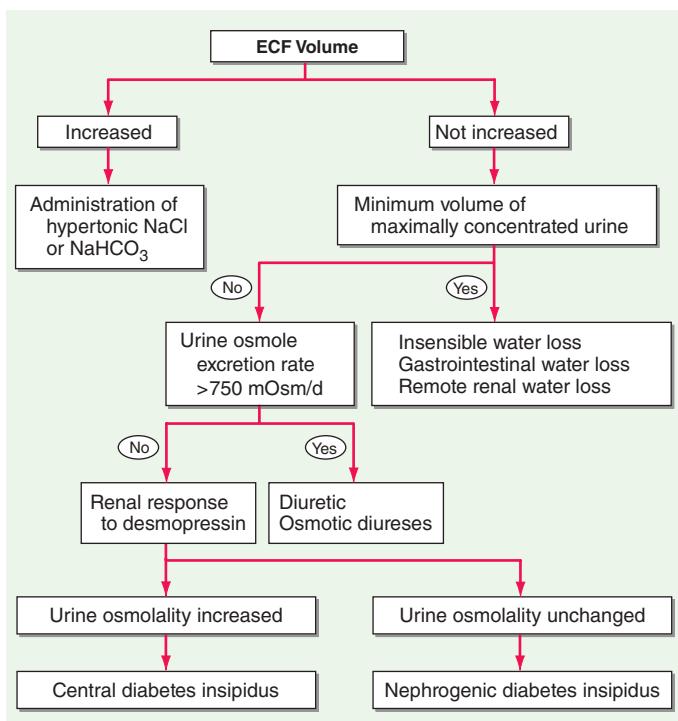


FIGURE 49-6 The diagnostic approach to hypernatremia. ECF, extracellular fluid.

typically results in isotonic stool and thus hypovolemia with or without hypovolemic hyponatremia.

Common causes of renal water loss include osmotic diuresis secondary to hyperglycemia, excess urea, postobstructive diuresis, or mannitol; these disorders share an increase in urinary solute excretion and urinary osmolality (see “Diagnostic Approach,” below). Hypernatremia due to a water diuresis occurs in central or nephrogenic DI (NDI).

NDI is characterized by renal resistance to AVP, which can be partial or complete (see “Diagnostic Approach,” below). Genetic causes include loss-of-function mutations in the X-linked V₂ receptor; mutations in the AVP-responsive aquaporin-2 water channel can cause autosomal recessive and autosomal dominant NDI, whereas recessive deficiency of the aquaporin-1 water channel causes a more modest concentrating defect (Fig. 49-2). Hypercalcemia can also cause polyuria and NDI; calcium signals directly through the calcium-sensing receptor to downregulate Na⁺, K⁺, and Cl⁻ transport by the TALH and water transport in principal cells, thus reducing renal concentrating ability in hypercalcemia. Another common acquired cause of NDI is hypokalemia, which inhibits the renal response to AVP and downregulates aquaporin-2 expression. Several drugs can cause acquired NDI, in particular lithium, ifosfamide, and several antiviral agents. Lithium causes NDI by multiple mechanisms, including direct inhibition of renal glycogen synthase kinase-3 (GSK3), a kinase thought to be the pharmacologic target of lithium in bipolar disease; GSK3 is required for the response of principal cells to AVP. The entry of lithium through the amiloride-sensitive Na⁺ channel ENaC (Fig. 49-4) is required for the effect of the drug on principal cells, such that combined therapy within lithium and amiloride can mitigate lithium-associated NDI. However, lithium causes chronic tubulointerstitial scarring and chronic kidney disease after prolonged therapy, such that patients may have a persistent NDI long after stopping the drug, with a reduced therapeutic benefit from amiloride.

Finally, gestational DI is a rare complication of late-term pregnancy wherein increased activity of a circulating placental protease with “vasopressinase” activity leads to reduced circulating AVP and polyuria, often accompanied by hypernatremia. DDAVP is an effective therapy for this syndrome, given its resistance to the vasopressinase enzyme.

Clinical Features Hypernatremia increases osmolality of the ECF, generating an osmotic gradient between the ECF and ICF, an efflux of intracellular water, and cellular shrinkage. As in hyponatremia, the symptoms of hypernatremia are predominantly neurologic. Altered mental status is the most frequent manifestation, ranging from mild confusion and lethargy to deep coma. The sudden shrinkage of brain cells in acute hypernatremia may lead to parenchymal or subarachnoid hemorrhages and/or subdural hematomas; however, these vascular complications are primarily encountered in pediatric and neonatal patients. Osmotic damage to muscle membranes can also lead to hypernatremic rhabdomyolysis. Brain cells accommodate to a chronic increase in ECF osmolality (>48 h) by activating membrane transporters that mediate influx and intracellular accumulation of organic osmolytes (creatinine, betaine, glutamate, myoinositol, and taurine); this results in an increase in ICF water and normalization of brain parenchymal volume. In consequence, patients with *chronic* hypernatremia are less likely to develop severe neurologic compromise. However, the cellular response to chronic hypernatremia predisposes these patients to the development of cerebral edema and seizures during overly rapid hydration (overcorrection of plasma Na⁺ concentration by >10 mM/d).

Diagnostic Approach The history should focus on the presence or absence of thirst, polyuria, and/or an extrarenal source for water loss, such as diarrhea. The physical examination should include a detailed neurologic exam and an assessment of the ECFV; patients with a particularly large water deficit and/or a combined deficit in electrolytes and water may be hypovolemic, with reduced JVP and orthostasis. Accurate documentation of daily fluid intake and daily urine output is also critical for the diagnosis and management of hypernatremia.

Laboratory investigation should include a measurement of serum and urine osmolality, in addition to urine electrolytes. The appropriate response to hypernatremia and a serum osmolality >295 mOsm/kg is an increase in circulating AVP and the excretion of low volumes (<500 mL/d) of maximally concentrated urine, i.e., urine with osmolality >800 mOsm/kg; should this be the case, then an extrarenal source of water loss is primarily responsible for the generation of hypernatremia. Many patients with hypernatremia are polyuric; should an osmotic diuresis be responsible, with excessive excretion of Na⁺-Cl⁻, glucose, and/or urea, then daily solute excretion will be >750–1000 mOsm/d (>15 mOsm/kg body water per day) (Fig. 49-6). More commonly, patients with hypernatremia and polyuria will have a predominant water diuresis, with excessive excretion of hypotonic, dilute urine.

Adequate differentiation between nephrogenic and central causes of DI requires the measurement of the response in urinary osmolality to DDAVP, combined with measurement of circulating AVP in the setting of hypertonicity. By definition, patients with baseline hypernatremia are hypertonic, with an adequate stimulus for AVP by the posterior pituitary. Therefore, in contrast to polyuric patients with a normal or reduced baseline plasma Na⁺ concentration and osmolality, a water deprivation test (Chap. 48) is unnecessary in hypernatremia; indeed, water deprivation is absolutely contraindicated in this setting, given the risk for worsening the hypernatremia. Patients with NDI will fail to respond to DDAVP, with a urine osmolality that increases by <50% or <150 mOsm/kg from baseline, in combination with a normal or high circulating AVP level; patients with central DI will respond to DDAVP, with a reduced circulating AVP. Patients may exhibit a partial response to DDAVP, with a >50% rise in urine osmolality that nonetheless fails to reach 800 mOsm/kg; the level of circulating AVP will help differentiate the underlying cause, i.e., NDI versus central DI. In pregnant patients, AVP assays should be drawn in tubes containing the protease inhibitor 1,10-phenanthroline, to prevent in vitro degradation of AVP by placental vasopressinase.

For patients with hypernatremia due to renal loss of water, it is critical to quantify *ongoing* daily losses using the calculated electrolyte-free water clearance, in addition to calculation of the baseline water deficit (the relevant formulas are discussed in Table 49-3). This requires daily measurement of urine electrolytes, combined with accurate measurement of daily urine volume.

Hypernatremia

The underlying cause of hypernatremia should be withdrawn or corrected, be it drugs, hyperglycemia, hypercalcemia, hypokalemia, or diarrhea. The approach to the correction of hypernatremia is outlined in Table 49-3. It is imperative to correct hypernatremia slowly to avoid cerebral edema, typically replacing the calculated free water deficit over 48 h. Notably, the plasma Na^+ concentration should be corrected by no $>10 \text{ m}\bar{\text{M}}/\text{d}$, which may take longer than 48 h in patients with severe hypernatremia ($>160 \text{ mM}$). A rare exception is patients with acute hypernatremia ($<48 \text{ h}$) due to sodium loading, who can safely be corrected rapidly at a rate of 1 mM/h .

Water should ideally be administered by mouth or by nasogastric tube, as the most direct way to provide free water, i.e., water without electrolytes. Alternatively, patients can receive free water in dextrose-containing IV solutions, such as 5% dextrose ($D_5\text{W}$); blood glucose should be monitored in case hyperglycemia occurs. Depending on the history, blood pressure, or clinical volume status, it may be appropriate to initially treat with hypotonic saline solutions (1/4 or 1/2 normal saline); normal saline is usually inappropriate in the absence of very severe hypernatremia, where normal saline is proportionally more hypotonic relative to plasma, or frank hypotension. Calculation of urinary electrolyte-free water clearance (Table 49-3) is required to estimate daily, ongoing loss of free water in patients with NDI or central DI, which should be replenished daily.

Additional therapy may be feasible in specific cases. Patients with central DI should respond to the administration of intravenous, intranasal, or oral DDAVP. Patients with NDI due to lithium may reduce their polyuria with amiloride (2.5–10 mg/d), which decreases entry of lithium into principal cells by inhibiting ENaC (see above); in practice, however, most patients with lithium-associated DI are able to compensate for their polyuria by simply increasing their daily water intake. Thiazides may reduce polyuria due to NDI, ostensibly by inducing hypovolemia and increasing proximal tubular water reabsorption. Occasionally, nonsteroidal anti-inflammatory drugs (NSAIDs) have been used to treat polyuria associated with NDI, reducing the negative effect of intrarenal prostaglandins on urinary concentrating mechanisms; however, this assumes the risks of NSAID-associated gastric and/or renal toxicity. Furthermore, it must be emphasized that thiazides, amiloride, and NSAIDs are only appropriate for *chronic* management of polyuria from NDI and have *no* role in the acute management of associated hypernatremia, where the focus is on replacing free water deficits and ongoing free water loss.

POTASSIUM DISORDERS

Homeostatic mechanisms maintain plasma K^+ concentration between 3.5 and 5.0 mM, despite marked variation in dietary K^+ intake. In a healthy individual at steady state, the entire daily intake of potassium is excreted, ~90% in the urine and 10% in the stool; thus, the kidney plays a dominant role in potassium homeostasis. However, >98% of total-body potassium is intracellular, chiefly in muscle; buffering of extracellular K^+ by this large intracellular pool plays a crucial role in the regulation of plasma K^+ concentration. Changes in the exchange and distribution of intra- and extracellular K^+ can thus lead to marked hypo- or hyperkalemia. A corollary is that massive necrosis and the attendant release of tissue K^+ can cause severe hyperkalemia, particularly in the setting of acute kidney injury and reduced excretion of K^+ .

Changes in whole-body K^+ content are primarily mediated by the kidney, which *reabsorbs* filtered K^+ in hypokalemic, K^+ -deficient states and *secretes* K^+ in hyperkalemic, K^+ -replete states. Although K^+ is transported along the entire nephron, it is the principal cells of the connecting segment (CNT) and cortical CD that play a dominant role in renal K^+ secretion, whereas alpha-intercalated cells of the outer medullary CD function in renal tubular reabsorption of filtered K^+ in K^+ -deficient states. In principal cells, apical Na^+ entry via the amiloride-sensitive ENaC generates a lumen-negative potential difference, which drives

passive K^+ exit through apical K^+ channels (Fig. 49-4). Two major K^+ channels mediate distal tubular K^+ secretion: the secretory K^+ channel ROMK (renal outer medullary K^+ channel; also known as Kir1.1 or KcnJ1) and the flow-sensitive “big potassium” (BK) or maxi- K^+ channel. ROMK is thought to mediate the bulk of constitutive K^+ secretion, whereas increases in distal flow rate and/or genetic absence of ROMK activate K^+ secretion via the BK channel.

An appreciation of the relationship between ENaC-dependent Na^+ entry and distal K^+ secretion (Fig. 49-4) is required for the bedside interpretation of potassium disorders. For example, decreased distal delivery of Na^+ , as occurs in hypovolemic, prerenal states, tends to blunt the ability to excrete K^+ , leading to hyperkalemia; on the other hand, an *increase* in distal delivery of Na^+ and distal flow rate, as occurs after treatment with thiazide and loop diuretics, can enhance K^+ secretion and lead to hypokalemia. Hyperkalemia is also a predictable consequence of drugs that directly inhibit ENaC, due to the role of this Na^+ channel in generating a lumen-negative potential difference. Aldosterone in turn has a major influence on potassium excretion, increasing the activity of ENaC channels and thus amplifying the driving force for K^+ secretion across the luminal membrane of principal cells. Abnormalities in the renin-angiotensin-aldosterone system can thus cause both hypokalemia and hyperkalemia. Notably, however, potassium excess and potassium restriction have opposing, aldosterone-independent effects on the density and activity of apical K^+ channels in the distal nephron, i.e., factors other than aldosterone modulate the renal capacity to secrete K^+ . In addition, potassium restriction and hypokalemia activates aldosterone-independent distal *reabsorption* of filtered K^+ , activating apical H^+/K^+ -ATPase activity in intercalated cells within the outer medullary CD. Reflective perhaps of this physiology, changes in plasma K^+ concentration are not universal in disorders associated with changes in aldosterone activity.

HYPOKALEMIA

Hypokalemia, defined as a plasma K^+ concentration of $<3.5 \text{ mM}$, occurs in up to 20% of hospitalized patients. Hypokalemia is associated with a tenfold increase in in-hospital mortality, due to adverse effects on cardiac rhythm, blood pressure, and cardiovascular morbidity. Mechanistically, hypokalemia can be caused by redistribution of K^+ between tissues and the ECF or by renal and nonrenal loss of K^+ (Table 49-4). Systemic hypomagnesemia can also cause treatment-resistant hypokalemia, due to a combination of reduced cellular uptake of K^+ and exaggerated renal secretion. Spurious hypokalemia or “pseudohypokalemia” can occasionally result from in vitro cellular uptake of K^+ after venipuncture, for example, due to profound leukocytosis in acute leukemia.

Redistribution and Hypokalemia Insulin, β_2 -adrenergic activity, thyroid hormone, and alkalosis promote Na^+/K^+ -ATPase-mediated cellular uptake of K^+ , leading to hypokalemia. Inhibition of the passive efflux of K^+ can also cause hypokalemia, albeit rarely; this typically occurs in the setting of systemic inhibition of K^+ channels by toxic barium ions. Exogenous insulin can cause iatrogenic hypokalemia, particularly during the management of K^+ -deficient states such as diabetic ketoacidosis. Alternatively, the stimulation of endogenous insulin can provoke hypokalemia, hypomagnesemia, and/or hypophosphatemia in malnourished patients given a carbohydrate load. Alterations in the activity of the endogenous sympathetic nervous system can cause hypokalemia in several settings, including alcohol withdrawal, hyperthyroidism, acute myocardial infarction, and severe head injury. β_2 agonists, including both bronchodilators and tocolytics (ritodrine), are powerful activators of cellular K^+ uptake; “hidden” sympathomimetics, such as pseudoephedrine and ephedrine in cough syrup or dieting agents, may also cause unexpected hypokalemia. Finally, xanthine-dependent activation of cAMP-dependent signaling, downstream of the β_2 receptor, can lead to hypokalemia, usually in the setting of overdose (theophylline) or marked overingestion (dietary caffeine).

Redistributive hypokalemia can also occur in the setting of hyperthyroidism, with periodic attacks of hypokalemic paralysis (thyrotoxic periodic paralysis [TPP]). Similar episodes of hypokalemic weakness in the absence of thyroid abnormalities occur in *familial* hypokalemic

TABLE 49-4 Causes of Hypokalemia

I.	Decreased intake
A.	Starvation
B.	Clay ingestion
II.	Redistribution into cells
A.	Acid-base
1.	Metabolic alkalosis
B.	Hormonal
1.	Insulin
2.	Increased β_2 -adrenergic sympathetic activity: post-myocardial infarction, head injury
3.	β_2 -Adrenergic agonists—bronchodilators, tocolytics
4.	α -Adrenergic antagonists
5.	Thyrotoxic periodic paralysis
6.	Downstream stimulation of Na^+/K^+ -ATPase: theophylline, caffeine
C.	Anabolic state
1.	Vitamin B_{12} or folic acid administration (red blood cell production)
2.	Granulocyte-macrophage colony-stimulating factor (white blood cell production)
3.	Total parenteral nutrition
D.	Other
1.	Pseudohypokalemia
2.	Hypothermia
3.	Familial hypokalemic periodic paralysis
4.	Barium toxicity: systemic inhibition of “leak” K^+ channels
III.	Increased loss
A.	Nonrenal
1.	Gastrointestinal loss (diarrhea)
2.	Integumentary loss (sweat)
B.	Renal
1.	Increased distal flow and distal Na^+ delivery: diuretics, osmotic diuresis, salt-wasting nephropathies
2.	Increased secretion of potassium
a.	Mineralocorticoid excess: primary hyperaldosteronism (aldosterone-producing adenomas, primary or unilateral adrenal hyperplasia, idiopathic hyperaldosteronism due to bilateral adrenal hyperplasia, and adrenal carcinoma), genetic hyperaldosteronism (familial hyperaldosteronism types I/II/III, congenital adrenal hyperplasias), secondary hyperaldosteronism (malignant hypertension, renin-secreting tumors, renal artery stenosis, hypovolemia), Cushing's syndrome, Bartter's syndrome, Gitelman's syndrome
b.	Apparent mineralocorticoid excess: genetic deficiency of 11β -dehydrogenase-2 (syndrome of apparent mineralocorticoid excess), inhibition of 11β -dehydrogenase-2 (glycyrrhetic acid and/or carbenoxolone; licorice, food products, drugs), Liddle's syndrome (genetic activation of epithelial Na^+ channels)
c.	Distal delivery of nonreabsorbed anions: vomiting, nasogastric suction, proximal renal tubular acidosis, diabetic ketoacidosis, glue-sniffing (toluene abuse), penicillin derivatives (penicillin, nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin)
3.	Magnesium deficiency

periodic paralysis, usually caused by missense mutations of voltage sensor domains within the α_1 subunit of L-type calcium channels or the skeletal Na^+ channel; these mutations generate an abnormal gating pore current activated by hyperpolarization. TPP develops more frequently in patients of Asian or Hispanic origin; this shared predisposition has been linked to genetic variation in Kir2.6, a muscle-specific, thyroid hormone-responsive K^+ channel. Patients with TPP typically present with weakness of the extremities and limb girdles, with paralytic episodes that occur most frequently between 1 and 6 A.M. Signs and symptoms of hyperthyroidism are not invariably present. Hypokalemia is usually profound and almost invariably accompanied by hypophosphatemia and hypomagnesemia. The hypokalemia in TPP is also attributed to both direct and indirect activation of the

Na^+/K^+ -ATPase, resulting in increased uptake of K^+ by muscle and other tissues. Increases in β -adrenergic activity play an important role in that high-dose propranolol (3 mg/kg) rapidly reverses the associated hypokalemia, hypophosphatemia, and paralysis.

Nonrenal Loss of Potassium The loss of K^+ in sweat is typically low, except under extremes of physical exertion. Direct gastric losses of K^+ due to vomiting or nasogastric suctioning are also minimal; however, the ensuing hypochloremic alkalosis results in persistent kaliuresis due to secondary hyperaldosteronism and bicarbonaturia, i.e., a *renal* loss of K^+ . Diarrhea is a globally important cause of hypokalemia, given the worldwide prevalence of infectious diarrheal disease. Noninfectious gastrointestinal processes such as celiac disease, ileostomy, villous adenomas, inflammatory bowel disease, colonic pseudo-obstruction (Ogilvie's syndrome), VIPomas, and chronic laxative abuse can also cause significant hypokalemia; an exaggerated intestinal secretion of potassium by upregulated colonic BK channels has been directly implicated in the pathogenesis of hypokalemia in many of these disorders.

Renal Loss of Potassium Drugs can increase renal K^+ excretion by a variety of different mechanisms. Diuretics are a particularly common cause, due to associated increases in distal tubular Na^+ delivery and distal tubular flow rate, in addition to secondary hyperaldosteronism. Thiazides have a greater effect on plasma K^+ concentration than loop diuretics, despite their lesser natriuretic effect. The diuretic effect of thiazides is largely due to inhibition of the Na^+-Cl^- cotransporter NCC in DCT cells. This leads to a direct increase in the delivery of luminal Na^+ to the principal cells immediately downstream in the CNT and cortical CD, which augments Na^+ entry via ENaC, increases the lumen-negative potential difference, and amplifies K^+ secretion. The higher propensity of thiazides to cause hypokalemia may also be secondary to thiazide-associated hypocalcioruria, versus the *hypercalcioruria* seen with loop diuretics; the increases in downstream luminal calcium in response to loop diuretics inhibit ENaC in principal cells, thus reducing the lumen-negative potential difference and attenuating distal K^+ excretion. High doses of penicillin-related antibiotics (nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin) can increase obligatory K^+ excretion by acting as nonreabsorbable anions in the distal nephron. Finally, several renal tubular toxins cause renal K^+ and magnesium wasting, leading to hypokalemia and hypomagnesemia; these drugs include aminoglycosides, amphotericin, fosfomycin, cisplatin, and ifosfamide (see also “Magnesium Deficiency and Hypokalemia,” below).

Aldosterone activates the ENaC channel in principal cells via multiple synergistic mechanisms, thus increasing the driving force for K^+ excretion. In consequence, increases in aldosterone bioactivity and/or gains in function of aldosterone-dependent signaling pathways are associated with hypokalemia. Increases in circulating aldosterone (hyperaldosteronism) may be primary or secondary. Increased levels of circulating renin in secondary forms of hyperaldosteronism lead to increased angiotensin II and thus aldosterone; renal artery stenosis is perhaps the most frequent cause (Table 49-4). Primary hyperaldosteronism may be genetic or acquired. Hypertension and hypokalemia, due to increases in circulating 11-deoxycorticosterone, occur in patients with congenital adrenal hyperplasia caused by defects in either steroid 11β -hydroxylase or steroid 17α -hydroxylase; deficient 11β -hydroxylase results in associated virilization and other signs of androgen excess, whereas reduced sex steroids in 17α -hydroxylase deficiency lead to hypogonadism.

The major forms of *isolated* primary genetic hyperaldosteronism are familial hyperaldosteronism type I (FH-I, also known as glucocorticoid-remediable hyperaldosteronism [GRA]) and familial hyperaldosteronism types II and III (FH-II and FH-III), in which aldosterone production is not repressible by exogenous glucocorticoids. FH-I is caused by a chimeric gene duplication between the homologous 11β -hydroxylase (*CYP11B1*) and aldosterone synthase (*CYP11B2*) genes, fusing the adrenocorticotrophic hormone (ACTH)-responsive 11β -hydroxylase promoter to the coding region of aldosterone synthase; this chimeric gene is under the control of ACTH and thus repressible by glucocorticoids. FH-III is caused by mutations in the *KCNJ5* gene, which encodes

the G-protein-activated inward rectifier K⁺ channel 4 (GIRK4); these mutations lead to the acquisition of sodium permeability in the mutant GIRK4 channels, causing an exaggerated membrane depolarization in adrenal glomerulosa cells and the activation of voltage-gated calcium channels. The resulting calcium influx is sufficient to produce aldosterone secretion and cell proliferation, leading to adrenal adenomas and hyperaldosteronism.

Acquired causes of primary hyperaldosteronism include aldosterone-producing adenomas (APAs), primary or unilateral adrenal hyperplasia (PAH), idiopathic hyperaldosteronism (IHA) due to bilateral adrenal hyperplasia, and adrenal carcinoma; APA and IHA account for close to 60 and 40%, respectively, of diagnosed hyperaldosteronism. Acquired somatic mutations in KCNJ5 or less frequently in the ATP1A1 (an Na⁺/K⁺ ATPase α subunit) and ATP2B3 (a Ca²⁺ ATPase) genes can be detected in APAs; as in FH-III (see above), the exaggerated depolarization of adrenal glomerulosa cells caused by these mutations is implicated in the excessive adrenal proliferation and the exaggerated release of aldosterone.

Random testing of plasma renin activity (PRA) and aldosterone is a helpful screening tool in hypokalemic and/or hypertensive patients, with an aldosterone:PRA ratio of >50 suggestive of primary hyperaldosteronism. Hypokalemia and multiple antihypertensive drugs may alter the aldosterone:PRA ratio by suppressing aldosterone or increasing PRA, leading to a ratio of <50 in patients who do in fact have primary hyperaldosteronism; therefore, the clinical context should always be considered when interpreting these results.

The glucocorticoid cortisol has equal affinity for the MLR to that of aldosterone, with resultant "mineralocorticoid-like" activity. However, cells in the aldosterone-sensitive distal nephron are protected from this "illicit" activation by the enzyme 11 β -hydroxysteroid dehydrogenase-2 (11 β HSD-2), which converts cortisol to cortisone; cortisone has minimal affinity for the MLR. Recessive loss-of-function mutations in the 11 β HSD-2 gene are thus associated with cortisol-dependent activation of the MLR and the syndrome of apparent mineralocorticoid excess (SAME), encompassing hypertension, hypokalemia, hypercalcemia, and metabolic alkalosis, with suppressed PRA and suppressed aldosterone. A similar syndrome is caused by biochemical inhibition of 11 β HSD-2 by glycyrrhetic/glycyrrhizic acid and/or carbenoxolone. Glycyrrhetic acid is a natural sweetener found in licorice root, typically encountered in licorice and its many guises or as a flavoring agent in tobacco and food products.

Finally, hypokalemia may also occur with systemic increases in glucocorticoids. In Cushing's syndrome caused by increases in pituitary ACTH (Chap. 379), the incidence of hypokalemia is only 10%, whereas it is 60–100% in patients with ectopic secretion of ACTH, despite a similar incidence of hypertension. Indirect evidence suggests that the activity of renal 11 β HSD-2 is reduced in patients with ectopic ACTH compared with Cushing's syndrome, resulting in SAME.

Finally, defects in multiple renal tubular transport pathways are associated with hypokalemia. For example, loss-of-function mutations in subunits of the acidifying H⁺-ATPase in alpha-intercalated cells cause hypokalemic distal renal tubular acidosis, as do many acquired disorders of the distal nephron. Liddle's syndrome is caused by autosomal dominant gain-in-function mutations of ENaC subunits. Disease-associated mutations either activate the channel directly or abrogate aldosterone-inhibited retrieval of ENaC subunits from the plasma membrane; the end result is increased expression of activated ENaC channels at the plasma membrane of principal cells. Patients with Liddle's syndrome classically manifest severe hypertension with hypokalemia, unresponsive to spironolactone yet sensitive to amiloride. Hypertension and hypokalemia are, however, variable aspects of the Liddle's phenotype; more consistent features include a blunted aldosterone response to ACTH and reduced urinary aldosterone excretion.

Loss of the transport functions of the TALH and DCT nephron segments causes hereditary hypokalemic alkalosis, Bartter's syndrome (BS) and Gitelman's syndrome (GS), respectively. Patients with classic BS typically suffer from polyuria and polydipsia, due to the reduction in renal concentrating ability. They may have an increase in urinary

calcium excretion, and 20% are hypomagnesemic. Other features include marked activation of the renin-angiotensin-aldosterone axis. Patients with antenatal BS suffer from a severe systemic disorder characterized by marked electrolyte wasting, polyhydramnios, and hypercalciuria with nephrocalcinosis; renal prostaglandin synthesis and excretion are significantly increased, accounting for much of the systemic symptoms. There are five disease genes for BS, all of them functioning in some aspect of regulated Na⁺, K⁺, and Cl⁻ transport by the TALH. In contrast, GS is genetically homogeneous, caused almost exclusively by loss-of-function mutations in the thiazide-sensitive Na⁺-Cl⁻ cotransporter of the DCT. Patients with GS are uniformly hypomagnesemic and exhibit marked hypocalcemia, rather than the hypercalcemia typically seen in BS; urinary calcium excretion is thus a critical diagnostic test in GS. GS is a milder phenotype than BS; however, patients with GS may suffer from chondrocalcinosis, an abnormal deposition of calcium pyrophosphate dihydrate (CPPD) in joint cartilage (Chap. 309).

Magnesium Deficiency and Hypokalemia Magnesium depletion has inhibitory effects on muscle Na⁺/K⁺-ATPase activity, reducing influx into muscle cells and causing a secondary kaliuresis. In addition, magnesium depletion causes exaggerated K⁺ secretion by the distal nephron; this effect is attributed to a reduction in the magnesium-dependent, intracellular block of K⁺ efflux through the secretory K⁺ channel of principal cells (ROMK; Fig. 49-4). In consequence, hypomagnesemic patients are clinically refractory to K⁺ replacement in the absence of Mg²⁺ repletion. Notably, magnesium deficiency is also a common concomitant of hypokalemia because many disorders of the distal nephron may cause both potassium and magnesium wasting (Chap. 309).

Clinical Features Hypokalemia has prominent effects on cardiac, skeletal, and intestinal muscle cells. In particular, hypokalemia is a major risk factor for both ventricular and atrial arrhythmias. Hypokalemia predisposes to digoxin toxicity by a number of mechanisms, including reduced competition between K⁺ and digoxin for shared binding sites on cardiac Na⁺/K⁺-ATPase subunits. Electrocardiographic changes in hypokalemia include broad flat T waves, ST depression, and QT prolongation; these are most marked when serum K⁺ is <2.7 mmol/L. Hypokalemia can thus be an important precipitant of arrhythmia in patients with additional genetic or acquired causes of QT prolongation. Hypokalemia also results in hyperpolarization of skeletal muscle, thus impairing the capacity to depolarize and contract; weakness and even paralysis may ensue. It also causes a skeletal myopathy and predisposes to rhabdomyolysis. Finally, the paralytic effects of hypokalemia on intestinal smooth muscle may cause intestinal ileus.

The functional effects of hypokalemia on the kidney can include Na⁺-Cl⁻ and HCO₃⁻ retention, polyuria, phosphaturia, hypocitraturia, and an activation of renal ammoniagenesis. Bicarbonate retention and other acid-base effects of hypokalemia can contribute to the generation of metabolic alkalosis. Hypokalemic polyuria is due to a combination of central polydipsia and an AVP-resistant renal concentrating defect. Structural changes in the kidney due to hypokalemia include a relatively specific vacuolizing injury to proximal tubular cells, interstitial nephritis, and renal cysts. Hypokalemia also predisposes to acute kidney injury and can lead to end-stage renal disease (ESRD) in patients with long-standing hypokalemia due to eating disorders and/or laxative abuse.

Hypokalemia and/or reduced dietary K⁺ are implicated in the pathophysiology and progression of hypertension, heart failure, and stroke. For example, short-term K⁺ restriction in healthy humans and patients with essential hypertension induces Na⁺-Cl⁻ retention and hypertension. Correction of hypokalemia is particularly important in hypertensive patients treated with diuretics, in whom blood pressure improves with potassium supplementation and the establishment of normokalemia.

Diagnostic Approach The cause of hypokalemia is usually evident from history, physical examination, and/or basic laboratory tests. The history should focus on medications (e.g., laxatives, diuretics,

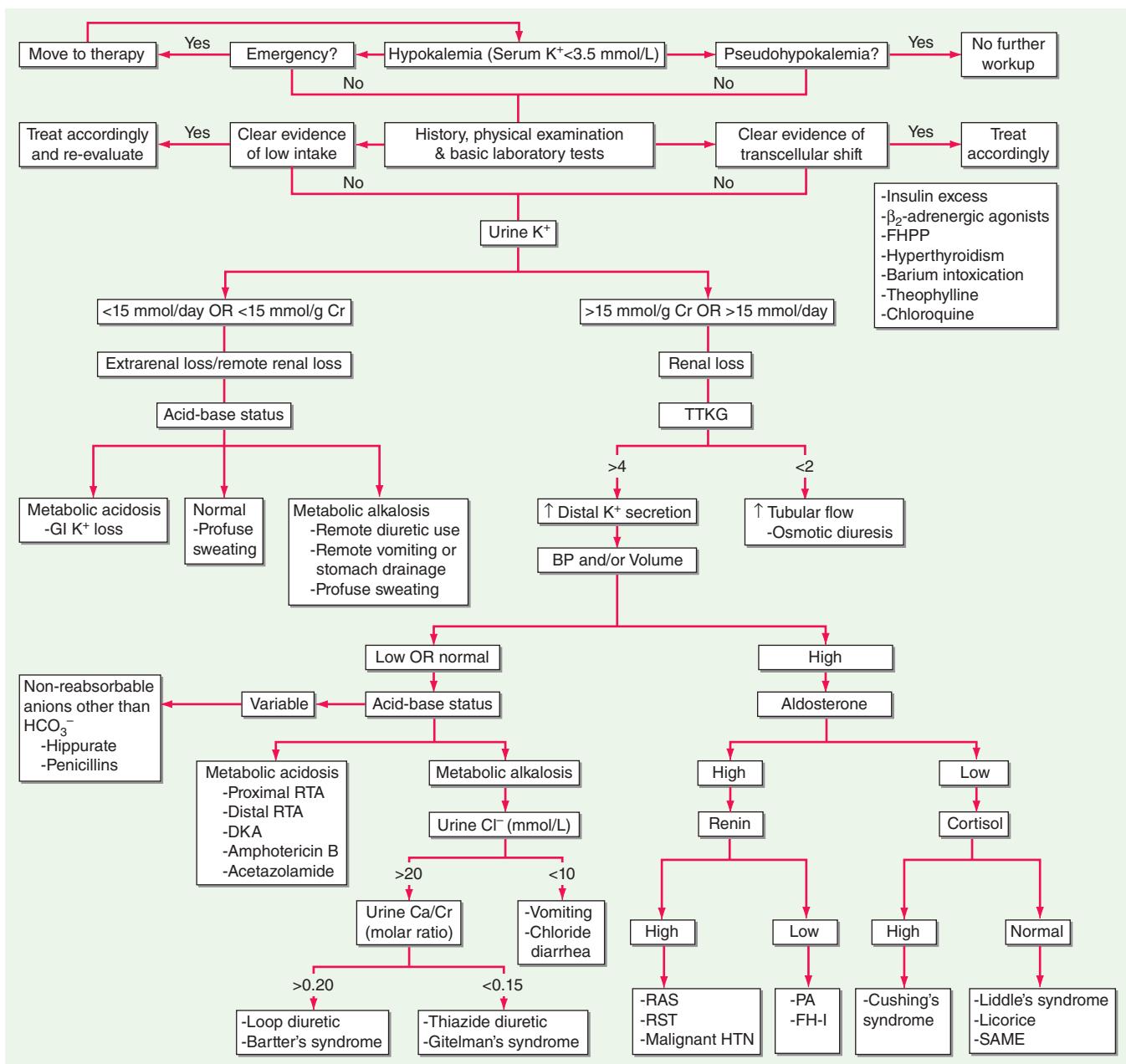


FIGURE 49-7 The diagnostic approach to hypokalemia. See text for details. AME, apparent mineralocorticoid excess; BP, blood pressure; CCD, cortical collecting duct; DKA, diabetic ketoacidosis; FH-I, familial hyperaldosteronism type I; FHPP, familial hypokalemic periodic paralysis; GI, gastrointestinal; GRA, glucocorticoid remivable aldosteronism; HTN, hypertension; PA, primary aldosteronism; RAS, renal artery stenosis; RST, renin-secreting tumor; RTA, renal tubular acidosis; SAME, syndrome of apparent mineralocorticoid excess; TTKG, transtubular potassium gradient. (Used with permission from DB Mount, K Zandi-Nejad K: Disorders of potassium balance, in Brenner and Rector's The Kidney, 8th ed, BM Brenner [ed]. Philadelphia, W.B. Saunders & Company, 2008, pp 547-587.)

antibiotics), diet and dietary habits (e.g., licorice), and/or symptoms that suggest a particular cause (e.g., periodic weakness, diarrhea). The physical examination should pay particular attention to blood pressure, volume status, and signs suggestive of specific hypokalemic disorders, e.g., hyperthyroidism and Cushing's syndrome. Initial laboratory evaluation should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+} , Ca^{2+} , a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes (Fig. 49-7). The presence of a non-anion gap acidosis suggests a distal, hypokalemic renal tubular acidosis or diarrhea; calculation of the urinary anion gap can help differentiate these two diagnoses. Renal K^+ excretion can be assessed with a 24-h urine collection; a 24-h K^+ excretion of <15 mmol is indicative of an extrarenal cause of hypokalemia (Fig. 49-7). If only a random, spot urine sample is available, serum and urine osmolality can be used to calculate the transtubular K^+ gradient (TTKG), which should be <3 in the presence of hypokalemia (see also "Hyperkalemia"). Alternatively, a urinary K^+ -to-creatinine ratio of >13 mmol/g creatinine (>1.5 mmol/mmol creatinine) is compatible

with excessive renal K^+ excretion. Urine Cl^- is usually decreased in patients with hypokalemia from a nonreabsorbable anion, such as antibiotics or HCO_3^- . The most common causes of chronic hypokalemic alkalosis are surreptitious vomiting, diuretic abuse, and GS; these can be distinguished by the pattern of urinary electrolytes. Hypokalemic patients with vomiting due to bulimia will thus typically have a urinary $Cl^- < 10$ mmol/L; urine Na^+ , K^+ , and Cl^- are persistently elevated in GS, due to loss of function in the thiazide-sensitive Na^+-Cl^- cotransporter, but less elevated in diuretic abuse and with greater variability. Urine diuretic screens for loop diuretics and thiazides may be necessary to further exclude diuretic abuse.

Other tests, such as urinary Ca^{2+} , thyroid function tests, and/or PRA and aldosterone levels, may also be appropriate in specific cases. A plasma aldosterone:PRA ratio of >50 , due to suppression of circulating renin and an elevation of circulating aldosterone, is suggestive of hyperaldosteronism. Patients with hyperaldosteronism or apparent mineralocorticoid excess may require further testing, for example

adrenal vein sampling (**Chap. 379**) or the clinically available testing for specific genetic causes (e.g., FH-I, SAME, Liddle's syndrome). Patients with primary aldosteronism should thus be tested for the chimeric FH-I/GRA gene (see above) if they are younger than 20 years of age or have a family history of primary aldosteronism or stroke at a young age (<40 years). Preliminary differentiation of Liddle's syndrome due to mutant ENaC channels from SAME due to mutant 11 β HSD-2 (see above), both of which cause hypokalemia and hypertension with aldosterone suppression, can be made on a clinical basis and then confirmed by genetic analysis; patients with Liddle's syndrome should respond to amiloride (ENaC inhibition) but not spironolactone, whereas patients with SAME will respond to spironolactone.

TREATMENT

Hypokalemia

The goals of therapy in hypokalemia are to prevent life-threatening and/or serious chronic consequences, to replace the associated K⁺ deficit, and to correct the underlying cause and/or mitigate future hypokalemia. The urgency of therapy depends on the severity of hypokalemia, associated clinical factors (e.g., cardiac disease, digoxin therapy), and the rate of decline in serum K⁺. Patients with a prolonged QT interval and/or other risk factors for arrhythmia should be monitored by continuous cardiac telemetry during repletion. Urgent but cautious K⁺ replacement should be considered in patients with severe redistributive hypokalemia (plasma K⁺ concentration <2.5 mM) and/or when serious complications ensue; however, this approach has a risk of rebound hyperkalemia following acute resolution of the underlying cause. When excessive activity of the sympathetic nervous system is thought to play a dominant role in redistributive hypokalemia, as in TPP, theophylline overdose, and acute head injury, high-dose propranolol (3 mg/kg) should be considered; this nonspecific β -adrenergic blocker will correct hypokalemia without the risk of rebound hyperkalemia.

Oral replacement with K⁺-Cl⁻ is the mainstay of therapy in hypokalemia. Potassium phosphate, oral or IV, may be appropriate in patients with combined hypokalemia and hypophosphatemia. Potassium bicarbonate or potassium citrate should be considered in patients with concomitant metabolic acidosis. Notably, hypomagnesemic patients are refractory to K⁺ replacement alone, such that concomitant Mg²⁺ deficiency should *always* be corrected with oral or intravenous repletion. The deficit of K⁺ and the rate of correction should be estimated as accurately as possible; renal function, medications, and comorbid conditions such as diabetes should also be considered, so as to gauge the risk of overcorrection. In the absence of abnormal K⁺ redistribution, the total deficit correlates with serum K⁺, such that serum K⁺ drops by ~0.27 mM for every 100-mmol reduction in total-body stores; loss of 400–800 mmol of total-body K⁺ results in a reduction in serum K⁺ by ~2.0 mM. Notably, given the delay in redistributing potassium into intracellular compartments, this deficit must be replaced gradually over 24–48 h, with frequent monitoring of plasma K⁺ concentration to avoid transient overrepletion and transient hyperkalemia.

The use of intravenous administration should be limited to patients unable to use the enteral route or in the setting of severe complications (e.g., paralysis, arrhythmia). Intravenous K⁺-Cl⁻ should always be administered in saline solutions, rather than dextrose, because the dextrose-induced increase in insulin can acutely exacerbate hypokalemia. The peripheral intravenous dose is usually 20–40 mmol of K⁺-Cl⁻ per liter; higher concentrations can cause localized pain from chemical phlebitis, irritation, and sclerosis. If hypokalemia is severe (<2.5 mmol/L) and/or critically symptomatic, intravenous K⁺-Cl⁻ can be administered through a central vein with cardiac monitoring in an intensive care setting, at rates of 10–20 mmol/h; higher rates should be reserved for acutely life-threatening complications. The absolute amount of administered K⁺ should be restricted (e.g., 20 mmol in 100 mL of saline solution) to prevent inadvertent infusion of a large dose. Femoral veins are

preferable, because infusion through internal jugular or subclavian central lines can acutely increase the local concentration of K⁺ and affect cardiac conduction.

Strategies to minimize K⁺ losses should also be considered. These measures may include minimizing the dose of non-K⁺-sparing diuretics, restricting Na⁺ intake, and using clinically appropriate combinations of non-K⁺-sparing and K⁺-sparing medications (e.g., loop diuretics with ACE inhibitors).

HYPERKALEMIA

Hyperkalemia is defined as a plasma potassium level of 5.5 mM, occurring in up to 10% of hospitalized patients; severe hyperkalemia (>6.0 mM) occurs in ~1%, with a significantly increased risk of mortality. Although redistribution and reduced tissue uptake can acutely cause hyperkalemia, a decrease in renal K⁺ excretion is the most frequent underlying cause (**Table 49-5**). Excessive intake of K⁺ is a rare cause, given the adaptive capacity to increase renal secretion; however, dietary intake can have a major effect in susceptible patients, e.g., diabetics with hyporeninemic hypoaldosteronism and chronic kidney disease. Drugs that impact on the renin-angiotensin-aldosterone axis are also a major cause of hyperkalemia.

Pseudohyperkalemia Hyperkalemia should be distinguished from factitious hyperkalemia or “pseudohyperkalemia,” an artifactual increase in serum K⁺ due to the release of K⁺ during or after venipuncture. Pseudohyperkalemia can occur in the setting of excessive muscle activity during venipuncture (e.g., fist clenching), a marked increase in cellular elements (thrombocytosis, leukocytosis, and/or erythrocytosis) with in vitro efflux of K⁺, and acute anxiety during venipuncture with respiratory alkalosis and redistributive hyperkalemia. Cooling of blood following venipuncture is another cause, due to reduced cellular uptake; the converse is the increased uptake of K⁺ by cells at high ambient temperatures, leading to normal values for hyperkalemic patients and/or to spurious hypokalemia in normokalemic patients. Finally, there are multiple genetic subtypes of hereditary pseudohyperkalemia, caused by increases in the passive K⁺ permeability of erythrocytes. For example, causative mutations have been described in the red cell anion exchanger (AE1, encoded by the *SLC4A1* gene), leading to reduced red cell anion transport, hemolytic anemia, the acquisition of a novel AE1-mediated K⁺ leak, and pseudohyperkalemia.

Redistribution and Hyperkalemia Several different mechanisms can induce an efflux of intracellular K⁺ and hyperkalemia. Acidemia is associated with cellular uptake of H⁺ and an associated efflux of K⁺; it is thought that this effective K⁺-H⁺ exchange serves to help maintain extracellular pH. Notably, this effect of acidosis is limited to non-anion gap causes of metabolic acidosis and, to a lesser extent, respiratory causes of acidosis; hyperkalemia due to an acidosis-induced shift of potassium from the cells into the ECF does *not* occur in the anion gap acidoses lactic acidosis and ketoacidosis. Hyperkalemia due to hypertonic mannitol, hypertonic saline, and intravenous immune globulin is generally attributed to a “solvent drag” effect, as water moves out of cells along the osmotic gradient. Diabetics are also prone to osmotic hyperkalemia in response to intravenous hypertonic glucose, when given without adequate insulin. Cationic amino acids, specifically lysine, arginine, and the structurally related drug epsilon-aminocaproic acid, cause efflux of K⁺ and hyperkalemia, through an effective cation-K⁺ exchange of unknown identity and mechanism. Digoxin inhibits Na⁺/K⁺-ATPase and impairs the uptake of K⁺ by skeletal muscle, such that digoxin overdose predictably results in hyperkalemia. Structurally related glycosides are found in specific plants (e.g., yellow oleander, foxglove) and in the cane toad, *Bufo marinus* (bufadienolide); ingestion of these substances and extracts thereof can also cause hyperkalemia. Finally, fluoride ions also inhibit Na⁺/K⁺-ATPase, such that fluoride poisoning is typically associated with hyperkalemia.

Succinylcholine depolarizes muscle cells, causing an efflux of K⁺ through acetylcholine receptors (AChRs). The use of this agent is contraindicated in patients who have sustained thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization.

TABLE 49-5 Causes of Hyperkalemia

I. Pseudohyperkalemia
A. Cellular efflux; thrombocytosis, erythrocytosis, leukocytosis, in vitro hemolysis
B. Hereditary defects in red cell membrane transport
II. Intra- to extracellular shift
A. Acidosis
B. Hyperosmolality; radiocontrast, hypertonic dextrose, mannitol
C. β_2 -Adrenergic antagonists (noncardioselective agents)
D. Digoxin and related glycosides (yellow oleander, foxglove, bufadienolide)
E. Hyperkalemic periodic paralysis
F. Lysine, arginine, and ϵ -aminocaproic acid (structurally similar, positively charged)
G. Succinylcholine; thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization
H. Rapid tumor lysis
III. Inadequate excretion
A. Inhibition of the renin-angiotensin-aldosterone axis; ↑ risk of hyperkalemia when used in combination
1. Angiotensin-converting enzyme (ACE) inhibitors
2. Renin inhibitors; aliskiren (in combination with ACE inhibitors or angiotensin receptor blockers [ARBs])
3. Angiotensin receptor blockers (ARBs)
4. Blockade of the mineralocorticoid receptor: spironolactone, eplerenone, drospirenone
5. Blockade of the epithelial sodium channel (ENaC): amiloride, triamterene, trimethoprim, pentamidine, nafamostat
B. Decreased distal delivery
1. Congestive heart failure
2. Volume depletion
C. Hyporeninemic hypoaldosteronism
1. Tubulointerstitial diseases: systemic lupus erythematosus (SLE), sickle cell anemia, obstructive uropathy
2. Diabetes, diabetic nephropathy
3. Drugs: nonsteroidal anti-inflammatory drugs (NSAIDs), cyclooxygenase 2 (COX2) inhibitors, β -blockers, cyclosporine, tacrolimus
4. Chronic kidney disease, advanced age
5. Pseudohypoaldosteronism type II: defects in WNK1 or WNK4 kinases, Kelch-like 3 (KLHL3), or Cullin 3 (CUL3)
D. Renal resistance to mineralocorticoid
1. Tubulointerstitial diseases: SLE, amyloidosis, sickle cell anemia, obstructive uropathy, post-acute tubular necrosis
2. Hereditary: pseudohypoaldosteronism type I; defects in the mineralocorticoid receptor or the epithelial sodium channel (ENaC)
E. Advanced renal insufficiency
1. Chronic kidney disease
2. End-stage renal disease
3. Acute oliguric kidney injury
F. Primary adrenal insufficiency
1. Autoimmune: Addison's disease, polyglandular endocrinopathy
2. Infectious: HIV, cytomegalovirus, tuberculosis, disseminated fungal infection
3. Infiltrative: amyloidosis, malignancy, metastatic cancer
4. Drug-associated: heparin, low-molecular-weight heparin
5. Hereditary: adrenal hypoplasia congenita, congenital lipoid adrenal hyperplasia, aldosterone synthase deficiency
6. Adrenal hemorrhage or infarction, including in antiphospholipid syndrome

These disorders share a marked increase and redistribution of AChRs at the plasma membrane of muscle cells; depolarization of these upregulated AChRs by succinylcholine leads to an exaggerated efflux of K^+ through the receptor-associated cation channels, resulting in acute hyperkalemia.

Hyperkalemia Caused by Excess Intake or Tissue Necrosis

Increased intake of even small amounts of K^+ may provoke severe hyperkalemia in patients with predisposing factors; hence, an assessment of dietary intake is crucial. Foods rich in potassium include tomatoes, bananas, and citrus fruits; occult sources of K^+ , particularly K^+ -containing salt substitutes, may also contribute significantly. Iatrogenic causes include simple overreplacement with K^+-Cl^- or the administration of a potassium-containing medication (e.g., K^+ -penicillin) to a susceptible patient. Red cell transfusion is a well-described cause of hyperkalemia, typically in the setting of massive transfusions. Finally, severe tissue necrosis, as in acute tumor lysis syndrome and rhabdomyolysis, will predictably cause hyperkalemia from the release of intracellular K^+ .

Hypoaldosteronism and Hyperkalemia Aldosterone release from the adrenal gland may be reduced by hyporeninemic hypoaldosteronism, medications, primary hypoaldosteronism, or isolated deficiency of ACTH (secondary hypoaldosteronism). Primary hypoaldosteronism may be genetic or acquired (Chap. 379) but is commonly caused by autoimmunity, either in Addison's disease or in the context of a polyglandular endocrinopathy. HIV has surpassed tuberculosis as the most important infectious cause of adrenal insufficiency. The adrenal involvement in HIV disease is usually subclinical; however, adrenal insufficiency may be precipitated by stress, drugs such as ketoconazole that inhibit steroidogenesis, or the acute withdrawal of steroid agents such as megestrol.

Hyporeninemic hypoaldosteronism is a very common predisposing factor in several overlapping subsets of hyperkalemic patients: diabetics, the elderly, and patients with renal insufficiency. Classically, patients should have suppressed PRA and aldosterone; ~50% have an associated acidosis, with a reduced renal excretion of NH_4^+ , a positive urinary anion gap, and urine pH <5.5. Most patients are volume expanded, with secondary increases in circulating atrial natriuretic peptide (ANP) that inhibit both renal renin release and adrenal aldosterone release.

Renal Disease and Hyperkalemia Chronic kidney disease and end-stage kidney disease are very common causes of hyperkalemia, due to the associated deficit or absence of functioning nephrons. Hyperkalemia is more common in oliguric acute kidney injury; distal tubular flow rate and Na^+ delivery are less limiting factors in nonoliguric patients. Hyperkalemia out of proportion to GFR can also be seen in the context of tubulointerstitial disease that affects the distal nephron, such as amyloidosis, sickle cell anemia, interstitial nephritis, and obstructive uropathy.

Heredity renal causes of hyperkalemia have overlapping clinical features with hypoaldosteronism, hence the diagnostic label *pseudohypoaldosteronism* (PHA). PHA type I (PHA-I) has both an autosomal recessive and an autosomal dominant form. The autosomal dominant form is due to loss-of-function mutations in the MLR; the recessive form is caused by various combinations of mutations in the three subunits of ENaC, resulting in impaired Na^+ channel activity in principal cells and other tissues. Patients with recessive PHA-I suffer from lifelong salt wasting, hypotension, and hyperkalemia, whereas the phenotype of autosomal dominant PHA-I due to MLR dysfunction improves in adulthood. PHA type II (PHA-II; also known as *hereditary hypertension with hyperkalemia*) is in every respect the mirror image of GS caused by loss of function in NCC, the thiazide-sensitive Na^+-Cl^- cotransporter (see above); the clinical phenotype includes hypertension, hyperkalemia, hyperchloremic metabolic acidosis, suppressed PRA and aldosterone, hypercalciuria, and reduced bone density. PHA-II thus behaves like a gain of function in NCC, and treatment with thiazides results in resolution of the entire clinical phenotype. However, the NCC gene is not directly involved in PHA-II, which is caused by mutations in the WNK1 and WNK4 serine-threonine kinases or the upstream Kelch-like 3 (KLHL3) and Cullin 3 (CUL3) proteins, two components of an E3 ubiquitin ligase complex that regulates these kinases; these proteins collectively regulate NCC activity, with PHA-II-associated activation of the transporter.

Medication-Associated Hyperkalemia Most medications associated with hyperkalemia cause inhibition of some component of the renin-angiotensin-aldosterone axis. ACE inhibitors, angiotensin receptor blockers, renin inhibitors, and MRAs are predictable and common causes of hyperkalemia, particularly when prescribed in combination. The oral contraceptive agent Yasmin-28 contains the progestin drospirenone, which inhibits the MRA and can cause hyperkalemia in susceptible patients. Cyclosporine, tacrolimus, NSAIDs, and cyclooxygenase 2 (COX2) inhibitors cause hyperkalemia by multiple mechanisms, but share the ability to cause hyporeninemic hypoaldosteronism. Notably, most drugs that affect the renin-angiotensin-aldosterone axis also block the local adrenal response to hyperkalemia, thus attenuating the *direct* stimulation of aldosterone release by increased plasma K⁺ concentration.

Inhibition of apical ENaC activity in the distal nephron by amiloride and other K⁺-sparing diuretics results in hyperkalemia, often with a voltage-dependent hyperchloremic acidosis and/or hypovolemic hyponatremia. Amiloride is structurally similar to the antibiotics TMP and pentamidine, which also block ENaC; risk factors for TMP-associated hyperkalemia include the administered dose, renal insufficiency, and hyporeninemic hypoaldosteronism. Indirect inhibition of ENaC at the plasma membrane is also a cause of drug-associated hyperkalemia; nafamostat, a protease inhibitor used in some countries for anticoagulation and for the management of pancreatitis, inhibits aldosterone-induced renal proteases that activate ENaC by proteolytic cleavage.

Clinical Features Hyperkalemia is a medical emergency due to its effects on the heart. Cardiac arrhythmias associated with hyperkalemia include sinus bradycardia, sinus arrest, slow idioventricular rhythms, ventricular tachycardia, ventricular fibrillation, and asystole. Mild increases in extracellular K⁺ affect the repolarization phase of the cardiac action potential, resulting in changes in T-wave morphology; further increase in plasma K⁺ concentration depresses intracardiac conduction, with progressive prolongation of the PR and QRS intervals. Severe hyperkalemia results in loss of the P wave and a progressive widening of the QRS complex; development of a sine-wave sinoventricular rhythm suggests impending ventricular fibrillation or asystole. Hyperkalemia can also cause a type I Brugada pattern in the electrocardiogram (ECG), with a pseudo-right bundle branch block and persistent coved ST segment elevation in at least two precordial leads. This hyperkalemic Brugada's sign occurs in critically ill patients with severe hyperkalemia and can be differentiated from genetic Brugada's syndrome by an absence of P waves, marked QRS widening, and an abnormal QRS axis. Classically, the electrocardiographic manifestations in hyperkalemia progress from tall peaked T waves (5.5–6.5 mM), to a loss of P waves (6.5–7.5 mM) to a widened QRS complex (7.0–8.0 mM), and, ultimately, a to a sine wave pattern (>8.0 mM). However, these changes are notoriously insensitive, particularly in patients with chronic kidney disease or ESRD.

Hyperkalemia from a variety of causes can also present with ascending paralysis, denoted *secondary hyperkalemic paralysis* to differentiate it from familial hyperkalemic periodic paralysis (HYPP). The presentation may include diaphragmatic paralysis and respiratory failure. Patients with familial HYPP develop myopathic weakness during hyperkalemia induced by increased K⁺ intake or rest after heavy exercise. Depolarization of skeletal muscle by hyperkalemia unmasks an inactivation defect in skeletal Na⁺ channel; autosomal dominant mutations in the SCN4A gene encoding this channel are the predominant cause.

Within the kidney, hyperkalemia has negative effects on the ability to excrete an acid load, such that hyperkalemia per se can contribute to metabolic acidosis. This defect appears to be due in part to competition between K⁺ and NH₄⁺ for reabsorption by the TALH and subsequent countercurrent multiplication, ultimately reducing the medullary gradient for NH₃/NH₄ excretion by the distal nephron. Regardless of the underlying mechanism, restoration of normokalemia can, in many instances, correct hyperkalemic metabolic acidosis.

Diagnostic Approach The first priority in the management of hyperkalemia is to assess the need for emergency treatment, followed

by a comprehensive workup to determine the cause (Fig. 49-8). History and physical examination should focus on medications, diet and dietary supplements, risk factors for kidney failure, reduction in urine output, blood pressure, and volume status. Initial laboratory tests should include electrolytes, BUN, creatinine, serum osmolality, Mg²⁺ and Ca²⁺, a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes. A urine Na⁺ concentration of <20 mM indicates that distal Na⁺ delivery is a limiting factor in K⁺ excretion; volume repletion with 0.9% saline or treatment with furosemide may be effective in reducing plasma K⁺ concentration. Serum and urine osmolality are required for calculation of the transtubular K⁺ gradient (TTKG) (Fig. 49-8). The expected values of the TTKG are largely based on historical data, and are <3 in the presence of hypokalemia and >7–8 in the presence of hyperkalemia.

$$\text{TTKG} = \frac{[\text{K}^+]_{\text{urine}} \times \text{Osm}_{\text{serum}}}{[\text{K}^+]_{\text{serum}} \times \text{Osm}_{\text{urine}}}$$

TREATMENT

Hyperkalemia

Electrocardiographic manifestations of hyperkalemia should be considered a medical emergency and treated urgently. However, patients with significant hyperkalemia (plasma K⁺ concentration ≥6.5 mM) in the absence of ECG changes should also be aggressively managed, given the limitations of ECG changes as a predictor of cardiac toxicity. Urgent management of hyperkalemia includes admission to the hospital, continuous cardiac monitoring, and immediate treatment. The treatment of hyperkalemia is divided into three stages:

1. *Immediate antagonism of the cardiac effects of hyperkalemia.* Intravenous calcium serves to protect the heart, whereas other measures are taken to correct hyperkalemia. Calcium raises the action potential threshold and reduces excitability, without changing the resting membrane potential. By restoring the difference between resting and threshold potentials, calcium reverses the depolarization blockade due to hyperkalemia. The recommended dose is 10 mL of 10% calcium gluconate (3–4 mL of calcium chloride), infused intravenously over 2–3 min with cardiac monitoring. The effect of the infusion starts in 1–3 min and lasts 30–60 min; the dose should be repeated if there is no change in ECG findings or if they recur after initial improvement. Hypercalcemia potentiates the cardiac toxicity of digoxin; hence, intravenous calcium should be used with extreme caution in patients taking this medication; if judged necessary, 10 mL of 10% calcium gluconate can be added to 100 mL of 5% dextrose in water and infused over 20–30 min to avoid acute hypercalcemia.
2. *Rapid reduction in plasma K⁺ concentration by redistribution into cells.* Insulin lowers plasma K⁺ concentration by shifting K⁺ into cells. The recommended dose is 10 units of intravenous regular insulin followed immediately by 50 mL of 50% dextrose (D₅₀W, 25 g of glucose total); the effect begins in 10–20 min, peaks at 30–60 min, and lasts for 4–6 h. Bolus D₅₀W without insulin is *never* appropriate, given the risk of acutely worsening hyperkalemia due to the osmotic effect of hypertonic glucose. Hypoglycemia is common with insulin plus glucose; hence, this should be followed by an infusion of 10% dextrose at 50–75 mL/h, with close monitoring of plasma glucose concentration. In hyperkalemic patients with glucose concentrations of ≥200–250 mg/dL, insulin should be administered *without* glucose, again with close monitoring of glucose concentrations.

β₂-agonists, most commonly albuterol, are effective but underused agents for the acute management of hyperkalemia. Albuterol and insulin with glucose have an additive effect on plasma K⁺ concentration; however, ~20% of patients with ESRD are resistant to the effect of β₂-agonists; hence, these drugs should not be used without insulin. The recommended dose for inhaled

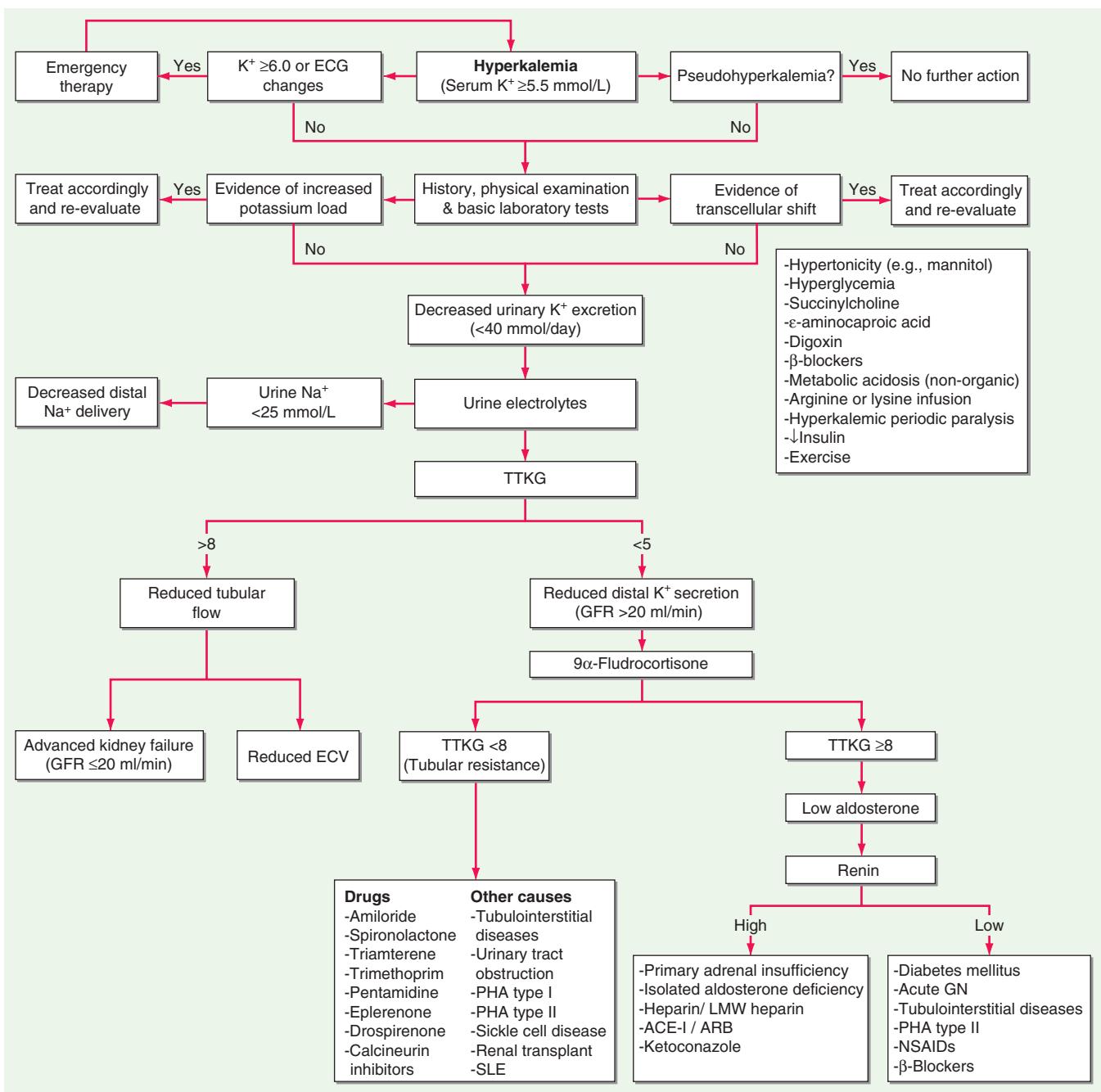


FIGURE 49-8 The diagnostic approach to hyperkalemia. See text for details. ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; CCD, cortical collecting duct; ECG, electrocardiogram; ECV, effective circulatory volume; GFR, glomerular filtration rate; GN, glomerulonephritis; HIV, human immunodeficiency virus; LMW heparin, low-molecular-weight heparin; NSAIDs, nonsteroidal anti-inflammatory drugs; PHA, pseudohypoaldosteronism; SLE, systemic lupus erythematosus; TTKG, transtubular potassium gradient. (Used with permission from DB Mount, K Zandi-Nejad K: Disorders of potassium balance, in Brenner and Rector's The Kidney, 8th ed, BM Brenner [ed]. Philadelphia, W.B. Saunders & Company, 2008, pp 547-587.)

albuterol is 10–20 mg of nebulized albuterol in 4 mL of normal saline, inhaled over 10 min; the effect starts at about 30 min, reaches its peak at about 90 min, and lasts for 2–6 h. Hyperglycemia is a side effect, along with tachycardia. β_2 -Agonists should be used with caution in hyperkalemic patients with known cardiac disease.

Intravenous bicarbonate has no role in the acute treatment of hyperkalemia, but may slowly attenuate hyperkalemia with sustained administration over several hours. It should not be given repeatedly as a hypertonic intravenous bolus of undiluted ampules, given the risk of associated hypernatremia, but should instead be infused in an isotonic or hypotonic fluid (e.g., 150 mEqu in 1 L of D_5W). In patients with metabolic acidosis, a delayed drop in plasma K^+ concentration can be seen after 4–6 h of isotonic bicarbonate infusion.

3. **Removal of potassium.** This is typically accomplished using cation exchange resins, diuretics, and/or dialysis. The cation exchange resin sodium polystyrene sulfonate (SPS) exchanges Na^+ for K^+ in the gastrointestinal tract and increases the fecal excretion of K^+ ; alternative calcium-based resins, when available, may be more appropriate in patients with an increased ECFV. The recommended dose of SPS is 15–30 g of powder, almost always given in a premade suspension with 33% sorbitol. The effect of SPS on plasma K^+ concentration is slow; the full effect may take up to 24 h and usually requires repeated doses every 4–6 h. Intestinal necrosis, typically of the colon or ileum, is a rare but usually fatal complication of SPS. Intestinal necrosis is more common in patients administered SPS via enema and/or in patients with reduced intestinal motility (e.g., in the postoperative state or after treatment with opioids). The coadministration of SPS with

sorbitol appears to increase the risk of intestinal necrosis; however, this complication can also occur with SPS alone. The low but real risk of intestinal necrosis with SPS, which can sometimes be the only available or appropriate therapy for the removal of potassium, must be weighed against the delayed onset of efficacy. Whenever possible, alternative therapies for the acute management of hyperkalemia (i.e., aggressive redistributive therapy, isotonic bicarbonate infusion, diuretics, and/or hemodialysis) should be used instead of SPS.

Novel intestinal potassium binders have recently become available for the management of hyperkalemia. These agents appear to lack the intestinal toxicity of SPS. Patiromer is a non-absorbed polymer provided as a powder for suspension, which binds K^+ in exchange for Ca^{2+} . In healthy adults, patiromer causes a decrease in urinary potassium, magnesium, and sodium excretion, suggesting the binding of the polymer to these cations in the intestine; notably, a side-effect of the medication is hypomagnesemia. ZS-9 is an inorganic, nonabsorbable crystalline compound that exchanges both Na^+ and H^+ ions in exchange for K^+ and NH_4^+ in the intestine. These agents promise to revolutionize the management of both chronic and acute hyperkalemia. In particular, the availability of safe, well-tolerated potassium binders is expected to allow for greater intensity of RAAS inhibition in both renal and cardiac disease.

Therapy with intravenous saline may be beneficial in hypovolemic patients with oliguria and decreased distal delivery of Na^+ , with the associated reductions in renal K^+ excretion. Loop and thiazide diuretics can be used to reduce plasma K^+ concentration in volume-replete or hypervolemic patients with sufficient renal function for a diuretic response; this may need to be combined with intravenous saline or isotonic bicarbonate to achieve or maintain euolemia.

Hemodialysis is the most effective and reliable method to reduce plasma K^+ concentration; peritoneal dialysis is considerably less effective. Patients with acute kidney injury require temporary, urgent venous access for hemodialysis, with the attendant risks; in contrast, patients with ESRD or advanced chronic kidney disease may have a preexisting venous access. The amount of K^+ removed during hemodialysis depends on the relative distribution of K^+ between ICF and ECF (potentially affected by prior therapy for hyperkalemia), the type and surface area of the dialyzer used, dialysate and blood flow rates, dialysate flow rate, dialysis duration, and the plasma-to-dialysate K^+ gradient.

FURTHER READING

- CHOI M et al: K^+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science* 331:768, 2011.
- MOUNT DB, ZANDI-NEJAD K: Disorders of potassium balance, in *Brenner and Rector's The Kidney*, 10th ed, K Skorecki et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp 559–600.
- PACKHAM DK et al: Sodium zirconium cyclosilicate in hyperkalemia. *N Engl J Med* 372:222, 2015.
- PERIANAYAGAM A et al: DDAVP is effective in preventing and reversing inadvertent overcorrection of hyponatremia. *Clin J Am Soc Nephrol* 3:331, 2008.
- SCHRIER RW: Decreased effective blood volume in edematous disorders: what does this mean? *J Am Soc Nephrol* 18:2028, 2007.
- SOOD L et al: Hypertonic saline and desmopressin: a simple strategy for safe correction of severe hyponatremia. *Am J Kidney Dis* 61:571, 2013.
- SOUPIART A et al: Efficacy and tolerance of urea compared with vaptans for long-term treatment of patients with SIADH. *Clin J Am Soc Nephrol* 7:742, 2012.
- WEIR MR et al: Patiromer in patients with kidney disease and hyperkalemia receiving RAAS inhibitors. *N Engl J Med* 372:211, 2015.

50

Hypercalcemia and Hypocalcemia

Sundeep Khosla



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 [$1,25(OH)_2D$]. These feedback mechanisms are orchestrated by integrating signals between the parathyroid glands, kidney, intestine, and bone (Fig. 50-1; Chap. 402). 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. 403 for a detailed discussion of this topic.

HYPERCALCEMIA

ETIOLOGY

The causes of hypercalcemia can be understood and classified based on derangements in the normal feedback mechanisms that regulate serum calcium (Table 50-1). 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

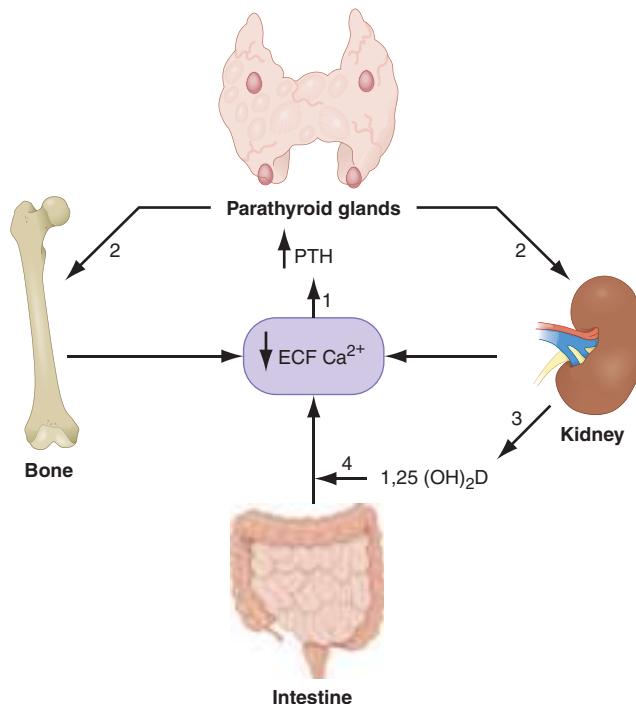


FIGURE 50-1 Feedback mechanisms maintaining extracellular calcium concentrations within a narrow, physiologic range (8.9–10.1 mg/dL [2.2–2.5 mM]). A decrease in extracellular (ECF) calcium (Ca^{2+}) triggers an increase in parathyroid hormone (PTH) secretion (1) via the calcium sensor receptor on parathyroid 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 $1,25(OH)_2D$ production (3). $1,25(OH)_2D$, in turn, acts principally on the intestine to increase calcium absorption (4). Collectively, these homeostatic mechanisms serve to restore serum calcium levels to normal.

TABLE 50-1 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) ₂ D 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.

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.

■ CLINICAL MANIFESTATIONS

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, 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 mM (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 50-1). 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. 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 50-1.

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. Zoledronic acid (e.g., 4 mg intravenously over ~30 min), pamidronate (e.g., 60–90 mg intravenously over 2–4 h), and ibandronate (2 mg intravenously over 2 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. An alternative to the bisphosphonates is gallium nitrate (200 mg/m² intravenously daily for 5 days), which is also effective, but has potential nephrotoxicity. More recently, the potent inhibitor of bone resorption, denosumab (120 mg sc on days 1, 8, 15, and 29, and then every 4 weeks), has also been shown to be effective in treating hypercalcemia refractory to bisphosphonates. 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)₂D-mediated hypercalcemia, glucocorticoids are the preferred therapy, as they decrease 1,25(OH)₂D 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)₂D 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 50-2) (Chap. 403). 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

TABLE 50-2 Causes of Hypocalcemia

Low Parathyroid Hormone Levels (Hypoparathyroidism)

- Parathyroid agenesis
 - Isolated
 - DiGeorge's syndrome
- 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)

Abbreviations: CaSR, calcium sensor receptor; PTH, parathyroid hormone.

all four glands during thyroid or parathyroid gland surgery. Hypoparathyroidism is a cardinal feature of autoimmune endocrinopathies (Chap. 381); 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.

Vitamin D deficiency, impaired 1,25(OH)₂D 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)₂D 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 D₂ or D₃ (25,000–100,000 U daily) or calcitriol [1,25(OH)₂D, 0.25–2 µg/d]. Other vitamin D metabolites (dihydrotachysterol, alfacalcidol) are now used less frequently.

Importantly, PTH (1-84) (Natpara) has recently been approved by the FDA 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 U, 2–3 times per week for several months), whereas vitamin D deficiency due to malabsorption may require much higher doses (100,000 U/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 exposure and poor dietary vitamin D intake.

■ FURTHER READING

- EASTELL R et al: Diagnosis of asymptomatic primary hyperparathyroidism: Proceedings of the 4th International Workshop. *J Clin Endocrinol Metab* 99:3570, 2014.
- KIM ES, KEATING GM: Recombinant human parathyroid hormone (1-84): A review in hypoparathyroidism. *Drugs* 75:1293, 2015.
- MAYR B et al: Genetics in endocrinology: Gain and loss of function mutations of the calcium-sensing receptor and associated proteins: Current treatment concepts. *Eur J Endocrinol* 174:R189, 2016.
- MINISOLA S et al: The diagnosis and management of hypercalcemia. *BMJ* 350:h2723, 2015.
- THAKKER RV: The calcium-sensing receptor: And its involvement in parathyroid pathology. *Ann Endocrinol* 76:81, 2015.

DIAGNOSIS OF GENERAL TYPES OF DISTURBANCES

The most common clinical disturbances are simple acid-base disorders, that is, metabolic acidosis or alkalosis or respiratory acidosis or alkalosis.

■ SIMPLE ACID-BASE DISORDERS

Primary respiratory disturbances (primary changes in Paco_2) invoke compensatory metabolic responses (secondary changes in $[\text{HCO}_3^-]$), 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 51-1**. 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) lowers extracellular fluid $[\text{HCO}_3^-]$ and decreases extracellular pH. This stimulates the medullary chemoreceptors to increase ventilation and to return the ratio of $[\text{HCO}_3^-]$ to Paco_2 , and thus pH, toward, but not to, normal. The degree of respiratory compensation expected in a metabolic acidosis can be predicted from the relationship: $\text{Paco}_2 = (1.5 \times [\text{HCO}_3^-]) + 8 \pm 2$. Thus, a patient with metabolic acidosis and $[\text{HCO}_3^-]$ of 12 mmol/L would be expected to have a Paco_2 of ~26 mmHg. Values for $\text{Paco}_2 < 24$ or > 28 mmHg define a mixed disturbance (metabolic acidosis and respiratory alkalosis or metabolic acidosis and respiratory acidosis, respectively). Compensatory responses for primary metabolic disorders move the Paco_2 in the same direction as the change in $[\text{HCO}_3^-]$, whereas, conversely, compensation for primary respiratory disorders moves the $[\text{HCO}_3^-]$ in the same direction as the primary change in Paco_2 (Table 51-1). Therefore, changes in Paco_2 and $[\text{HCO}_3^-]$ in **opposite directions** (i.e., Paco_2 or $[\text{HCO}_3^-]$ is increased, whereas the other value is decreased) indicate a **mixed acid-base disturbance**. Another way to judge the appropriateness of the response in $[\text{HCO}_3^-]$ or Paco_2 is to use an acid-base nomogram (**Fig. 51-1**). While the shaded areas of the

TABLE 51-1 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	$\text{Paco}_2 = (1.5 \times \text{HCO}_3^-) + 8 \pm 2$ or Paco_2 will \downarrow 1.25 mmHg per mmol/L \downarrow in $[\text{HCO}_3^-]$ or $\text{Paco}_2 = [\text{HCO}_3^-] + 15$	Low	Low	Low
Metabolic alkalosis	Paco_2 will \uparrow 0.75 mmHg per mmol/L \uparrow in $[\text{HCO}_3^-]$ or Paco_2 will \uparrow 6 mmHg per 10 mmol/L \uparrow in $[\text{HCO}_3^-]$ or $\text{Paco}_2 = [\text{HCO}_3^-] + 15$	High	High	High
Respiratory alkalosis		High	Low	Low
Acute	$[\text{HCO}_3^-]$ will \downarrow 0.2 mmol/L per mmHg \downarrow in Paco_2			
Chronic	$[\text{HCO}_3^-]$ will \downarrow 0.4 mmol/L per mmHg \downarrow in Paco_2			
Respiratory acidosis		Low	High	High
Acute	$[\text{HCO}_3^-]$ will \uparrow 0.1 mmol/L per mmHg \uparrow in Paco_2			
Chronic	$[\text{HCO}_3^-]$ will \uparrow 0.4 mmol/L per mmHg \uparrow in Paco_2			

51

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:

$$\text{pH} = 6.1 + \log \frac{\text{HCO}_3^-}{\text{PaCO}_2 \times 0.03001}$$

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 $[\text{HCO}_3^-]$.

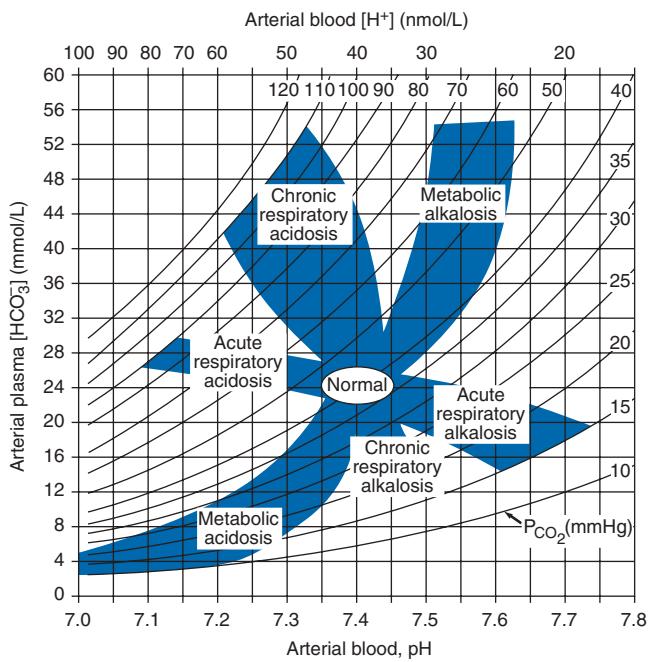


FIGURE 51-1 Acid-base nomogram. Shown are the 90% confidence limits (range of values) of the normal respiratory and metabolic compensations for primary acid-base disturbances. (From TD DuBose Jr: Acid-Base Disorders, in Brenner and Rector's *The Kidney*, 10th ed, K Skorecki, GM Chertow, PA Marsden, MW Taal, and Alan SL Yu [eds]. Philadelphia, Saunders, 2016, p. 522; with permission.)

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 51-1.

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 51-2). The diagnosis of mixed acid-base disorders requires consideration of the anion gap (AG), and requires the presence of or correction to a normal serum albumin of 4.5 g/dL. A patient with diabetic ketoacidosis (metabolic acidosis) may develop an independent respiratory problem (e.g., pneumonia) leading to a superimposed respiratory acidosis or alkalosis. Patients with underlying pulmonary disease (e.g., chronic obstructive pulmonary disease) may not respond to metabolic acidosis with an appropriate ventilatory response because of insufficient respiratory reserve. 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, an incongruity in the ΔAG (prevailing 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 renal dysfunction resulting in simultaneous metabolic acidosis. 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 51-2 Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory

Metabolic acidosis—respiratory alkalosis

Key: High- or normal-AG metabolic acidosis; prevailing Paco_2 below predicted value (Table 51-1)

Example: Na^+ , 140; K^+ , 4.0; Cl^- , 106; HCO_3^- , 14; AG, 20; Paco_2 , 24; pH, 7.39 (lactic acidosis, sepsis in ICU)

Metabolic acidosis—respiratory acidosis

Key: High- or normal-AG metabolic acidosis; prevailing Paco_2 above predicted value (Table 51-1)

Example: Na^+ , 140; K^+ , 4.0; Cl^- , 102; HCO_3^- , 18; AG, 20; Paco_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; Paco_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; Paco_2 , 67; pH, 7.42 (COPD on diuretics)

Mixed Metabolic Disorders

Metabolic acidosis—metabolic alkalosis

Key: Only detectable with high-AG acidosis; $\Delta\text{AG} >> \Delta\text{HCO}_3^-$

Example: Na^+ , 140; K^+ , 3.0; Cl^- , 95; HCO_3^- , 25; AG, 20; Paco_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; Paco_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

Acid-Base Disorders

A stepwise approach to the diagnosis of acid-base disorders follows (Table 51-3). Blood for electrolytes and arterial blood gases should be drawn simultaneously prior to therapy. An increase in $[\text{HCO}_3^-]$ occurs with either metabolic alkalosis or respiratory acidosis. Conversely, a decrease in $[\text{HCO}_3^-]$ 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 $[\text{HCO}_3^-]$ is calculated from the Henderson-Hasselbalch equation. This calculated value should be compared with the measured $[\text{HCO}_3^-]$ (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 identified.

TABLE 51-3 Steps in Acid-Base Diagnosis

1. Obtain arterial blood gas (ABG) and electrolytes simultaneously.
2. Compare $[\text{HCO}_3^-]$ on ABG and electrolytes to verify accuracy.
3. Calculate anion gap (AG), but correct to a normal albumin concentration of 4.5 g/dL.
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 51-1).
7. Compare ΔAG and ΔHCO_3^- .
8. Compare change in $[\text{Cl}^-]$ with change in $[\text{Na}^+]$.

CALCULATE THE ANION GAP

All evaluations of acid-base disorders should include a simple calculation of the AG. The AG is calculated as follows: $AG = \text{Na}^+ - (\text{Cl}^- + \text{HCO}_3^-)$. In the United States, the value for plasma $[\text{K}^+]$ is typically omitted from the calculation of the AG. The “normal” value for the AG reported by clinical laboratories has declined with improved methodology for measuring plasma electrolytes, and ranges from 6 to 12 mmol/L, with an average of ~10 mmol/L. The clinician is encouraged to be aware of the normal value for the AG in their clinical chemistry laboratory. 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 and, 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. 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 ($2.5 \times 2 = 5$; $5 + 15 = \text{corrected AG of } 20 \text{ mmol/L}$). The clinical disorders that cause a high-AG acidosis are displayed in Table 51-3.

A high AG is usually due to accumulation of non-chloride-containing acids that contain inorganic (phosphate, sulfate), organic (ketoads, lactate, uremic organic anions), exogenous (salicylate or ingested toxins with organic acid production), or unidentified anions. The high AG is significant clinically even if the $[\text{HCO}_3^-]$ or pH is normal. Simultaneous metabolic acidosis of the high-AG variety plus either chronic respiratory acidosis or metabolic alkalosis represents such a situation in which $[\text{HCO}_3^-]$ may be normal or even high (Table 51-3). In cases of high-AG metabolic acidosis it is valuable to compare the decline in $[\text{HCO}_3^-]$ (ΔHCO_3^- : $25 - \text{patient's } [\text{HCO}_3^-]$) with the increase in the AG ($\Delta\text{AG}: \text{patient's AG} - 10$).

Similarly, normal values for $[\text{HCO}_3^-]$, 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, $[\text{HCO}_3^-]$ of 40 mmol/L, $[\text{Na}^+]$ of 135, $[\text{Cl}^-]$ of 80, and $[\text{K}^+]$ 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 $[\text{HCO}_3^-]$ 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. 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 ($25 - 25 \text{ mmol/L}$), but the ΔAG is 20 ($30 - 10 \text{ mmol/L}$). Therefore, 20 mmol/L is unaccounted for in the Δ/Δ value (ΔAG to ΔHCO_3^-).

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 [CKD]). Metabolic acidosis has profound effects on the respiratory, cardiac, and nervous systems. The fall in blood pH is accompanied by a

TABLE 51-4 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)

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 vasoconstriction can 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 51-3 and **Table 51-4**). The presence of metabolic acidosis, a normal AG, and hyperchloremia denotes the presence of a normal AG metabolic acidosis.

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 $[\text{HCO}_3^-]$ can be estimated from the increment (Δ) in the AG ($\Delta\text{AG} = \text{patient's AG} - 10$), only if the acid anion that has accumulated in plasma is metabolizable (i.e., β -hydroxybutyrate, acetoacetate, and lactate). Conversely non-metabolizable anions that may accumulate in advanced stage CKD or after toxin ingestion are not metabolizable and do not represent “potential” HCO_3^- . With acute CKD improvement in kidney function to replenish the $[\text{HCO}_3^-]$ deficit is a slow and often unpredictable process. Consequently, patients with a normal AG acidosis (hyperchloremic acidosis) or an AG attributable to a non-metabolizable anion due to advanced kidney failure should receive alkali therapy, either PO (NaHCO_3 or Shohl’s solution) or IV (NaHCO_3), in an amount necessary to slowly increase the plasma $[\text{HCO}_3^-]$ to a target value of 22 mmol/L. Nevertheless, overcorrection should be avoided.

Controversy exists in regard to the use of alkali in patients with a pure AG acidosis owing to accumulation of a metabolizable organic acid anion (ketoacidosis or lactic acidosis). In general, severe acidemia (pH < 7.10) in an adult patient (especially the elderly and patients with severe heart disease) warrants the IV administration of 50 meq of NaHCO_3 diluted in 300 mL of sterile water over 30–45 min, during the initial 1–2 h of therapy. Provision of such modest quantities of alkali in this situation seems to provide an added measure of safety. Administration of alkali requires careful monitoring of plasma electrolytes, especially the plasma $[\text{K}^+]$, during the course of therapy. A reasonable initial goal is to increase the $[\text{HCO}_3^-]$ to 10–12 mmol/L and the pH to ~7.20, but clearly not to increase these values to normal. Estimation of the “bicarbonate deficit” by calculation of the volume of distribution of bicarbonate is often taught but is unnecessary and may result in administration of excessive amounts of alkali.

HIGH-ANION GAP ACIDOSSES

APPROACH TO THE PATIENT

There are four principal causes of a high-AG acidosis: (1) lactic acidosis, (2) ketoacidosis, (3) ingested toxins, and (4) acute and chronic renal failure (Table 51-4). 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 diabetes mellitus is present (diabetic ketoacidosis); (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); 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 (EG), 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-oxyproline. D-Lactic acid acidosis, which may be associated with jejunileal bypass, short bowel syndrome, or intestinal obstruction, is due to formation of D-lactate by gut bacteria.

APPROACH TO THE PATIENT

L-Lactic Acid Acidosis

The underlying condition that disrupts lactate metabolism should be corrected preemptively, if possible; tissue perfusion must be restored when inadequate, but vasoconstrictors should be avoided, if possible, because they may worsen tissue perfusion. Alkali therapy is generally advocated for acute, severe acidemia ($\text{pH} < 7.00$) 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 $[\text{HCO}_3^-]$ to normal by administration of exogenous NaHCO_3 are deleterious. A reasonable approach is to infuse sufficient NaHCO_3 to raise the arterial pH to no more than 7.2 or the $[\text{HCO}_3^-]$ to no more than 12, over 30–40 min.

NaHCO_3 therapy can cause fluid overload 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. When 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 excess NaHCO_3 has been administered excessively.

Ketoacidosis • DIABETIC KETOACIDOSIS (DKA) This condition is caused by increased fatty acid metabolism and the accumulation of ketoacids (acetacetate 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. The accumulation of ketoacids accounts for the increment in the AG and is accompanied most often by hyperglycemia (glucose $> 17 \text{ mmol/L}$ [300 mg/dL]). The relationship between the ΔAG and ΔHCO_3^- is usually 1:1 in DKA. It should be noted that because insulin prevents production of ketones, bicarbonate

therapy is rarely needed except with extreme acidemia ($\text{pH} < 7.10$), and then in only limited amounts. Patients with DKA are typically volume depleted and require fluid resuscitation with isotonic saline. Volume overexpansion with IV isotonic fluid administration is not uncommon, however, and contributes to the development of a hyperchloremic acidosis during treatment of DKA. The mainstay for treatment of this condition is IV regular insulin and is described in Chap. 396 in more detail.

ALCOHOLIC KETOACIDOSIS (AKA) Chronic alcoholics can develop ketoacidosis when alcohol consumption is abruptly curtailed and nutrition is poor. AKA is usually associated with binge drinking, vomiting, abdominal pain, starvation, and volume depletion. The glucose concentration is variable, and acidosis may be severe because of elevated ketones, predominantly β -hydroxybutyrate. Hypoperfusion may enhance lactic acid production, chronic respiratory alkalosis may accompany liver disease, and metabolic alkalosis can result from vomiting (refer to the relationship between ΔAG and ΔHCO_3^-). Thus, mixed acid-base disorders are common in AKA. As the circulation is restored by administration of isotonic saline, the preferential accumulation of β -hydroxybutyrate is then shifted to acetacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction (ketones) as the patient improves. The nitroprusside ketone reaction (Acetest) can detect acetacetate acid but not β -hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with therapy, but can be underestimated initially. 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 in the urine and, therefore, may have a relatively normal AG and a discrepancy in the $\Delta\text{AG}/\Delta\text{HCO}_3^-$ relationship.

TREATMENT

Alcoholic Ketoacidosis

Extracellular fluid deficits almost always accompany AKA and should be repleted by IV administration 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 typically emerges 12–24 h after admission, may be exacerbated by glucose infusion, and, if severe, may induce rhabdomyolysis or even respiratory arrest. Upper gastrointestinal hemorrhage, pancreatitis, and pneumonia may accompany this disorder.

DRUG- AND TOXIN-INDUCED ACIDOSIS

Salicylates (See also Chap. 449) 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). In the acidotic patient, to facilitate removal of salicylate, IV NaHCO_3 is administered in amounts adequate to alkalinize the urine and to maintain urine output (urine $\text{pH} > 7.5$), because raising the urine pH from 6.5 to 7.5 increases salicylate clearance fivefold. Patients with coexisting respiratory alkalosis should also receive NaHCO_3 , but with caution 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, but this drug can 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 can be performed against a bicarbonate-containing dialysate.

ALCOHOLS Under most physiologic conditions, sodium, urea, and glucose generate the osmotic pressure of blood. Plasma osmolality is calculated according to the following expression: $P_{\text{osm}} = 2\text{Na}^+ + \text{Glu} + \text{BUN}$ (all in mmol/L), or, using conventional laboratory values in which glucose and BUN are expressed in milligrams per deciliter: $P_{\text{osm}} = 2\text{Na}^+ + \text{Glu}/18 + \text{BUN}/2.8$. The calculated and determined osmolality should agree within 10–15 mmol/kg H₂O. When the measured osmolality exceeds the calculated osmolality by >10–15 mmol/kg H₂O, 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, EG, propylene glycol, methanol, and 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: EG, 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. 449) Ingestion of EG (commonly used in antifreeze) leads to a metabolic acidosis and severe damage to the CNS, heart, lungs, and kidneys. The combination of a high AG and high osmolar gap is highly suspicious for EG or methanol intoxication. The increased AG and osmolar gap in EG intoxication are attributable to EG and its metabolites, oxalic acid, glycolic acid, and other organic acids. Lactic acid production increases secondary to inhibition of the tricarboxylic acid cycle and altered intracellular redox state. In addition to the presence of elevated osmolar and AGs, the diagnosis is further enabled by recognition of oxalate crystals in the urine. Use of a Wood's lamp to visualize the fluorescent additive to commercial antifreeze in the urine of patients with EG ingestion, has been reported, but is not reliable. 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. Treatment should not be delayed while awaiting measurement of EG levels in this setting.

TREATMENT

Ethylene Glycol-Induced Acidosis

This includes the prompt institution of a saline or osmotic diuresis, thiamine and pyridoxine supplements, fomepizole, and usually, hemodialysis. The IV administration of the alcohol dehydrogenase inhibitor fomepizole (4-methylpyrazole; 15 mg/kg as a loading dose) 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. If used, ethanol IV should be infused to achieve a blood level of 22 mmol/L (100 mg/dL). Both fomepizole and ethanol reduce toxicity because they compete with EG for metabolism by alcohol dehydrogenase. Hemodialysis is indicated when the arterial pH is <7.3 or the osmolar gap exceeds 20 mOsm/kg.

METHANOL (See also Chap. 449) 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 usually present.

TREATMENT

Methanol-Induced Acidosis

This is similar to that for EG intoxication, including general supportive measures, fomepizole, and hemodialysis (as above).

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, most often in the setting of the intensive care unit in patients receiving frequent or continuous therapy. 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 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 occasionally hemodialysis for prolonged coma, hemodynamic instability, or levels >400 mg/dL.

PYROGLUTAMIC ACID Acetaminophen-induced high-AG metabolic acidosis is uncommon but is being recognized more often 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 the drug. Additionally, sodium bicarbonate IV should be given. Although N-acetylcysteine has been suggested, it is not known if it hastens the metabolism of 5-oxoproline by increasing intracellular glutathione concentrations in this setting.

Chronic Kidney Disease (See also Chap. 305) The hyperchloremic acidosis of moderate 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 NH₄⁺ production and excretion. Alkaline salts from bone buffer the acid retained in chronic kidney disease. Despite significant retention of acid (up to 20 mmol/d), the serum [HCO₃⁻] 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 increases urinary calcium excretion, proportional to cumulative acid retention, and contributes significantly to muscle wasting.

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 ESRD and the non-AG metabolic acidosis of stages 3 and 4 CKD require oral alkali replacement to maintain the $[HCO_3^-]$ to approximately the normal value (25 mmol/L). This can be accomplished with relatively modest amounts of alkali (1.0–1.5 mmol/kg body weight per day). Either NaHCO₃ tablets (650-mg tablets contain 7.8 meq) or sodium citrate (Shohl’s solution) is effective.

NON-ANION GAP METABOLIC ACIDOSSES

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 51-5), reciprocal changes in $[Cl^-]$ and $[HCO_3^-]$ result in a normal AG. In pure 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.

Stool contains a higher concentration of HCO₃⁻ and decomposed HCO₃⁻ 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 excretion of NH₄⁺, 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 NH₄⁺ excretion is typically low in RTA and high with diarrhea. Urinary NH₄⁺ levels can be estimated by calculating the urine AG (UAG): UAG = $[Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This indicates 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 low, suggesting a renal cause of the acidosis.

Proximal RTA (type 2 RTA) (Chap. 309) 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 HCO₃⁻ is >20 mmol/L. Because HCO₃⁻ is not reabsorbed normally in the proximal tubule, therapy with NaHCO₃ will enhance delivery of HCO₃⁻ 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 NH₄⁺ excretion (positive UAG, low urine $[NH_4^+]$), and inappropriately high urine pH ($pH > 5.5$). Most patients have hypocitraturia and hypercalciuria, so 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, for example, due to diabetic nephropathy, obstructive uropathy, or chronic tubulointerstitial disease.

Hyporeninemic hypaldosteronism typically causes non-AG metabolic acidosis, most commonly in older adults with diabetes mellitus or tubulointerstitial disease and CKD. Patients usually have mild to moderate CKD (GFR, 20–50 mL/min) and acidosis, with elevation in serum $[K^+]$ (5.2–6.0 mmol/L), concurrent hypertension, and congestive heart failure. Both the metabolic acidosis and the hyperkalemia are out

TABLE 51-5 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. Mineralocorticoid deficiency
 - b. Mineralocorticoid resistance (PHA I, autosomal dominant)
 - c. Voltage defect (PHA I, autosomal recessive, and PHA II)
 - d. 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.

of proportion to impairment in GFR. Nonsteroidal anti-inflammatory drugs, trimethoprim, pentamidine, angiotensin-converting enzyme (ACE) inhibitors, and aldosterone receptor blockers (ARBs), can also increase the risk for a hyperkalemia and a non-AG metabolic acidosis in patients with CKD (Table 51-5).

TREATMENT

Non-Anion Gap Metabolic Acidoses

For non-renal causes of non-AG acidosis due to gastrointestinal losses of bicarbonate, NaHCO₃ 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 potassium. In patients with proximal RTA (type 1), potassium administration is typically required. An oral solution of a combination 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 and is available commercially as Virtrate-3. The syrup preparation is not recommended for chronic administration. In classical distal RTA (type 2), potassium should be administered in the acutely acidotic patient

with hypokalemia. For chronic therapy, most patients respond to replacement with either sodium citrate (Shohl's solution) or NaHCO₃ tablets (650-mg tablets contain 7.8 meq) with the goal of correcting the serum [HCO₃⁻] to normal. These 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 cRTA, but in addition, effort toward correcting the plasma [K⁺] is necessary. This latter goal deserves emphasis because restoration of normokalemia increases urinary net acid excretion and in that way can greatly improve the metabolic acidosis. Chronic administration of oral sodium polystyrene sulfonate (15 g of powder prepared as an oral solution, and without sorbitol, once daily 2–3 times per week) is sometimes used. Additionally, the diet should be low in potassium-containing foods, all potassium-retaining medications should be discontinued, and a loop diuretic may be administered. The recent release of a new a non-absorbed, calcium-potassium cation exchange polymer, patiromer, may prove to be very useful for type 4 RTA patients with significant hyperkalemia. However, patiromer has not yet been investigated in this population of patients. Finally, patients with demonstrated adrenal insufficiency should also receive fludocortisolone, but the dose varies with the cause of the hormone deficiency, and should be assiduously avoided in patients with hyporeninemic-hypoaldosteronism.

METABOLIC ALKALOSIS

Metabolic alkalosis is established by an elevated arterial pH, an increase in the serum [HCO₃⁻], and an increase in Paco₂ as a result of compensatory alveolar hypoventilation (Table 51-1). It is often accompanied by hypochloremia and hypokalemia. The arterial pH establishes the diagnosis, because it is increased in metabolic alkalosis and decreased in respiratory acidosis. Metabolic alkalosis frequently occurs as a mixed acid base disorder in association with either respiratory acidosis, respiratory alkalosis, or metabolic acidosis.

PATHOGENESIS

Metabolic alkalosis occurs as a result of net gain of [HCO₃⁻] or loss of nonvolatile acid (usually HCl by vomiting) from the extracellular fluid. When vomiting causes loss of HCl from the stomach, HCO₃⁻ secretion cannot be initiated in the small bowel and thus HCO₃⁻ is added to the extracellular fluid. Thus, vomiting or nasogastric (NG) suction is an example of the *generation stage*, in which the loss of acid typically causes alkalosis. Upon cessation of vomiting, the *maintenance stage*, typically ensues because secondary factors prevent the kidneys from compensating by excreting HCO₃⁻.

Maintenance of metabolic alkalosis, therefore, represents a failure of the kidneys to eliminate excess HCO₃⁻ 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; or (2) hypokalemia exists because of autonomous hyperaldosteronism. In the first example, alkalosis is corrected by 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 administration.

DIFFERENTIAL DIAGNOSIS

To establish the cause of metabolic alkalosis (Table 51-6), it is necessary to assess the status of the extracellular fluid volume (ECFV), the recumbent and upright blood pressure (to determine if orthostasis is present), the serum [K⁺], and in some circumstances, an assessment of 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 normal values for both the urine [Na⁺] and [Cl⁻], in a patient who is not taking diuretics, suggest primary mineralocorticoid excess. The combination of hypokalemia

TABLE 51-6 Causes of Metabolic Alkalosis

- I. Exogenous HCO₃⁻ 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 chlорidorrhea
 - 4. Gastrocystoplasty
 - 5. Villous adenoma
 - B. Renal origin
 - 1. Diuretics
 - 2. Posthypercapnic state
 - 3. Hypercalcemia/hypoparathyroidism
 - 4. Recovery from lactic acidosis or ketoacidosis
 - 5. Nonreabsorbable anions including penicillin, carbenicillin
 - 6. Mg²⁺ 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.

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 is 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 and has low concentrations of Na⁺, K⁺, and Cl⁻, the most likely possibilities are prior vomiting, the posthypercapnic state, or prior diuretic ingestion. If, on the other hand, neither the urine sodium, potassium, nor chloride concentrations are 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 because of hypocaliuria in the latter disorder.

Alkali Administration Chronic administration of alkali to individuals with normal renal function rarely causes alkalosis. However,

in patients with coexistent hemodynamic disturbances associated with effective ECF volume depletion, alkalosis can develop because the normal capacity to excrete HCO_3^- is diminished or there may be enhanced reabsorption of HCO_3^- . Such patients include those who receive NaHCO_3 (PO or IV), citrate loads (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 (an often overlooked source of alkali loads) 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 will cause excretion of alkaline urine that is high in potassium. 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. 252) 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 $[\text{HCO}_3^-]$ increases because the reduced ECFV “contracts” around the $[\text{HCO}_3^-]$ in the 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. 309.

NONREABSORBABLE ANIONS AND MAGNESIUM DEFICIENCY Administration of large quantities of the penicillin derivatives carbenicillin or tricarboxylic acid cause their nonreabsorbable anions to appear in the urine. 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 potentiates the development of hypokalemic alkalosis by enhancing distal acidification through stimulation of renin and hence aldosterone secretion.

POTASSIUM DEPLETION Chronic K^+ depletion may cause metabolic alkalosis by increasing urinary acid excretion. The renal generation of NH_4^+ (ammoniogenesis) is upregulated directly by hypokalemia. Chronic K^+ deficiency also upregulates the renal H^+ , K^+ -ATPase to increase K^+ absorption at the expense of enhanced 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 concomitant 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 by treatment of the underlying disorder, such as correction shock or severe volume depletion by volume restoration, or with insulin therapy, respectively, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^- . Exogenous sources of HCO_3^- will be additive with that amount generated by organic anion metabolism to

create a surfeit of HCO_3^- . Acidosis-induced contraction of the ECFV and K^+ deficiency act in concert to sustain the 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 effect of the persistently elevated $[\text{HCO}_3^-]$ when the elevated Paco_2 is abruptly returned toward normal.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV EXPANSION, HYPERTENSION, AND HYPERALDOSTERONISM

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 is due to upregulation of the epithelial Na^+ channels in the collecting tubule to aldosterone, as a result of the associated ECFV expansion, causes hypertension. The kaliuresis persists because of mineralocorticoid excess and distal Na^+ absorption causing enhanced K^+ excretion, continued K^+ depletion with polydipsia, inability to concentrate the urine, and polyuria.

Liddle's syndrome (Chap. 309) results from an inherited gain of function mutation of genes that regulate the collecting duct Na^+ channel (ENaC). Liddle's is a rare monogenic form of hypertension due to volume expansion manifested as hypokalemic alkalosis and normal aldosterone levels.

Symptoms With metabolic alkalosis, changes in CNS and peripheral nervous system function are similar to those of hypocalcemia (Chap. 402); symptoms include mental confusion; obtundation; and a predisposition to seizures, paresthesia, muscular cramping, tetany, aggravation of arrhythmias, and hypoxemia in chronic obstructive pulmonary disease. Related electrolyte abnormalities include hypokalemia and hypophosphatemia.

TREATMENT

Metabolic Alkalosis

This is primarily directed at correcting the underlying stimulus for HCO_3^- generation. If primary aldosteronism or Cushing's syndrome is present, correction of the underlying cause, when successful, will reverse the hypokalemia and alkalosis. $[\text{H}^+]$ 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 remove the 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 preclude infusion of saline, renal HCO_3^- loss can be accelerated by administration of acetazolamide, a carbonic anhydrase inhibitor (125–250 mg IV), which is usually effective in patients with adequate renal function but can worsen K^+ losses. Dilute hydrochloric acid (0.1 N HCl) has been advocated historically in extreme cases, but can cause hemolysis, and must be delivered slowly in a central vein. This preparation is not available generally and must be mixed by the pharmacist. Because serious errors or harm may occur, its use is not recommended.

RESPIRATORY ACIDOSIS

Respiratory acidosis can be due to 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 51-7). In acute respiratory acidosis, there is a compensatory elevation (due to cellular buffering mechanisms) in HCO_3^- , which increases 1 mmol/L for every 10-mmHg increase in Paco_2 . In chronic respiratory acidosis (>24 h), renal adaptation increases the $[\text{HCO}_3^-]$ 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 51-7 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
1.	Emphysema
2.	Pneumoconiosis
3.	Bronchitis
4.	Adult respiratory distress syndrome
5.	Barotrauma
D.	Neuromuscular
1.	Polioymelitis
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 may cause anxiety, dyspnea, confusion, psychosis, and hallucinations and may progress to coma. Lesser degrees of dysfunction in chronic hypercapnia include sleep disturbances; 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 due to vasoconstriction secondary to loss of the vasodilator effects of CO_2 .

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. 290 and 291). 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 and supervised, may result in respiratory acidosis, particularly if CO_2 production suddenly rises (because of fever, agitation, sepsis, or overfeeding) or alveolar ventilation falls 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. 279). Permissive hypercapnia may be used to minimize intrinsic positive end-expiratory pressure in acute lung injury/acute respiratory distress syndrome and severe obstructive lung disease. The respiratory acidosis associated with permissive hypercapnia may require administration of NaHCO_3 to increase the arterial pH to ~7.15–7.20, but correction of the acidemia to a normal arterial pH is deleterious.

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. 279), 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 pulmonary disease and chronic CO_2 retention who are breathing spontaneously (Chap. 286). 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 measures aimed at improving lung function (Chap. 286) should be the primary focus of treatment.

RESPIRATORY ALKALOSIS

Alveolar hyperventilation decreases Paco_2 and increases the $\text{HCO}_3^-/\text{Paco}_2$ ratio, thus increasing pH (Table 51-7). 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 $[\text{HCO}_3^-]$ appear to vary proportionately with Paco_2 over a range from 40–15 mmHg. The relationship between arterial $[\text{H}^+]$ concentration and Paco_2 is $\sim 0.7 \text{ mmol/L per mmHg}$ (or 0.01 pH unit/mmHg), and that for plasma $[\text{HCO}_3^-]$ is $0.2 \text{ mmol/L per mmHg}$. 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 $[\text{HCO}_3^-]$ and a 0.3-mmol/L decrease (or 0.003 increase in pH) in $[\text{H}^+]$.

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 $[\text{Ca}^{2+}]$ 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 the most common cause of drug-induced respiratory alkalosis as a result of direct stimulation of the medullary chemoreceptor (Chap. 449). 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 $[\text{K}^+]$ is often reduced and the $[\text{Cl}^-]$ 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 . If the hypocapnia persists for >3 –5 days, chronic respiratory alkalosis is present, and the decline in Paco_2 reduces the serum $[\text{HCO}_3^-]$ by 4–5 mmol/L for each 10-mmHg decrease in Paco_2 . It is unusual to observe a plasma $\text{HCO}_3^- < 12 \text{ mmol/L}$ as a result of a pure respiratory alkalosis. Moreover, the compensatory reduction in the plasma HCO_3^- concentration is so

effective in chronic respiratory alkalosis that the pH does not decline significantly from the normal value. In this regard, chronic respiratory alkalosis is the only acid-base disorder that may return the pH to the normal value.

When a 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, tidal volume, and frequency 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.

FURTHER READING

- BEREND K, et al: Physiological approach to assessment of acid-base disturbances. *N Engl J Med* 371:1434, 2014.
- DUBOSE TD: Disorders of acid-base balance. In *Brenner and Rector's The Kidney*, 10th ed. Skorecki K, et al. (eds). Philadelphia, Elsevier, 2016, pp. 511–558.
- DUBOSE TD: Etiologic causes of metabolic acidosis I: The high anion gap acidosis, In *Metabolic Acidosis*. Wesson DE (ed). New York, Springer, 2016, pp. 17–26.
- 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.
- KURTZ I, et al: Acid-base analysis: A critique of the Stewart and bicarbonate-centered approaches. *Am J Physiol Renal Physiol* 294: F1009, 2008.
- PALMER BF, Clegg DJ: Electrolyte and acid-base disturbances in patients with diabetes mellitus. *N Engl J Med* 373:548, 2015.

Section 8 Alterations in the Skin

52

Approach to the Patient with a Skin Disorder

Kim B. Yancey, Thomas J. Lawley



The challenge of examining the skin lies in distinguishing normal from abnormal findings, distinguishing significant findings from trivial ones, and integrating pertinent signs and symptoms into an appropriate differential diagnosis. The fact that the largest organ in the body is visible is both an advantage and a disadvantage to those who examine it. It is advantageous because no special instrumentation is necessary and because the skin can be biopsied with little morbidity. However, the casual observer can be misled by a variety of stimuli and overlook important, subtle signs of skin or systemic disease. For instance, the sometimes minor differences in color and shape that distinguish a melanoma (Fig. 52-1) from a benign nevomelanocytic nevus (Fig. 52-2) can be difficult to recognize. A variety of descriptive terms have been developed that characterize cutaneous lesions (Tables 52-1, 52-2, and 52-3; Fig. 52-3), thereby aiding in their interpretation and in



FIGURE 52-1 Superficial spreading melanoma. This is the most common type of melanoma. Such lesions usually demonstrate asymmetry, border irregularity, color variegation (black, blue, brown, pink, and white), a diameter >6 mm, and a history of change (e.g., an increase in size or development of associated symptoms such as pruritus or pain).



FIGURE 52-2 Nevomelanocytic nevus. Nevi are benign proliferations of nevomelanocytes characterized by regularly shaped hyperpigmented macules or papules of a uniform color.

TABLE 52-2 Description of Secondary Skin Lesions

Lichenification: A distinctive thickening of the skin that is characterized by accentuated skin-fold markings.

Scale: Excessive accumulation of stratum corneum.

Crust: Dried exudate of body fluids that may be either yellow (i.e., serous crust) or red (i.e., hemorrhagic crust).

Erosion: Loss of epidermis without an associated loss of dermis.

Ulcer: Loss of epidermis and at least a portion of the underlying dermis.

Excoriation: Linear, angular erosions that may be covered by crust and are caused by scratching.

Atrophy: An acquired loss of substance. In the skin, this may appear as a depression with intact epidermis (i.e., loss of dermal or subcutaneous tissue) or as sites of shiny, delicate, wrinkled lesions (i.e., epidermal atrophy).

Scar: A change in the skin secondary to trauma or inflammation. Sites may be erythematous, hypopigmented, or hyperpigmented depending on their age or character. Sites on hair-bearing areas may be characterized by destruction of hair follicles.

TABLE 52-3 Common Dermatologic Terms

Alopecia: Hair loss, partial or complete.

Annular: Ring-shaped.

Cyst: A soft, raised, encapsulated lesion filled with semisolid or liquid contents.

Herpetiform: In a grouped configuration.

Lichenoid eruption: Violaceous to purple, polygonal lesions that resemble those seen in lichen planus.

Milia: Small, firm, white papules filled with keratin.

Morbilliform rash: Generalized, small erythematous macules and/or papules that resemble lesions seen in measles.

Nummular: Coin-shaped.

Poikiloderma: Skin that displays variegated pigmentation, atrophy, and telangiectases.

Polycyclic lesions: A configuration of skin lesions formed from coalescing rings or incomplete rings.

Pruritus: A sensation that elicits the desire to scratch. Pruritus is often the predominant symptom of inflammatory skin diseases (e.g., atopic dermatitis, allergic contact dermatitis); it is also commonly associated with xerosis and aged skin. Systemic conditions that can be associated with pruritus include chronic renal disease, cholestasis, pregnancy, malignancy, thyroid disease, polycythemia vera, and delusions of parasitosis.

TABLE 52-1 Description of Primary Skin Lesions

Macule: A flat, colored lesion, <2 cm in diameter, not raised above the surface of the surrounding skin. A “freckle,” or ephelid, is a prototypical pigmented macule.

Patch: A large (>2 cm) flat lesion with a color different from the surrounding skin. This differs from a macule only in size.

Papule: A small, solid lesion, <0.5 cm in diameter, raised above the surface of the surrounding skin and thus palpable (e.g., a closed comedone, or whitehead, in acne).

Nodule: A larger (0.5 to 5.0 cm), firm lesion raised above the surface of the surrounding skin. This differs from a papule only in size (e.g., a large dermal nevomelanocytic nevus).

Tumor: A solid, raised growth >5 cm in diameter.

Plaque: A large (>1 cm), flat-topped, raised lesion; edges may either be distinct (e.g., in psoriasis) or gradually blend with surrounding skin (e.g., in eczematous dermatitis).

Vesicle: A small, fluid-filled lesion, <0.5 cm in diameter, raised above the plane of surrounding skin. Fluid is often visible, and the lesions are translucent (e.g., vesicles in allergic contact dermatitis caused by *Toxicodendron* [poison ivy]).

Pustule: A vesicle filled with leukocytes. Note: The presence of pustules does not necessarily signify the existence of an infection.

Bulla: A fluid-filled, raised, often translucent lesion >0.5 cm in diameter.

Wheal: A raised, erythematous, edematous papule or plaque, usually representing short-lived vasodilation and vasopermeability.

Telangiectasia: A dilated, superficial blood vessel.

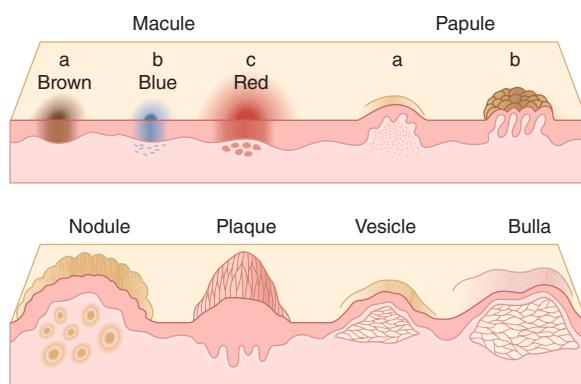


FIGURE 52-3 A schematic representation of several common primary skin lesions (see Table 52-1).

TABLE 52-4 Selected Common Dermatologic Conditions

DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY	DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY
Acne vulgaris	Face, upper back, chest	Open and closed comedones, erythematous papules, pustules, cysts	Seborrheic keratosis	Trunk, face, extremities	Brown plaques with adherent, greasy scale; “stuck on” appearance
Rosacea	Blush area of cheeks, nose, forehead, chin	Erythema, telangiectases, papules, pustules	Folliculitis Impetigo	Any hair-bearing area Anywhere	Follicular pustules Papules, vesicles, pustules, often with honey-colored crusts
Seborrheic dermatitis	Scalp, eyebrows, perinasal areas	Erythema with greasy yellow-brown scale	Herpes simplex	Lips, genitalia	Grouped vesicles progressing to crusted erosions
Atopic dermatitis	Antecubital and popliteal fossae; may be widespread	Patches and plaques of erythema, scaling, and lichenification; pruritus	Herpes zoster	Dermatomal, usually trunk but may be anywhere	Vesicles limited to a dermatome (often painful)
Stasis dermatitis	Ankles, lower legs over medial malleoli	Patches of erythema and scaling on background of hyperpigmentation associated with signs of venous insufficiency	Varicella	Face, trunk, relative sparing of extremities	Lesions arise in crops and quickly progress from erythematous macules, to papules, to vesicles, to pustules, to crusted sites.
Dyshidrotic eczema	Palms, soles, sides of fingers and toes	Deep vesicles	Pityriasis rosea	Trunk (Christmas tree pattern); herald patch followed by multiple smaller lesions	Symmetric erythematous papules and plaques with a collarette of scale
Allergic contact dermatitis	Anywhere	Localized erythema, vesicles, scale, and pruritus (e.g., fingers, earlobes—nickel; dorsal aspect of foot—shoe; exposed surfaces—poison ivy)	Tinea versicolor	Chest, back, abdomen, proximal extremities	Scaly hyper- or hypopigmented macules
Psoriasis	Elbows, knees, scalp, lower back, fingernails (may be generalized)	Papules and plaques covered with silvery scale; nails have pits	Candidiasis	Groin, beneath breasts, vagina, oral cavity	Erythematous macerated areas with satellite pustules; white, friable patches on mucous membranes
Lichen planus	Wrists, ankles, mouth (may be widespread)	Violaceous flat-topped papules and plaques	Dermatophytosis	Feet, groin, beard, or scalp	Varies with site (e.g., tinea corporis—scaly annular plaque)
Keratosis pilaris	Extensor surfaces of arms and thighs, buttocks	Keratotic follicular papules with surrounding erythema	Scabies	Groin, axillae, between fingers and toes, beneath breasts	Excoriated papules, burrows, pruritus
Melasma	Forehead, cheeks, temples, upper lip	Tan to brown patches	Insect bites	Anywhere	Erythematous papules with central puncta
Vitiligo	Periorificial, trunk, extensor surfaces of extremities, flexor wrists, axillae	Chalk-white macules	Cherry angioma Keloid Dermatofibroma	Trunk Anywhere (site of previous injury) Anywhere	Red, blood-filled papules Firm tumor, pink, purple, or brown Firm red to brown nodule that shows dimpling of overlying skin with lateral compression
Actinic keratosis	Sun-exposed areas	Skin-colored or red-brown macule or papule with dry, rough, adherent scale	Acrochordons (skin tags)	Groin, axilla, neck	Fleshy papules
Basal cell carcinoma	Face	Papule with pearly, telangiectatic border on sun-damaged skin	Urticaria	Anywhere	Wheals, sometimes with surrounding flare; pruritus
Squamous cell carcinoma	Face, especially lower lip, ears	Indurated and possibly hyperkeratotic lesions often showing ulceration and/or crusting	Transient acantholytic dermatosis Xerosis	Trunk, especially anterior chest Extensor extremities, especially legs	Erythematous papules Dry, erythematous, scaling patches; pruritus

the formulation of a differential diagnosis (**Table 52-4**). For example, the finding of scaling papules, which are present in psoriasis or atopic dermatitis, places the patient in a different diagnostic category than would hemorrhagic papules, which may indicate vasculitis or sepsis (**Figs. 52-4 and 52-5, respectively**). It is also important to differentiate primary from secondary skin lesions. If the examiner focuses on linear erosions overlying an area of erythema and scaling, he or she may incorrectly assume that the erosion is the primary lesion and that the redness and scale are secondary, whereas the correct interpretation would be that the patient has a pruritic eczematous dermatitis with erosions caused by scratching.

APPROACH TO THE PATIENT

Skin Disorder

In examining the skin it is usually advisable to assess the patient before taking an extensive history. This approach ensures that the entire cutaneous surface will be evaluated, and objective findings can be integrated with relevant historical data. Four basic features of a skin lesion must be noted and considered during a physical examination: the *distribution* of the eruption, the *types* of primary and secondary lesions, the *shape* of individual lesions, and the *arrangement* of the lesions. An ideal skin examination includes evaluation



FIGURE 52-4 Necrotizing vasculitis. Palpable purpuric papules on the lower legs are seen in this patient with cutaneous small-vessel vasculitis. (Courtesy of Robert Swerlick, MD; with permission.)

of the skin, hair, and nails as well as the mucous membranes of the mouth, eyes, nose, nasopharynx, and anogenital region. In the initial examination, it is important that the patient be disrobed as completely as possible to minimize chances of missing important individual skin lesions and permit accurate assessment of the distribution of the eruption. The patient should first be viewed from a distance of about 1.5–2 m (4–6 ft) so that the general character of the skin and the distribution of lesions can be evaluated. Indeed, the distribution of lesions often correlates highly with diagnosis (Fig. 52-6). For example, a hospitalized patient with a generalized erythematous exanthem is more likely to have a drug eruption than is a patient with a similar rash limited to the sun-exposed portions of the face. Once the distribution of the lesions has been established, the nature of the primary lesion must be determined. Thus, when lesions are distributed on elbows, knees, and scalp, the most likely possibility based solely on distribution is psoriasis or dermatitis herpetiformis (Figs. 52-7 and 52-8, respectively). The primary lesion in psoriasis is a scaly papule that soon forms erythematous plaques covered with a white scale, whereas that of dermatitis herpetiformis is an urticarial papule that quickly becomes a small vesicle. In this manner, identification of the primary lesion directs the examiner toward the proper diagnosis. Secondary changes in skin can also be quite helpful. For example, scale represents excessive epidermis, while



FIGURE 52-5 Meningococcemia. An example of fulminant meningococcemia with extensive angular purpuric patches. (Courtesy of Stephen E. Gellis, MD; with permission.)

crust is the result of a discontinuous epithelial cell layer. Palpation of skin lesions can yield insight into the character of an eruption. Thus, red papules on the lower extremities that blanch with pressure can be a manifestation of many different diseases, but hemorrhagic red papules that do not blanch with pressure indicate palpable purpura characteristic of necrotizing vasculitis (Fig. 52-4).

The shape of lesions is also an important feature. Flat, round, erythematous papules and plaques are common in many cutaneous diseases. However, target-shaped lesions that consist in part of erythematous plaques are specific for erythema multiforme (Fig. 52-9). Likewise, the arrangement of individual lesions is important. Erythematous papules and vesicles can occur in many conditions, but their arrangement in a specific linear array suggests an external etiology such as allergic contact dermatitis (Fig. 52-10) or primary irritant dermatitis. In contrast, lesions with a generalized arrangement are common and suggest a systemic etiology.

As in other branches of medicine, a complete history should be obtained to emphasize the following features:

1. Evolution of lesions
 - a. Site of onset
 - b. Manner in which the eruption progressed or spread
 - c. Duration
 - d. Periods of resolution or improvement in chronic eruptions
2. Symptoms associated with the eruption
 - a. Itching, burning, pain, numbness
 - b. What, if anything, has relieved symptoms
 - c. Time of day when symptoms are most severe
3. Current or recent medications (prescribed as well as over-the-counter)
4. Associated systemic symptoms (e.g., malaise, fever, arthralgias)
5. Ongoing or previous illnesses
6. History of allergies
7. Presence of photosensitivity
8. Review of systems
9. Family history (particularly relevant for patients with melanoma, atopy, psoriasis, or acne)
10. Social, sexual, or travel history

■ DIAGNOSTIC TECHNIQUES

Many skin diseases can be diagnosed on the basis of gross clinical appearance, but sometimes relatively simple diagnostic procedures can yield valuable information. In most instances, they can be performed at the bedside with a minimum of equipment.

Skin Biopsy A skin biopsy is a straightforward minor surgical procedure; however, it is important to biopsy a lesion that is most likely to yield diagnostic findings. This decision may require expertise in skin diseases and knowledge of superficial anatomic structures in selected areas of the body. In this procedure, a small area of skin is anesthetized with 1% lidocaine with or without epinephrine. The skin lesion in question can be excised or saucerized with a scalpel or removed by punch biopsy. In the latter technique, a punch is pressed against the surface of the skin and rotated with downward pressure until it penetrates to the subcutaneous tissue. The circular biopsy is then lifted with forceps, and the bottom is cut with iris scissors. Biopsy sites may or may not need suture closure, depending on size and location.

KOH Preparation A potassium hydroxide (KOH) preparation is performed on scaling skin lesions where a fungal infection is suspected. The edge of such a lesion is scraped gently with a no. 15 scalpel blade. The removed scale is collected on a glass microscope slide and then treated with 1 or 2 drops of a solution of 10–20% KOH. KOH dissolves keratin and allows easier visualization of fungal elements. Brief heating of the slide accelerates dissolution of keratin. When the preparation is viewed under the microscope, the refractile hyphae are seen more easily when the light intensity is reduced and the condenser is lowered. This technique can be used to identify hyphae in dermatophyte infections, pseudohyphae and budding yeasts in *Candida* infections,

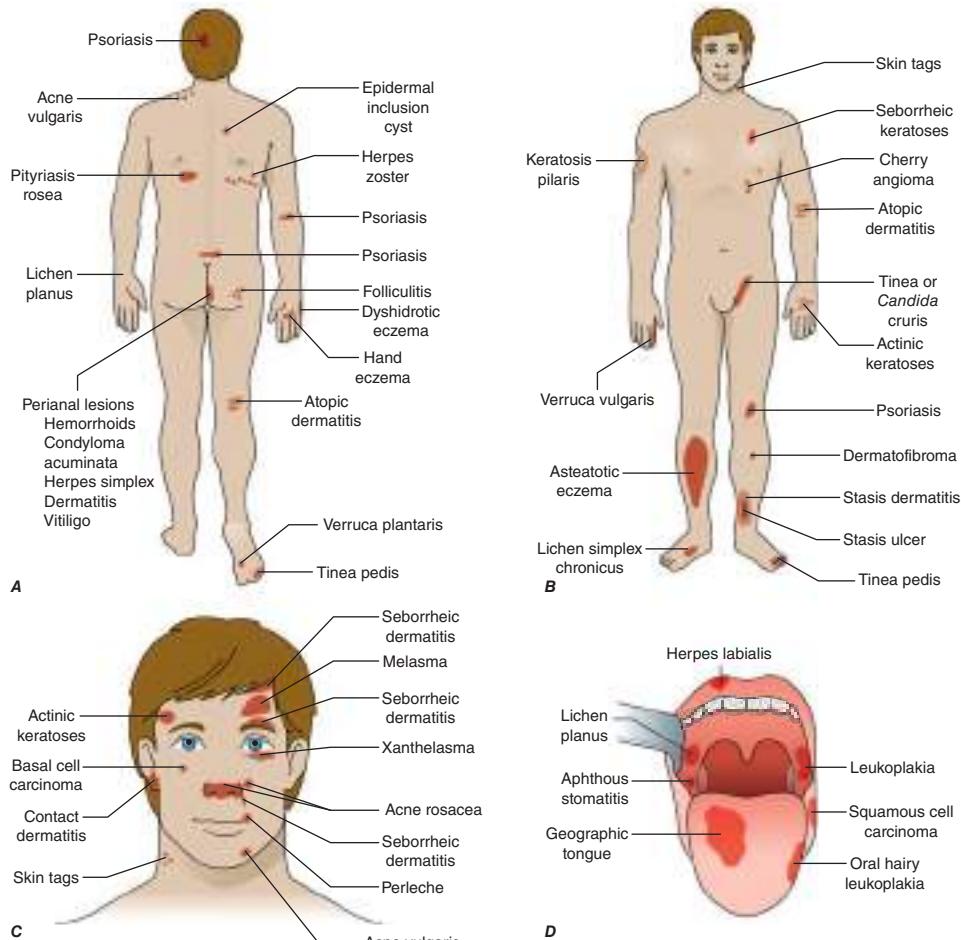


FIGURE 52-6 Distribution of some common dermatologic diseases and lesions.

and “spaghetti and meatballs” yeast forms in tinea versicolor. The same sampling technique can be used to obtain scale for culture of selected pathogenic organisms.

Tzanck Smear A Tzanck smear is a cytologic technique most often used in the diagnosis of herpesvirus infections (herpes simplex virus [HSV] or varicella zoster virus [VZV]) (see Figs. 188-1 and 188-3). An early vesicle, not a pustule or crusted lesion, is unroofed, and the base of the lesion is scraped gently with a scalpel blade. The material is placed on a glass slide, air-dried, and stained with Giemsa or Wright’s stain. Multinucleated epithelial giant cells suggest the presence of HSV or VZV; culture, immunofluorescence microscopy, or genetic testing must be performed to identify the specific virus.

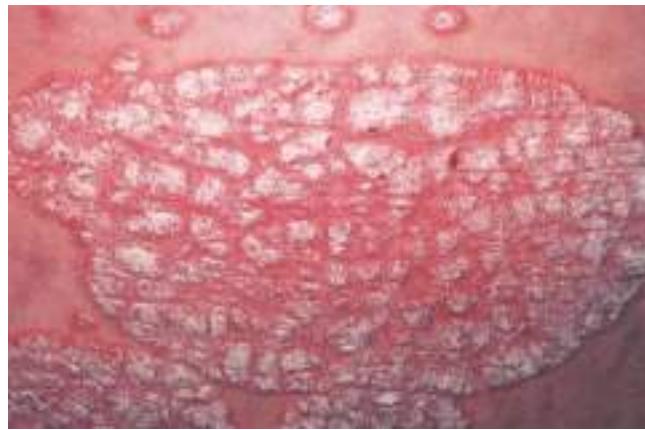


FIGURE 52-7 Psoriasis. This papulosquamous skin disease is characterized by small and large erythematous papules and plaques with overlying adherent silvery scale.

Diascopy Diascopy is designed to assess whether a skin lesion will blanch with pressure as, for example, in determining whether a red lesion is hemorrhagic or simply blood-filled. Urticaria (Fig. 52-11) will blanch with pressure, whereas a purpuric lesion caused by necrotizing vasculitis (Fig. 52-4) will not. Diascopy is performed by pressing a microscope slide or magnifying lens against a lesion and noting the amount of blanching that occurs. Granulomas often have an opaque to transparent, brown-pink “apple jelly” appearance on diascopy.

Wood’s Light A Wood’s lamp generates 360-nm ultraviolet (“black”) light that can be used to aid the evaluation of certain skin disorders. For example, a Wood’s lamp will cause erythrasma (a superficial, intertriginous infection caused by *Corynebacterium minutissimum*)



FIGURE 52-8 Dermatitis herpetiformis. This disorder typically displays pruritic, grouped papulovesicles on elbows, knees, buttocks, and posterior scalp. Vesicles are often excoriated due to associated pruritus.



FIGURE 52-9 Erythema multiforme. This eruption is characterized by multiple erythematous plaques with a target or iris morphology. It usually represents a hypersensitivity reaction to drugs (e.g., sulfonamides) or infections (e.g., HSV). (Courtesy of the Yale Resident's Slide Collection; with permission.)



FIGURE 52-10 Allergic contact dermatitis (ACD). **A.** An example of ACD in its acute phase, with sharply demarcated, weeping, eczematous plaques in a perioral distribution. **B.** ACD in its chronic phase, with an erythematous, lichenified, weeping plaque on skin chronically exposed to nickel in a metal snap. (**B.** Courtesy of Robert Swerlick, MD; with permission.)



FIGURE 52-11 Urticaria. Discrete and confluent, edematous, erythematous papules and plaques are characteristic of this whealing eruption.



FIGURE 52-12 Vitiligo. Characteristic lesions display an acral distribution and striking depigmentation as a result of loss of melanocytes.

to show a characteristic coral pink color, and wounds colonized by *Pseudomonas* will appear pale blue. Tinea capitis caused by certain dermatophytes (e.g., *Microsporum canis* or *M. audouinii*) exhibits a yellow fluorescence. Pigmented lesions of the epidermis such as freckles are accentuated, while dermal pigment such as postinflammatory hyperpigmentation fades under a Wood's light. Vitiligo (Fig. 52-12) appears totally white under a Wood's lamp, and previously unsuspected areas of involvement often become apparent. A Wood's lamp may also aid in the demonstration of tinea versicolor, sites of depigmentation within and/or surrounding melanomas, and in recognition of ash leaf spots in patients with tuberous sclerosis.

Patch Tests Patch testing is designed to document sensitivity to a specific antigen. In this procedure, a battery of suspected allergens is applied to the patient's back under occlusive dressings and allowed to remain in contact with the skin for 48 h. The dressings are removed, and the area is examined for evidence of delayed hypersensitivity reactions (e.g., erythema, edema, or papulovesicles). This test is best performed by physicians with special expertise in patch testing and is often helpful in the evaluation of patients with chronic dermatitis.

FURTHER READING

- BOLOGNA JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.
GOLDSMITH LA et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 8th ed. New York, McGraw-Hill, 2012.
JAMES WD: *Andrews' Diseases of the Skin: Clinical Dermatology*, 12th ed. Philadelphia, Elsevier, 2016.

53

Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders

Leslie P. Lawley, Calvin O. McCall,
Thomas J. Lawley



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.

TABLE 53-1 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

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 53-1.

 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.

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 pattern is typified by dermatitis of flexural skin, particularly in the antecubital and popliteal fossae (Fig. 53-1). 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



FIGURE 53-1 Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. (Courtesy of Robert Swerlick, MD; with permission.)

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.

TREATMENT

Atopic Dermatitis

Therapy for AD should include avoidance of cutaneous irritants, adequate moisturizing 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. Two nonglucocorticoid anti-inflammatory agents are available: tacrolimus ointment and pimecrolimus cream. These agents are macrolide immunosuppressants that are approved by the U.S. Food and Drug Administration (FDA) for topical use in AD. Reports of broader effectiveness appear in the literature. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. However, concerns have emerged regarding the potential for lymphomas in patients treated with these agents. Thus, caution should be exercised when these agents are considered. Currently, they are also 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, because 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), or 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 H₂ 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. 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 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. 52, Fig. 52-10).

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 much 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, vesication, 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 those 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. Allergic contact dermatitis 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, intensely pruritic, small papules and vesicles on the thenar and hypothenar eminences and the sides of the fingers (Fig. 53-2). Lesions tend to occur in crops that slowly form crusts and then heal.

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.



FIGURE 53-2 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.

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 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.

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 a large number of 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. 53-3](#)).



FIGURE 53-3 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.

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.

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. Secondarily infected lesions should be treated with appropriate oral antibiotics, but it should be noted that all ulcers will become colonized with bacteria, and the purpose of antibiotic therapy should not be to clear all bacterial growth. Care must be taken to exclude treatable causes of leg ulcers (hypercoagulation, vasculitis) 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. 53-4](#)). 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.

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 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



FIGURE 53-4 Seborrheic dermatitis. Central facial erythema with overlying greasy, yellowish scale is seen in this patient. (Courtesy of Jean Bolognia, MD; with permission.)

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 53-2)

■ 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 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. 52, Fig. 52-7), 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 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).

TABLE 53-2 Papulosquamous Disorders

	CLINICAL FEATURES	OTHER NOTABLE FEATURES	HISTOLOGIC FEATURES
Psoriasis	Sharply demarcated, erythematous plaques with micaceous 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.

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 53-3). 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 is a new oral agent that 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 53-4). Experience with some of these biologic agents is limited, and information regarding combination therapy and adverse events continues to emerge. 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 TB, 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.

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. 53-5). 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 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.

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. 53-6). 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.

CUTANEOUS INFECTIONS (TABLE 53-5)

IMPETIGO, ECTHYMA, AND FURUNCULOSIS

Impetigo is a common superficial bacterial infection of skin caused most often by *S. aureus* (Chap. 142) and in some cases by group A β -hemolytic streptococci (Chap. 143). The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust (see Chap. 143, Fig. 143-3). Lesions may occur on normal skin (primary infection) or in areas already affected by another

TABLE 53-3 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

Abbreviation: FDA, Food and Drug Administration.

^aInitial test dose is required. ^bInitial dose escalation is required.

TABLE 53-4 FDA-Approved Biologics for Psoriasis or Psoriatic Arthritis

		ADMINISTRATION		FREQUENCY	WARNINGS, SELECTED
AGENT	MECHANISM OF ACTION	INDICATION	ROUTE		
Etanercept	Anti-TNF- α	Ps, PsA	SC	Once or twice weekly ^a	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Adalimumab	Anti-TNF- α	Ps, PsA	SC	Every other week ^a	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Infliximab	Anti-TNF- α	Ps, PsA	IV	Every 8 weeks ^a	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Golimumab	Anti-TNF- α	PsA	SC	Every 4 or 8 weeks	Serious infections, hepatotoxicity, CHF, hypersensitivity reactions, neurologic events, potential for increased malignancies
Ustekinumab	Anti-IL-12 and anti-IL-23	Ps, PsA	SC	Every 12 weeks ^a	Serious infections, neurologic events, potential for increased malignancies
Certolizumab pegol	Anti-TNF- α	PsA	SC	Every 2 or 4 weeks ^a	Serious infections, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies, hepatotoxicity
Secukinumab	Anti-IL-17	Ps, PsA	SC	Every 4 weeks ^a	Serious infections, hypersensitivity reaction, inflammatory bowel disease
Ixekizumab	Anti-IL-17	Ps	SC	Every 4 weeks ^a	Serious infections, hypersensitivity reaction, inflammatory bowel disease

^aInitial 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 .

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 nonbullos 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, which is facilitated by the use of 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 last decade 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. 124.

■ DERMATOPHYTOSIS

Dermatophytes are fungi that infect skin, hair, and nails and include members of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton* (Chap. 214). *Tinea corporis*, or infection of the relatively hairless skin of the body (glabrous skin), may have a variable appearance depending



FIGURE 53-5 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.)



FIGURE 53-6 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.

TABLE 53-5 Common Skin Infections

	CLINICAL FEATURES	ETOLOGIC AGENT	TREATMENT
Impetigo	Honey-colored crusted papules, plaques, or bullae	Group A <i>Streptococcus</i> and <i>Staphylococcus aureus</i>	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	<i>Trichophyton</i> , <i>Epidermophyton</i> , or <i>Microsporum</i> 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	<i>Candida albicans</i> and other <i>Candida</i> 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	<i>Malassezia furfur</i>	Topical selenium sulfide lotion or azoles

Abbreviation: KOH, potassium hydroxide.

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 vesication. 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 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*.

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 of 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 newer 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. 211). 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. 197). 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, 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 wetness 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 most 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, non verrucous 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. 85). 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. 72), 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 a majority of 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 new 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. 187.

■ HERPES ZOSTER

See Chap. 188.

■ 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 after 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 (*Propionibacterium 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 large 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. 53-7). 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.

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 oral



FIGURE 53-7 Acne vulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. (Courtesy of Kalman Watsky, MD; with permission.)

contraceptive pills, 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 *P. 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 azelaic acid, erythromycin, clindamycin, or dapsone) are also useful adjuncts to therapy. Benzoyl peroxide products should be used in combination with topical antibiotics (erythromycin and clindamycin) 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 tetracycline in doses of 250–500 mg bid or doxycycline in doses of 100 mg bid. Minocycline is also useful. Such antibiotics appear to have anti-inflammatory effects independent of their antibacterial effects. If the patient is not showing appropriate response within 3 months, changes in the plan should be considered. 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.

Patients with severe nodulocystic acne unresponsive to the therapies discussed above may benefit from treatment with the synthetic retinoid isotretinoin. Its dose is based on the patient's weight, and it is given once daily for 5 months. 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 extremely 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.



FIGURE 53-8 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.)

■ 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, telangiectases, and superficial pustules (Fig. 53-8) but is not associated with the presence of comedones. Rosacea rarely involves the chest or back.

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 telangiectases 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 metronidazole, sodium sulfacetamide, azelaic acid, topical ivermectin, or topical brimonidine. More severe disease requires oral tetracyclines: tetracycline, 250–500 mg bid; doxycycline, 100 mg bid; or minocycline, 50–100 mg bid. 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. S2).

FURTHER READING

- BOLOGNA JL, JORIZZO JL, SCHAFER JV (eds): *Dermatology*, 3rd ed. Philadelphia, Saunders, 2012.
- GOLDSMITH LA et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 8th ed. New York, McGraw-Hill, 2012.
- JAMES WD, BERGER TG, ELSTON DM (eds): *Andrew's Diseases of the Skin Clinical Dermatology*, 12th ed. Philadelphia, Elsevier, 2016.
- WOLFF K, JOHNSON RA, SAAVEDRA AP (eds): *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 7th ed. New York, McGraw-Hill, 2013.

54

Skin Manifestations of Internal Disease

Jean L. Bolognia, Irwin M. Braverman



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 54-1) 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

TABLE 54-1 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

^aDiscussed in detail in **Chap. 53**; cardiovascular disease and the metabolic syndrome are comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus. ^bAssociated with chronic sun exposure more often than exposure to arsenic; usually one or a few lesions. ^cSee also Red Lesions in "Papulonodular Skin Lesions." ^dAlso cutaneous lesions of HTLV-1-associated adult T cell leukemia/lymphoma. ^eSee also Red-Brown Lesions in "Papulonodular Skin Lesions."

Abbreviation: HIV, human immunodeficiency virus.

diseases—*tinea*, *psoriasis*, *pityriasis rosea*, and *lichen planus*—are primary cutaneous disorders (**Chap. 53**). 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, HIV or streptococcal infections, and a rapid taper of systemic glucocorticoids are known to exacerbate psoriasis; despite being used to treat psoriasis, TNF- α inhibitors can also induce psoriatic lesions. Comorbidities in patients with psoriasis include cardiovascular disease and metabolic syndrome.

Whenever the 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, antimarialls, quinidine, beta blockers, TNF- α inhibitors, anti-PD-1/PD-L1 Ab, and ACE inhibitors. In some populations, 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 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 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 annular plaques on the face, nonscarring alopecia, condyloma lata (broad-based and moist), and mucous patches as well as lymphadenopathy, 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 54-2) *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 54-3**); (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

TABLE 54-2 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
4. Idiopathic (usually older men)

^aDiscussed in detail in **Chap. 53**.

TABLE 54-3 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, arthritis, pustules, SAPHO syndrome ^b	Skin biopsy	Topical glucocorticoids, vitamin D; UV-B (narrowband) > PUVA; oral retinoid; MTX, cyclosporine, anti-TNF agents, apremilast, anti-IL-12/23 Ab, anti-IL-17A or -IL-17 receptor Ab
Dermatitis^a					
Atopic	Acute: Erythema, fine scale, crust, indistinct borders, excoriations Chronic: Lichenification (increased skin markings), excoriations	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 <i>Staphylococcus aureus</i> or HSV Exclude superimposed irritant or allergic contact dermatitis	Skin biopsy	Topical glucocorticoids, tacrolimus, pimecrolimus, tar, and antipruritics; oral antihistamines; open wet dressings; UV-B ± UV-A > PUVA; oral/IM glucocorticoids (short-term); MTX; mycophenolate mofetil; azathioprine; cyclosporine; anti-IL-4/13 Ab 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 re-challenge 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; perhaps anti-IL-12/23 Ab, anti-TNF agents, anti-IL-17 Ab

^aDiscussed in detail in **Chap. 53.** ^bSAPHO 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 + ultraviolet A irradiation; SAPHO, synovitis, acne, pustulosis, hyperostosis, and osteitis (a subtype is chronic recurrent multifocal osteomyelitis); TNF, tumor necrosis factor; UV-A, ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular psoriasis*.

Drug-induced erythroderma (exfoliative dermatitis) may begin as an exanthematous (morbilliform) eruption (**Chap. 56**) or may arise as diffuse erythema. A number of drugs can produce an erythroderma, including penicillins, sulfonamides, 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 reaction* (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. There have been isolated case reports of erythroderma secondary to some solid tumors—lung, liver, prostate, thyroid, and colon—but it is primarily during a late stage of the disease.

ALOPECIA

(**Table 54-4**) The two major forms of alopecia are scarring and non-scarring. *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.

TABLE 54-4 Causes of Alopecia

I. Nonscarring alopecia	
A. Primary cutaneous disorders	
1. Androgenetic alopecia	
2. Telogen effluvium	
3. Alopecia areata	
4. Tinea capitis	
5. Traumatic alopecia ^a	
6. Psoriasisiform alopecia, including TNF- α inhibitor-induced	
B. Drugs	
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. Linear morphea (linear scleroderma) ^c	
B. Systemic diseases	
1. Discoid lesions in the setting of systemic lupus erythematosus ^b	
2. Sarcoidosis	
3. Cutaneous metastases	

^aMost patients with trichotillomania or early stages of traction alopecia and some patients with pressure-induced alopecia. ^bWhile the majority of patients with discoid lesions have only cutaneous disease, these lesions do represent one of the 11 American College of Rheumatology criteria (1982) for systemic lupus erythematosus. ^cCan 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.

In *nonscarring alopecia*, the hair shafts are absent or miniaturized, but the hair follicles are preserved, explaining the reversible nature of nonscarring alopecia.

The most common causes of nonscarring alopecia include *androgenetic alopecia*, *telogen effluvium*, *alopecia areata*, *tinea capitis*, and the early phase of *traumatic alopecia* (Table 54-5). 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 enlarged clitoris, the possibility of an ovarian or adrenal gland tumor should be considered.

Exposure to various drugs can also cause diffuse hair loss, usually by inducing a telogen effluvium. An exception is the anagen effluvium observed with antimitotic 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 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 54-4).

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) and *cutaneous metastases*.

In the early phases of discoid lupus, lichen planus, and folliculitis decalvans, there are circumscribed areas of alopecia. Fibrosis and

TABLE 54-5 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 Increased levels of circulating androgens (ovarian or adrenal source in women)	If no evidence of hyperandrogenemia, then topical minoxidil; finasteride ^a ; spironolactone (women); hair transplant
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 (<i>kerion</i>) ^b	Invasion of hairs by dermatophytes, most commonly <i>Trichophyton tonsurans</i>	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	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

^aTo date, Food and Drug Administration-approved for men. ^bScarring alopecia can occur at sites of kerions. ^cMay also be scarring, especially late-stage traction alopecia.

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 54-6) 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. 53 and 55). 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*.

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).

TABLE 54-6 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
 - 1. *Sarcoidosis*
 - 2. *Subacute cutaneous lupus erythematosus*, LE tumidus
 - 3. Annular erythema of *Sjögren's syndrome*
 - 4. *Secondary syphilis* (especially the face)
 - 5. Cutaneous T cell lymphoma (especially *mycosis fungoïdes*)
 - 6. Interstitial granulomatous dermatitis^b

^aMigratory erythema with erosions; favors lower extremities and girdle area.

^bUnderlying diseases include rheumatoid arthritis, LE, and granulomatosis with polyangiitis.

Abbreviations: CDC, Centers for Disease Control and Prevention; LE, lupus erythematosus.

TABLE 54-7 Causes of Acneiform Eruptions

- I. Primary cutaneous disorders
 - A. *Acne vulgaris*
 - B. *Acne rosacea*
- II. Drugs, e.g., anabolic steroids, glucocorticoids, lithium, EGFR inhibitors, 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. *Behcet's disease*

Abbreviation: EGFR, epidermal growth factor receptor; MEK, MAP (mitogen activated protein) kinase.

ACNE

(Table 54-7) In addition to *acne vulgaris* and *acne rosacea*, the two major forms of acne (Chap. 53), there are drugs and systemic diseases that can lead to acneiform eruptions.

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 54-8) 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.

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

TABLE 54-8 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.	Cuticular/periungual
1.	Lupus erythematosus
2.	Systemic sclerosis (scleroderma)
3.	Dermatomyositis
4.	Hereditary hemorrhagic telangiectasia
E.	Papular
1.	Hereditary hemorrhagic telangiectasia
F.	Spider angioma
1.	Cirrhosis

^aBecoming less common.

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, often referred to as the CREST (calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) variant (Chap. 353), 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.

Cuticular 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, 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 54-9) 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.

The differential diagnosis of *localized hypomelanosis* includes the following primary cutaneous disorders: *idiopathic guttate hypomelanosis*, *postinflammatory hypopigmentation*, *pityriasis (tinea) versicolor*, *vitiligo*, *chemical- or drug-induced leukoderma*, *nevus depigmentosus* (see below),

TABLE 54-9 Causes of Hypopigmentation

I.	Primary cutaneous disorders
A.	Diffuse
1.	Generalized vitiligo ^a
B.	Localized
1.	Idiopathic guttate hypomelanosis
2.	Postinflammatory
3.	Pityriasis (tinea) versicolor
4.	Vitiligo ^a
5.	Chemical- or drug-induced leukoderma, e.g., topical imiquimod, oral imatinib
6.	Nevus depigmentosus
7.	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)
2.	Melanoma-associated leukoderma, spontaneous or immunotherapy-induced
3.	Vogt-Koyanagi-Harada syndrome
4.	Onchocerciasis
5.	Sarcoidosis
6.	Cutaneous T cell lymphoma (especially mycosis fungoides)
7.	Tuberculoid and indeterminate leprosy
8.	Linear nevoid hypopigmentation (hypomelanosis of Ito) ^e
9.	Incontinentia pigmenti (stage IV)
10.	Tuberous sclerosis
11.	Waardenburg syndrome and Shah-Waardenburg syndrome

^aAbsence of melanocytes in areas of leukoderma. ^bNormal number of melanocytes. ^cPlatelet storage defect and restrictive lung disease secondary to deposits of ceroid-like material or immunodeficiency; due to mutations in β or γ subunit of adaptor-related protein complex 3 as well as subunits of biogenesis of lysosome-related organelles complex (BLOC)-1, 2, and 3. ^dGiant lysosomal granules and recurrent infections. ^eMinority of patients in a nonreferral setting have systemic abnormalities (musculoskeletal, central nervous system, ocular).

TABLE 54-10 Hypopigmentation (Primary Cutaneous Disorders, Localized)

	CLINICAL CHARACTERISTICS	WOOD'S LAMP EXAMINATION (UV-A; PEAK = 365 NM)	SKIN BIOPSY SPECIMEN	PATHOGENESIS	TREATMENT
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
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
Pityriasis (tinea) versicolor	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 <i>Malassezia</i> . Yeast is lipophilic and produces C ₉ and C ₁₁ dicarboxylic acids, which <i>in vitro</i> 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
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 <i>KIT</i> protooncogene that encodes the tyrosine kinase receptor for stem cell growth factor (kit ligand)	None; occasionally transplants

Abbreviations: MBEH, monobenzylether of hydroquinone; UV-B, ultraviolet B irradiation; PUVA, psoralens + ultraviolet A irradiation.

and piebaldism (Table 54-10). 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 patients with vitiligo. Circulating autoantibodies are often found, and the most common ones are antithyroglobulin, antimicrosomal, and antithyroid-stimulating hormone receptor antibodies.

There are four systemic diseases that should be considered in a patient with skin findings suggestive of vitiligo—Vogt-Koyanagi-Harada

syndrome, systemic sclerosis, onchocerciasis, and melanoma-associated leukoderma. A history of aseptic meningitis, nontraumatic uveitis, tinnitus, hearing loss, and/or dysacusia 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. 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 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 ipilimumab, 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.

There are two systemic disorders (neurocristopathies) that may have the cutaneous findings of piebaldism (Table 54-9). 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 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 fair-skinned 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 54-10) 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 54-11**) 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 seborrheic keratoses and acanthosis nigricans belong to the first group. *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

TABLE 54-11 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
 - 3. Increased pigment production
 - a. Ephelide (freckle)
 - b. Café au lait macule
 - c. Postinflammatory hyperpigmentation
 - d. Melasma
 - 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. Seborrheic keratoses (sign of Leser-Trélat)
 - b. Acanthosis nigricans (insulin resistance, other endocrine disorders, paraneoplastic)
 - 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, McCune-Albright syndrome^b)
 - b. Urticaria pigmentosa^c
 - 4. Dermal pigmentation
 - a. Incontinentia pigmenti (stage III)
 - b. Dyskeratosis congenita
 - 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 B₁₂, folate deficiency
 - d. Pellagra
 - e. Malabsorption, including Whipple's disease
 - 3. Melanosis secondary to metastatic melanoma
 - 4. Autoimmune
 - a. Biliary cirrhosis
 - b. Systemic sclerosis (scleroderma)
 - c. POEMS syndrome
 - d. Eosinophilia-myalgia syndrome^d
 - 5. Drugs (e.g. cyclophosphamide) and metals (e.g. silver)

^aAlso lentigines. ^bPolyostotic fibrous dysplasia. ^cSee also "Papulonodular Skin Lesions." ^dLate 1980s.

Abbreviations: LAMB, lentigines, atrial myomas, mucocutaneous myomas, and blue nevi; LEOPARD, lentigines, ECG abnormalities, ocular hypertelorism, pulmonary stenosis and subaortic valvar stenosis, abnormal genitalia, retardation of growth, and deafness (sensorineural); NAME, nevi, atrial myoma, myxoid neurofibroma, and ephelides (freckles); POEMS, polyneuropathy, organomegaly, endocrinopathies, M-protein, and skin changes.

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. *Acanthosis nigricans* can also 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).

A proliferation of melanocytes results in the following pigmented lesions: *lentigo*, *melanocytic nevus*, and *melanoma* (Chap. 72). In an adult, the majority of lentigines are related to sun exposure, which explains their distribution. However, in the Peutz-Jeghers and LEOPARD (lentigines; ECG abnormalities, primarily conduction defects; ocular hypertelorism; pulmonary stenosis and subaortic valvular stenosis; abnormal genitalia [cryptorchidism, hypospadias]; retardation of growth; and deafness [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* (lentigines, atrial myxomas, mucocutaneous myxomas, and blue nevi) syndrome or *NAME* (nevi, atrial myxoma, myxoid neurofibroma, and ephelides [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] 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—swirled 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 re-administration 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 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.

The metabolic causes of hyperpigmentation include *porphyria cutanea tarda* (PCT), *hemochromatosis*, *vitamin B₁₂ 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 B₆ 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, *biliary cirrhosis* 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 biliary cirrhosis, 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 (polyneuropathy; organomegaly [liver, spleen, lymph nodes]; endocrinopathies [impotence, gynecomastia]; M-protein; and skin 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 54-12**) 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. 55**).

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms (**Chap. 53**). 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. 56**). Clinically, phototoxic eruptions resemble an exaggerated sunburn with diffuse erythema and bullae in sun-exposed areas. The most commonly associated drugs are doxycycline, quinolones, thiazides, NSAIDs, voriconazole, 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. Severe acute graft-versus-host disease (grade 4), vancomycin-induced LABD, and flares of 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 pneumoniae*, or, occasionally, drugs. Involvement of the mucous membranes (ocular, nasal, oral, and genital) is seen more commonly in the latter form. 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, phenytoin, barbiturates, lamotrigine, aminopenicillins, nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine), and carbamazepine. 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. 187**), varicella-zoster virus (**Chap. 188**), and *S. aureus* (**Chap. 142**).

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), 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,

TABLE 54-12 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}
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
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. Staphylococcal scalded-skin syndrome ^{a,g}
5. Bullous impetigo ^a
II. Systemic diseases
A. Autoimmune
1. Paraneoplastic pemphigus ^a
B. Infections
1. Cutaneous emboli ^b
C. Metabolic
1. Diabetic bullae ^{a,b}
2. Porphyria cutanea tarda ^b
3. Porphyria variegata ^b
4. Pseudoporphyria ^b
5. Bullous dermatosis of hemodialysis ^b
D. Ischemia
1. Coma bullae
E. Secondary blistering diseases
1. Toxic epidermal necrolysis ^e (respiratory and GI tracts can be involved)

^aIntraepidermal. ^bSubepidermal. ^cAssociated with gluten enteropathy. ^dAssociated with inflammatory bowel disease. ^eDegeneration of cells within the basal layer of the epidermis can give impression split is subepidermal. ^fAlso systemic.

^gIn adults, associated with renal failure and immunocompromised state.

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 inclusion 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. 55**).

EXANTHEMS

(**Table 54-13**) Exanthems are characterized by an acute generalized eruption. The most common presentation is erythematous macules and papules (morbilloform) and less often confluent blanching erythema (scarlatiniform). *Morbilloform* 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, and transient lymphadenopathy. 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.

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*, *dengue virus*, *Zika virus*, and *West Nile virus* infections. Detection of specific IgM antibodies or fourfold elevations in IgG antibodies often allows the proper diagnosis.

TABLE 54-13 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 virus, Zika virus, Chikungunya, and West Nile virus infections
5. HIV seroconversion exanthem (plus mucosal ulcerations)
C. Bacterial
1. Typhoid fever
2. Early secondary syphilis
3. Early <i>Rickettsia</i> 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

^aPrimary infection in infants and reactivation in the setting of immunosuppression.

Abbreviations: CMV, cytomegalovirus; HHV, human herpesvirus; HIV, human immunodeficiency virus.

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 is seen in toxic shock syndrome (TSS), in Kawasaki disease, and after severe febrile illnesses. Certain strains of staphylococci also produce an erythrogenic toxin 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. 142**), 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. 143), 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. 356) 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. 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 54-14) *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 common *physical urticarias* include dermatographism, solar urticaria, cold urticaria, and cholinergic urticaria. Patients with *dermatographism* exhibit linear wheals following minor pressure or scratching of the skin. 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 protoporphyrin. 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. Autosomally 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. 347), 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, *serum sickness*, and *serum sickness-like illnesses* (e.g., due to cefaclor, minocycline).

PAPULONODULAR SKIN LESIONS

(Table 54-15) 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.

■ 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 inclusion cysts, lipomas, rheumatoid nodules, neurofibromas, angiomyomas, neuromas, and adnexal tumors such as tricholemmomas. Both *epidermoid inclusion 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 inclusion 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, 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. 86).

TABLE 54-14 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
 - C. Serum sickness
 - D. Angioedema (hereditary and acquired)

^aA small minority develop anaphylaxis. ^bAlso systemic. ^cAcquired angioedema can be idiopathic, associated with a lymphoproliferative disorder, or due to a drug, e.g., angiotensin-converting enzyme (ACE) inhibitors.

TABLE 54-15 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)
D.	Neuromas (MEN syndrome, type 2b)
E.	Adnexal tumors
1.	Basal cell carcinomas (basal cell nevus syndrome)
2.	Tricholemmomas (Cowden disease)
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)
2.	Bacillary angiomatosis (primarily in AIDS)
B.	Papules/plaques
1.	Cutaneous lupus
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
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
IX.	Purple
A.	Kaposi's sarcoma
B.	Angiosarcoma
C.	Palpable purpura (see Table 54-16)
X.	Brown-black ^c
XI.	Any color
A.	Metastases

^aIf multiple with childhood onset, consider Gardner syndrome. ^bMay have darker hue in more darkly pigmented individuals. ^cSee also "Hyperpigmentation."

Abbreviation: MEN, multiple endocrine neoplasia.

In some patients, the neurofibromas are localized and unilateral due to somatic mosaicism.

Angiofibromas are firm pink 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. 86**.

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. 381**).

Adnexal tumors are derived from pluripotent cells of the epidermis that can differentiate toward hair, sebaceous, 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 polypsis. Keratoses of the palms, soles, and dorsal aspect of the hands are also seen.

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 polyarthritides 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 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. 197), 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-blue 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*). Electron photomicrographs of angiokeratomas and clinically normal skin demonstrate lamellar lipid deposits in fibroblasts, pericytes, and endothelial cells that are diagnostic of this disease. 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) cuticular 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—infiltred 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 disease, 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 leukemic 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., all-trans-retinoic acid, granulocyte colony-stimulating factor [G-CSF]). 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.

Common causes of erythematous subcutaneous nodules include inflamed epidermoid inclusion 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 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 *Chlamydophila 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. 356). 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. 360) 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, 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. 347). 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, 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.

■ VIOLOCACEOUS 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. 197) 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 angiokeratoma). 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 while 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 54-16) *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 petechiae and purpura are primary cutaneous disorders such as *trauma*, *solar (actinic) 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

TABLE 54-16 Causes of Purpura

- I. Primary cutaneous disorders
 - A. Nonpalpable
 - 1. Trauma
 - 2. Solar (actinic, senile) purpura
 - 3. Steroid purpura
 - 4. Capillaritis
 - 5. Livedoid vasculopathy in the setting of venous hypertension^a
 - B. Drugs (e.g. anti-platelet agents, anti-coagulants)
 - C. Systemic diseases
- II. Drugs (e.g. anti-platelet agents, anti-coagulants)
 - 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
 - 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
 - g. 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^b
 - a. Acute meningococcemia
 - b. Disseminated gonococcal infection
 - c. Rocky Mountain spotted fever
 - d. Ecthyma gangrenosum

^aAlso associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency. ^bBacterial (including rickettsial), fungal, or parasitic.

Abbreviation: ITP, idiopathic thrombocytopenic purpura.

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. 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.

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. 111), *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, 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 referred to as warfarin-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 clefs 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. 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.

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. 356](#)). 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 HSP.

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 54-17](#). Peripheral vascular diseases of the extremities are reviewed in [Chap. 275](#), as is Raynaud's phenomenon.

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, factor V Leiden ([Chaps. 113 and 350](#)).

TABLE 54-17 Causes of Mucocutaneous Ulcers

- I. Primary cutaneous disorders
 - A. Peripheral vascular disease (**Chap. 275**)
 - 1. Venous
 - 2. Arterial^a
 - B. Livedo 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. 143**)
 - 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. 94**)
 - 3. Cryoglobulinemia,^c cryofibrinogenemia
 - 4. Cholesterol emboli^{a,c}
 - 5. Necrobiosis lipoidica^d
 - 6. Antiphospholipid syndrome (**Chap. 112**)
 - 7. Neuropathic^e (**Chap. 396**)
 - 8. Panniculitis
 - 9. Kaposi's sarcoma, acral angiodermatitis
 - 10. Diffuse dermal angiomas
 - B. Hands and feet
 - 1. Raynaud's phenomenon (**Chap. 275**)
 - 2. Buerger disease
 - C. Generalized
 - 1. Pyoderma gangrenosum, but most commonly legs
 - 2. Calciphylaxis (**Chap. 403**)
 - 3. Infections, e.g., dimorphic fungi, leishmaniasis
 - 4. Lymphoma
 - D. Face, especially perioral, and anogenital
 - 1. Chronic herpes simplex^f
- III. Mucosal
 - A. Behcet's syndrome (**Chap. 357**)
 - B. Erythema multiforme major, Stevens-Johnson syndrome, TEN
 - C. Primary blistering disorders (**Chap. 55**)
 - D. Lupus erythematosus, lichen planus
 - E. Inflammatory bowel disease
 - F. Acute HIV infection
 - G. Reactive arthritis

^aUnderlying atherosclerosis. ^bAlso associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency, antiphospholipid antibodies. ^cReviewed in section on Purpura.

^dReviewed in section on Papulonodular Skin Lesions. ^eFavors plantar surface of the foot. ^fSign of immunosuppression.

Abbreviations: HIV, human immunodeficiency virus; TEN, toxic epidermal necrolysis.

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, hairy cell leukemia, myelofibrosis, or a monoclonal gammopathy, usually IgA. 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.

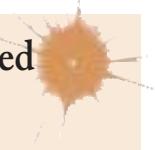
FURTHER READING

- BOLOGNA JL, SCHAFER 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.
- RIGOPOULOS D, LARIOS G, KATSAMBAS A: Skin signs of systemic diseases. *Clin Dermatol* 29:531, 2011.
- TAYLOR SC et al (eds): *Taylor and Kelly's Dermatology for Skin of Color*, 2nd ed. New York, McGraw-Hill, 2016.
- THIERS BH, SAHN RE, CALLEN JP: Cutaneous manifestations of internal malignancy. *CA: Cancer J Clin* 59:73, 2009.

55

Immunologically Mediated Skin Diseases

Kim B. Yancey, Thomas J. Lawley



A number of immunologically mediated skin diseases and immunologically mediated systemic disorders with cutaneous manifestations are now recognized as distinct entities with consistent clinical, histologic, and immunopathologic findings. Clinically, these disorders are characterized by morbidity (pain, pruritus, disfigurement) and, in some instances, result in death (largely due to loss of epidermal barrier function and/or secondary infection). The major features of the more common immunologically mediated skin diseases are summarized in this chapter (**Table 55-1**), as are autoimmune systemic disorders with cutaneous manifestations.

AUTOIMMUNE CUTANEOUS DISEASES

PEMPHIGUS VULGARIS

Pemphigus refers to a group of autoantibody-mediated intraepidermal blistering diseases characterized by loss of cohesion between epidermal cells (a process termed *acantholysis*). Manual pressure to the skin of these patients may elicit the separation of the epidermis (*Nikolsky's sign*). This finding, while characteristic of pemphigus, is not specific to this group of disorders and is also seen in toxic epidermal necrolysis, Stevens-Johnson syndrome, and a few other skin diseases.

Pemphigus vulgaris (PV) is a mucocutaneous blistering disease that predominantly occurs in patients >40 years of age. PV typically begins on mucosal surfaces and often progresses to involve the skin.

TABLE 55-1 Immunologically Mediated Blistering Diseases

DISEASE	CLINICAL MANIFESTATIONS	HISTOLOGY	IMMUNOPATHOLOGY	AUTOANTIGENS ^a
Pemphigus vulgaris	Flaccid blisters, denuded skin, oromucosal lesions	Acantholytic blister formed in suprabasal layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg3 (plus Dsg1 in patients with skin involvement)
Pemphigus foliaceus	Crusts and shallow erosions on scalp, central face, upper chest, and back	Acantholytic blister formed in superficial layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg1
Paraneoplastic pemphigus	Painful stomatitis with papulosquamous or lichenoid eruptions that may progress to blisters	Acantholysis, keratinocyte necrosis, and vacuolar interface dermatitis	Cell surface deposits of IgG and C3 on keratinocytes and (variably) similar immunoreactants in epidermal BMZ	Plakin protein family members and desmosomal cadherins (see text for details)
Bullous pemphigoid	Large tense blisters on flexor surfaces and trunk	Subepidermal blister with eosinophil-rich infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	BPAG1, BPAG2
Pemphigoid gestationis	Pruritic, urticarial plaques rimmed by vesicles and bullae on the trunk and extremities	Teardrop-shaped, subepidermal blisters in dermal papillae; eosinophil-rich infiltrate	Linear band of C3 in epidermal BMZ	BPAG2 (plus BPAG1 in some patients)
Dermatitis herpetiformis	Extremely pruritic small papules and vesicles on elbows, knees, buttocks, and posterior neck	Subepidermal blister with neutrophils in dermal papillae	Granular deposits of IgA in dermal papillae	Epidermal transglutaminase
Linear IgA disease	Pruritic small papules on extensor surfaces; occasionally larger, arciform blisters	Subepidermal blister with neutrophil-rich infiltrate	Linear band of IgA in epidermal BMZ	BPAG2 (see text for specific details)
Epidermolysis bullosa acquisita	Blisters, erosions, scars, and milia on sites exposed to trauma; widespread, inflammatory, tense blisters may be seen initially	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	Type VII collagen
Mucous membrane pemphigoid	Erosive and/or blistering lesions of mucous membranes and possibly the skin; scarring of some sites	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG, IgA, and/or C3 in epidermal BMZ	BPAG2, laminin-332, or others

^aAutoantigens bound by these patients' autoantibodies are defined as follows: Dsg1, desmoglein 1; Dsg3, desmoglein 3; BPAG1, bullous pemphigoid antigen 1; BPAG2, bullous pemphigoid antigen 2.

Abbreviation: BMZ, basement membrane zone.

This disease is characterized by fragile, flaccid blisters that rupture to produce extensive denudation of mucous membranes and skin (**Fig. 55-1**). The mouth, scalp, face, neck, axilla, groin, and trunk are typically involved. PV may be associated with severe skin pain; some patients experience pruritus as well. Lesions usually heal without scarring except at sites complicated by secondary infection or mechanically induced dermal wounds. Postinflammatory hyperpigmentation is usually present for some time at sites of healed lesions.

Biopsies of early lesions demonstrate intraepidermal vesicle formation secondary to loss of cohesion between epidermal cells (i.e., acantholytic blisters). Blister cavities contain acantholytic epidermal cells, which appear as round homogeneous cells containing hyperchromatic nuclei. Basal keratinocytes remain attached to the epidermal basement membrane; hence, blister formation takes place within the suprabasal portion of the epidermis. Lesional skin may contain focal collections of intraepidermal eosinophils within blister cavities; dermal alterations are slight, often limited to an eosinophil-predominant leukocytic infiltrate. Direct immunofluorescence microscopy of lesional or intact patient skin shows deposits of IgG on the surface of keratinocytes; deposits of complement components are typically found in lesional but not in uninvolved skin. Deposits of IgG on keratinocytes are derived from circulating autoantibodies to cell-surface autoantigens. Such circulating autoantibodies can be demonstrated in 80–90% of PV patients by indirect immunofluorescence microscopy; monkey esophagus is the optimal substrate for these studies. Patients with PV have IgG autoantibodies to *desmogleins* (Dsgs), transmembrane desmosomal glycoproteins that belong to the cadherin family of calcium-dependent adhesion molecules. Such autoantibodies can be precisely quantitated by enzyme-linked immunosorbent assay (ELISA). Patients with early PV (i.e., mucosal disease) have IgG autoantibodies to Dsg3; patients with advanced PV (i.e., mucocutaneous disease) have IgG autoantibodies to both Dsg3 and Dsg1. Experimental studies have shown that autoantibodies from patients with PV are pathogenic (i.e., responsible for blister formation) and that their titer correlates with disease activity. Recent studies have shown that the anti-Dsg autoantibody profile in

these patients' sera as well as the tissue distribution of Dsg3 and Dsg1 determine the site of blister formation in patients with PV. Coexpression of Dsg3 and Dsg1 by epidermal cells protects against pathogenic IgG antibodies to either of these cadherins but not against pathogenic autoantibodies to both.

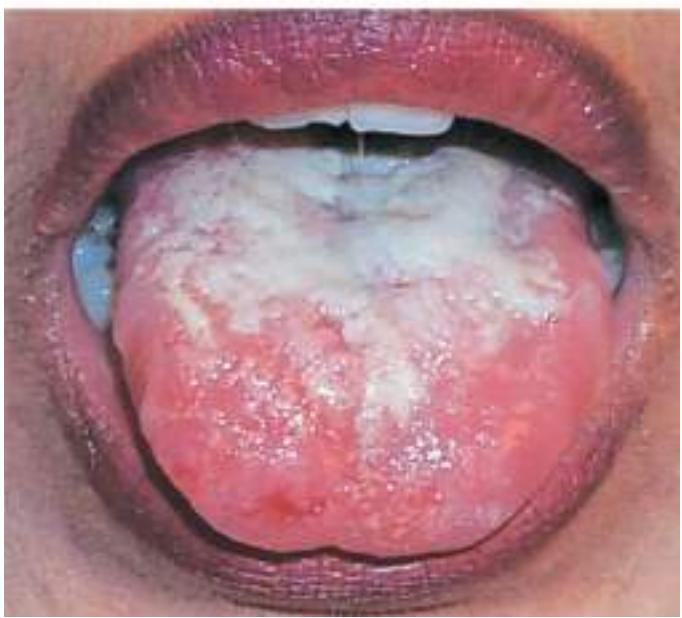
PV can be life-threatening. Prior to the availability of glucocorticoids, mortality rates ranged from 60% to 90%; the current figure is ~5%. Common causes of morbidity and death are infection and complications of treatment. Bad prognostic factors include advanced age, widespread involvement, and the requirement for high doses of glucocorticoids (with or without other immunosuppressive agents) for control of disease. The course of PV in individual patients is variable and difficult to predict. Some patients experience remission, while others may require long-term treatment or succumb to complications of their disease or its treatment. The mainstay of treatment is systemic glucocorticoids. Patients with moderate to severe PV are usually started on prednisone at 1 mg/kg per day. If new lesions continue to appear after 1–2 weeks of treatment, the dose may need to be increased and/or prednisone may need to be combined with other immunosuppressive agents such as azathioprine (2–2.5 mg/kg per day), mycophenolate mofetil (20–35 mg/kg per day), rituximab (375 mg/m² per week × 4, or 1000 mg on days 1 and 15), or cyclophosphamide (1–2 mg/kg per day). Patients with severe, treatment-resistant disease may derive benefit from plasmapheresis (six high-volume exchanges [i.e., 2–3 L per exchange] over ~2 weeks) and/or IV immunoglobulin (IVIg) (2 g/kg over 3–5 days every 6–8 weeks). It is important to bring severe or progressive disease under control quickly in order to lessen the severity and/or duration of this disorder. Increasingly, rituximab and daily glucocorticoids are used early in PV patients to avert the development of advanced and/or treatment-resistant disease.

■ PEMPHIGUS FOLIACEUS

Pemphigus foliaceus (PF) is distinguished from PV by several features. In PF, acantholytic blisters are located high within the epidermis, usually just beneath the stratum corneum. Hence, PF is a more superficial



A



B

FIGURE 55-1 Pemphigus vulgaris. **A.** Flaccid bullae are easily ruptured, resulting in multiple erosions and crusted plaques. **B.** Involvement of the oral mucosa, which is almost invariable, may present with erosions on the gingiva, buccal mucosa, palate, posterior pharynx, or tongue. (B, Courtesy of Robert Swerlick, MD; with permission.)

blistering disease than PV. The distribution of lesions in the two disorders is much the same, except that in PF mucous membranes are almost always spared. Patients with PF rarely have intact blisters but rather exhibit shallow erosions associated with erythema, scale, and crust formation. Mild cases of PF can resemble severe seborrheic dermatitis; severe PF may cause extensive exfoliation. Sun exposure (ultraviolet irradiation) may be an aggravating factor.

PF has immunopathologic features in common with PV. Specifically, direct immunofluorescence microscopy of perilesional skin demonstrates IgG on the surface of keratinocytes. Similarly, patients with PF have circulating IgG autoantibodies directed against the surface of keratinocytes. In PF, autoantibodies are directed against Dsg1, a 160-kDa desmosomal cadherin. These autoantibodies can be quantitated by ELISA. As noted for PV, the autoantibody profile in patients with PF (i.e., anti-Dsg1 IgG) and the tissue distribution of this autoantigen (i.e., expression in oral mucosa that is compensated by coexpression of Dsg3) are thought to account for the distribution of lesions in this disease.

Endemic forms of PF are found in south-central rural Brazil, where the disease is known as *fogo salvagem* (FS), as well as in selected sites in Latin America and Tunisia. Endemic PF, like other forms of this disease, is mediated by IgG autoantibodies to Dsg1. Clusters of FS overlap with those of leishmaniasis, a disease transmitted by bites of the sand fly *Lutzomyia longipalis*. Recent studies have shown that sand-fly salivary antigens (specifically, the LJM11 salivary protein) are recognized by IgG autoantibodies from FS patients (as well as by monoclonal antibodies to Dsg1 derived from these patients). Moreover, mice immunized with LJM11 produce antibodies to Dsg1. Thus, these findings suggest that insect bites may deliver salivary antigens that initiate a cross-reactive humoral immune response, which may lead to FS in genetically susceptible individuals.

Although pemphigus has been associated with several autoimmune diseases, its association with thymoma and/or myasthenia gravis is particularly notable. To date, >30 cases of thymoma and/or myasthenia gravis have been reported in association with pemphigus, usually with PF. Patients may also develop pemphigus as a consequence of drug exposure; drug-induced pemphigus usually resembles PF rather than PV. Drugs containing a thiol group in their chemical structure (e.g., penicillamine, captopril, enalapril) are most commonly associated with drug-induced pemphigus. Nonthiol drugs linked to pemphigus include penicillins, cephalosporins, and piroxicam. Some cases of drug-induced pemphigus are durable and require treatment with systemic glucocorticoids and/or immunosuppressive agents.

PF is generally a less severe disease than PV and usually carries a better prognosis. Localized disease can sometimes be treated with topical or intralesional glucocorticoids; more active cases can usually be controlled with systemic glucocorticoids either alone or in combination with other immunosuppressive agents. Patients with severe, treatment-resistant disease may require more aggressive interventions, as described above for patients with PV.

■ PARANEOPLASTIC PEMPHIGUS

Paraneoplastic pemphigus (PNP) is an autoimmune acantholytic mucocutaneous disease associated with an occult or confirmed neoplasm. Patients with PNP typically have painful stomatitis in association with papulosquamous and/or lichenoid eruptions that often progress to blisters. Palm and sole involvement are common in these patients and raise the possibility that prior reports of neoplasia-associated erythema multiforme actually may have represented unrecognized cases of PNP. Biopsies of lesional skin from these patients show varying combinations of acantholysis, keratinocyte necrosis, and vacuolar-interface dermatitis. Direct immunofluorescence microscopy of a patient's skin shows deposits of IgG and complement on the surface of keratinocytes and (variably) similar immunoreactants in the epidermal basement membrane zone. Patients with PNP have IgG autoantibodies to cytoplasmic proteins that are members of the plakin family (e.g., desmoplakins I and II, bullous pemphigoid antigen [BPAG]1, envoplakin, periplakin, and plectin) and to cell-surface proteins that are members of the cadherin family (e.g., Dsg1 and Dsg3). Passive transfer studies have shown that autoantibodies from patients with PNP are pathogenic in animal models.

The predominant neoplasms associated with PNP are non-Hodgkin's lymphoma, chronic lymphocytic leukemia, thymoma, spindle cell tumors, Waldenström's macroglobulinemia, and Castleman's disease; the last-mentioned neoplasm is particularly common among children with PNP. Rare cases of seronegative PNP have been reported in patients with B cell malignancies previously treated with rituximab. In addition to severe skin lesions, many patients with PNP develop life-threatening bronchiolitis obliterans. PNP is generally resistant to conventional therapies (i.e., those used to treat PV); rarely, a patient's disease may ameliorate or even remit following ablation or removal of underlying neoplasms.

■ BULLOUS PEMPHIGOID

Bullous pemphigoid (BP) is a polymorphic autoimmune subepidermal blistering disease usually seen in the elderly. Initial lesions may consist of urticarial plaques; most patients eventually display tense blisters on



FIGURE 55-2 **Bullous pemphigoid** with tense vesicles and bullae on erythematous, urticarial bases. (Courtesy of the Yale Resident's Slide Collection; with permission.)

either normal-appearing or erythematous skin (**Fig. 55-2**). The lesions are usually distributed over the lower abdomen, groin, and flexor surface of the extremities; oral mucosal lesions are found in some patients. Pruritus may be nonexistent or severe. As lesions evolve, tense blisters tend to rupture and be replaced by erosions with or without surmounting crust. Nontraumatized blisters heal without scarring. The major histocompatibility complex class II allele HLA-DQ β 1*0301 is prevalent in patients with BP. Despite isolated reports, several studies have shown that patients with BP do not have a higher incidence of malignancy than appropriately age- and gender-matched controls.

Biopsies of early lesional skin demonstrate subepidermal blisters and histologic features that roughly correlate with the clinical character of the particular lesion under study. Lesions on normal-appearing skin generally contain a sparse perivascular leukocytic infiltrate with some eosinophils; conversely, biopsies of inflammatory lesions typically show an eosinophil-rich infiltrate at sites of vesicle formation and in perivascular areas. In addition to eosinophils, cell-rich lesions also contain mononuclear cells and neutrophils. It is not possible to distinguish BP from other subepidermal blistering diseases by routine histologic studies alone.

Direct immunofluorescence microscopy of normal-appearing perilesional skin from patients with BP shows linear deposits of IgG and/or C3 in the epidermal basement membrane. The sera of ~70% of these patients contain circulating IgG autoantibodies that bind the epidermal basement membrane of normal human skin in indirect immunofluorescence microscopy. IgG from an even higher percentage of patients reacts with the epidermal side of 1 M NaCl split skin (an alternative immunofluorescence microscopy test substrate used to distinguish circulating IgG autoantibodies to the basement membrane in patients with BP from those in patients with similar, yet different, subepidermal blistering diseases; see below). In BP, circulating autoantibodies recognize 230- and 180-kDa hemidesmosome-associated proteins in basal keratinocytes (i.e., BPAG1 and BPAG2, respectively). Autoantibodies to BPAG2 are thought to deposit *in situ*, activate complement, produce dermal mast-cell degranulation, and generate granulocyte-rich infiltrates that cause tissue damage and blister formation.

BP may persist for months to years, with exacerbations or remissions. Extensive involvement may result in widespread erosions and compromise cutaneous integrity; elderly and/or debilitated patients may die. The mainstay of treatment is systemic glucocorticoids. Local or minimal disease can sometimes be controlled with topical glucocorticoids alone; more extensive lesions generally respond to systemic glucocorticoids either alone or in combination with other immunosuppressive agents. Patients usually respond to prednisone (0.75–1 mg/kg per day).

In some instances, azathioprine (2–2.5 mg/kg per day), mycophenolate mofetil (20–35 mg/kg per day), or rituximab (375 mg/m² per week × 4, or 1000 mg on days 1 and 15) are necessary adjuncts.

■ PEMPHIGOID GESTATIONIS

Pemphigoid gestationis (PG), also known as *herpes gestationis*, is a rare, nonviral, subepidermal blistering disease of pregnancy and the puerperium. PG may begin during any trimester of pregnancy or present shortly after delivery. Lesions are usually distributed over the abdomen, trunk, and extremities; mucous membrane lesions are rare. Skin lesions in these patients may be quite polymorphic and consist of erythematous urticarial papules and plaques, vesiculopapules, and/or frank bullae. Lesions are almost always extremely pruritic. Severe exacerbations of PG frequently follow delivery, typically within 24–48 h. PG tends to recur in subsequent pregnancies, often beginning earlier during such gestations. Brief flare-ups of disease may occur with resumption of menses and may develop in patients later exposed to oral contraceptives. Occasionally, infants of affected mothers have transient skin lesions.

Biopsies of early lesional skin show teardrop-shaped subepidermal vesicles forming in dermal papillae in association with an eosinophil-rich leukocytic infiltrate. Differentiation of PG from other subepidermal bullous diseases by light microscopy is difficult. However, direct immunofluorescence microscopy of perilesional skin from PG patients reveals the immunopathologic hallmark of this disorder: linear deposits of C3 in the epidermal basement membrane. These deposits develop as a consequence of complement activation produced by low-titer IgG anti-basement membrane autoantibodies directed against BPAG2, the same hemidesmosome-associated protein that is targeted by autoantibodies in patients with BP—a subepidermal bullous disease that resembles PG clinically, histologically, and immunopathologically.

The goals of therapy in patients with PG are to prevent the development of new lesions, relieve intense pruritus, and care for erosions at sites of blister formation. Many patients require treatment with moderate doses of daily glucocorticoids (i.e., 20–40 mg of prednisone) at some point in their course. Mild cases (or brief flare-ups) may be controlled by vigorous use of potent topical glucocorticoids. Infants born of mothers with PG appear to be at increased risk of being born slightly premature or “small for dates.” Current evidence suggests that there is no difference in the incidence of uncomplicated live births between PG patients treated with systemic glucocorticoids and those managed more conservatively. If systemic glucocorticoids are administered, newborns are at risk for development of reversible adrenal insufficiency.

■ DERMATITIS HERPETIFORMIS

Dermatitis herpetiformis (DH) is an intensely pruritic, papulovesicular skin disease characterized by lesions symmetrically distributed over extensor surfaces (i.e., elbows, knees, buttocks, back, scalp, and posterior neck) (**see Fig. 52-8**). Primary lesions in this disorder consist of papules, papulovesicles, or urticarial plaques. Because pruritus is prominent, patients may present with excoriations and crusted papules but no observable primary lesions. Patients sometimes report that their pruritus has a distinctive burning or stinging component; the onset of such local symptoms reliably heralds the development of distinct clinical lesions 12–24 h later. Almost all DH patients have associated, usually subclinical, gluten-sensitive enteropathy (**Chap. 318**), and >90% express the HLA-B8/DRw3 and HLA-DQw2 haplotypes. DH may present at any age, including in childhood; onset in the second to fourth decades is most common. The disease is typically chronic.

Biopsy of early lesional skin reveals neutrophil-rich infiltrates within dermal papillae. Neutrophils, fibrin, edema, and microvesicle formation at these sites are characteristic of early disease. Older lesions may demonstrate nonspecific features of a subepidermal bulla or an excoriated papule. Because the clinical and histologic features of this disease can be variable and resemble those of other subepidermal blistering disorders, the diagnosis is confirmed by direct immunofluorescence microscopy of normal-appearing perilesional skin. Such studies demonstrate granular deposits of IgA (with or without complement

components) in the papillary dermis and along the epidermal basement membrane zone. IgA deposits in the skin are unaffected by control of disease with medication; however, these immunoreactants diminish in intensity or disappear in patients maintained for long periods on a strict gluten-free diet (see below). Patients with DH have granular deposits of IgA in their epidermal basement membrane zone and should be distinguished from individuals with linear IgA deposits at this site (see below).

Although most DH patients do not report overt gastrointestinal symptoms or have laboratory evidence of malabsorption, biopsies of the small bowel usually reveal blunting of intestinal villi and a lymphocytic infiltrate in the lamina propria. As is true for patients with celiac disease, this gastrointestinal abnormality can be reversed by a gluten-free diet. Moreover, if maintained, this diet alone may control the skin disease and eventuate in clearance of IgA deposits from these patients' epidermal basement membrane zones. Subsequent gluten exposure in such patients alters the morphology of their small bowel, elicits a flare-up of their skin disease, and is associated with the reappearance of IgA in their epidermal basement membrane zones. As in patients with celiac disease, dietary gluten sensitivity in patients with DH is associated with IgA anti-endomysial autoantibodies that target tissue transglutaminase. Studies indicate that patients with DH also have high-avidity IgA autoantibodies to epidermal transglutaminase and that the latter is co-localized with granular deposits of IgA in the papillary dermis of DH patients. Patients with DH also have an increased incidence of thyroid abnormalities, achlorhydria, atrophic gastritis, and autoantibodies to gastric parietal cells. These associations likely relate to the high frequency of the HLA-B8/DRw3 haplotype in these patients, since this marker is commonly linked to autoimmune disorders. The mainstay of treatment of DH is dapsone, a sulfone. Patients respond rapidly (24–48 h) to dapsone (50–200 mg/d), but require careful pretreatment evaluation and close follow-up to ensure that complications are avoided or controlled. All patients taking dapsone at >100 mg/d will have some hemolysis and methemoglobinemia, which are expected pharmacologic side effects of this agent. Gluten restriction can control DH and lessen dapsone requirements; this diet must rigidly exclude gluten to be of maximal benefit. Many months of dietary restriction may be necessary before a beneficial result is achieved. Good dietary counseling by a trained dietitian is essential.

■ LINEAR IgA DISEASE

Linear IgA disease, once considered a variant form of DH, is actually a separate and distinct entity. Clinically, patients with linear IgA disease may resemble individuals with DH, BP, or other subepidermal blistering diseases. Lesions typically consist of papulovesicles, bullae, and/or urticarial plaques that develop predominantly on central or flexural sites. Oral mucosal involvement occurs in some patients. Severe pruritus resembles that seen in patients with DH. Patients with linear IgA disease do not have an increased frequency of the HLA-B8/DRw3 haplotype or an associated enteropathy and therefore are not candidates for treatment with a gluten-free diet.

Histologic alterations in early lesions may be virtually indistinguishable from those in DH. However, direct immunofluorescence microscopy of normal-appearing perilesional skin reveals a linear band of IgA (and often C3) in the epidermal basement membrane zone. Most patients with linear IgA disease have circulating IgA anti-basement membrane autoantibodies directed against neoepitopes in the proteolytically processed extracellular domain of BPAG2. These patients generally respond to treatment with dapsone (50–200 mg/d).

■ EPIDERMOLYSIS BULLOSA ACQUISITA

Epidermolysis bullosa acquisita (EBA) is a rare, noninherited, polymorphic, chronic, subepidermal blistering disease. (**The inherited form is discussed in Chap. 406.**) Patients with classic or noninflammatory EBA have blisters on noninflamed skin, atrophic scars, milia, nail dystrophy, and oral lesions. Because lesions generally occur at sites exposed to minor trauma, classic EBA is considered a mechanobullous disease. Other patients with EBA have widespread inflammatory scarring and bullous lesions that resemble severe BP. Inflammatory EBA may

evolve into the classic, noninflammatory form of this disease. Rarely, patients present with lesions that predominate on mucous membranes. The HLA-DR2 haplotype is found with increased frequency in EBA patients. Studies suggest that EBA is sometimes associated with inflammatory bowel disease (especially Crohn's disease).

The histology of lesional skin varies with the character of the lesion being studied. Noninflammatory bullae are subepidermal, feature a sparse leukocytic infiltrate, and resemble the lesions in patients with porphyria cutanea tarda. Inflammatory lesions consist of neutrophil-rich subepidermal blisters. EBA patients have continuous deposits of IgG (and frequently C3) in a linear pattern within the epidermal basement membrane zone. Ultrastructurally, these immunoreactants are found in the sublamina densa region in association with anchoring fibrils. Approximately 50% of EBA patients have demonstrable circulating IgG anti-basement membrane autoantibodies directed against type VII collagen—the collagen species that makes up anchoring fibrils. Such IgG autoantibodies bind the dermal side of 1 M NaCl split skin (in contrast to IgG autoantibodies in patients with BP). Studies have shown that passive transfer of experimental or patient IgG against type VII collagen can produce lesions in mice that clinically, histologically, and immunopathologically resemble those in patients with EBA.

Treatment of EBA is generally unsatisfactory. Some patients with inflammatory EBA may respond to systemic glucocorticoids, either alone or in combination with immunosuppressive agents. Other patients (especially those with neutrophil-rich inflammatory lesions) may respond to dapsone. The chronic, noninflammatory form of EBA is largely resistant to treatment, although some patients may respond to cyclosporine, azathioprine, IVIg, or rituximab.

■ MUCOUS MEMBRANE PEMPHIGOID

Mucous membrane pemphigoid (MMP) is a rare, acquired, subepithelial immunobullous disease characterized by erosive lesions of mucous membranes and skin that result in scarring of at least some sites of involvement. Common sites include the oral mucosa (especially the gingiva) and conjunctiva; other sites that may be affected include the nasopharyngeal, laryngeal, esophageal, and anogenital mucosa. Skin lesions (present in about one-third of patients) tend to predominate on the scalp, face, and upper trunk and generally consist of a few scattered erosions or tense blisters on an erythematous or urticarial base. MMP is typically a chronic and progressive disorder. Serious complications may arise as a consequence of ocular, laryngeal, esophageal, or anogenital lesions. Erosive conjunctivitis may result in shortened fornices, symblepharon, ankyloblepharon, entropion, corneal opacities, and (in severe cases) blindness. Similarly, erosive lesions of the larynx may cause hoarseness, pain, and tissue loss that, if unrecognized and untreated, may eventuate in complete destruction of the airway. Esophageal lesions may result in stenosis and/or strictures that could place patients at risk for aspiration. Strictures may also complicate anogenital involvement.

Biopsies of lesional tissue generally show subepithelial vesiculobullae and a mononuclear leukocytic infiltrate. Neutrophils and eosinophils may be seen in biopsies of early lesions; older lesions may demonstrate a scant leukocytic infiltrate and fibrosis. Direct immunofluorescence microscopy of perilesional tissue typically reveals deposits of IgG, IgA, and/or C3 in the epidermal basement membrane. Because many patients with MMP exhibit no evidence of circulating anti-basement membrane autoantibodies, testing of perilesional skin is important diagnostically. Although MMP was once thought to be a single nosologic entity, it is now largely regarded as a disease phenotype that may develop as a consequence of an autoimmune reaction to a variety of molecules in the epidermal basement membrane (e.g., BPAG2, laminin-332, type VII collagen, $\alpha_6\beta_4$ integrin) and other antigens yet to be completely defined. Studies suggest that MMP patients with autoantibodies to laminin-332 have an increased relative risk for cancer. Treatment of MMP is largely dependent upon the sites of involvement. Due to potentially severe complications, patients with ocular, laryngeal, esophageal, and/or anogenital involvement require aggressive systemic treatment with dapsone, prednisone, or the latter

in combination with another immunosuppressive agent (e.g., azathioprine, mycophenolate mofetil, cyclophosphamide, or rituximab), or IVIg. Less threatening forms of the disease may be managed with topical or intralesional glucocorticoids.

AUTOIMMUNE SYSTEMIC DISEASES WITH PROMINENT CUTANEOUS FEATURES

■ DERMATOMYOSITIS

The cutaneous manifestations of dermatomyositis ([Chap. 358](#)) are often distinctive but at times may resemble those of systemic lupus erythematosus (SLE) ([Chap. 349](#)), scleroderma ([Chap. 353](#)), or other overlapping connective tissue diseases ([Chap. 353](#)). The extent and severity of cutaneous disease may or may not correlate with the extent and severity of the myositis. The cutaneous manifestations of dermatomyositis are similar, whether the disease appears in children or in the elderly, except that calcification of subcutaneous tissue is a common late sequela in childhood dermatomyositis.

The cutaneous signs of dermatomyositis may precede or follow the development of myositis by weeks to years. Cases lacking muscle involvement (i.e., *dermatomyositis sine myositis* or *amyopathic dermatomyositis*) have also been reported. The most common manifestation is a purple-red discoloration of the upper eyelids, sometimes associated with scaling ("heliotrope" erythema; [Fig. 55-3](#)) and periorbital edema. Erythema on the cheeks and nose in a "butterfly" distribution may resemble the malar eruption of SLE. Erythematous or violaceous scaling patches are common on the upper anterior chest, posterior neck, scalp, and the extensor surfaces of the arms, legs, and hands. Erythema and scaling may be particularly prominent over the elbows, knees, and dorsal interphalangeal joints. Approximately one-third of patients have violaceous, flat-topped papules over the dorsal interphalangeal joints that are pathognomonic of dermatomyositis (Gottron's papules) ([Fig. 55-4](#)). Thin violaceous papules and plaques on the elbows and knees of patients with dermatomyositis are referred to as *Gottron's sign* ([Fig. 55-4](#)). These lesions can be contrasted with the erythema and scaling on the dorsum of the fingers that spares the skin over the interphalangeal joints of some SLE patients. Periungual telangiectases and edema may be prominent in patients with dermatomyositis. Lacy or reticulated erythema may be associated with fine scaling on the extensor and lateral surfaces of the thighs and upper arms. Other patients, particularly those with long-standing disease, develop areas of hypopigmentation, hyperpigmentation, mild atrophy, and telangiectasia known as *poikiloderma*. Poikiloderma is rare in both SLE and scleroderma and thus can serve as a clinical sign that distinguishes dermatomyositis from these two diseases. Cutaneous changes may be similar in dermatomyositis and various overlap syndromes where



FIGURE 55-3 Dermatomyositis. Periorbital violaceous erythema characterizes the classic heliotrope rash. (Courtesy of James Krell, MD; with permission.)



FIGURE 55-4 Gottron's papules. Dermatomyositis often involves the hands as erythematous flat-topped papules over the knuckles. Periungual telangiectases are also evident.

thickening and binding down of the skin of the hands (*sclerodactyly*) as well as Raynaud's phenomenon can be seen. However, the presence of severe muscle disease, Gottron's papules, heliotrope erythema, and poikiloderma serve to distinguish patients with dermatomyositis. Skin biopsy of the erythematous, scaling lesions of dermatomyositis may reveal only mild nonspecific inflammation, but sometimes may show changes indistinguishable from those found in cutaneous lupus erythematosus (LE), including epidermal atrophy, hydropic degeneration of basal keratinocytes, and dermal changes consisting of edema of the upper dermis, interstitial mucin deposition, and a mild mononuclear cell infiltrate. Direct immunofluorescence microscopy of lesional skin is usually negative, although granular deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone have been described in some patients. Treatment should be directed at the systemic disease. Topical glucocorticoids are sometimes useful; patients should avoid exposure to ultraviolet irradiation and aggressively use photoprotective measures, including broad-spectrum sunscreens.

■ LUPUS ERYTHEMATOSUS

The cutaneous manifestations of LE ([Chap. 349](#)) can be divided into acute, subacute, and chronic or discoid types. *Acute cutaneous LE* is characterized by erythema of the nose and malar eminences in a "butterfly" distribution ([Fig. 55-5A](#)). The erythema is often sudden in onset, accompanied by edema and fine scale, and correlated with systemic involvement. Patients may have widespread involvement of the face as well as erythema and scaling of the extensor surfaces of the extremities and upper chest ([Fig. 55-5B](#)). These acute lesions, while sometimes evanescent, usually last for days and are often associated with exacerbations of systemic disease. Skin biopsy of acute lesions typically shows hydropic degeneration of basal keratinocytes, dermal edema, and (in some cases) a sparse infiltrate of mononuclear cells in the upper dermis as well as dermal mucin. Direct immunofluorescence microscopy of lesional skin frequently reveals deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone. Treatment is aimed at control of systemic disease. Photoprotection is very important in this as well as in other forms of LE.

Subacute cutaneous lupus erythematosus (SCLE) is characterized by a widespread photosensitive, nonscarring eruption. In most patients, renal and central nervous system involvement is mild or absent. SCLE may present as a papulosquamous eruption that resembles psoriasis or as annular polycyclic lesions. In the papulosquamous form, discrete erythematous papules arise on the back, chest, shoulders, extensor surfaces of the arms, and dorsum of the hands; lesions are uncommon on the central face and the flexor surfaces of the arms as well as below the waist. These slightly scaling papules tend to merge into large plaques, some with a reticulate appearance. The annular form involves the same areas and presents with erythematous papules that evolve into oval, circular, or polycyclic lesions. The lesions of SCLE are more widespread but have less tendency for scarring than lesions of discoid LE. In many

**A****B**

FIGURE 55-5 Acute cutaneous lupus erythematosus (LE). **A.** Acute cutaneous LE on the face, showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common. **B.** Acute cutaneous LE on the upper chest, demonstrating brightly erythematous and slightly edematous papules and plaques. (*B*, Courtesy of Robert Swerlick, MD; with permission.)

patients with SCLE, drugs (e.g., hydrochlorothiazide, calcium channel blockers, proton pump inhibitors) may induce or exacerbate disease. Skin biopsy typically reveals epidermal changes that include atrophy, hydropic degeneration of basal keratinocytes, and apoptosis accompanied by an infiltrate of mononuclear cells in the upper dermis. Direct immunofluorescence microscopy of lesional skin reveals deposits of immunoglobulin(s) in the epidermal basement membrane zone in about one-half of these cases. A particulate pattern of IgG deposition throughout the epidermis has been associated with SCLE. Most SCLE patients have anti-Ro autoantibodies. Local therapy alone is usually unsuccessful. Most patients require treatment with aminoquinoline antimalarial drugs. Low-dose therapy with oral glucocorticoids is sometimes necessary. Photoprotective measures against both ultraviolet B and ultraviolet A wavelengths are very important.

Discoid lupus erythematosus (DLE, also called *chronic cutaneous LE*) is characterized by discrete lesions, most often found on the face, scalp, and/or external ears. The lesions are erythematous papules or plaques with a thick, adherent scale that occludes hair follicles (follicular plugging). When the scale is removed, its underside shows small excrescences that correlate with the openings of hair follicles (so-called “carpet tacking”), a finding relatively specific for DLE. Long-standing lesions develop central atrophy, scarring, and hypopigmentation but frequently have erythematous, sometimes raised borders (Fig. 55-6). These lesions persist for years and tend to expand slowly. Up to 15% of patients with DLE eventually meet the American College



FIGURE 55-6 Discoid (chronic cutaneous) lupus erythematosus (LE). Violaceous, hyperpigmented, atrophic plaques, follicular plugging, and scarring are typical features of chronic cutaneous LE.

of Rheumatology criteria for SLE. Typical discoid lesions are frequently seen in patients with SLE. Biopsy of DLE lesions shows hyperkeratosis, follicular plugging, atrophy of the epidermis, hydropic degeneration of basal keratinocytes, thickening of the epidermal basement membrane zone, and a mononuclear cell infiltrate adjacent to epidermal, adnexal, and microvascular basement membranes. Direct immunofluorescence microscopy demonstrates immunoglobulin(s) and complement deposits at the basement membrane zone in ~90% of cases. Treatment is focused on control of local cutaneous disease and consists mainly of photoprotection and topical or intralesional glucocorticoids. If local therapy is ineffective, use of aminoquinoline antimalarial agents may be indicated.

SCLERODERMA AND MORPHEA

The skin changes of scleroderma (Chap. 353) usually begin on the fingers, hands, toes, feet, and face, with episodes of recurrent nonpitting edema. Sclerosis of the skin commences distally on the fingers (sclerodactyly) and spreads proximally, usually accompanied by resorption of bone of the fingertips, which may have punched out ulcers, stellate scars, or areas of hemorrhage (Fig. 55-7). The fingers may actually shrink and become sausage-shaped, and, because the fingernails are usually unaffected, they may curve over the end of the fingertips. Periungual telangiectases are usually present, but periungual erythema is rare. In advanced cases, the extremities show contractures and calcinosis cutis. Facial involvement includes a smooth, unwrinkled brow, taut skin over the nose, shrinkage of tissue around the mouth, and perioral



FIGURE 55-7 Scleroderma showing acral sclerosis and focal digital ulcers.



FIGURE 55-8 Scleroderma often eventuates in development of an expressionless, masklike facies.

radial furrowing (Fig. 55-8). Matlike telangiectases are often present, particularly on the face and hands. Involved skin feels indurated, smooth, and bound to underlying structures; hyper- and hypopigmentation are common as well. *Raynaud's phenomenon* (i.e., cold-induced blanching, cyanosis, and reactive hyperemia) is documented in almost all patients and can precede development of scleroderma by many years. *Linear scleroderma* is a limited form of disease that presents in a linear, bandlike distribution and tends to involve deep as well as superficial layers of skin. The combination of calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectases has been termed as the *CREST syndrome*. Anti-centromere autoantibodies have been reported in a very high percentage of patients with CREST syndrome but in only a small minority of patients with scleroderma. Skin biopsy reveals thickening of the dermis, homogenization of collagen bundles, atrophic pilosebaceous and eccrine glands, and a sparse mononuclear cell infiltrate in the dermis and subcutaneous fat. Direct immunofluorescence microscopy of lesional skin is usually negative.

Morphea is characterized by localized thickening and sclerosis of skin; it dominates on the trunk. This disorder may affect children or adults. Morphea begins as erythematous or flesh-colored plaques that become sclerotic, develop central hypopigmentation, and have an erythematous border. In most cases, patients have one or a few lesions, and the disease is termed *localized morphea*. In some patients, widespread cutaneous lesions may occur without systemic involvement (*generalized morphea*). Many adults with generalized morphea have concomitant rheumatic or other autoimmune disorders. Skin biopsy of morphea is generally indistinguishable from that of scleroderma. Scleroderma and morphea are usually quite resistant to therapy. For this reason, physical therapy to prevent joint contractures and to maintain function is employed and is often helpful. Treatment options for early, rapidly progressive disease include phototherapy (UVA1 [ultraviolet A1 irradiation] or PUVA [psoralens + ultraviolet A irradiation]) or methotrexate (15–20 mg/week) alone or in combination with daily glucocorticoids.

Diffuse fasciitis with eosinophilia is a clinical entity that can sometimes be confused with scleroderma. There is usually a sudden onset of swelling, induration, and erythema of the extremities, frequently following significant physical exertion. The proximal portions of the extremities (upper arms, forearms, thighs, calves) are more often involved than are the hands and feet. While the skin is indurated, it usually displays a woody, dimpled, or "pseudocellulite" appearance rather than being bound down as in scleroderma; contractures may occur early secondary to fascial involvement. The latter may also cause muscle groups to be separated and veins to appear depressed (i.e., the "groove sign"). These skin findings are accompanied by peripheral-blood eosinophilia, increased erythrocyte sedimentation rate,

and sometimes hypergammaglobulinemia. Deep biopsy of affected areas of skin reveals inflammation and thickening of the deep fascia overlying muscle. An inflammatory infiltrate composed of eosinophils and mononuclear cells is usually found. Patients with eosinophilic fasciitis appear to be at increased risk for developing bone marrow failure or other hematologic abnormalities. While the ultimate course of eosinophilic fasciitis is uncertain, many patients respond favorably to treatment with prednisone in doses of 40–60 mg/d.

The *eosinophilia-myalgia syndrome*, a disorder with epidemic numbers of cases reported in 1989 and linked to ingestion of L-tryptophan manufactured by a single company in Japan, is a multisystem disorder characterized by debilitating myalgias and absolute eosinophilia in association with varying combinations of arthralgias, pulmonary symptoms, and peripheral edema. In a later phase (3–6 months after initial symptoms), these patients often develop localized scleroderma-like skin changes, weight loss, and/or neuropathy (Chap. 353). The precise cause of this syndrome, which may resemble other sclerotic skin conditions, is unknown. However, the implicated lots of L-tryptophan contained the contaminant 1,1-ethylened bis[tryptophan]. This contaminant may be pathogenic or may be a marker for another substance that provokes the disorder.

FURTHER READING

- BOLOGNA JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.
GOLDSMITH LA et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 8th ed. New York, McGraw-Hill, 2012.
HAMMERS CM, STANLEY JR: Mechanisms of disease: Pemphigus and bullous pemphigoid. *Annu Rev Pathol* 11:175, 2016.
SCHMIDT E, ZILLIKENS D: Pemphigoid diseases. *Lancet* 381:320, 2013.

56

Cutaneous Drug Reactions

Robert G. Micheletti, Misha Rosenbach,
Bruce U. Wintroub, Kanade Shinkai



Cutaneous reactions are among the most frequent adverse reactions to drugs. 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 3 billion prescriptions for >60,000 drug products, which include >2000 different active agents, 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 each year in the United States. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several large cohort studies established that acute cutaneous reactions to drugs affect about 3% of hospitalized patients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

Many drugs of common use are associated with a 1–2% rate of rashes during premarketing clinical trials. The risk is often higher when medications are used in general, unselected populations. The rate may reach 3–7% for amoxicillin, sulfamethoxazole, many anticonvulsants, and anti-HIV agents.

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, but neither their incidence nor their impact on public health has been evaluated.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions. Severe reactions are too rare to be detected in such cohorts. Although rare, severe cutaneous reactions to drugs have an important impact on health because of significant sequelae, including mortality. Adverse drug rashes are responsible for hospitalization, increase the duration of hospital stay, and can 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 immunocompromise or immune dysregulation. Individuals with advanced HIV disease (e.g., CD4 T lymphocyte count <200 cells/ μ L) have a 40- to 50-fold increased risk of adverse reactions to sulfamethoxazole (Chap. 197) and increased risk of severe hypersensitivity reactions.

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.

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 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. Specific clones are generated by medications that are frequently a cause of drug eruptions: penicillin G, amoxicillin, cephalosporins, sulfamethoxazole, phenobarbital, carbamazepine, and lamotrigine. Both CD4 and CD8 clones have been obtained; however, their specific roles in drug allergy have not been elucidated. 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 final 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 56-1).

Immediate Reactions 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. 346). 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.

TABLE 56-1 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 $T_{H}1$ cells	Tuberculin skin test, contact dermatitis
Type IVb	T lymphocyte-mediated eosinophil inflammation	IL-4, IL-5, IL-13 $T_{H}2$ 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.

Nonsteroidal anti-inflammatory drugs (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. 356). 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. 356). Cephalosporin and other medications, including monoclonal antibodies such as infliximab, rituximab, and omalizumab, may be associated with clinically similar "serum sickness-like" reactions. The mechanism of this reaction is unknown but is unrelated to immune complex formation and complement activation.

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 56-1). 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.

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 particular 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, highlighting a role for differential pharmacokinetic or pharmacodynamic effects. The value of routine screening of P450 enzymes has not been determined, though its cost-effectiveness in certain populations (e.g., patients with seizure disorder) has been suggested.

Associations between drug hypersensitivities and HLA haplotypes suggest a key role for immune mechanisms. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B*57:01 ([Chap. 197](#)). In Taiwan, within a homogeneous Han Chinese population, a 100% association was observed between SJS/TEN (but not DIHS) related to carbamazepine and HLA-B*15:02. In the same population, another 100% 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, the strong associations found in Taiwan have not been observed in other countries with more heterogeneous populations.

■ 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.

Several investigators have proposed that specific HLA haplotypes associated with drug hypersensitivity indeed play a pathogenic role; stimulation of carbamazepine-specific cytotoxic T lymphocytes (CTLs) in the context of HLA-B*15:02 results in production of a putative mediator of keratinocyte necrosis in TEN. Other studies have identified 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. 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. 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. 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 anti-neutrophil cytoplasmic antibody (p-ANCA) (minocycline) antibodies. Subacute lupus erythematosus (SCLE) can be induced by a growing list of drugs, including thiazide diuretics, TNF-inhibitors, terbinafine, and minocycline. IFN and TNF-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 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, and the *FLT3*-inhibitor class of drugs 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 small 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, and the extent of exposure to the activating wavelengths of ultraviolet (UV) light ([Chap. 57](#)).

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. 56-1). Lesions are sharply demarcated, erythematous, or purpuric, and may progress to form large, hemorrhagic bullae with necrosis and eschar formation.

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 areas adjacent to the injection site or more distant sites if infused.

Warfarin-induced cutaneous necrosis is treated with vitamin K, heparin, surgical debridement, and intensive wound care. Treatment with protein C concentrates may also be helpful. Newer anticoagulants such as dabigatran etexilate may avoid warfarin necrosis in high-risk patients.

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.

PARONYCHIA Paronychia and multiple pyogenic granuloma 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.

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 have 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. 56-2). 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.

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 palmar/plantar, flexural sites or areas of frictional pressure. BRAF inhibitors are associated with photosensitivity, palmar/plantar 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



FIGURE 56-1 Warfarin necrosis involving the breasts.



FIGURE 56-2 Toxic erythema of chemotherapy.

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.

■ IMMUNE CUTANEOUS REACTIONS: COMMON

Maculopapular Eruptions Morbilliform or maculopapular eruptions (Fig. 56-3) 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 should raise concern for a more severe reaction. Involvement of mucous membranes is rare and should prompt consideration of SJS. 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. Skin biopsy often shows nonspecific inflammatory changes.

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,



FIGURE 56-3 Morbilliform drug eruption.

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.

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-like reactions usually occur 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,



FIGURE 56-4 Allergic contact dermatitis (bullous) due to adhesive tape.

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 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. 56-4). Harsh disinfectant skin cleansers may lead to localized irritant dermatitis.

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. 56-5). 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 56-5 Fixed drug eruption.

■ 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; since eosinophilia is not always present, the term DIHS is now 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. 56-6). 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 involvements are more common with minocycline, gastrointestinal involvement is almost exclusively seen with abacavir, and some medications typically lack 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. Reactivation of herpes viruses, in particular human herpesviruses 6 and 7, EBV, and cytomegalovirus (CMV), has been frequently reported in this syndrome, although the causal role of viral infection has been debated. Recent research suggests that inciting drugs may reactivate quiescent herpes viruses, resulting in expansion of viral-specific CD8+ T lymphocytes and subsequent end-organ damage.



FIGURE 56-6 Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS).

368 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 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).

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. 56-7). The term *Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) overlap* is used to describe cases with 10–30% epidermal detachment (Fig. 56-8), and *TEN* is used to describe cases with >30% detachment (Figs. 56-9 and 56-10).

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* infection in children causes a clinically distinct presentation with prominent mucositis and limited cutaneous involvement. The name *Mycoplasma*-induced rash and mucositis 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. 56-11). 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. 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. Issues such as fluid management, atraumatic wound care, infection prevention and treatment, and ophthalmologic and respiratory support are critical. Systemic glucocorticoid therapy (prednisone 1–2 mg/kg)



FIGURE 56-7 Stevens-Johnson syndrome (SJS).



FIGURE 56-8 SJS-TEN overlap.

may be useful early in disease evolution; however, long-term or late systemic glucocorticoid use has been associated with increased mortality. After initial enthusiasm for the use of intravenous immunoglobulin (IVIG) in the treatment of *SJS/TEN*, more recent data question whether it is beneficial. There are emerging data to support treatment with cyclosporine and etanercept. 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. 56-12). Patients typically present with diffuse erythema or erythroderma, as well as high spiking fevers, and



FIGURE 56-9 Toxic epidermal necrolysis, hand.



FIGURE 56-10 Toxic epidermal necrolysis.

leukocytosis. 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; 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.

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, though 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.



FIGURE 56-11 Target-like lesion in SJS.



FIGURE 56-12 Acute generalized exanthematous pustulosis.

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. 56-13) (Chap. 356). 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.

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 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?
3. Which drug or drugs are suspected, and should they be withdrawn?
4. What recommendation can be made for future medication use?



FIGURE 56-13 Cutaneous small-vessel vasculitis (CSVV, leukocytoclastic vasculitis).

■ 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 56-2** lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. **Table 56-3** lists the most important of these reactions, along

TABLE 56-2 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/ μ L
- Lymphocytosis with atypical lymphocytes
- Abnormal liver or kidney function tests

Source: Adapted from JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. *N Engl J Med* 331:1272, 1994.

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.

■ 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 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 in the population 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

TABLE 56-3 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	Urticular 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

^aOverlap of SJS and TEN have features of both, and attachment of 10–30% of body surface area may occur.

Source: Adapted from JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. N Engl J Med 331:1272, 1994.

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. Approximately 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics.

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 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. 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 limited 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.

- BELUM VR: Characterisation and management of dermatologic adverse events to agents targeting the PD-1 receptor. *Eur J Cancer* 60:12, 2016.
- CORNEJO-GARCIA JA et al: The genetics of drug hypersensitivity reactions. *J Investig Allergol Clin Immunol* 26:222, 2016.
- CREAMER D et al: U.K. guidelines for the management of Stevens-Johnson syndrome/toxic epidermal necrolysis in adults 2016. *Br J Dermatol* 174:1194, 2016.
- HARP JL et al: Severe cutaneous adverse reactions: impact of immunology, genetics, and pharmacology. *Semin Cutan Med Surg* 33:17, 2014.
- 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.
- LACOUTURE ME et al: Ipilimumab in patients with cancer and the management of dermatologic adverse events. *J Am Acad Dermatol* 71:161, 2014.
- MAYORGA C et al: In vitro tests for drug hypersensitivity reactions: An ENDA/EAACI Drug Allergy Interest Group position paper. *Allergy* 71:1103, 2016.
- OUESSALAH A et al: Genetic variants associated with drug-induced immediate hypersensitivity reactions: A PRISMA-compliant systematic review. *Allergy* 71:443, 2016.
- 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*, 2016 ePub ahead of print. Accessed September 28, 2016.
- 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.
- WHITE KD et al: Evolving models of the immunopathogenesis of T cell-mediated drug allergy: The role of host, pathogens, and drug response. *J Allergy Clin Immunol* 136:219, 2015.
- WOLVERTON SE: Practice gaps: Drug reactions. *Dermatol Clin* 34:311, 2016.

57

Photosensitivity and Other Reactions to Light

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 are capable of affecting 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 Epidermal DNA—predominantly in keratinocytes and in Langerhans cells, which are 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 most basal cell and squamous cell carcinomas (BCCs and SCCs, respectively). 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), 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 chemical components 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. 409), porphyrins are released into the circulation in increased amounts from the bone marrow and 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 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 the treatment of thin actinic keratoses. Red light has a longer wavelength, 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 mechanisms.

Genetic studies have revealed additional genes that influence skin color variation in humans, such as the gene for tyrosinase (*TYR*) and the genes *APBA2/OCA2*, *SLC45A2*, and *SLC24A5*. The human *MC1R* gene encodes a G protein-coupled receptor that binds

TABLE 57-1 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

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 in 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 57-1).

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 κ B-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 also has been shown to activate phospholipase A₂, resulting in increases in eicosanoids such as prostaglandin E₂, 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 D₃, which then undergoes temperature-dependent isomerization to form the stable hormone vitamin D₃. 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 D₃. 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 D₃, 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. On the basis of 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 nonmelanoma skin cancer (NMSC). The two most common types of NMSC are BCC and SCC (Chap. 72). 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 malignant lesions, a process thought to require additional genetic alterations.

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. 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 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 that are critical for their formation. BCCs harbor inactivating mutations particularly in the tumor-suppressor gene *patched* or activating mutations in the oncogene *smoothened*, which results in the constitutive activation of the sonic hedgehog signaling pathway and increased cell proliferation. New evidence links 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. Thus, 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 and SCCs) 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.

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. 72**). 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.

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 melanogenesis, which is consistent with findings in mouse models showing that pheomelanin can promote melanoma formation 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 novel therapeutic approach for melanoma, termed immune checkpoint blockade, targets inhibitors of T cell activation (such as CTLA-4 or PD-1) that in a subset of patients has resulted in a durable and potent immune destruction of melanoma cells, resulting in prolonged survival of patients with metastatic melanoma. It has recently been shown that a high mutational load in melanomas correlated indeed 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 immunosuppression (inhibition of immune responses to antigens applied at the irradiated site) and systemic immunosuppression (inhibition of immune responses to antigens applied at remote, unirradiated sites). For example, human skin exposure to modest doses of UV-B can deplete the epidermal antigen-presenting cells known as Langerhans cells, thereby reducing the degree of allergic sensitization to 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 distant from the irradiated site. Various immunomodulatory factors and immune cells have been implicated in UVR-induced systemic immunosuppression, including tumor necrosis factor α , interleukin 4, interleukin 10, *cis*-urocanic acid, and eicosanoids. Experimental evidence suggests that prostaglandin E₂ signaling through prostaglandin E receptor subtype 4 mediates UVR-induced systemic immunosuppression by elevating the number of regulatory T cells, and this effect can be inhibited with NSAIDs.

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. 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 catalyzing 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.

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 interleukin 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 is a major molecular trigger of this immunosuppressive effect.

Perhaps the most graphic demonstration of the role of 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 cancers are the most common types of malignancies arising in these patients. Rates of BCC and SCC increase with the duration and degree of immunosuppression. These patients 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.

■ 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. Fragrances such as musk ambrette that were previously present in numerous cosmetic products are also potent photosensitizers.

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 57-2**). 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.

Polymorphous Light Eruption The most common type of photosensitivity disease is *polymorphous light eruption* (PMLE). Many affected individuals never seek medical attention because the condition

TABLE 57-2 Classification of Photosensitivity Diseases

TYPE	DISEASE
Genetic	Erythropoietic porphyria
	Erythropoietic protoporphyrina
	Porphyria cutanea tarda—familial
	Variegate porphyria
	Hepatoerythropoietic porphyria
	Albinism
	Xeroderma pigmentosum
	Rothmund-Thomson syndrome
	Bloom syndrome
	Cockayne syndrome
	Kindler syndrome
	Phenylketonuria
Metabolic	Porphyria cutanea tarda—sporadic
	Hartnup disease
	Kwashiorkor
	Pellagra
Phototoxic	Carcinoid syndrome
Internal	
	Drugs
External	
	Drugs, plants, food
Photoallergic	
	Solar urticaria
	Drug photoallergy
Neoplastic and degenerative	Persistent light reaction/chronic actinic dermatitis
	Photoaging
	Actinic keratosis
Idiopathic	Melanoma and nonmelanoma skin cancer
	Polymorphous light eruption
	Hydroa aestivale
Photoaggravated	Actinic prurigo
	Lupus erythematosus
	Systemic
	Subacute cutaneous
	Discoid
	Dermatomyositis
	Herpes simplex
	Lichen planus actinicus
	Acne vulgaris (aestivale)

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 seriously involved. 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 erythema 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.

TABLE 57-3 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		+

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, a few of which are listed in **Table 57-3**. 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.

Photoallergy is much less common and is distinct in that it is an immunopathologic process. The excited-state photosensitizer may create highly unstable haptene free radicals that bind covalently to macromolecules to form a functional antigen capable of evoking a delayed-type hypersensitivity response. Some drugs and chemicals that can produce photoallergy are listed in **Table 57-4**. 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*.

A very 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. 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,

TABLE 57-4 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		+

determining the minimal erythema 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 suberythema 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 risks, it may be necessary to employ an immunosuppressive drug such as azathioprine, cyclophosphamide, cyclosporine, or mycophenolate mofetil.

Porphyria The porphyrias (**Chap. 409**) 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 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 either causing (1) chronic blistering photosensitivity or (2) acute nonblistering photosensitivity. Chronic cutaneous porphyrias include porphyria cutanea tarda (PCT), congenital erythropoietic porphyria (CEP), hepatoerythropoietic porphyria (HEP), hereditary coproporphyrina (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 protoporphyrina (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 protoporphyrinia (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.

The diagnosis is confirmed by demonstration of elevated levels of free erythrocyte protoporphyrin. Detection of increased plasma protoporphyrin helps distinguish EPP from lead poisoning and iron-deficiency anemia, in both of which erythrocyte protoporphyrin levels are elevated in the absence of cutaneous photosensitivity and elevated plasma protoporphyrin levels.

Rigorous sunlight protection is essential in the management of EPP. Therapies that may increase sunlight tolerance in patients with EPP may be helpful as well, such as oral administration of β -carotene, which is an effective scavenger of free radicals. Notably, a recent clinical trial showed that a synthetic peptide analogue of α -MSH, afamelanotide, increased skin pigmentation through melanogenesis and thereby enhanced tolerance to sunlight in patients with EPP. Patients treated with afamelanotide tolerated sun exposure without pain for longer periods of time and had an improved quality of life as compared to untreated patients. Interestingly, initial studies suggest that afamelanotide may also be beneficial when combined with NB-UVB in the treatment of patients with vitiligo (in patients with skin phototypes IV–VI).

An algorithm for managing patients with photosensitivity is presented in **Fig. 57-1**.

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

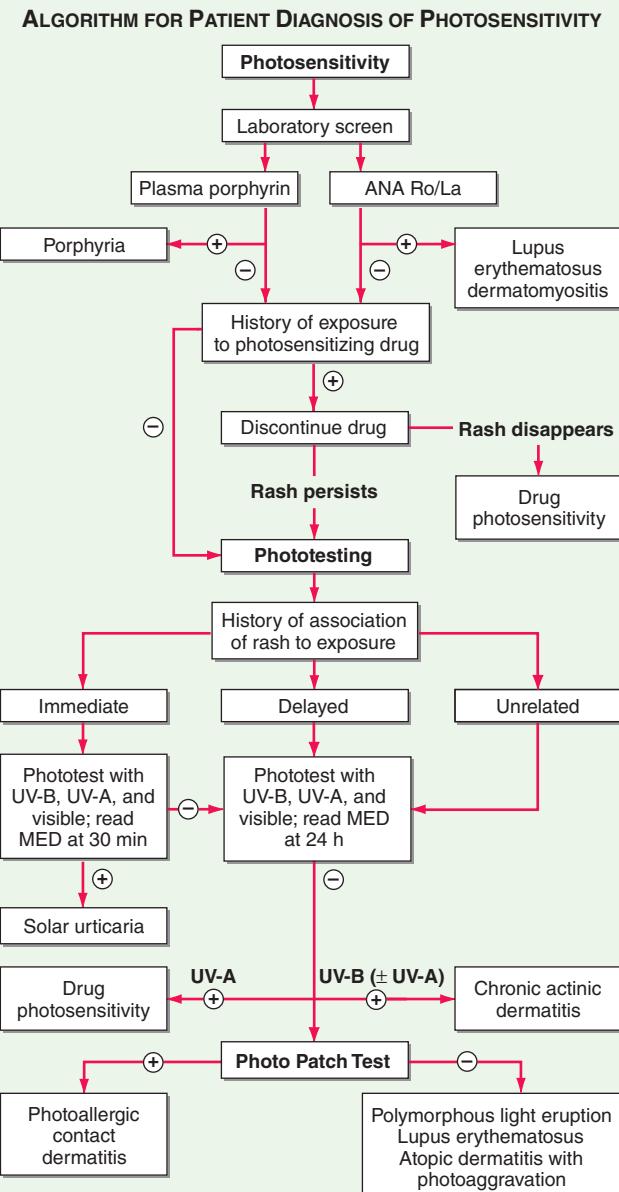


FIGURE 57-1 Algorithm for the diagnosis of a patient with photosensitivity. ANA, antinuclear antibody; MED, minimal erythema dose; UV-A and UV-B, ultraviolet spectrum segments including wavelengths of 320–400 nm and 290–320 nm, respectively.

sleeves, and trousers all reduce direct exposure. Sunscreens are now considered over-the-counter drugs, and a monograph from the U.S. Food and Drug Administration (FDA) has recognized category I ingredients as safe and effective. Those ingredients are listed in Table 57-5. 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+).

Broad-spectrum sunscreens contain both UV-B-absorbing and UV-A-absorbing chemicals, the latter including avobenzone and ecamulse (terephthalylidene dicamphor sulfonic acid). These chemicals absorb UVR and transfer the absorbed energy to surrounding cells. In contrast, physical UV blockers (zinc oxide and titanium dioxide) scatter or reflect UVR.

In addition to light absorption, a critical determinant of the sustained photoprotective effect of sunscreens is their water resistance. The FDA monograph has defined strict testing criteria for sunscreens

TABLE 57-5 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
Sulisobenzene (benzophenone-4)	10
Titanium dioxide	25
Trolamine salicylate	12
Zinc oxide	25

Abbreviation: FDA, U.S. Food and Drug Administration.

that claim to possess a high degree of water resistance. 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 and atopic dermatitis. 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-methoxysoralen 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

- FELL GL et al: Skin beta-endorphin mediates addiction to UV light. *Cell* 157:1527, 2014.
- JANSEN R et al: Photoprotection: Part II. Sunscreen: development, efficacy, and controversies. *J Am Acad Dermatol* 69:867, 2013.
- MARTINCORENA I et al: Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348:880, 2015.
- SANCHEZ-DANES A et al: Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature* 536:298, 2016.
- VAN ALLEN EM et al: Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350:207, 2015.

lymphocyte nucleus may be microcytic; those larger than the small lymphocyte nucleus may be macrocytic. Macrocytic cells also tend to be more oval than spherical in shape and are sometimes called macroovalocytes. The automated mean corpuscular volume (MCV) can assist in making a classification. However, some patients may have both iron and vitamin B₁₂ deficiency, which will produce an MCV in the normal range but wide variation in red cell size. When the red cells vary greatly in size, *anisocytosis* is said to be present. When the red cells vary greatly in shape, *poikilocytosis* is said to be present. The electronic cell counter provides an independent assessment of variability in red cell size. It measures the range of red cell volumes and reports the results as "red cell distribution width" (RDW). This value is calculated from the MCV; thus, cell width is not being measured but cell volume is. The term is derived from the curve displaying the frequency of cells at each volume, also called the distribution. The width of red cell volume distribution curve is what determines the RDW. The RDW is calculated as follows: RDW = (standard deviation of MCV ÷ mean MCV) × 100. In the presence of morphologic anisocytosis, RDW (normally 11–14%) increases to 15–18%. The RDW is useful in at least two clinical settings. In patients with microcytic anemia, the differential diagnosis is generally between iron deficiency and thalassemia. In thalassemia, the small red cells are generally of uniform size with a normal small RDW. In iron deficiency, the size variability and the RDW are large. In addition, a large RDW can suggest a dimorphic anemia when a chronic atrophic gastritis can produce both vitamin B₁₂ malabsorption to produce macrocytic anemia and blood loss to produce iron deficiency. In such settings, RDW is also large. An elevated RDW also has been reported as a risk factor for all-cause mortality in population-based studies, a finding that is unexplained currently.

After red cell size is assessed, one examines the hemoglobin content of the cells. They are either normal in color (*normochromic*) or pale in color (*hypochromic*). They are never "hyperchromic." If more than the normal amount of hemoglobin is made, the cells get larger—they do not become darker. In addition to hemoglobin content, the red cells are examined for inclusions. Red cell inclusions are the following:

- Basophilic stippling*—diffuse fine or coarse blue dots in the red cell usually representing RNA residue—especially common in lead poisoning
- Howell-Jolly bodies*—dense blue circular inclusions that represent nuclear remnants—their presence implies defective splenic function
- Nuclei*—red cells may be released or pushed out of the marrow prematurely before nuclear extrusion—often implies a myelophthisic process or a vigorous narrow response to anemia, usually hemolytic anemia
- Parasites*—red cell parasites include malaria and babesia (**Chap. A6**)
- Polychromatophilia*—the red cell cytoplasm has a bluish hue, reflecting the persistence of ribosomes still actively making hemoglobin in a young red cell

Vital stains are necessary to see precipitated hemoglobin called *Heinz bodies*.

Red cells can take on a variety of different shapes. All abnormally shaped red cells are *poikilocytes*. Small red cells without the central pallor are *spherocytes*; they can be seen in hereditary spherocytosis, hemolytic anemias of other causes, and clostridial sepsis. *Dacrocytes* are teardrop-shaped cells that can be seen in hemolytic anemias, severe iron deficiency, thalassemias, myelofibrosis, and myelodysplastic syndromes. *Schistocytes* are helmet-shaped cells that reflect microangiopathic hemolytic anemia or fragmentation on an artificial heart valve. *Echinocytes* are spiculated red cells with the spikes evenly spaced; they can represent an artifact of abnormal drying of the blood smear or reflect changes in stored blood. They also can be seen in renal failure and malnutrition and are often reversible. *Acanthocytes* are spiculated red cells with the spikes irregularly distributed. This process tends to be irreversible and reflects underlying renal disease, abetalipoproteinemia, or splenectomy. *Elliptocytes* are elliptical-shaped red cells that can reflect an inherited defect in the red cell membrane, but they also are seen in iron deficiency, myelodysplastic syndromes, megaloblastic anemia, and thalassemias. *Stomatocytes* are red cells in which the area

Section 9 Hematologic Alterations

58

Interpreting Peripheral Blood Smears

Dan L. Longo



Some of the relevant findings in peripheral blood, enlarged lymph nodes, and bone marrow are illustrated in this chapter. Systematic histologic examination of the bone marrow and lymph nodes is beyond the scope of a general medicine textbook. However, every internist should know how to examine a peripheral blood smear.

The examination of a peripheral blood smear is one of the most informative exercises a physician can perform. Although advances in automated technology have made the examination of a peripheral blood smear by a physician seem less important, the technology is not a completely satisfactory replacement for a blood smear interpretation by a trained medical professional who also knows the patient's clinical history, family history, social history, and physical findings. It is useful to ask the laboratory to generate a Wright's-stained peripheral blood smear and examine it.

The best place to examine blood cell morphology is the feathered edge of the blood smear where red cells lie in a single layer, side by side, just barely touching one another but not overlapping. The author's approach is to look at the smallest cellular elements, the platelets, first and work his way up in size to red cells and then white cells.

Using an oil immersion lens that magnifies the cells 100-fold, one counts the platelets in five to six fields, averages the number per field, and multiplies by 20,000 to get a rough estimate of the platelet count. The platelets are usually 1–2 µm in diameter and have a blue granulated appearance. There is usually 1 platelet for every 20 or so red cells. Of course, the automated counter is much more accurate, but gross disparities between the automated and manual counts should be assessed. Large platelets may be a sign of rapid platelet turnover, as young platelets are often larger than old ones; alternatively, certain rare inherited syndromes can produce large platelets. If the platelet count is low, the absence of large (young) platelets may be an indicator of marrow production problems. Platelet clumping visible on the smear can be associated with falsely low automated platelet counts. Clumping may be caused by the anticoagulant into which the blood is drawn. Similarly, neutrophil fragmentation can be a source of falsely elevated automated platelet counts. The absence of platelet granules may be an artifact of the handling of the blood or may indicate marrow disease or a rare congenital anomaly, gray platelet syndrome. Elevated platelet counts usually signify a myeloproliferative disorder or a reaction to systemic inflammation.

Next one examines the red blood cells. One can gauge their size by comparing the red cell to the nucleus of a small lymphocyte. Both are normally about 8-µm wide. Red cells that are smaller than the small

of central pallor takes on the morphology of a slit instead of the usual round shape. Stomatocytes can indicate an inherited red cell membrane defect and also can be seen in alcoholism. *Target cells* have an area of central pallor that contains a dense center, or bull's-eye. These cells are seen classically in thalassemia, but they are also present in iron deficiency, cholestatic liver disease, and some hemoglobinopathies. They also can be generated artifactually by improper slide making.

One last feature of the red cells to assess before moving to the white blood cells is the distribution of the red cells on the smear. In most individuals, the cells lie side by side in a single layer. Some patients have red cell clumping (called *agglutination*) in which the red cells pile upon one another; it is seen in certain paraproteinemias and autoimmune hemolytic anemias. Another abnormal distribution involves red cells lying in single cell rows on top of one another like stacks of coins. This is called *rouleaux formation* and reflects abnormal serum protein levels.

Finally, one examines the white blood cells. Three types of granulocytes are usually present: neutrophils, eosinophils, and basophils, in decreasing frequency. Neutrophils are generally the most abundant white cell. They are round, are 10–14 μm wide, and contain a lobulated nucleus with two to five lobes connected by a thin chromatin thread. Bands are immature neutrophils that have not completed nuclear condensation and have a U-shaped nucleus. Bands reflect a left shift in neutrophil maturation in an effort to make more cells more rapidly. Neutrophils can provide clues to a variety of conditions. Vacuolated

neutrophils may be a sign of bacterial sepsis. The presence of 1- to 2- μm blue cytoplasmic inclusions, called *Döhle bodies*, can reflect infections, burns, or other inflammatory states. If the neutrophil granules are larger than normal and stain a darker blue, “toxic granulations” are said to be present, and they also suggest a systemic inflammation. The presence of neutrophils with more than five nuclear lobes suggests megaloblastic anemia. Large misshapen granules may reflect the inherited Chédiak-Higashi syndrome.

Eosinophils are slightly larger than neutrophils, have bilobed nuclei, and contain large red granules. Diseases of eosinophils are associated with too many of them rather than any morphologic or qualitative change. They normally total less than one-thirtieth the number of neutrophils. Basophils are even more rare than eosinophils in the blood. They have large dark blue granules and may be increased as part of chronic myeloid leukemia.

Lymphocytes can be present in several morphologic forms. Most common in healthy individuals are small lymphocytes with a small dark nucleus and scarce cytoplasm. In the presence of viral infections, more of the lymphocytes are larger, about the size of neutrophils, with abundant cytoplasm and a less condensed nuclear chromatin. These cells are called *reactive lymphocytes*. About 1% of lymphocytes are larger and contain blue granules in a light blue cytoplasm; they are called *large granular lymphocytes*. In chronic lymphoid leukemia, the small lymphocytes are increased in number, and many of them are ruptured

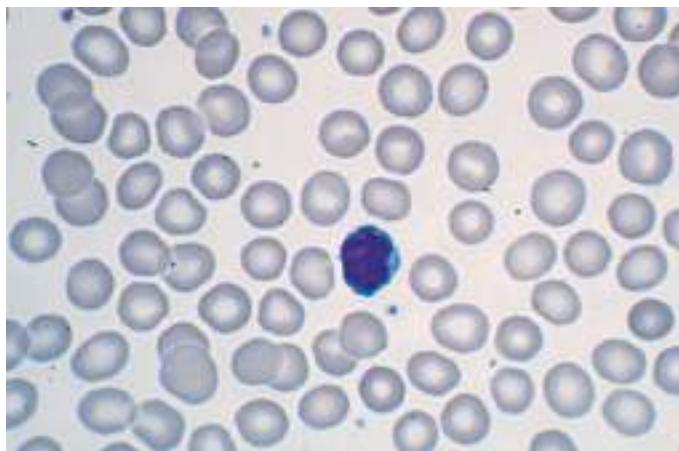


FIGURE 58-1 Normal peripheral blood smear. Small lymphocyte in center of field. Note that the diameter of the red blood cell is similar to the diameter of the small lymphocyte nucleus.

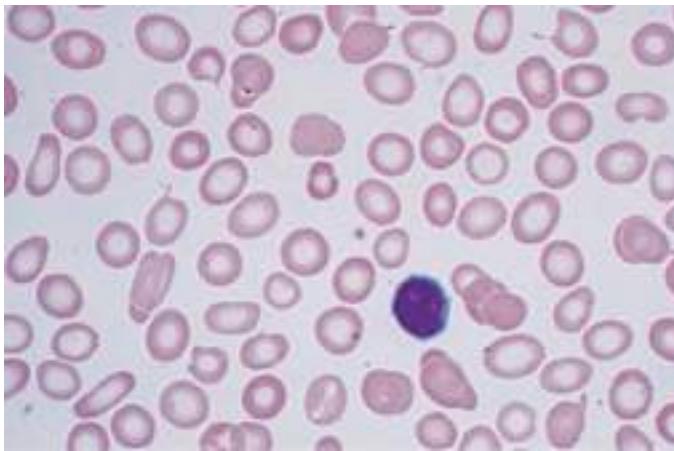


FIGURE 58-3 Hypochromic microcytic anemia of iron deficiency. Small lymphocyte in field helps assess the red blood cell size.

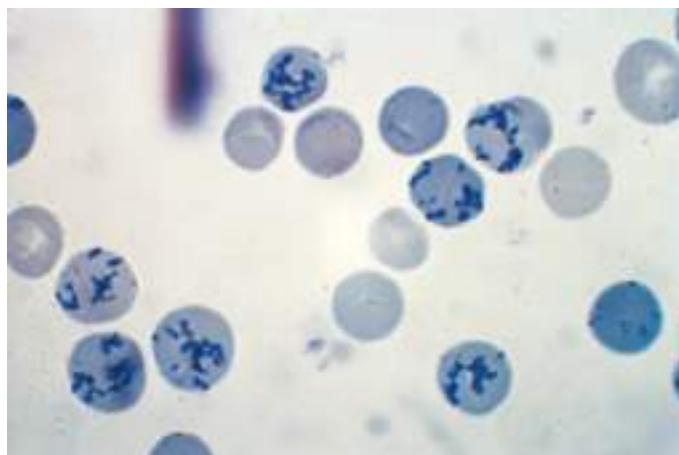


FIGURE 58-2 Reticulocyte count preparation. This new methylene blue-stained blood smear shows large numbers of heavily stained reticulocytes (the cells containing the dark blue-staining RNA precipitates).

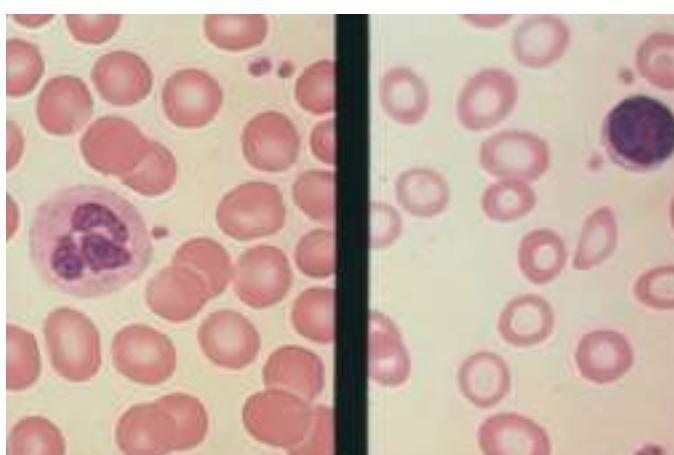


FIGURE 58-4 Iron deficiency anemia next to normal red blood cells. Microcytes (right panel) are smaller than normal red blood cells (cell diameter <7 μm) and may or may not be poorly hemoglobinized (hypochromic).

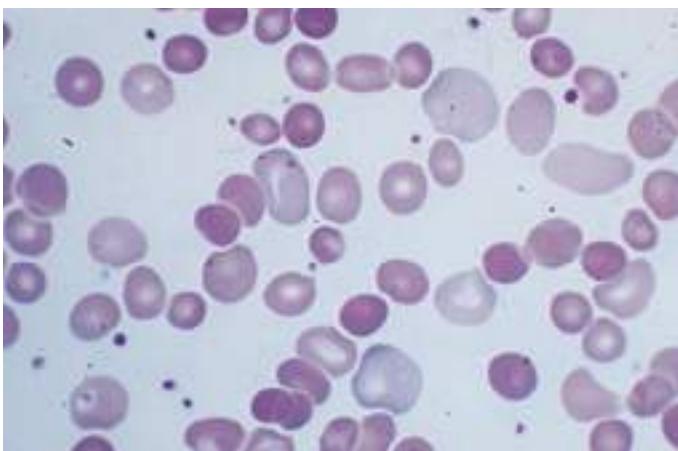


FIGURE 58-5 Polychromatophilia. Note large red cells with light purple coloring.

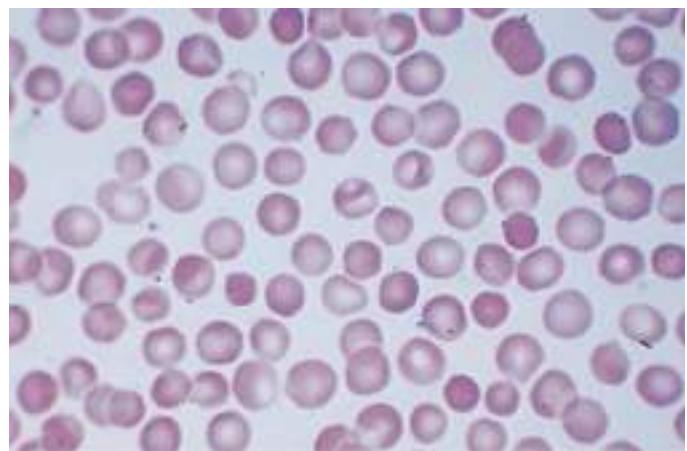


FIGURE 58-8 Spherocytosis. Note small hyperchromatic cells without the usual clear area in the center.

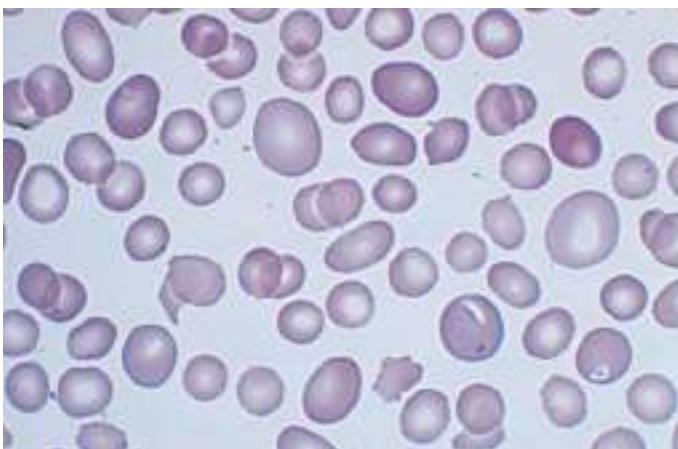


FIGURE 58-6 Macrocytosis. These cells are both larger than normal (mean corpuscular volume >100) and somewhat oval in shape. Some morphologists call these cells macroovalocytes.

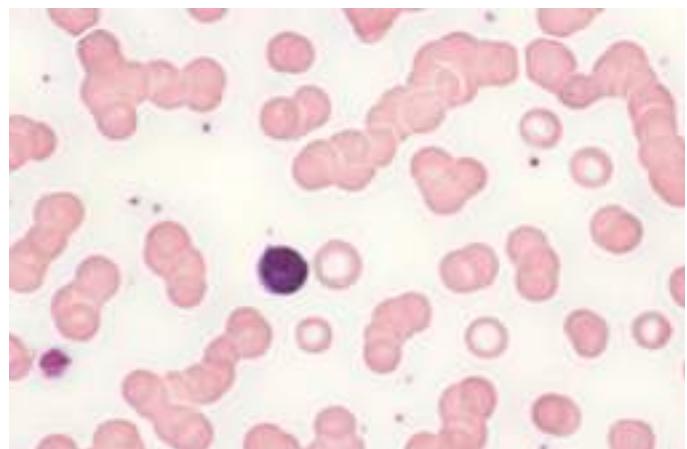


FIGURE 58-9 Rouleaux formation. Small lymphocyte in center of field. These red cells align themselves in stacks and are related to increased serum protein levels.

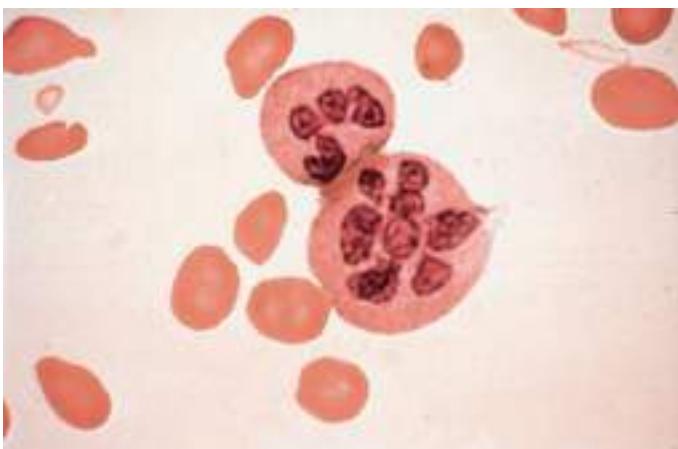


FIGURE 58-7 Hypersegmented neutrophils. Hypersegmented neutrophils (multilobed polymorphonuclear leukocytes) are larger than normal neutrophils with five or more segmented nuclear lobes. They are commonly seen with folic acid or vitamin B₁₂ deficiency.

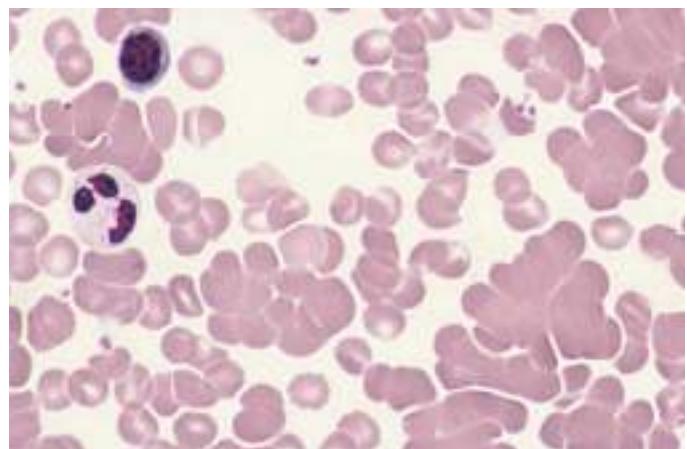


FIGURE 58-10 Red cell agglutination. Small lymphocyte and segmented neutrophil in upper left center. Note irregular collections of aggregated red cells.

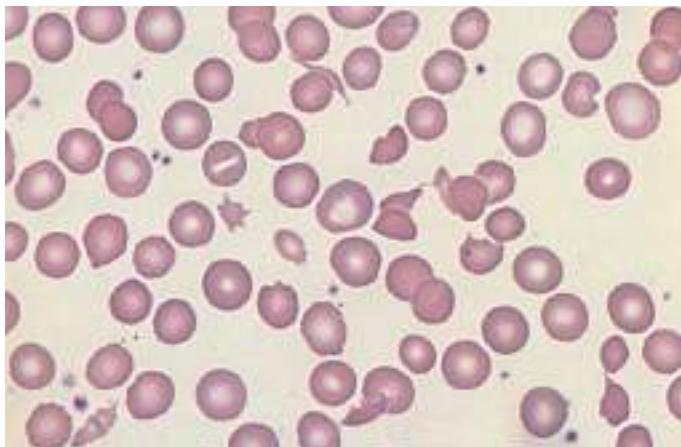


FIGURE 58-11 Fragmented red cells. Heart valve hemolysis.

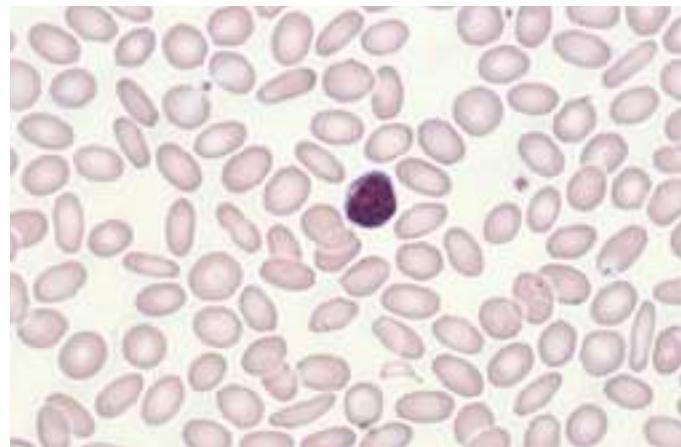


FIGURE 58-14 Elliptocytosis. Small lymphocyte in center of field. Elliptical shape of red cells related to weakened membrane structure, usually due to mutations in spectrin.

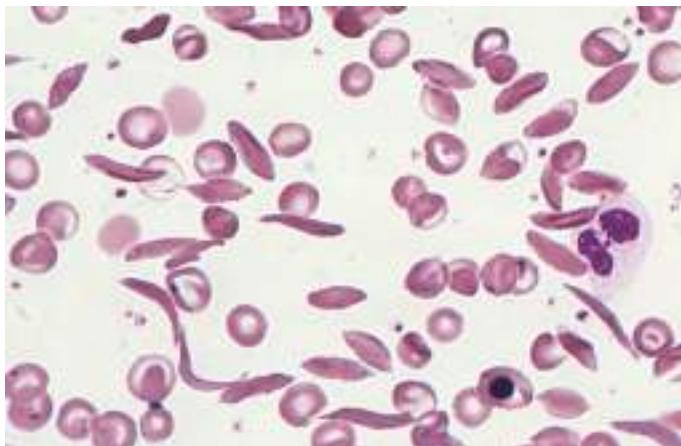


FIGURE 58-12 Sickled cells. Homozygous sickle cell disease. A nucleated red cell and neutrophil are also in the field.

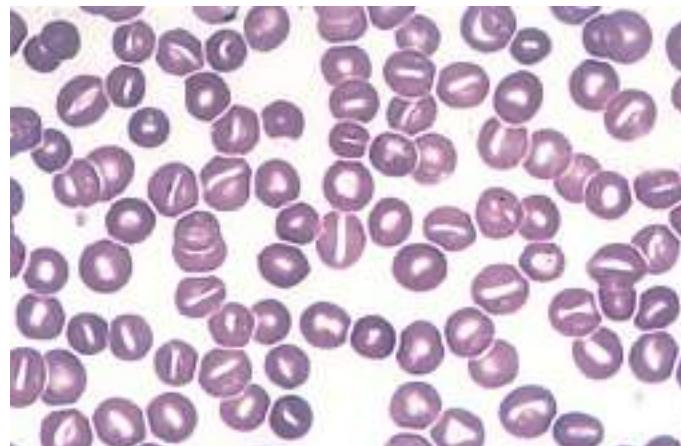


FIGURE 58-15 Stomatocytosis. Red cells characterized by a wide transverse slit or stoma. This often is seen as an artifact in a dehydrated blood smear. These cells can be seen in hemolytic anemias and in conditions in which the red cell is overhydrated or dehydrated.

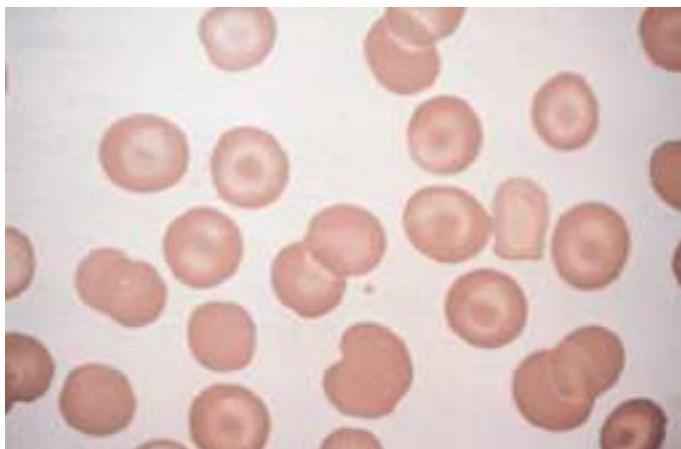


FIGURE 58-13 Target cells. Target cells are recognized by the bull's-eye appearance of the cell. Small numbers of target cells are seen with liver disease and thalassemia. Larger numbers are typical of hemoglobin C disease.

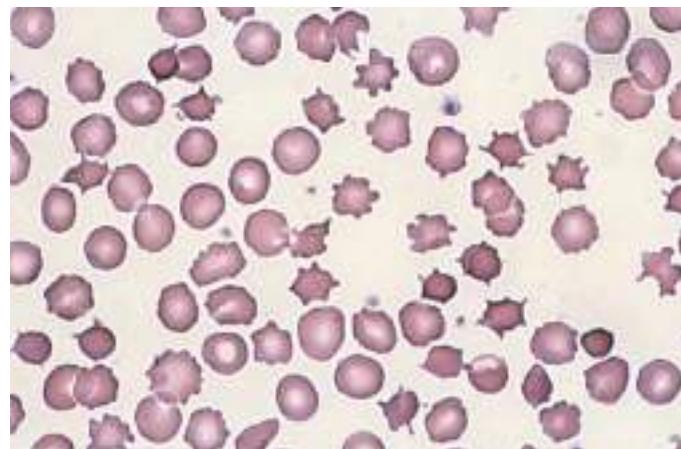


FIGURE 58-16 Acanthocytosis. Spiculated red cells are of two types: acanthocytes are contracted dense cells with irregular membrane projections that vary in length and width; echinocytes have small, uniform, and evenly spaced membrane projections. Acanthocytes are present in severe liver disease, in patients with abetalipoproteinemia, and in rare patients with McLeod blood group. Echinocytes are found in patients with severe uremia, in glycolytic red cell enzyme defects, and in microangiopathic hemolytic anemia.



FIGURE 58-17 Howell-Jolly bodies. Howell-Jolly bodies are tiny nuclear remnants that normally are removed by the spleen. They appear in the blood after splenectomy (defect in removal) and with maturation/dysplastic disorders (excess production).

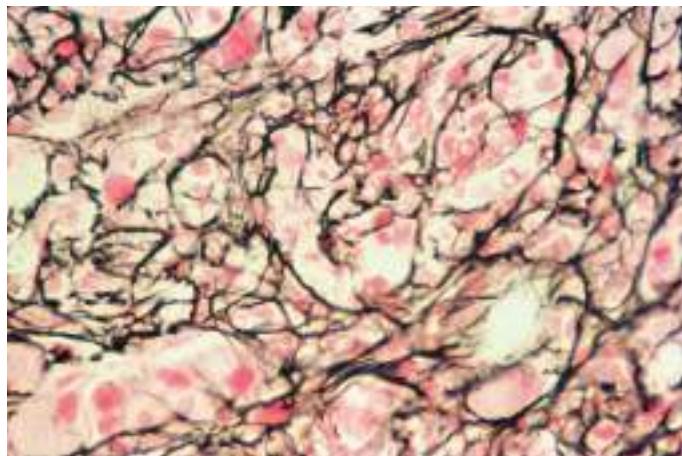


FIGURE 58-20 Reticulin stain of marrow myelofibrosis. Silver stain of a myelofibrotic marrow showing an increase in reticulin fibers (black-staining threads).

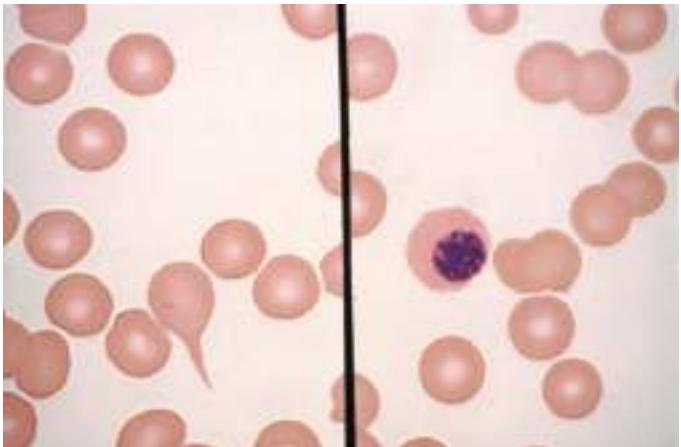


FIGURE 58-18 Teardrop cells and nucleated red blood cells characteristic of myelofibrosis. A teardrop-shaped red blood cell (left panel) and a nucleated red blood cell (right panel) as typically seen with myelofibrosis and extramedullary hematopoiesis.

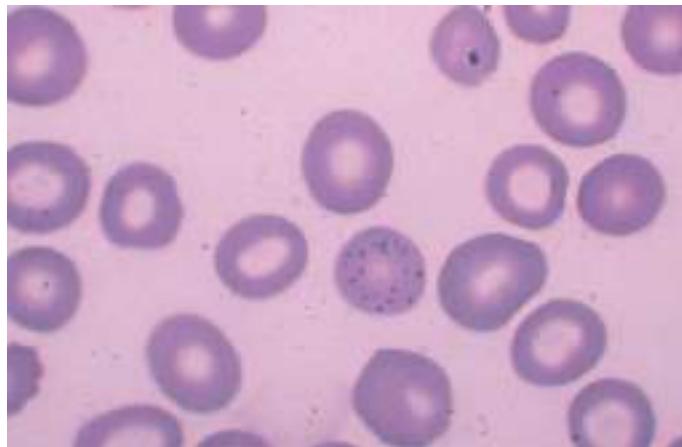


FIGURE 58-21 Stippled red cell in lead poisoning. Mild hypochromia. Coarsely stippled red cell.

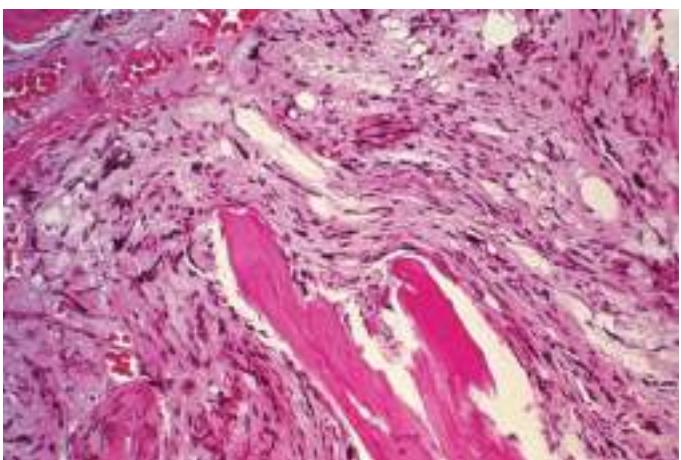


FIGURE 58-19 Myelofibrosis of the bone marrow. Total replacement of marrow precursors and fat cells by a dense infiltrate of reticulin fibers and collagen (H&E stain).

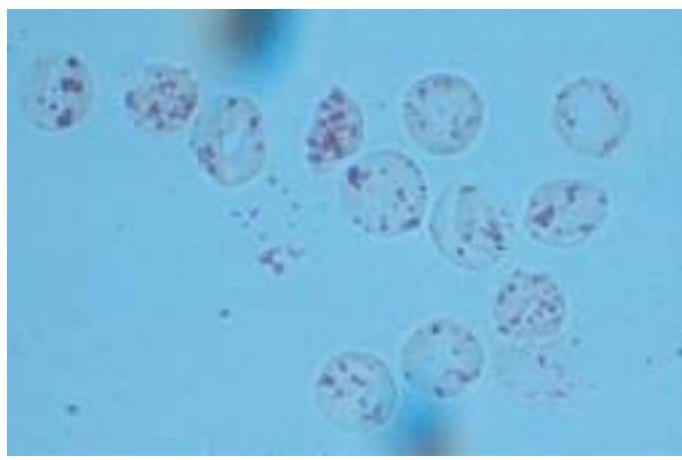


FIGURE 58-22 Heinz bodies. Blood mixed with hypotonic solution of crystal violet. The stained material is precipitates of denatured hemoglobin within cells.

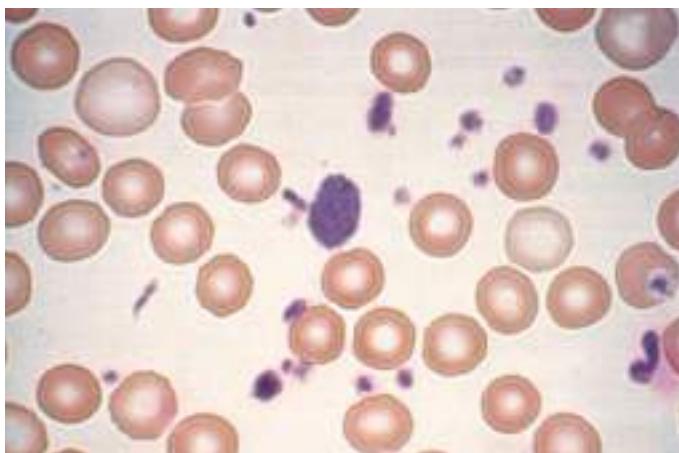


FIGURE 58-23 Giant platelets. Giant platelets, together with a marked increase in the platelet count, are seen in myeloproliferative disorders, especially primary thrombocythemia.

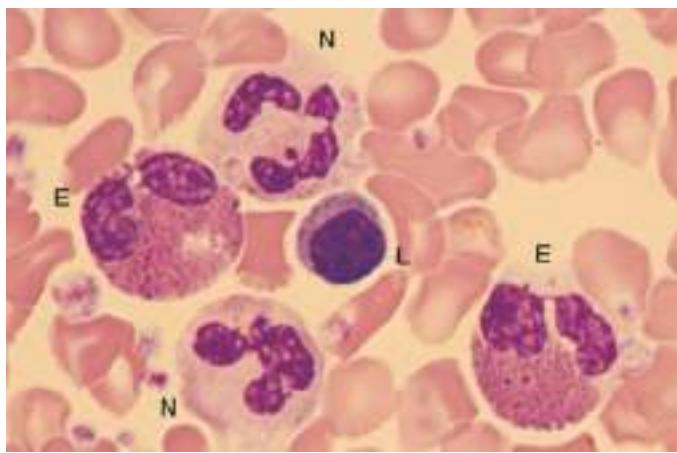


FIGURE 58-26 Normal eosinophils. The film was prepared from the buffy coat of the blood from a normal donor. E, eosinophil; L, lymphocyte; N, neutrophil.

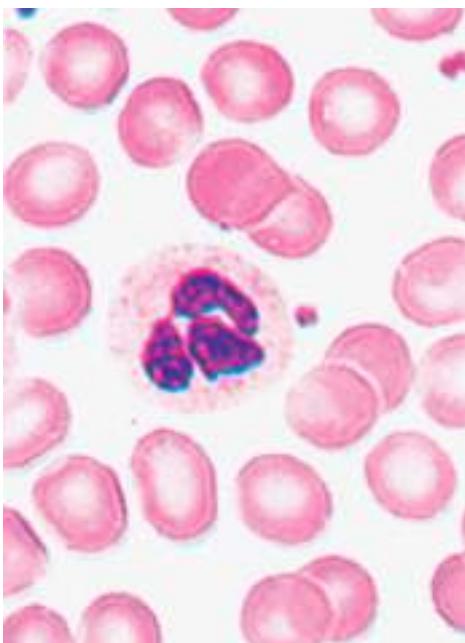


FIGURE 58-24 Normal granulocytes. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

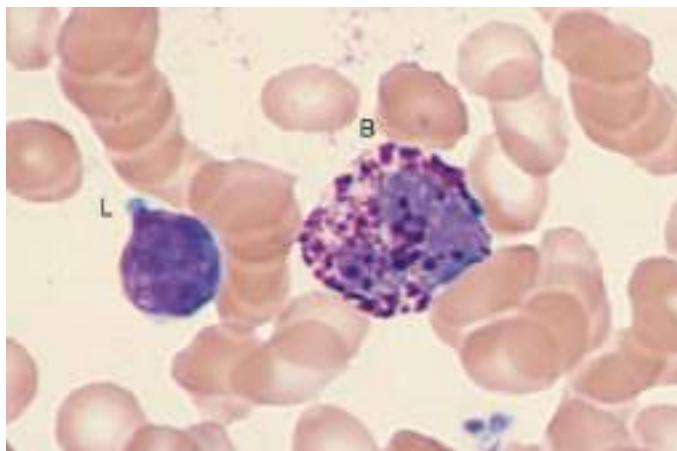


FIGURE 58-27 Normal basophil. The film was prepared from the buffy coat of the blood from a normal donor. B, basophil; L, lymphocyte.

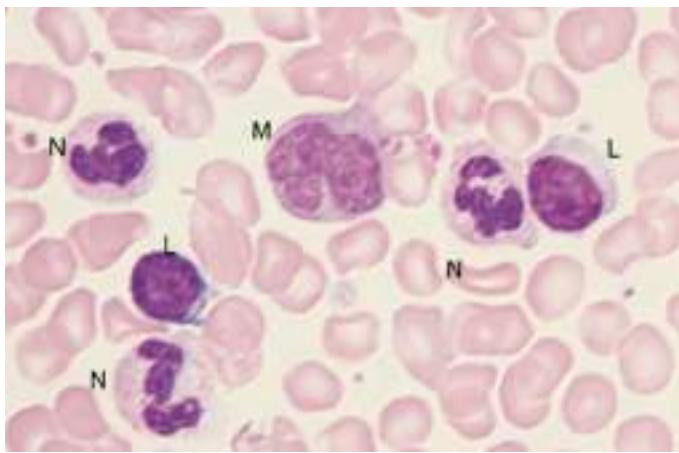


FIGURE 58-25 Normal monocytes. The film was prepared from the buffy coat of the blood from a normal donor. L, lymphocyte; M, monocyte; N, neutrophil.

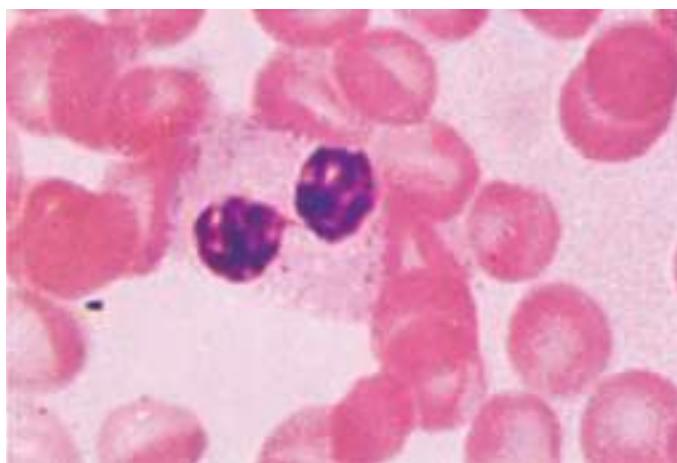


FIGURE 58-28 Pelger-Huet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or "pince-nez," configuration.

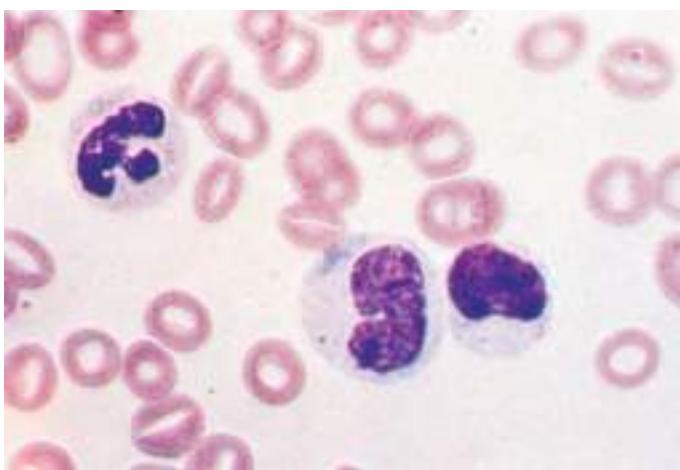


FIGURE 58-29 Döhle body. Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

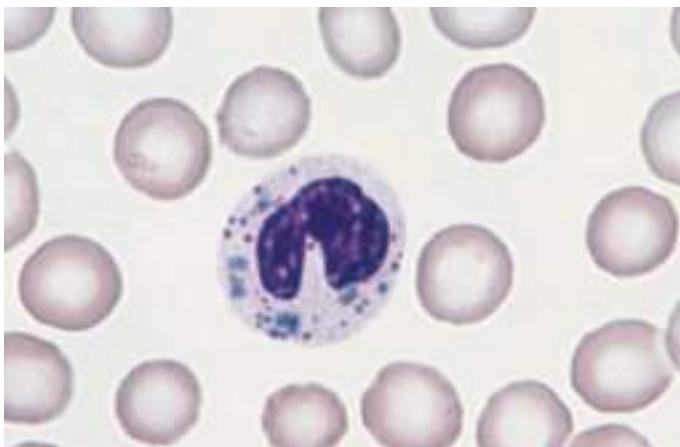


FIGURE 58-30 Chédiak-Higashi disease. Note giant granules in neutrophil.

in making the blood smear, leaving a smudge of nuclear material without a surrounding cytoplasm or cell membrane; they are called *smudge cells* and are rare in the absence of chronic lymphoid leukemia.

Monocytes are the largest white blood cells, ranging from 15 to 22 μm in diameter. The nucleus can take on a variety of shapes but usually appears to be folded; the cytoplasm is gray.

Abnormal cells may appear in the blood. Most often the abnormal cells originate from neoplasms of bone marrow-derived cells, including lymphoid cells, myeloid cells, and occasionally red cells. More rarely, other types of tumors can get access to the bloodstream, and rare epithelial malignant cells may be identified. The chances of seeing such abnormal cells are increased by examining blood smears made from buffy coats, the layer of cells that is visible on top of sedimenting red cells when blood is left in the test tube for an hour. Smears made from finger sticks may include rare endothelial cells.

ACKNOWLEDGMENT

Figures in this chapter were borrowed from *Williams Hematology*, 7th edition, M Lichtman et al (eds). New York, McGraw-Hill, 2005; *Hematology in General Practice*, 4th edition, RS Hillman, KA Ault, New York, McGraw-Hill, 2005.

59

Anemia and Polycythemia

John W. Adamson, Dan L. Longo



HEMATOPOIESIS AND THE PHYSIOLOGIC BASIS OF RED CELL PRODUCTION

Hematopoiesis is the process by which the formed elements of blood are produced. The process is regulated through a series of steps beginning with the hematopoietic stem cell. Stem cells are capable of producing red cells, all classes of granulocytes, monocytes, platelets, and the cells of the immune system. The precise molecular mechanism by which the stem cell becomes committed to a given lineage is not fully defined. However, experiments in mice suggest that erythroid cells come from a common erythroid/megakaryocyte progenitor that does not develop in the absence of expression of the GATA-1 and FOG-1 (friend of GATA-1) transcription factors (Chap. 92). Following lineage commitment, hematopoietic progenitor and precursor cells come increasingly under the regulatory influence of growth factors and hormones. For red cell production, erythropoietin (EPO) is the primary regulatory hormone. EPO is required for the maintenance of committed erythroid progenitor cells that, in the absence of the hormone, undergo programmed cell death (*apoptosis*). The regulated process of red cell production is *erythropoiesis*, and its key elements are illustrated in Fig. 59-1.

In the bone marrow, the first morphologically recognizable erythroid precursor is the pronormoblast. This cell can undergo four to five cell divisions, which result in the production of 16–32 mature red cells. With increased EPO production, or the administration of EPO as a drug, early progenitor cell numbers are amplified and, in turn, give rise to increased numbers of erythrocytes. The regulation of EPO production itself is linked to tissue oxygenation.

In mammals, O_2 is transported to tissues bound to the hemoglobin contained within circulating red cells. The mature red cell is 8 μm in diameter, anucleate, discoid in shape, and extremely pliable in order to traverse the microcirculation successfully; its membrane integrity is maintained by the intracellular generation of ATP. Normal red cell production results in the daily replacement of 0.8–1% of all circulating red cells in the body, since the average red cell lives 100–120 days. The organ responsible for red cell production is called the *erythron*. The erythron is a dynamic organ made up of a rapidly proliferating pool of marrow erythroid precursor cells and a large mass of mature circulating red blood cells. The size of the red cell mass reflects the balance of red cell production and destruction. The physiologic basis of red cell production and destruction provides an understanding of the mechanisms that can lead to anemia.

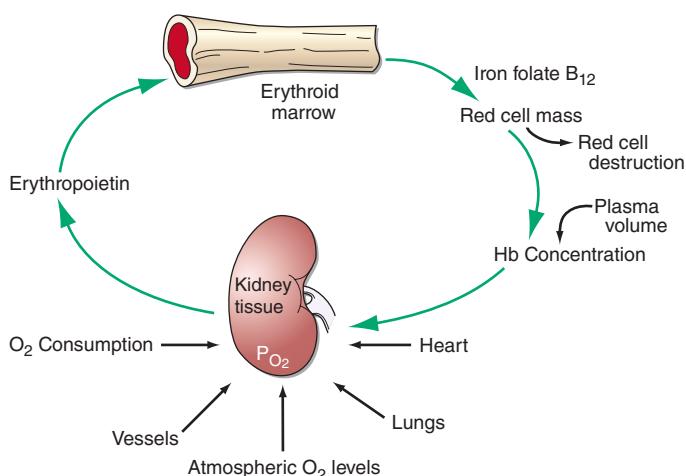


FIGURE 59-1 The physiologic regulation of red cell production by tissue oxygen tension. Hb, hemoglobin.

The physiologic regulator of red cell production, the glycoprotein hormone EPO, is produced and released by peritubular capillary lining cells within the kidney. These cells are highly specialized epithelial-like cells. A small amount of EPO is produced by hepatocytes. The fundamental stimulus for EPO production is the availability of O_2 for tissue metabolic needs. Key to EPO gene regulation is hypoxia-inducible factor (HIF)-1 α . In the presence of O_2 , HIF-1 α is hydroxylated at a key proline, allowing HIF-1 α to be ubiquitinylated and degraded via the proteasome pathway. If O_2 becomes limiting, this critical hydroxylation step does not occur, allowing HIF-1 α to partner with other proteins, translocate to the nucleus, and upregulate the expression of the EPO gene, among others.

Impaired O_2 delivery to the kidney can result from a decreased red cell mass (*anemia*), impaired O_2 loading of the hemoglobin molecule or a high O_2 affinity mutant hemoglobin (*hypoxemia*), or, rarely, impaired blood flow to the kidney (renal artery stenosis). EPO governs the day-to-day production of red cells, and ambient levels of the hormone can be measured in the plasma by sensitive immunoassays—the normal level being 10–25 U/L. When the hemoglobin concentration falls below 100–120 g/L (10–12 g/dL), plasma EPO levels increase in proportion to the severity of the anemia (Fig. 59-2). In circulation, EPO has a half-clearance time of 6–9 h. EPO acts by binding to specific receptors on the surface of marrow erythroid precursors, inducing them to proliferate and to mature. With EPO stimulation, red cell production can increase four- to fivefold within a 1- to 2-week period, but only in the presence of adequate nutrients, especially iron. The functional capacity of the erythron, therefore, requires normal renal production of EPO, a functioning erythroid marrow, and an adequate supply of substrates for hemoglobin synthesis. A defect in any of these key components can lead to anemia. Generally, anemia is recognized in the laboratory when a patient's hemoglobin level or hematocrit is reduced below an expected value (the normal range). The likelihood and severity of anemia are defined based on the deviation of the patient's hemoglobin/hematocrit from values expected for age- and sex-matched normal subjects. The hemoglobin concentration in adults has a Gaussian distribution. The mean hematocrit value for adult males is 47% (standard deviation, $\pm 7\%$) and that for adult females is 42% ($\pm 5\%$). Any single hematocrit or hemoglobin value carries with it a likelihood of associated anemia. Thus, a hematocrit of <39% in an adult male or <35% in an adult female has only about a 25% chance of being normal. Hematocrit levels are less useful than hemoglobin levels in assessing anemia because they are calculated rather than measured directly. Suspected low hemoglobin or hematocrit values are more easily interpreted if previous values for the same patient are known for comparison. The World Health Organization (WHO) defines anemia as a hemoglobin level <130 g/L (13 g/dL) in men and <120 g/L (12 g/dL) in women.

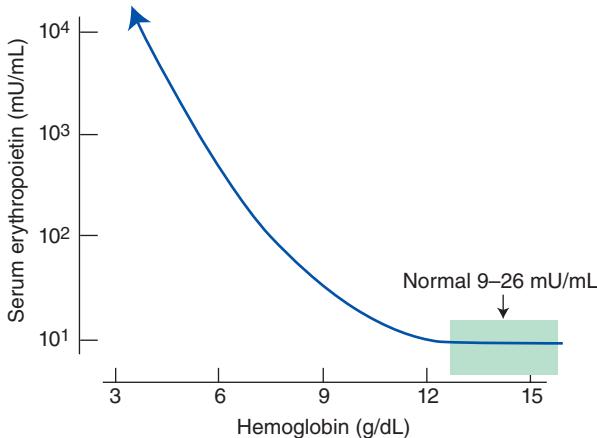


FIGURE 59-2 Erythropoietin (EPO) levels in response to anemia. When the hemoglobin level falls to 120 g/L (12 g/dL), plasma EPO levels increase logarithmically. In the presence of chronic kidney disease or chronic inflammation, EPO levels are typically lower than expected for the degree of anemia. As individuals age, the level of EPO needed to sustain normal hemoglobin levels appears to increase. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

The critical elements of erythropoiesis—EPO production, iron availability, the proliferative capacity of the bone marrow, and effective maturation of red cell precursors—are used for the initial classification of anemia (see below).

ANEMIA

■ CLINICAL PRESENTATION OF ANEMIA

Signs and Symptoms Anemia is most often recognized by abnormal screening laboratory tests. Patients less commonly present with advanced anemia and its attendant signs and symptoms. Acute anemia is due to blood loss or hemolysis. If blood loss is mild, enhanced O_2 delivery is achieved through changes in the O_2 -hemoglobin dissociation curve mediated by a decreased pH or increased CO_2 (*Bohr effect*). With acute blood loss, hypovolemia dominates the clinical picture, and the hematocrit and hemoglobin levels do not reflect the volume of blood lost. Signs of vascular instability appear with acute losses of 10–15% of the total blood volume. In such patients, the issue is not anemia but hypotension and decreased organ perfusion. When >30% of the blood volume is lost suddenly, patients are unable to compensate with the usual mechanisms of vascular contraction and changes in regional blood flow. The patient prefers to remain supine and will show postural hypotension and tachycardia. If the volume of blood lost is >40% (i.e., >2 L in the average-sized adult), signs of hypovolemic shock including confusion, dyspnea, diaphoresis, hypotension, and tachycardia appear (Chap. 97). Such patients have significant deficits in vital organ perfusion and require immediate volume replacement.

With acute hemolysis, the signs and symptoms depend on the mechanism that leads to red cell destruction. Intravascular hemolysis with release of free hemoglobin may be associated with acute back pain, free hemoglobin in the plasma and urine, and renal failure. Symptoms associated with more chronic or progressive anemia depend on the age of the patient and the adequacy of blood supply to critical organs. Symptoms associated with moderate anemia include fatigue, loss of stamina, breathlessness, and tachycardia (particularly with physical exertion). However, because of the intrinsic compensatory mechanisms that govern the O_2 -hemoglobin dissociation curve, the gradual onset of anemia—particularly in young patients—may not be associated with signs or symptoms until the anemia is severe (hemoglobin <70–80 g/L [7–8 g/dL]). When anemia develops over a period of days or weeks, the total blood volume is normal to slightly increased, and changes in cardiac output and regional blood flow help compensate for the overall loss in O_2 -carrying capacity. Changes in the position of the O_2 -hemoglobin dissociation curve account for some of the compensatory response to anemia. With chronic anemia, intracellular levels of 2,3-bisphosphoglycerate rise, shifting the dissociation curve to the right and facilitating O_2 unloading. This compensatory mechanism can only maintain normal tissue O_2 delivery in the face of a 20–30 g/L (2–3 g/dL) deficit in hemoglobin concentration. Finally, further protection of O_2 delivery to vital organs is achieved by the shunting of blood away from organs that are relatively rich in blood supply, particularly the kidney, gut, and skin.

Certain disorders are commonly associated with anemia. Chronic inflammatory states (e.g., infection, rheumatoid arthritis, cancer) are associated with mild to moderate anemia, whereas lymphoproliferative disorders, such as chronic lymphocytic leukemia and certain other B cell neoplasms, may be associated with autoimmune hemolysis.

APPROACH TO THE PATIENT

Anemia

The evaluation of the patient with anemia requires a careful history and physical examination. Nutritional history related to drugs or alcohol intake and family history of anemia should always be assessed. Certain geographic backgrounds and ethnic origins are associated with an increased likelihood of an inherited disorder of the hemoglobin molecule or intermediary metabolism. Glucose-6-phosphate

dehydrogenase (G6PD) deficiency and certain hemoglobinopathies are seen more commonly in those of Middle Eastern or African origin, including African Americans who have a high frequency of G6PD deficiency. Other information that may be useful includes exposure to certain toxic agents or drugs and symptoms related to other disorders commonly associated with anemia. These include symptoms and signs such as bleeding, fatigue, malaise, fever, weight loss, night sweats, and other systemic symptoms. Clues to the mechanisms of anemia may be provided on physical examination by findings of infection, blood in the stool, lymphadenopathy, splenomegaly, or petechiae. Splenomegaly and lymphadenopathy suggest an underlying lymphoproliferative disease, whereas petechiae suggest platelet dysfunction. Past laboratory measurements are helpful to determine a time of onset.

In the anemic patient, physical examination may demonstrate a forceful heartbeat, strong peripheral pulses, and a systolic "flow" murmur. The skin and mucous membranes may be pale if the hemoglobin is <80–100 g/L (8–10 g/dL). This part of the physical examination should focus on areas where vessels are close to the surface such as the mucous membranes, nail beds, and palmar creases. If the palmar creases are lighter in color than the surrounding skin when the hand is hyperextended, the hemoglobin level is usually <80 g/L (8 g/dL).

LABORATORY EVALUATION

Table 59-1 lists the tests used in the initial workup of anemia. A routine complete blood count (CBC) is required as part of the evaluation and includes the hemoglobin, hematocrit, and red cell indices: the mean cell volume (MCV) in femtoliters, mean cell hemoglobin (MCH) in picograms per cell, and mean concentration

TABLE 59-2 Red Blood Cell Indices

INDEX	NORMAL VALUE
Mean cell volume (MCV) = (hematocrit × 10)/ (red cell count × 10 ⁶)	90 ± 8 fL
Mean cell hemoglobin (MCH) = (hemoglobin × 10)/(red cell count × 10 ⁶)	30 ± 3 pg
Mean cell hemoglobin concentration = (hemoglobin × 10)/hematocrit, or MCH/MCV	33 ± 2%

of hemoglobin per volume of red cells (MCHC) in grams per liter (non-SI: grams per deciliter). The MCH is the least useful of the indices; it tends to track with the MCV. The red cell indices are calculated as shown in **Table 59-2**, and the normal variations in the hemoglobin and hematocrit with age are shown in **Table 59-3**. A number of physiologic factors affect the CBC, including age, sex, pregnancy, smoking, and altitude. High-normal hemoglobin values may be seen in men and women who live at altitude or smoke heavily. Hemoglobin elevations due to smoking reflect normal compensation due to the displacement of O₂ by CO in hemoglobin binding. Other important information is provided by the reticulocyte count and measurements of iron supply including serum iron, total iron-binding capacity (TIBC; an indirect measure of serum transferrin), and serum ferritin. Marked alterations in the red cell indices usually reflect disorders of maturation or iron deficiency. A careful evaluation of the peripheral blood smear is important, and clinical laboratories often provide a description of both the red and white cells, a white cell differential count, and the platelet count. In patients with severe anemia and abnormalities in red blood cell morphology and/or low reticulocyte counts, a bone marrow aspirate or biopsy can assist in the diagnosis. Other tests of value in the diagnosis of specific anemias are discussed in chapters on specific disease states.

The components of the CBC also help in the classification of anemia. *Microcytosis* is reflected by a lower than normal MCV (<80), whereas high values (>100) reflect *macrocytosis*. The MCHC reflect defects in hemoglobin synthesis (*hypochromia*). Automated cell counters describe the red cell volume distribution width (RDW). The MCV (representing the peak of the distribution curve) is insensitive to the appearance of small populations of macrocytes or microcytes. An experienced laboratory technician will be able to identify minor populations of large or small cells or hypochromic cells before the red cell indices change.

Peripheral Blood Smear The peripheral blood smear provides important information about defects in red cell production (**Chap. 58**). As a complement to the red cell indices, the blood smear also reveals variations in cell size (*anisocytosis*) and shape (*poikilocytosis*). The degree of anisocytosis usually correlates with increases in the RDW or the range of cell sizes. Poikilocytosis suggests a defect in the maturation of red cell precursors in the bone marrow or fragmentation of circulating red cells. The blood smear may also reveal *polychromasia*—red cells that are slightly larger than normal and grayish blue in color on the Wright-Giemsa stain. These cells

TABLE 59-1 Laboratory Tests in Anemia Diagnosis

- I. Complete blood count (CBC)
 - A. Red blood cell count
 - 1. Hemoglobin
 - 2. Hematocrit
 - 3. Reticulocyte count
 - B. Red blood cell indices
 - 1. Mean cell volume (MCV)
 - 2. Mean cell hemoglobin (MCH)
 - 3. Mean cell hemoglobin concentration (MCHC)
 - 4. Red cell distribution width (RDW)
 - C. White blood cell count
 - 1. Cell differential
 - 2. Nuclear segmentation of neutrophils
 - D. Platelet count
 - E. Cell morphology
 - 1. Cell size
 - 2. Hemoglobin content
 - 3. Anisocytosis
 - 4. Poikilocytosis
 - 5. Polychromasia
- II. Iron supply studies
 - A. Serum iron
 - B. Total iron-binding capacity
 - C. Serum ferritin
- III. Marrow examination
 - A. Aspirate
 - 1. M/E ratio^a
 - 2. Cell morphology
 - 3. Iron stain
 - B. Biopsy
 - 1. Cellularity
 - 2. Morphology

^aM/E ratio, ratio of myeloid to erythroid precursors.

TABLE 59-3 Changes in Normal Hemoglobin/Hematocrit Values with Age, Sex, and Pregnancy

AGE/SEX	HEMOGLOBIN, g/dL	HEMATOCRIT, %
At birth	17	52
Childhood	12	36
Adolescence	13	40
Adult man	16 (±2)	47 (±6)
Adult woman (menstruating)	13 (±2)	40 (±6)
Adult woman (postmenopausal)	14 (±2)	42 (±6)
During pregnancy	12 (±2)	37 (±6)

Source: From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.

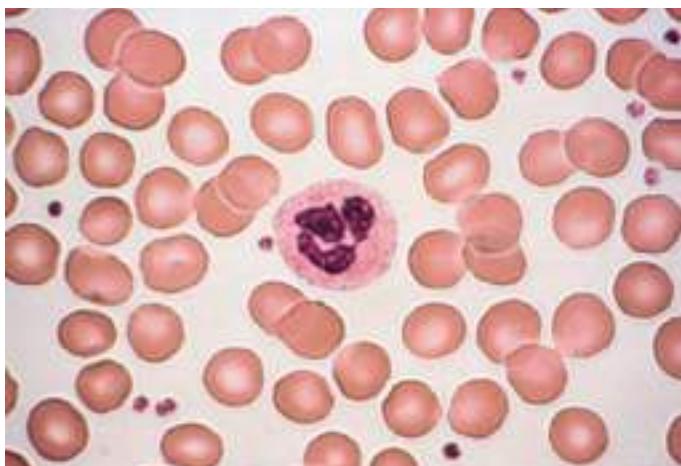


FIGURE 59-3 Normal blood smear (Wright stain). High-power field showing normal red cells, a neutrophil, and a few platelets. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

are reticulocytes that have been prematurely released from the bone marrow, and their color represents residual amounts of ribosomal RNA. These cells appear in circulation in response to EPO stimulation or to architectural damage of the bone marrow (fibrosis, infiltration of the marrow by malignant cells, etc.) that results in their disordered release from the marrow. The appearance of nucleated red cells, Howell-Jolly bodies, target cells, sickle cells, and others may provide clues to specific disorders (Figs. 59-3 to 59-11).

Reticulocyte Count An accurate reticulocyte count is key to the initial classification of anemia. Reticulocytes are red cells that have been recently released from the bone marrow. They are identified by staining with a supravital dye that precipitates the ribosomal RNA (Fig. 59-12). These precipitates appear as blue or black punctate spots and can be counted manually or, currently, by fluorescent emission of dyes that bind to RNA. This residual RNA is metabolized over the first 24–36 h of the reticulocyte's life span in circulation. Normally, the reticulocyte count ranges from 1 to 2% and reflects the daily replacement of 0.8–1.0% of the circulating red cell population. A corrected reticulocyte percentage or the absolute number of reticulocytes provides a reliable measure of effective red cell production.

In the initial classification of anemia, the patient's reticulocyte count is compared with the expected reticulocyte response. In general, if the EPO and erythroid marrow responses to moderate anemia [hemoglobin <100 g/L (10 g/dL)] are intact, the red cell production

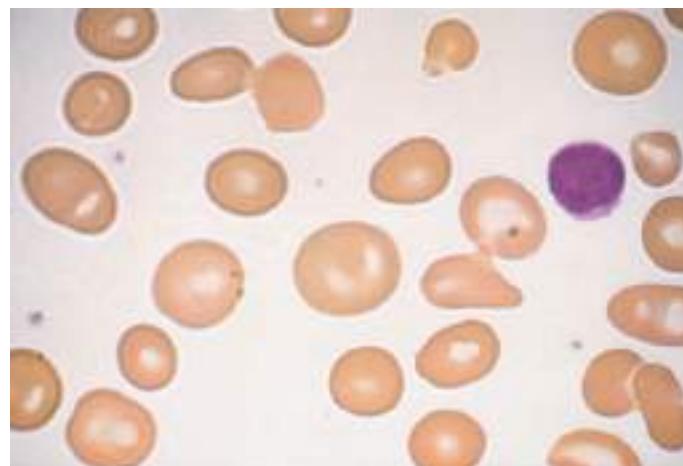


FIGURE 59-5 Macrocytosis. Red cells are larger than a small lymphocyte and well hemoglobinized. Often macrocytes are oval shaped (macro-ovalocytes).



FIGURE 59-6 Howell-Jolly bodies. In the absence of a functional spleen, nuclear remnants are not culled from the red cells and remain as small homogeneously staining blue inclusions on Wright stain. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

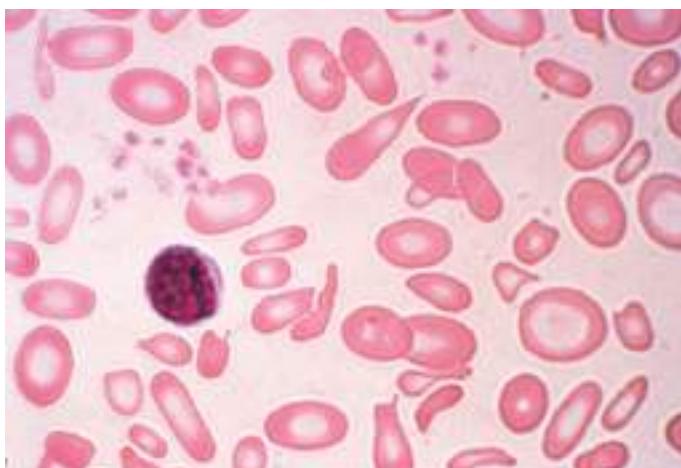


FIGURE 59-4 Severe iron-deficiency anemia. Microcytic and hypochromic red cells smaller than the nucleus of a lymphocyte associated with marked variation in size (anisocytosis) and shape (poikilocytosis). (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

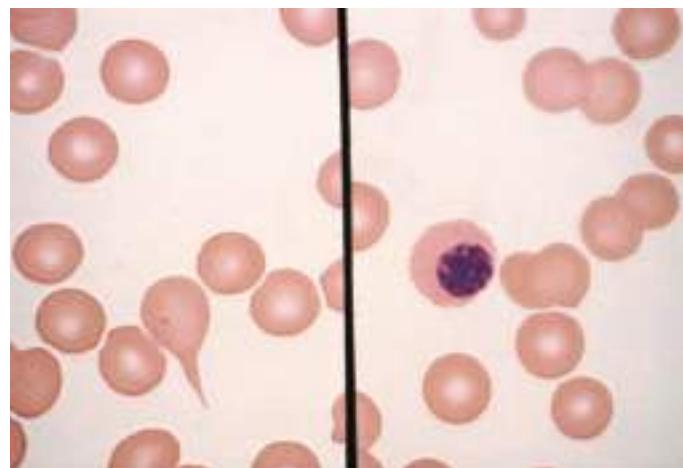


FIGURE 59-7 Red cell changes in myelofibrosis. The left panel shows a teardrop-shaped cell. The right panel shows a nucleated red cell. These forms can be seen in myelofibrosis. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

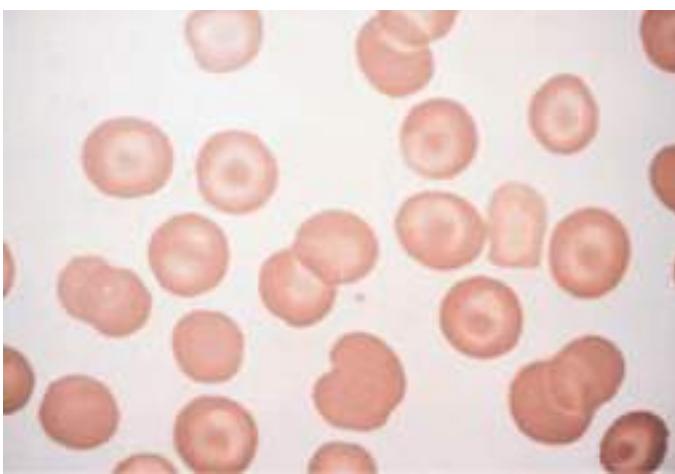


FIGURE 59-8 Target cells. Target cells have a bull's-eye appearance and are seen in thalassemia and in liver disease. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

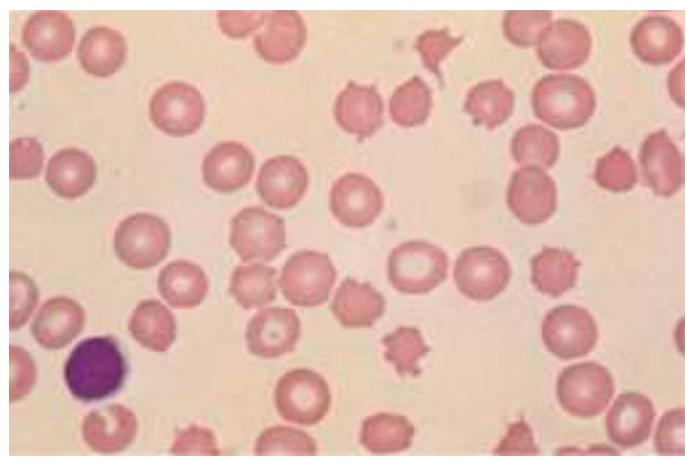


FIGURE 59-11 Spur cells. Spur cells are recognized as distorted red cells containing several irregularly distributed thorn-like projections. Cells with this morphologic abnormality are also called acanthocytes. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

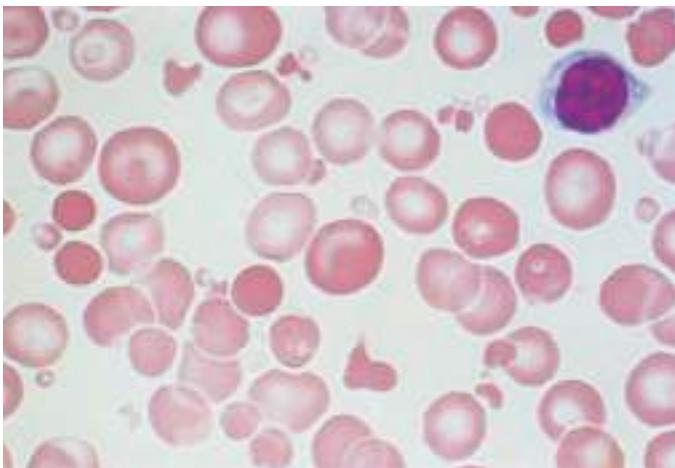


FIGURE 59-9 Red cell fragmentation. Red cells may become fragmented in the presence of foreign bodies in the circulation, such as mechanical heart valves, or in the setting of thermal injury. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

rate increases to two to three times normal within 10 days following the onset of anemia. In the face of established anemia, a reticulocyte response less than two to three times normal indicates an inadequate marrow response.

To use the reticulocyte count to estimate marrow response, two corrections are necessary. The first correction adjusts the reticulocyte count based on the reduced number of circulating red cells. With anemia, the percentage of reticulocytes may be increased while the absolute number is unchanged. To correct for this effect, the reticulocyte percentage is multiplied by the ratio of the patient's hemoglobin or hematocrit to the expected hemoglobin/hematocrit for the age and sex of the patient (Table 59-4). This provides an estimate of the reticulocyte count corrected for anemia. To convert the corrected reticulocyte count to an index of marrow production, a further correction is required, depending on whether some of the reticulocytes in circulation have been released from the marrow prematurely. For this second correction, the peripheral blood smear is examined to see if there are polychromatophilic macrocytes present.

These cells, representing prematurely released reticulocytes, are referred to as "shift" cells, and the relationship between the degree of shift and the necessary shift correction factor is shown in Fig. 59-13. The correction is necessary because these prematurely released cells survive as reticulocytes in circulation for >1 day, thereby providing a falsely high estimate of daily red cell production. If polychromasia is increased, the reticulocyte count, already corrected for anemia,

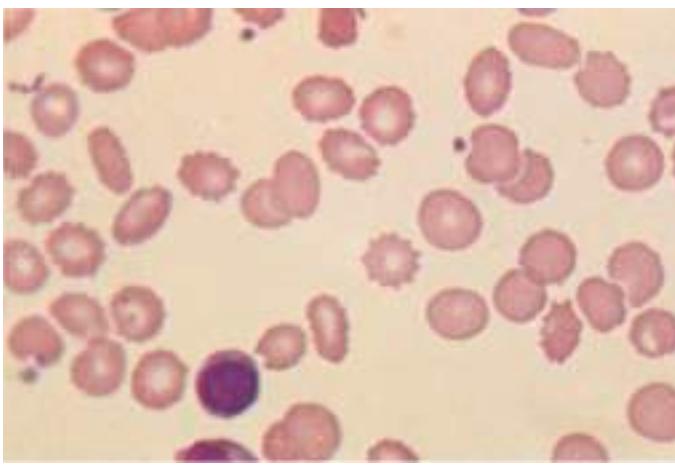


FIGURE 59-10 Uremia. The red cells in uremia may acquire numerous regularly spaced, small, spiny projections. Such cells, called burr cells or echinocytes, are readily distinguishable from irregularly spiculated acanthocytes shown in Fig. 59-11.

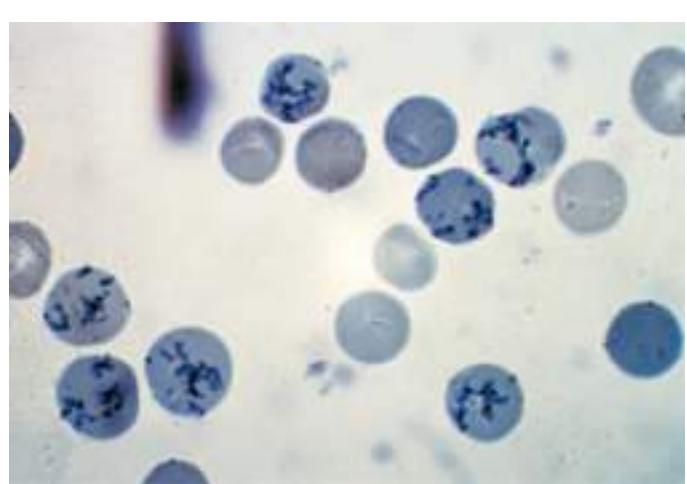


FIGURE 59-12 Reticulocytes. Methylene blue stain demonstrates residual RNA in newly made red cells. (From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.)

TABLE 59-4 Calculation of Reticulocyte Production Index**Correction #1 for Anemia:**

This correction produces the corrected reticulocyte count.

In a person whose reticulocyte count is 9%, hemoglobin 7.5 g/dL, and hematocrit 23%, the absolute reticulocyte count = $9 \times (7.5/15)$ [or $\times (23/45)$] = 4.5%

Note. This correction is not done if the reticulocyte count is reported in absolute numbers (e.g., 50,000/ μ L of blood)

Correction #2 for Longer Life of Prematurely Released Reticulocytes in the Blood:

This correction produces the reticulocyte production index.

In a person whose reticulocyte count is 9%, hemoglobin 7.5 gm/dL, and hematocrit 23%, the reticulocyte production index

$$= 9 \times \frac{(7.5/15)(\text{hemoglobin correction})}{2(\text{maturation time correction})} = 2.25$$

should be corrected again by 2 to account for the prolonged reticulocyte maturation time. The second correction factor varies from 1 to 3 depending on the severity of anemia. In general, a correction of 2 is simply used. An appropriate correction is shown in Table 59-4. If polychromatophilic cells are not seen on the blood smear, the second correction is not indicated. The now doubly corrected reticulocyte count is the *reticulocyte production index*, and it provides an estimate of marrow production relative to normal. In many hospital laboratories, the reticulocyte count is reported not only as a percentage but also in absolute numbers. If so, no correction for dilution is required. A summary of the appropriate marrow response to varying degrees of anemia is shown in Table 59-5.

Premature release of reticulocytes is normally due to increased EPO stimulation. However, if the integrity of the bone marrow release process is lost through tumor infiltration, fibrosis, or other disorders, the appearance of nucleated red cells or polychromatophilic macrocytes should still invoke the second reticulocyte correction. The shift correction should always be applied to a patient with anemia and a very high reticulocyte count to provide a true index of effective red cell production. Patients with severe chronic hemolytic anemia may increase red cell production as much as six- to sevenfold. This measure alone confirms the fact that the patient has an appropriate EPO response, a normally functioning bone marrow, and sufficient iron available to meet the demands for new red

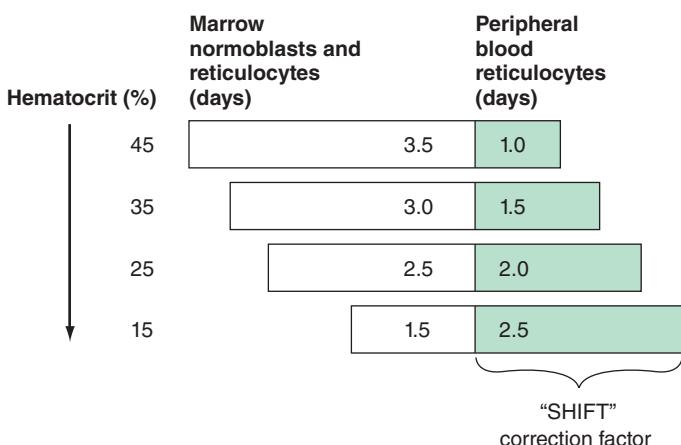


FIGURE 59-13 Correction of the reticulocyte count. To use the reticulocyte count as an indicator of effective red cell production, the reticulocyte number must be corrected based on the level of anemia and the circulating life span of the reticulocytes. Erythroid cells take ~4.5 days to mature. At a normal hemoglobin, reticulocytes are released to the circulation with ~1 day left as reticulocytes. However, with different levels of anemia, reticulocytes (and even earlier erythroid cells) may be released from the marrow prematurely. Most patients come to clinical attention with hematocrits in the mid-20s, and thus a correction factor of 2 is commonly used because the observed reticulocytes will live for 2 days in the circulation before losing their RNA.

TABLE 59-5 Normal Marrow Response to Anemia

HEMOGLOBIN	PRODUCTION INDEX	RETICULOCYTE COUNT
15 g/dL	1	50,000/ μ L
11 g/dL	2.0–2.5	100–150,000/ μ L
8 g/dL	3.0–4.0	300–400,000/ μ L

cell formation. If the reticulocyte production index is <2 in the face of established anemia, a defect in erythroid marrow proliferation or maturation must be present.

Tests of Iron Supply and Storage The laboratory measurements that reflect the availability of iron for hemoglobin synthesis include the serum iron, the TIBC, and the percent transferrin saturation. The percent transferrin saturation is derived by dividing the serum iron level ($\times 100$) by the TIBC. The normal serum iron ranges from 9 to 27 μ mol/L (50–150 μ g/dL), whereas the normal TIBC is 54–64 μ mol/L (300–360 μ g/dL); the normal transferrin saturation ranges from 25 to 50%. A diurnal variation in the serum iron leads to a variation in the percent transferrin saturation. The serum ferritin is used to evaluate total body iron stores. Adult males have serum ferritin levels that average ~100 μ g/L, corresponding to iron stores of ~1 g. Adult females have lower serum ferritin levels averaging 30 μ g/L, reflecting lower iron stores (~300 mg). A serum ferritin level of 10–15 μ g/L indicates depletion of body iron stores. However, ferritin is also an acute-phase reactant and, in the presence of acute or chronic inflammation, may rise several-fold above baseline levels. As a rule, a serum ferritin >200 μ g/L means there is at least some iron in tissue stores.

Bone Marrow Examination A bone marrow aspirate and smear or a needle biopsy can be useful in the evaluation of some patients with anemia. In patients with hypoproliferative anemia and normal iron status, a bone marrow is indicated. Marrow examination can diagnose primary marrow disorders such as myelofibrosis, a red cell maturation defect, or an infiltrative disease (Figs. 59-14 to 59-16). The increase or decrease of one cell lineage (myeloid vs erythroid) compared to another is obtained by a differential count of nucleated cells in a bone marrow smear (the myeloid/erythroid [M/E] ratio). A patient with a hypoproliferative anemia (see below) and a reticulocyte production index <2 will demonstrate an M/E ratio of 2 or 3:1. In contrast, patients with hemolytic disease and a production index >3 will have an M/E ratio of at least 1:1. Maturation disorders are identified from the discrepancy between the M/E ratio and the reticulocyte production index (see below). Either the marrow smear or biopsy can be stained for the presence of iron stores or iron in developing red cells. The storage iron is in the form of ferritin or

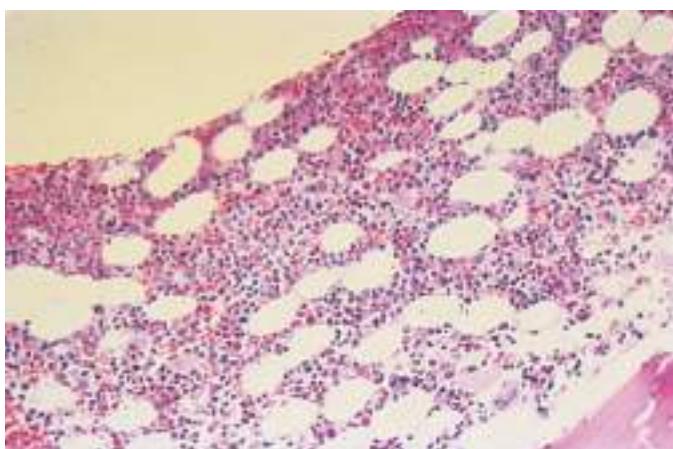


FIGURE 59-14 Normal bone marrow. This is a low-power view of a section of a normal bone marrow biopsy stained with hematoxylin and eosin (H&E). Note that the nucleated cellular elements account for ~40–50% and the fat (clear areas) accounts for ~50–60% of the area. (From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)

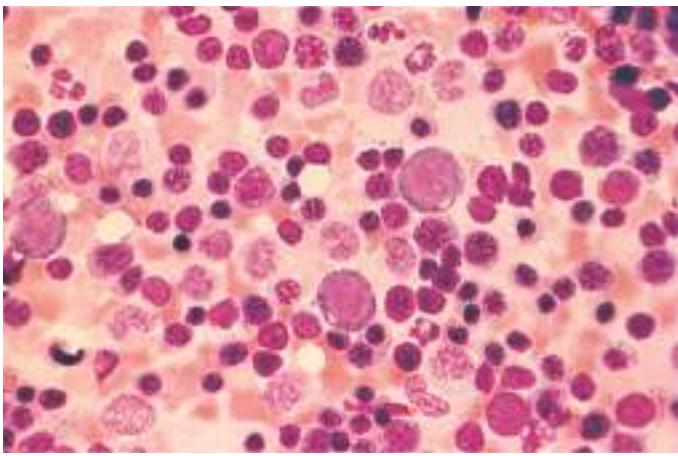


FIGURE 59-15 Erythroid hyperplasia. This marrow shows an increase in the fraction of cells in the erythroid lineage as might be seen when a normal marrow compensates for acute blood loss or hemolysis. The myeloid/erythroid (M/E) ratio is about 1:1. (From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)

hemosiderin. On carefully prepared bone marrow smears, small ferritin granules can normally be seen under oil immersion in 20–40% of developing erythroblasts. Such cells are called *sideroblasts*.

OTHER LABORATORY MEASUREMENTS

Additional laboratory tests may be of value in confirming specific diagnoses. **For details of these tests and how they are applied in individual disorders, see Chaps. 93 to 97.**

■ DEFINITION AND CLASSIFICATION OF ANEMIA

Initial Classification of Anemia The functional classification of anemia has three major categories. These are (1) marrow production defects (*hypoproliferation*), (2) red cell maturation defects (*ineffective erythropoiesis*), and (3) decreased red cell survival (*blood loss/hemolysis*). The classification is shown in Fig. 59-17. A hypoproliferative anemia is typically seen with a low reticulocyte production index together with little or no change in red cell morphology (a normocytic, normochromic anemia) (Chap. 93). Maturation disorders typically have a slight to moderately elevated reticulocyte production index that is accompanied by either macrocytic (Chap. 95) or microcytic (Chaps. 93, 94) red cell indices. Increased red blood cell destruction secondary to hemolysis results in an increase in the reticulocyte production index to at least

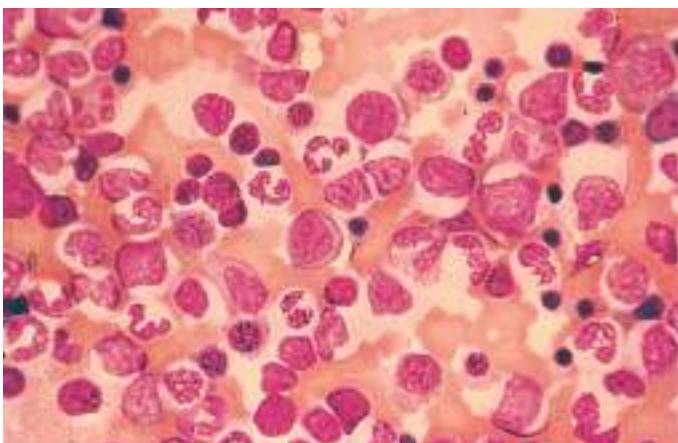


FIGURE 59-16 Myeloid hyperplasia. This marrow shows an increase in the fraction of cells in the myeloid or granulocytic lineage as might be seen in a normal marrow responding to infection. The myeloid/erythroid (M/E) ratio is >3:1. (From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)

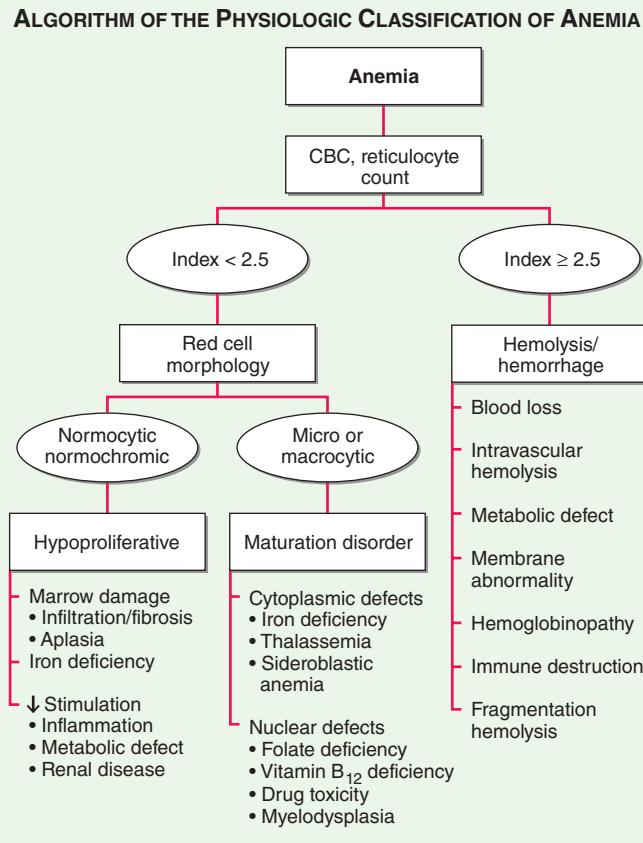


FIGURE 59-17 The physiologic classification of anemia. CBC, complete blood count.

three times normal (Chap. 96), provided sufficient iron is available. Hemolytic anemia does not typically result in production indices of more than 2.0–2.5 times normal because of the limitations placed on expansion of the erythroid marrow by iron availability (Chap. 97).

In the first branch point of the classification of anemia, a reticulocyte production index >2.5 indicates that hemolysis is most likely. A reticulocyte production index <2 indicates either a hypoproliferative anemia or maturation disorder. The latter two possibilities can often be distinguished by the red cell indices, by examination of the peripheral blood smear, or by a marrow examination. If the red cell indices are normal, the anemia is almost certainly hypoproliferative in nature. Maturation disorders are characterized by ineffective red cell production and a low reticulocyte production index. Bizarre red cell shapes—macrocytes or hypochromic microcytes—are seen on the peripheral blood smear. With a hypoproliferative anemia, no erythroid hyperplasia is noted in the marrow, whereas patients with ineffective red cell production have erythroid hyperplasia and an M/E ratio $<1:1$.

Hypoproliferative Anemias At least 75% of all cases of anemia are hypoproliferative in nature. A hypoproliferative anemia reflects absolute or relative marrow failure in which the erythroid marrow has not proliferated appropriately for the degree of anemia. The majority of hypoproliferative anemias are due to mild to moderate iron deficiency or inflammation. A hypoproliferative anemia can result from marrow damage, iron deficiency, or inadequate EPO stimulation. The last may reflect impaired renal function, suppression of EPO production by inflammatory cytokines such as interleukin 1, or reduced tissue needs for O₂ from metabolic disease such as hypothyroidism. Only occasionally is the marrow unable to produce red cells at a normal rate, and this is most prevalent in patients with renal failure. With diabetes mellitus or myeloma, the EPO deficiency may be more marked than would be predicted by the degree of renal insufficiency. In general, hypoproliferative anemias are characterized by normocytic, normochromic red cells, although microcytic, hypochromic cells may be observed with mild iron deficiency or long-standing chronic inflammatory disease.

The key laboratory tests in distinguishing between the various forms of hypoproliferative anemia include the serum iron and iron-binding capacity, evaluation of renal and thyroid function, a marrow biopsy or aspirate to detect marrow damage or infiltrative disease, and serum ferritin to assess iron stores. An iron stain of the marrow will determine the pattern of iron distribution. Patients with the anemia of acute or chronic inflammation show a distinctive pattern of serum iron (low), TIBC (normal or low), percent transferrin saturation (low), and serum ferritin (normal or high). These changes in iron values are brought about by hepcidin, the iron regulatory hormone that is produced by the liver and is increased in inflammation (Chap. 93). A distinct pattern of results is noted in mild to moderate iron deficiency (low serum iron, high TIBC, low percent transferrin saturation, low serum ferritin) (Chap. 93). Marrow damage by drugs, infiltrative disease such as leukemia or lymphoma, or marrow aplasia is diagnosed from the peripheral blood and bone marrow morphology. With infiltrative disease or fibrosis, a marrow biopsy is required.

Maturation Disorders The presence of anemia with an inappropriately low reticulocyte production index, macro- or microcytosis on smear, and abnormal red cell indices suggests a maturation disorder. Maturation disorders are divided into two categories: nuclear maturation defects, associated with macrocytosis, and cytoplasmic maturation defects, associated with microcytosis and hypochromia usually from defects in hemoglobin synthesis. The inappropriately low reticulocyte production index is a reflection of the ineffective erythropoiesis that results from the destruction within the marrow of developing erythroblasts. Bone marrow examination shows erythroid hyperplasia.

Nuclear maturation defects result from vitamin B₁₂ or folic acid deficiency, drug damage, or myelodysplasia. Drugs that interfere with cellular DNA synthesis, such as methotrexate or alkylating agents, can produce a nuclear maturation defect. Alcohol, alone, is also capable of producing macrocytosis and a variable degree of anemia, but this is usually associated with folic acid deficiency. Measurements of folic acid and vitamin B₁₂ are critical not only in identifying the specific vitamin deficiency but also because they reflect different pathogenetic mechanisms (Chap. 95).

Cytoplasmic maturation defects result from severe iron deficiency or abnormalities in globin or heme synthesis. Iron deficiency occupies an unusual position in the classification of anemia. If the iron-deficiency anemia is mild to moderate, erythroid marrow proliferation is blunted and the anemia is classified as hypoproliferative. However, if the anemia is severe and prolonged, the erythroid marrow will become hyperplastic despite the inadequate iron supply, and the anemia will be classified as ineffective erythropoiesis with a cytoplasmic maturation defect. In either case, an inappropriately low reticulocyte production index, microcytosis, and a classic pattern of iron values make the diagnosis clear and easily distinguish iron deficiency from other cytoplasmic maturation defects such as the thalassemias. Defects in heme synthesis, in contrast to globin synthesis, are less common and may be acquired or inherited (Chap. 409). Acquired abnormalities are usually associated with myelodysplasia, may lead to either a macro- or microcytic anemia, and are frequently associated with mitochondrial iron loading. In these cases, iron is taken up by the mitochondria of the developing erythroid cell but not incorporated into heme. The iron-encrusted mitochondria surround the nucleus of the erythroid cell, forming a ring. Based on the distinctive finding of so-called ringed sideroblasts on the marrow iron stain, patients are diagnosed as having a sideroblastic anemia—almost always reflecting myelodysplasia. Again, studies of iron parameters are helpful in the differential diagnosis of these patients.

Blood Loss/Hemolytic Anemia In contrast to anemias associated with an inappropriately low reticulocyte production index, hemolysis is associated with red cell production indices ≥ 2.5 times normal. The stimulated erythropoiesis is reflected in the blood smear by the appearance of increased numbers of polychromatophilic macrocytes. A marrow examination is rarely indicated if the reticulocyte production index is increased appropriately. The red cell indices are typically normocytic or slightly macrocytic, reflecting the increased number

of reticulocytes. Acute blood loss is not associated with an increased reticulocyte production index because of the time required to increase EPO production and, subsequently, marrow proliferation (Chap. 97). Subacute blood loss may be associated with modest reticulocytosis. Anemia from chronic blood loss presents more often as iron deficiency than with the picture of increased red cell production.

The evaluation of blood loss anemia is usually not difficult. Most problems arise when a patient presents with an increased red cell production index from an episode of acute blood loss that went unrecognized. The cause of the anemia and increased red cell production may not be obvious. The confirmation of a recovering state may require observations over a period of 2–3 weeks, during which the hemoglobin concentration will rise and the reticulocyte production index fall (Chap. 97).

Hemolytic disease, while dramatic, is among the least common forms of anemia. The ability to sustain a high reticulocyte production index reflects the ability of the erythroid marrow to compensate for hemolysis and, in the case of extravascular hemolysis, the efficient recycling of iron from the destroyed red cells to support red cell production. With intravascular hemolysis, such as paroxysmal nocturnal hemoglobinuria, the loss of iron may limit the marrow response. The level of response depends on the severity of the anemia and the nature of the underlying disease process.

Hemoglobinopathies, such as sickle cell disease and the thalassemias, present a mixed picture. The reticulocyte index may be high but is inappropriately low for the degree of marrow erythroid hyperplasia (Chap. 94).

Hemolytic anemias present in different ways. Some appear suddenly as an acute, self-limited episode of intravascular or extravascular hemolysis, a presentation pattern often seen in patients with autoimmune hemolysis or with inherited defects of the Embden-Meyerhof pathway or the glutathione reductase pathway. Patients with inherited disorders of the hemoglobin molecule or red cell membrane generally have a lifelong clinical history typical of the disease process. Those with chronic hemolytic disease, such as hereditary spherocytosis, may actually present not with anemia but with a complication stemming from the prolonged increase in red cell destruction such as symptomatic bilirubin gallstones or splenomegaly. Patients with chronic hemolysis are also susceptible to aplastic crises if an infectious process interrupts red cell production.

The differential diagnosis of an acute or chronic hemolytic event requires the careful integration of family history, the pattern of clinical presentation, and—whether the disease is congenital or acquired—careful examination of the peripheral blood smear. Precise diagnosis may require more specialized laboratory tests, such as hemoglobin electrophoresis or a screen for red cell enzymes. Acquired defects in red cell survival are often immunologically mediated and require a direct or indirect antiglobulin test or a cold agglutinin titer to detect the presence of hemolytic antibodies or complement-mediated red cell destruction (Chap. 96).

TREATMENT

Anemia

An overriding principle is to initiate treatment of mild to moderate anemia only when a specific diagnosis is made. Rarely, in the acute setting, anemia may be so severe that red cell transfusions are required before a specific diagnosis is available. Whether the anemia is of acute or gradual onset, the selection of the appropriate treatment is determined by the documented cause(s) of the anemia. Often, the cause of the anemia is multifactorial. For example, a patient with severe rheumatoid arthritis who has been taking anti-inflammatory drugs may have a hypoproliferative anemia associated with chronic inflammation as well as chronic blood loss associated with intermittent gastrointestinal bleeding. In every circumstance, it is important to evaluate the patient's iron status fully before and during the treatment of any anemia. **Transfusion is discussed in Chap. 109; iron therapy is discussed in Chap. 93;**

treatment of megaloblastic anemia is discussed in Chap. 95; treatment of other entities is discussed in their respective chapters (sickle cell anemia, Chap. 94; hemolytic anemias, Chap. 96; aplastic anemia and myelodysplasia, Chap. 98).

Therapeutic options for the treatment of anemias have expanded dramatically during the past 30 years. Blood component therapy is available and safe. Recombinant EPO as an adjunct to anemia management has transformed the lives of patients with chronic renal failure on dialysis and reduced transfusion needs of anemic cancer patients receiving chemotherapy. Eventually, patients with inherited disorders of globin synthesis or mutations in the globin gene, such as sickle cell disease, may benefit from the successful introduction of targeted genetic therapy (Chap. 458).

POLYCYTHEMIA

Polycythemia is defined as an increase in the hemoglobin above normal. This increase may be real or only apparent because of a decrease in plasma volume (spurious or relative polycythemia). The term *erythrocytosis* may be used interchangeably with polycythemia, but some draw a distinction between them: erythrocytosis implies documentation of increased red cell mass, whereas polycythemia refers to any increase in red cells. Often patients with polycythemia are detected through an incidental finding of elevated hemoglobin or hematocrit levels. Concern that the hemoglobin level may be abnormally high is usually triggered at 170 g/L (17 g/dL) for men and 150 g/L (15 g/dL) for women. Hematocrit levels >50% in men or >45% in women may be abnormal. Hematocrits >60% in men and >55% in women are almost invariably associated with an increased red cell mass. Given that the machine that quantitates red cell parameters actually measures hemoglobin concentrations and calculates hematocrits, hemoglobin levels may be a better index.

Features of the clinical history that are useful in the differential diagnosis include smoking history; current living at high altitude; or a history of diuretic use, congenital heart disease, sleep apnea, or chronic lung disease.

Patients with polycythemia may be asymptomatic or experience symptoms related to the increased red cell mass or the underlying disease process that leads to the increased red cell mass. The dominant symptoms from an increased red cell mass are related to hyperviscosity and thrombosis (both venous and arterial), because the blood viscosity increases logarithmically at hematocrits >55%. Manifestations include neurologic symptoms such as vertigo, tinnitus, headache, and visual disturbances. Hypertension is often present. Patients with *polycythemia vera* may have aquagenic pruritus, symptoms related to hepatosplenomegaly, easy bruising, epistaxis, or bleeding from the gastrointestinal tract. Peptic ulcer disease is common. Such patients also may present with digital ischemia, Budd-Chiari syndrome, hepatic or splenic/mesenteric vein thrombosis. Patients with hypoxemia may develop cyanosis on minimal exertion or have headache, impaired mental acuity, and fatigue.

The physical examination usually reveals a ruddy complexion. Splenomegaly favors polycythemia vera as the diagnosis (Chap. 99). The presence of cyanosis or evidence of a right-to-left shunt suggests congenital heart disease presenting in the adult, particularly tetralogy of Fallot or Eisenmenger's syndrome (Chap. 264). Increased blood viscosity raises pulmonary artery pressure; hypoxemia can lead to increased pulmonary vascular resistance. Together, these factors can produce cor pulmonale.

Polycythemia can be spurious (related to a decrease in plasma volume; Gaisböck's syndrome), primary, or secondary in origin. The secondary causes are all mediated by EPO: either a physiologically adapted appropriate level based on tissue hypoxia (lung disease, high altitude, CO poisoning, high-affinity hemoglobinopathy) or an abnormal overproduction (renal cysts, renal artery stenosis, tumors with ectopic EPO production). A rare familial form of polycythemia is associated with normal EPO levels but hyperresponsive EPO receptors due to mutations.

APPROACH TO THE PATIENT

Polycythemia

As shown in Fig. 59-18, the first step is to document the presence of an increased red cell mass using the principle of isotope dilution by administering ^{51}Cr -labeled autologous red blood cells to the patient and sampling blood radioactivity over a 2-h period. If the red cell mass is normal (<36 mL/kg in men, <32 mL/kg in women), the patient has spurious or relative polycythemia. If the red cell mass is increased (>36 mL/kg in men, >32 mL/kg in women), serum EPO levels should be measured. If EPO levels are low or unmeasurable, the patient most likely has polycythemia vera. A mutation in JAK2 (Val617Phe), a key member of the cytokine intracellular signaling pathway, can be found in 90–95% of patients with polycythemia vera. Many of those without this particular JAK2 mutation have mutations in exon 12. As a practical matter, few centers assess red cell mass in the setting of an increased hemoglobin level. The alternative workup is to measure EPO levels, check for JAK2 mutation(s), and perform an abdominal ultrasound to assess spleen size. Tests that support the diagnosis of polycythemia vera include elevated white blood cell count, increased absolute basophil count, and thrombocytosis.

If serum EPO levels are elevated, one needs to distinguish whether the elevation is a physiologic response to hypoxia or related to autonomous EPO production. Patients with low arterial O_2 saturation (<92%) should be further evaluated for the presence of heart or lung disease, if they are not living at high altitude. Patients with normal O_2 saturation who are smokers may have elevated EPO levels because of CO displacement of O_2 . If carboxyhemoglobin (COHb) levels are high, the diagnosis is "smoker's polycythemia." Such patients should be urged to stop smoking. Those who cannot stop smoking require phlebotomy to control their polycythemia. Patients with normal O_2 saturation who do not smoke either have an

AN APPROACH TO DIAGNOSING PATIENTS WITH POLYCYTHEMIA

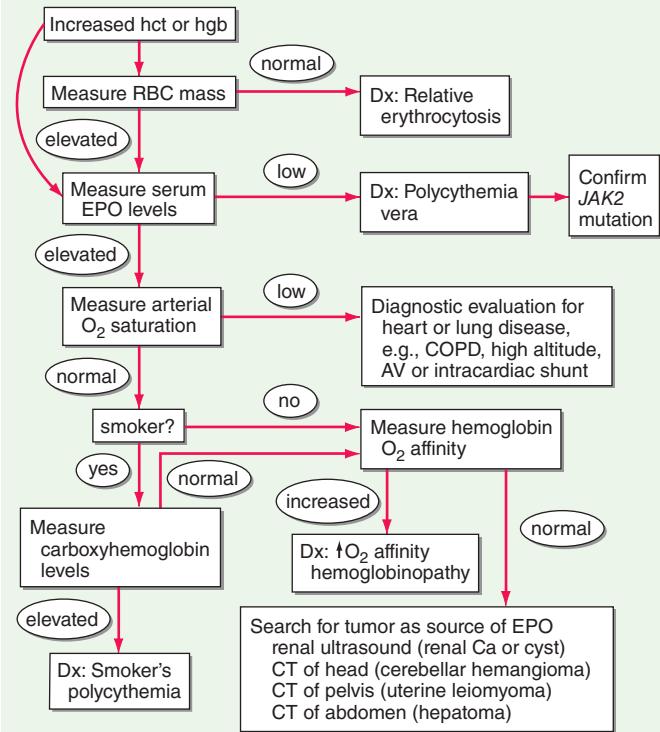


FIGURE 59-18 An approach to the differential diagnosis of patients with an elevated hemoglobin (possible polycythemia). AV, atrioventricular; COPD, chronic obstructive pulmonary disease; CT, computed tomography; EPO, erythropoietin; hct, hematocrit; hgb, hemoglobin; IVP, intravenous pyelogram; RBC, red blood cell.

abnormal hemoglobin that does not deliver O₂ to the tissues (evaluated by finding elevated O₂-hemoglobin affinity) or have a source of EPO production that is not responding to the normal feedback inhibition. Further workup is dictated by the differential diagnosis of EPO-producing neoplasms. Hepatoma, uterine leiomyoma, and renal cancer or cysts are all detectable with abdominopelvic computed tomography scans. Cerebellar hemangiomas may produce EPO, but they present with localizing neurologic signs and symptoms rather than polycythemia-related symptoms.

FURTHER READING

- HILLMAN RS et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.
- MCMULLIN MF et al: Guidelines for the diagnosis, investigation and management of polycythaemia/erythrocytosis. *Br J Haematol* 130:174, 2005.
- SANKARAN VG, WEISS MJ: Anemia: Progress in molecular mechanisms and therapies. *Nat Med* 21:221, 2015.

The blood delivers leukocytes to the various tissues from the bone marrow, where they are produced. Normal blood leukocyte counts are 4.3–10.8 × 10⁹/L, with neutrophils representing 45–74% of the cells, bands 0–4%, lymphocytes 16–45%, monocytes 4–10%, eosinophils 0–7%, and basophils 0–2%. Variation among individuals and among different ethnic groups can be substantial, with lower leukocyte numbers for certain African-American ethnic groups. The various leukocytes are derived from a common stem cell in the bone marrow. Three-fourths of the nucleated cells of bone marrow are committed to the production of leukocytes. Leukocyte maturation in the marrow is under the regulatory control of a number of different factors, known as colony-stimulating factors (CSFs) and interleukins (ILs). Because an alteration in the number and type of leukocytes is often associated with disease processes, total white blood cell (WBC) count (cells per μL) and differential counts are informative. This chapter focuses on neutrophils, monocytes, and eosinophils. **Lymphocytes and basophils are discussed in Chaps. 342 and 346, respectively.**

NEUTROPHILS

MATURATION

Important events in neutrophil life are summarized in Fig. 60-1. In normal humans, neutrophils are produced only in the bone marrow. The minimum number of stem cells necessary to support hematopoiesis is estimated to be 400–500 at any one time. Human blood monocytes, tissue macrophages, and stromal cells produce CSFs, hormones required for the growth of monocytes and neutrophils in the bone marrow. The hematopoietic system not only produces enough neutrophils (~1.3 × 10¹¹ cells per 80-kg person per day) to carry out physiologic functions but also has a large reserve stored in the marrow, which can be mobilized in response to inflammation or infection. An increase in the number of blood neutrophils is called *neutrophilia*, and the presence of immature cells is termed a *shift to the left*. A decrease in the number of blood neutrophils is called *neutropenia*.

Neutrophils and monocytes evolve from pluripotent stem cells under the influence of cytokines and CSFs (Fig. 60-2). The proliferation phase through the metamyelocyte takes about 1 week, while the maturation phase from metamyelocyte to mature neutrophil takes another week. The myeloblast is the first recognizable precursor cell and is followed by the *promyelocyte*. The promyelocyte evolves

60

Disorders of Granulocytes and Monocytes

Steven M. Holland, John I. Gallin

Leukocytes, the major cells comprising inflammatory and immune responses, include neutrophils, T and B lymphocytes, natural killer (NK) cells, monocytes, eosinophils, and basophils. These cells have specific functions, such as antibody production by B lymphocytes or destruction of bacteria by neutrophils, but in no single infectious disease is the exact role of the cell types completely established. Thus, whereas neutrophils are classically thought to be critical to host defense against bacteria, they may also play important roles in defense against viral infections.

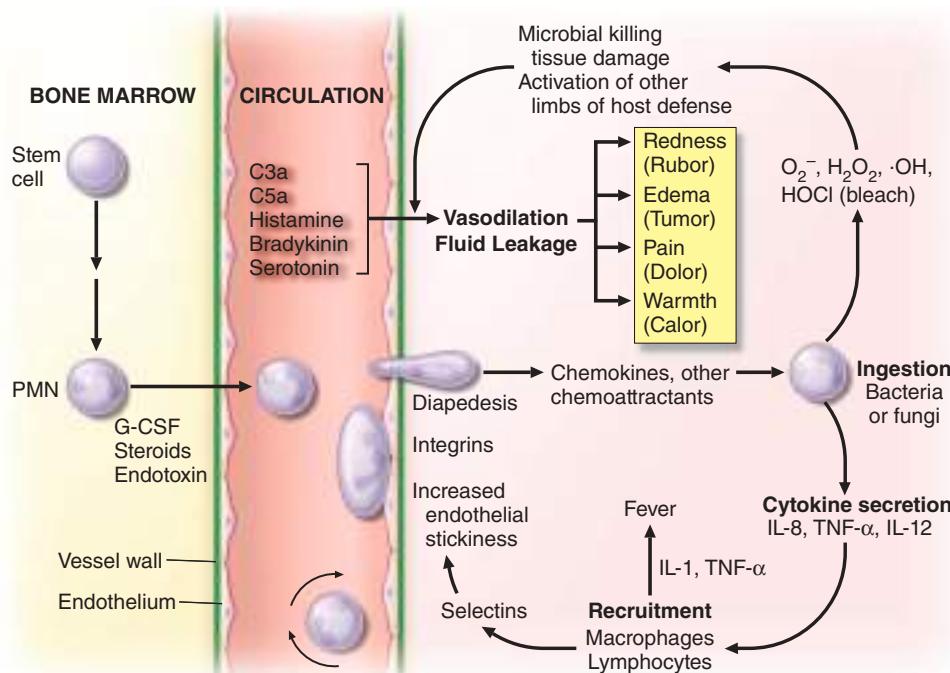


FIGURE 60-1 Schematic events in neutrophil production, recruitment, and inflammation. The four cardinal signs of inflammation (rubor, tumor, calor, dolor) are indicated, as are the interactions of neutrophils with other cells and cytokines. G-CSF, granulocyte colony-stimulating factor; IL, interleukin; PMN, polymorphonuclear leukocyte; TNF- α , tumor necrosis factor α .

Cell	Stage	Surface Markers ^a	Characteristics
	MYELOBLAST	CD33, CD13, CD15	Prominent nucleoli
	PROMYELOCYTE	CD33, CD13, CD15	Large cell Primary granules appear
	MYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Secondary granules appear
	METAMYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Kidney bean-shaped nucleus
	BAND FORM	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, band-shaped nucleus
	NEUTROPHIL	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, multilobed nucleus

^aCD = Cluster Determinant; ● Nucleolus; ● Primary granule; ● Secondary granule.

FIGURE 60-2 Stages of neutrophil development shown schematically. Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) are critical to this process. Identifying cellular characteristics and specific cell-surface markers are listed for each maturational stage.

when the classic lysosomal granules, called the *primary*, or *azurophilic granules* are produced. The primary granules contain hydrolases, elastase, myeloperoxidase, cathepsin G, cationic proteins, and bactericidal/permeability-increasing protein, which is important for killing gram-negative bacteria. Azurophilic granules also contain *defensins*, a family of cysteine-rich polypeptides with broad antimicrobial activity against bacteria, fungi and certain enveloped viruses. The promyelocyte divides to produce the *myelocyte*, a cell responsible for the synthesis of the *specific*, or *secondary, granules*, which contain unique (specific) constituents such as lactoferrin, vitamin B₁₂-binding protein, membrane components of the reduced nicotinamide-adenine dinucleotide phosphate (NADPH) oxidase required for hydrogen peroxide production, histaminase, and receptors for certain chemoattractants and adherence-promoting factors (CR3) as well as receptors for the basement membrane component, laminin. The secondary granules do not contain acid hydrolases and therefore are not classic lysosomes. Packaging of secondary granule contents during myelopoiesis is controlled by CCAAT/enhancer binding protein- ϵ . Secondary granule contents are readily released extracellularly, and their mobilization is important in modulating inflammation. During the final stages of maturation, no cell division occurs, and the cell passes through the metamyelocyte stage and then to the band neutrophil with a sausage-shaped nucleus (Fig. 60-3). As the band cell matures, the nucleus assumes a lobulated configuration. The nucleus of neutrophils normally contains up to four segments (Fig. 60-4). Excessive segmentation (>5 nuclear lobes) may be a manifestation of folate or vitamin B₁₂ deficiency or the congenital neutropenia syndrome of warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) described below. The Pelger-Hüet anomaly

(Fig. 60-5), an infrequent dominant benign inherited trait, results in neutrophils with distinctive bilobed nuclei that must be distinguished from band forms. Acquired bilobed nuclei, pseudo Pelger-Hüet anomaly, can occur with acute infections or in myelodysplastic syndromes. The physiologic role of the normal multi-lobed nucleus of neutrophils is unknown, but it may allow great deformation of neutrophils during migration into tissues at sites of inflammation.

In severe acute bacterial infection, prominent neutrophil cytoplasmic granules, called *toxic granulations*, are occasionally seen. Toxic granulations are immature or abnormally staining azurophil granules. Cytoplasmic inclusions, also called *Döhle bodies* (Fig. 60-3), can be seen during infection and are fragments of ribosome-rich endoplasmic reticulum. Large neutrophil vacuoles are often present in acute bacterial infection and probably represent pinocytosed (internalized) membrane.

Neutrophils are heterogeneous in function. Monoclonal antibodies have been developed that recognize only a subset of mature neutrophils. The meaning of neutrophil heterogeneity is not known.

The morphology of eosinophils and basophils is shown in Fig. 60-6.

MARROW RELEASE AND CIRCULATING COMPARTMENTS

Specific signals, including IL-1, tumor necrosis factor α (TNF- α), the CSFs, complement fragments, and chemokines, mobilize leukocytes from the bone marrow and deliver them to the blood in an unstimulated state. Under normal conditions, ~90% of the neutrophil pool is in the bone marrow, 2–3% in the circulation, and the remainder in the tissues (Fig. 60-7).

The circulating pool exists in two dynamic compartments: one freely flowing and one marginated. The freely flowing pool is about one-half the neutrophils in the basal state and is composed of those cells that are in the blood and not in contact with the endothelium. Marginated leukocytes are those that are in close physical contact with the endothelium (Fig. 60-8). In the

pulmonary circulation, where an extensive capillary bed (~1000 capillaries per alveolus) exists, margination occurs because the capillaries are about the same size as a mature neutrophil. Therefore, neutrophil fluidity and deformability are necessary to make the transit through the pulmonary bed. Increased neutrophil rigidity and decreased

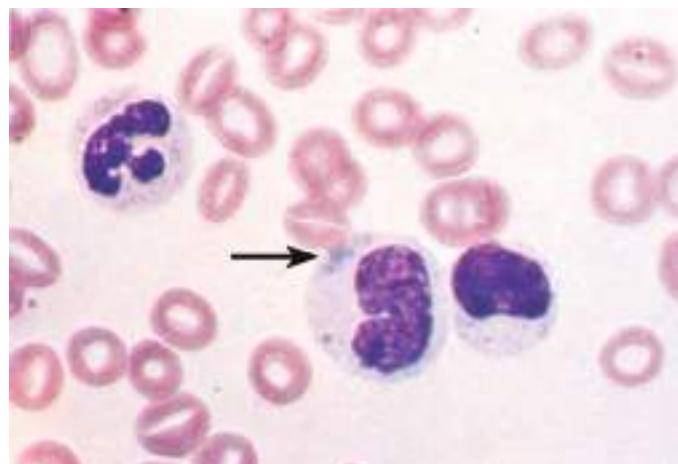


FIGURE 60-3 Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining, nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

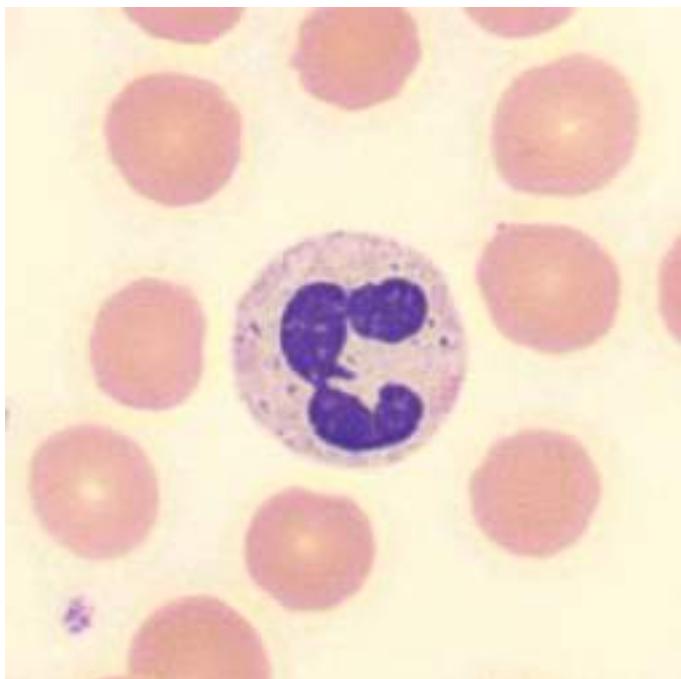


FIGURE 60-4 Normal granulocyte. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

deformability lead to augmented neutrophil trapping and margination in the lung. In contrast, in the systemic postcapillary venules, margination is mediated by the interaction of specific cell-surface molecules called *selectins*. Selectins are glycoproteins expressed on neutrophils and endothelial cells, among others, that cause a low-affinity interaction, resulting in “rolling” of the neutrophil along the endothelial surface. On neutrophils, the molecule L-selectin (cluster determinant [CD] 62L) binds to glycosylated proteins on endothelial cells (e.g., glycosylation-dependent cell adhesion molecule [GlyCAM1] and CD34). Glycoproteins on neutrophils, most importantly sialyl-Lewis^x (SLe^x, CD15s), are targets for binding of selectins expressed on endothelial cells (E-selectin [CD62E] and P-selectin [CD62P]) and other leukocytes. In response to chemotactic stimuli from injured tissues (e.g., complement product C5a, leukotriene B₄, IL-8) or bacterial products (e.g., N-formylmethionylleucylphenylalanine [f-met-leu-phe]), neutrophil adhesiveness increases through mobilization of intracellular adhesion proteins stored in specific granules to the cell surface, and the cells “stick” to the endothelium through *integrins*. The integrins are leukocyte glycoproteins that exist as complexes of a common

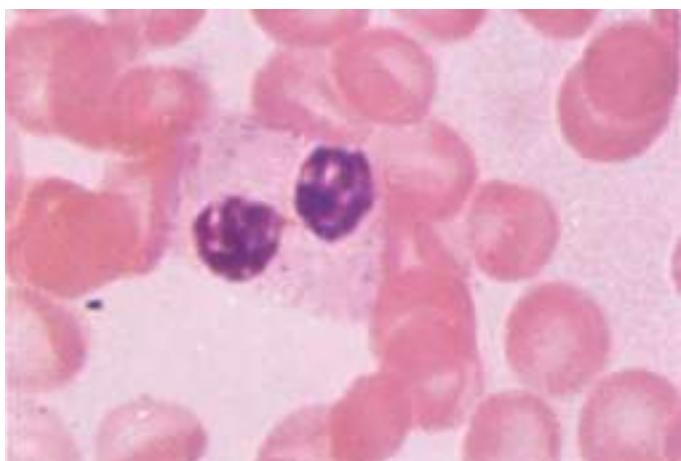


FIGURE 60-5 Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or “pince-nez,” configuration.

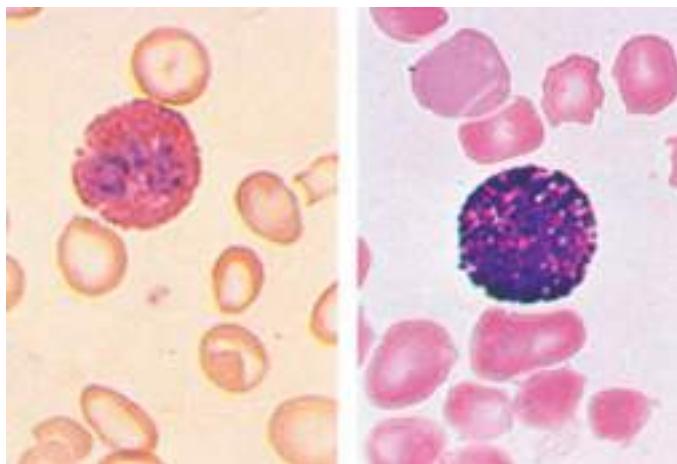


FIGURE 60-6 Normal eosinophil (left) and basophil (right). The eosinophil contains large, bright orange granules and usually a bilobed nucleus. The basophil contains large purple-black granules that fill the cell and obscure the nucleus.

CD18 β chain with CD11a (LFA-1), CD11b (called Mac-1, CR3, or the C3bi receptor), and CD11c (called p150,95 or CR4). CD11a/CD18 and CD11b/CD18 bind to specific endothelial receptors (intercellular adhesion molecules [ICAM] 1 and 2).

On cell stimulation, L-selectin is shed from neutrophils, and E-selectin increases in the blood, presumably because it is shed from endothelial cells; receptors for chemoattractants and opsonins are mobilized; and the phagocytes orient toward the chemoattractant source in the extravascular space, increase their motile activity (chemokinesis), and migrate directionally (chemotaxis) into tissues. The process of migration into tissues is called *diapedesis* and involves the crawling of neutrophils between postcapillary endothelial cells that open junctions between adjacent cells to permit leukocyte passage. Diapedesis involves platelet/endothelial cell adhesion molecule (PECAM) 1 (CD31), which is expressed on both the emigrating leukocyte and the endothelial cells. The endothelial responses (increased blood flow from increased vasodilation and permeability) are mediated by anaphylatoxins (e.g., C3a and C5a) as well as vasodilators such as histamine,

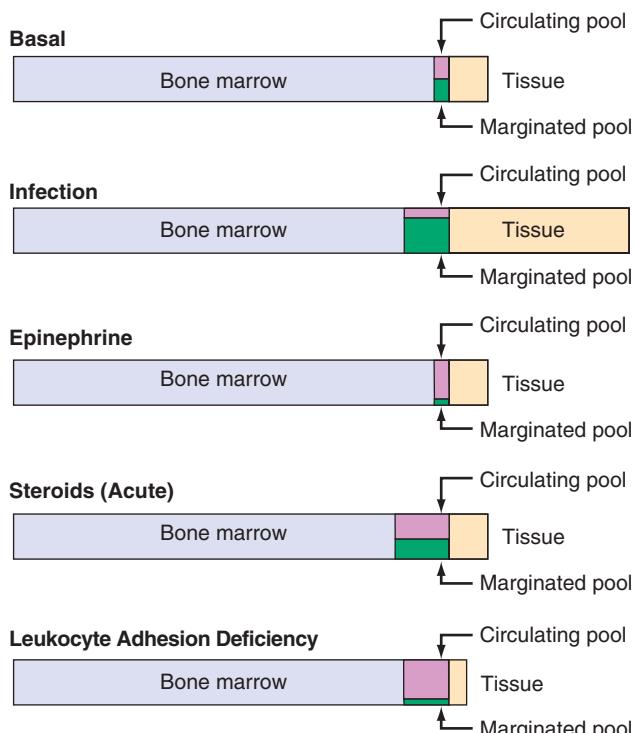


FIGURE 60-7 Schematic neutrophil distribution and kinetics between the different anatomic and functional pools.

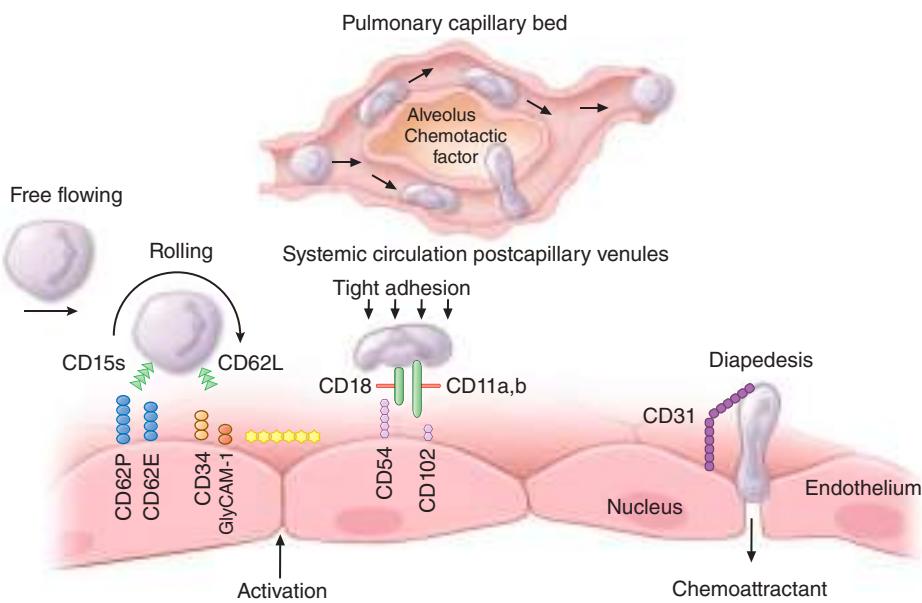


FIGURE 60-8 Neutrophil travel through the pulmonary capillaries is dependent on neutrophil deformability.

Neutrophil rigidity (e.g., caused by C5a) enhances pulmonary trapping and response to pulmonary pathogens in a way that is not so dependent on cell-surface receptors. Intraalveolar chemotactic factors, such as those caused by certain bacteria (e.g., *Streptococcus pneumoniae*), lead to diapedesis of neutrophils from the pulmonary capillaries into the alveolar space. Neutrophil interaction with the endothelium of the systemic postcapillary venules is dependent on molecules of attachment. The neutrophil "rolls" along the endothelium using selectins: neutrophil CD15s (sialyl-Lewis^x) binds to CD62E (E-selectin) and CD62P (P-selectin) on endothelial cells; CD62L (L-selectin) on neutrophils binds to CD34 and other molecules (e.g., GlyCAM-1) expressed on endothelium. Chemokines or other activation factors stimulate integrin-mediated "tight adhesion": CD11a/CD18 (LFA-1) and CD11b/CD18 (Mac-1, CR3) bind to CD54 (ICAM-1) and CD102 (ICAM-2) on the endothelium. Diapedesis occurs between endothelial cells: CD31 (PECAM-1) expressed by the emigrating neutrophil interacts with CD31 expressed at the endothelial cell-cell junction. CD, cluster determinant; GlyCAM, glycosylation-dependent cell adhesion molecule; ICAM, intercellular adhesion molecule; PECAM, platelet/endothelial cell adhesion molecule.

bradykinin, serotonin, nitric oxide, vascular endothelial growth factor (VEGF), and prostaglandins E and I. Cytokines regulate some of these processes (e.g., TNF- α induction of VEGF, interferon [IFN] γ inhibition of prostaglandin E).

In the healthy adult, most neutrophils leave the body by migration through the mucous membrane of the gastrointestinal tract. Normally, neutrophils spend a short time in the circulation (half-life, 6–7 h). Senescent neutrophils are cleared from the circulation by macrophages in the lung and spleen. Once in the tissues, neutrophils release enzymes, such as collagenase and elastase, which may help establish abscess cavities. Neutrophils ingest pathogenic materials that have been opsonized by IgG and C3b. Fibronectin and the tetrapeptide tuftsin also facilitate phagocytosis.

With phagocytosis comes a burst of oxygen consumption and activation of the hexose-monophosphate shunt. A membrane-associated NADPH oxidase, consisting of membrane and cytosolic components, is assembled and catalyzes the univalent reduction of oxygen to superoxide anion, which is then converted by superoxide dismutase to hydrogen peroxide and other toxic oxygen products (e.g., hydroxyl radical). Hydrogen peroxide + chloride + neutrophil myeloperoxidase generate hypochlorous acid (bleach), hypochlorite, and chlorine. These products oxidize and halogenate microorganisms and tumor cells and, when uncontrolled, can damage host tissue. Strongly cationic proteins, defensins, elastase, cathepsins, and probably nitric oxide also participate in microbial killing. Lactoferrin chelates iron, an important growth factor for microorganisms, especially fungi. Other enzymes, such as lysozyme and acid proteases, help digest microbial debris. After 1–4 days in tissues, neutrophils die. The apoptosis of neutrophils is also cytokine-regulated; granulocyte colony-stimulating factor (G-CSF) and IFN- γ prolong their life span. Under certain conditions, such as in delayed-type hypersensitivity, monocyte accumulation occurs within 6–12 h of initiation of inflammation. Neutrophils, monocytes, microorganisms in various states of digestion, and altered local tissue cells make up the inflammatory exudate, pus. Myeloperoxidase confers the characteristic

green color to pus and may participate in turning off the inflammatory process by inactivating chemoattractants and immobilizing phagocytic cells.

Neutrophils respond to certain cytokines (IFN- γ , granulocyte-macrophage colony-stimulating factor [GM-CSF], IL-8) and produce cytokines and chemotactic signals (TNF- α , IL-8, macrophage inflammatory protein [MIP] 1) that modulate the inflammatory response. In the presence of fibrinogen, f-met-leu-phe or leukotriene B₄ induce IL-8 production by neutrophils, providing autocrine amplification of inflammation. *Chemokines (chemoattractant cytokines)* are small proteins produced by many different cell types, including endothelial cells, fibroblasts, epithelial cells, neutrophils, and monocytes, that regulate neutrophil, monocyte, eosinophil, and lymphocyte recruitment and activation. Chemokines transduce their signals through heterotrimeric G protein-linked receptors that have seven cell membrane-spanning domains, the same type of cell-surface receptor that mediates the response to the classic chemoattractants f-met-leu-phe and C5a. Four major groups of chemokines are recognized based on the cysteine structure near the N terminus: C, CC, CXC, and CXXC. The CXC cytokines such as IL-8 mainly attract neutrophils; CC chemokines such as MIP-1 attract lymphocytes, monocytes, eosinophils, and basophils; the C chemokine lymphotactin

is T-cell tropic; the CXXC chemokine fractalkine attracts neutrophils, monocytes, and T cells. These molecules and their receptors not only regulate the trafficking and activation of inflammatory cells, but specific chemokine receptors also serve as co-receptors for HIV infection (**Chap. 197**) and have a role in other viral infections such as West Nile infection and atherosclerosis.

■ NEUTROPHIL ABNORMALITIES

Defects in the neutrophil life cycle can lead to dysfunction and compromised host defenses. Inflammation is often depressed, and the clinical result is often recurrent, severe bacterial and fungal infections. Aphthous ulcers of mucous membranes (gray ulcers without pus) and gingivitis and periodontal disease suggest a phagocytic cell disorder. Patients with congenital phagocyte defects can have infections within the first few days of life. Skin, ear, upper and lower respiratory tract, and bone infections are common. Sepsis and meningitis are rare. In some disorders, the frequency of infection is variable, and patients can go for months or even years without major infection. Aggressive management of these congenital diseases, including hematopoietic stem cell transplantation and gene therapy, has extended the life span of patients well into adulthood.

Neutropenia The consequences of absent neutrophils are dramatic. Susceptibility to infectious diseases increases sharply when neutrophil counts fall to <1000 cells/ μ L. When the absolute neutrophil count (ANC; band forms and mature neutrophils combined) falls to <500 cells/ μ L, control of endogenous microbial flora (e.g., mouth, gut) is impaired; when the ANC is <200/ μ L, the local inflammatory process is absent. Neutropenia can be due to depressed production, increased peripheral destruction, or excessive peripheral pooling. A falling neutrophil count or a significant decrease in the number of neutrophils below steady-state levels, together with a failure to increase neutrophil counts in the setting of infection or other challenge, requires investigation. Acute neutropenia, such as that caused by cancer chemotherapy,

TABLE 60-1 Causes of Neutropenia**Decreased Production**

Drug-induced—alkylating agents (nitrogen mustard, busulfan, chlorambucil, cyclophosphamide); antimetabolites (methotrexate, 6-mercaptopurine, 5-flucytosine); noncytotoxic agents (antibiotics [chloramphenicol, penicillins, sulfonamides], phenothiazines, tranquilizers [meprobamate], anticonvulsants [carbamazepine], antipsychotics [clozapine], certain diuretics, anti-inflammatory agents, antithyroid drugs, many others)

Hematologic diseases—idiopathic, cyclic neutropenia, Chédiak-Higashi syndrome, aplastic anemia, infantile genetic disorders (see text)

Tumor invasion, myelofibrosis

Nutritional deficiency—vitamin B₁₂, folate (especially alcoholics)

Infection—tuberculosis, typhoid fever, brucellosis, tularemia, measles, infectious mononucleosis, malaria, viral hepatitis, leishmaniasis, AIDS

Peripheral Destruction

Antineutrophil antibodies and/or splenic or lung trapping

Autoimmune disorders—Felty's syndrome, rheumatoid arthritis, lupus erythematosus

Drugs as haptens—aminopyrine, α-methyldopa, phenylbutazone, mercurial diuretics, some phenothiazines

Granulomatosis with polyangiitis (Wegener's)

Peripheral Pooling (Transient Neutropenia)

Overwhelming bacterial infection (acute endotoxemia)

Hemodialysis

Cardiopulmonary bypass

is more likely to be associated with increased risk of infection than neutropenia of long duration (months to years) that reverses in response to infection or carefully controlled administration of endotoxin (see "Laboratory Diagnosis and Management," below).

Some causes of inherited and acquired neutropenia are listed in **Table 60-1**. The most common neutropenias are iatrogenic, resulting from the use of cytotoxic or immunosuppressive therapies for malignancy or control of autoimmune disorders. These drugs cause neutropenia because they result in decreased production of rapidly growing progenitor (stem) cells of the marrow. Certain antibiotics such as chloramphenicol, trimethoprim-sulfamethoxazole, flucytosine, vidarabine, and the antiretroviral drug zidovudine may cause neutropenia by inhibiting proliferation of myeloid precursors. Azathioprine and 6-mercaptopurine are metabolized by the enzyme thiopurine methyltransferase (TMPT), hypofunctional polymorphisms that are found in 11% of whites and can lead to accumulation of 6-thioguanine and profound marrow toxicity. The marrow suppression is generally dose-related and dependent on continued administration of the drug. Cessation of the offending agent and recombinant human G-CSF usually reverse these forms of neutropenia.

Another important mechanism for iatrogenic neutropenia is the effect of drugs that serve as immune haptens and sensitize neutrophils or neutrophil precursors to immune-mediated peripheral destruction. This form of drug-induced neutropenia can be seen within 7 days of exposure to the drug; with previous drug exposure, resulting in preexisting antibodies, neutropenia may occur a few hours after administration of the drug. Although any drug can cause this form of neutropenia, the most frequent causes are commonly used antibiotics, such as sulfa-containing compounds, penicillins, and cephalosporins. Fever and eosinophilia may also be associated with drug reactions, but often these signs are not present. Drug-induced neutropenia can be severe, but discontinuation of the sensitizing drug is sufficient for recovery, which is usually seen within 5–7 days and is complete by 10 days. Readministration of the sensitizing drug should be avoided, because abrupt neutropenia will often result. For this reason, diagnostic challenge should be avoided.

Autoimmune neutropenias caused by circulating antineutrophil antibodies are another form of acquired neutropenia that results in increased destruction of neutrophils. Acquired neutropenia may also be seen with viral infections, including acute infection with HIV.

Acquired neutropenia may be cyclic in nature, occurring at intervals of several weeks. Acquired cyclic or stable neutropenia may be associated with an expansion of large granular lymphocytes (LGLs), which may be T cells, NK cells, or NK-like cells. Patients with large granular lymphocytosis may have moderate blood and bone marrow lymphocytosis, neutropenia, polyclonal hypergammaglobulinemia, splenomegaly, rheumatoid arthritis, and absence of lymphadenopathy. Such patients may have a chronic and relatively stable course. Recurrent bacterial infections are frequent. Benign and malignant forms of this syndrome occur. In some patients, a spontaneous regression has occurred even after 11 years, suggesting an immunoregulatory defect as the basis for at least one form of the disorder. Glucocorticoids, cyclosporine, and methotrexate are commonly used to manage these cytopenias.

Hereditary Neutropenias Hereditary neutropenias are rare and may manifest in early childhood as a profound constant neutropenia or agranulocytosis. Congenital forms of neutropenia include Kostmann's syndrome (neutrophil count <100/μL), which is often fatal and due to mutations in the antiapoptosis gene HAX-1; severe chronic neutropenia (neutrophil count of 300–1500/μL) due to mutations in neutrophil elastase (ELANE); hereditary cyclic neutropenia, or, more appropriately, cyclic hematopoiesis, also due to mutations in neutrophil elastase (ELANE); the cartilage-hair hypoplasia syndrome due to mutations in the mitochondrial RNA-processing endoribonuclease RMRP; Shwachman-Diamond syndrome associated with pancreatic insufficiency due to mutations in the Shwachman-Bodian-Diamond syndrome gene SBDS; the WHIM (*warts, hypogammaglobulinemia, infections, myelokathexis* [retention of WBCs in the marrow]) syndrome, characterized by neutrophil hypersegmentation and bone marrow myeloid arrest due to mutations in the chemokine receptor CXCR4; and neutropenias associated with other immune defects, such as X-linked agammaglobulinemia, Wiskott-Aldrich syndrome, and CD40 ligand deficiency. Mutations in the G-CSF receptor can develop in severe congenital neutropenia and are linked to leukemia. Absence of both myeloid and lymphoid cells is seen in reticular dysgenesis, due to mutations in the nuclear genome-encoded mitochondrial enzyme adenylate kinase-2 (AK2).

Maternal factors can be associated with neutropenia in the newborn. Transplacental transfer of IgG directed against antigens on fetal neutrophils can result in peripheral destruction. Drugs (e.g., thiazides) ingested during pregnancy can cause neutropenia in the newborn by either depressed production or peripheral destruction.

In Felty's syndrome—the triad of rheumatoid arthritis, splenomegaly, and neutropenia (**Chap. 351**)—spleen-produced antibodies can shorten neutrophil life span, while large granular lymphocytes can attack marrow neutrophil precursors. Splenectomy may increase the neutrophil count in Felty's syndrome and lower serum neutrophil-binding IgG. Some Felty's syndrome patients also have neutropenia associated with an increased number of LGLs. Splenomegaly with peripheral trapping and destruction of neutrophils is also seen in lysosomal storage diseases and in portal hypertension.

Neutrophilia Neutrophilia results from increased neutrophil production, increased marrow release, or defective margination (**Table 60-2**). The most important acute cause of neutrophilia is infection. Neutrophilia from acute infection represents both increased production and increased marrow release. Increased production is also associated with chronic inflammation and certain myeloproliferative diseases. Increased marrow release and mobilization of the marginated leukocyte pool are induced by glucocorticoids. Release of epinephrine, as with vigorous exercise, excitement, or stress, will demarginate neutrophils in the spleen and lungs and double the neutrophil count in minutes. Cigarette smoking can elevate neutrophil counts above the normal range. Leukocytosis with cell counts of 10,000–25,000/μL occurs in response to infection and other forms of acute inflammation and results from both release of the marginated pool and mobilization of marrow reserves. Persistent neutrophilia with cell counts of ≥30,000–50,000/μL is called a *leukemoid reaction*, a term often used to distinguish this degree of neutrophilia from leukemia. In a leukemoid reaction, the circulating neutrophils are usually mature and not clonally derived.

TABLE 60-2 Causes of Neutrophilia

Increased Production
Idiopathic
Drug-induced—glucocorticoids, G-CSF
Infection—bacterial, fungal, sometimes viral
Inflammation—thermal injury, tissue necrosis, myocardial and pulmonary infarction, hypersensitivity states, collagen vascular diseases
Myeloproliferative diseases—myelocytic leukemia, myeloid metaplasia, polycythemia vera
Increased Marrow Release
Glucocorticoids
Acute infection (endotoxin)
Inflammation—thermal injury
Decreased or Defective Margination
Drugs—epinephrine, glucocorticoids, nonsteroidal anti-inflammatory agents
Stress, excitement, vigorous exercise
Leukocyte adhesion deficiency type 1 (CD18); leukocyte adhesion deficiency type 2 (selectin ligand, CD15s); leukocyte adhesion deficiency type 3 (FERMT3)
Miscellaneous
Metabolic disorders—ketoacidosis, acute renal failure, eclampsia, acute poisoning
Drugs—lithium
Other—metastatic carcinoma, acute hemorrhage or hemolysis

Abbreviation: G-CSF, granulocyte colony-stimulating factor.

Abnormal Neutrophil Function Inherited and acquired abnormalities of phagocyte function are listed in **Table 60-3**. The resulting diseases are best considered in terms of the functional defects of adherence, chemotaxis, and microbial activity. The distinguishing features of the important inherited disorders of phagocyte function are shown in **Table 60-4**.

DISORDERS OF ADHESION Three main types of leukocyte adhesion deficiency (LAD) have been described. All are autosomal recessive and result in the inability of neutrophils to exit the circulation to sites of infection, leading to leukocytosis and increased susceptibility to infection (Fig. 60-8). Patients with LAD 1 have mutations in *CD18*, the common component of the integrins LFA-1, Mac-1, and p150,95, leading to a defect in tight adhesion between neutrophils and the endothelium. The heterodimer formed by *CD18/CD11b* (Mac-1) is also the receptor for the complement-derived opsonin C3bi (CR3). The *CD18* gene is located on distal chromosome 21q. The severity of the defect determines the severity of clinical disease. Complete lack of expression of the leukocyte integrins results in a severe phenotype in which inflammatory stimuli do not increase the expression of leukocyte integrins on neutrophils or activated T and B cells. Neutrophils (and monocytes)

from patients with LAD 1 adhere poorly to endothelial cells and protein-coated surfaces and exhibit defective spreading, aggregation, and chemotaxis. The inability of neutrophils to exit the vasculature to the tissue deprives the tissue macrophage of its expected neutrophil ingestion, leading to macrophage production of IL-23, which induces T-cell production of IL-17, a potent proinflammatory cytokine. These processes conspire to drive inflammation in LAD1. Patients with LAD 1 have recurrent bacterial infections involving the skin, oral and genital mucosa, and respiratory and intestinal tracts; persistent leukocytosis (resting neutrophil counts of 15,000–20,000/ μ L) because cells do not marginate; and, in severe cases, a history of delayed separation of the umbilical stump. Infections, especially of the skin, may become necrotic with progressively enlarging borders, slow healing, and development of dysplastic scars. The most common bacteria are *Staphylococcus aureus* and enteric gram-negative bacteria. LAD 2 is caused by an abnormality of fucosylation of SLe^x (CD15s), the ligand on neutrophils that interacts with selectins on endothelial cells and is responsible for neutrophil rolling along the endothelium. Infection susceptibility in LAD 2 appears to be less severe than in LAD 1. LAD 2 is also known as *congenital disorder of glycosylation IIc* (CDGIIc) due to mutation in a GDP-fucose transporter (*SLC35C1*). LAD 3 is characterized by infection susceptibility, leukocytosis, and petechial hemorrhage due to impaired integrin activation caused by mutations in the gene *FERMT3*.

DISORDERS OF NEUTROPHIL GRANULES The most common neutrophil defect is myeloperoxidase deficiency, a primary granule defect inherited as an autosomal recessive trait; the incidence is ~1 in 2000 persons. Isolated myeloperoxidase deficiency is not associated with clinically compromised defenses, presumably because other defense systems such as hydrogen peroxide generation are amplified. Microbicidal activity of neutrophils is delayed but not absent. Myeloperoxidase deficiency may make other acquired host defense defects more serious, and patients with myeloperoxidase deficiency and diabetes are more susceptible to *Candida* infections. An acquired form of myeloperoxidase deficiency occurs in myelomonocytic leukemia and acute myeloid leukemia.

Chédiak-Higashi syndrome (CHS) is a rare disease with autosomal recessive inheritance due to defects in the lysosomal transport protein LYST, encoded by the gene *CHS1* at 1q42. This protein is required for normal packaging and disbursement of granules. Neutrophils (and all cells containing lysosomes) from patients with CHS characteristically have large granules (Fig. 60-9), making it a systemic disease. Patients with CHS have nystagmus, partial oculocutaneous albinism, and an increased number of infections resulting from many bacterial agents. Some CHS patients develop an “accelerated phase” in childhood with a hemophagocytic syndrome and an aggressive lymphoma requiring bone marrow transplantation. CHS neutrophils and monocytes have impaired chemotaxis and abnormal rates of microbial killing due to slow rates of fusion of the lysosomal granules with phagosomes.

TABLE 60-3 Types of Granulocyte and Monocyte Disorders

FUNCTION	CAUSE OF INDICATED DYSFUNCTION		
	DRUG-INDUCED	ACQUIRED	INHERITED
Adherence-aggregation	Aspirin, colchicine, alcohol, glucocorticoids, ibuprofen, piroxicam	Neonatal state, hemodialysis	Leukocyte adhesion deficiency types 1, 2, and 3
Deformability		Leukemia, neonatal state, diabetes mellitus, immature neutrophils	
Chemokinesis-chemotaxis	Glucocorticoids (high dose), auranofin, colchicine (weak effect), phenylbutazone, naproxen, indomethacin, interleukin 2	Thermal injury, malignancy, malnutrition, periodontal disease, neonatal state, systemic lupus erythematosus, rheumatoid arthritis, diabetes mellitus, sepsis, influenza virus infection, herpes simplex virus infection, acrodermatitis enteropathica, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, hyper IgE–recurrent infection (Job's) syndrome (in some patients), Down's syndrome, α -mannosidase deficiency, leukocyte adhesion deficiencies, Wiskott-Aldrich syndrome
Microbicidal activity	Colchicine, cyclophosphamide, glucocorticoids (high dose), TNF- α -blocking antibodies	Leukemia, aplastic anemia, certain neutropenias, tuftsin deficiency, thermal injury, sepsis, neonatal state, diabetes mellitus, malnutrition, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, chronic granulomatous disease, defects in IFN γ /IL-12 axis

Abbreviations: IFN γ , interferon γ ; IL, interleukin; TNF- α , tumor necrosis factor alpha.

TABLE 60-4 Inherited Disorders of Phagocyte Function: Differential Features

CLINICAL MANIFESTATIONS	CELLULAR OR MOLECULAR DEFECTS	DIAGNOSIS
Chronic Granulomatous Diseases (70% X-Linked, 30% Autosomal Recessive)		
Severe infections of skin, ears, lungs, liver, and bone with catalase-positive microorganisms such as <i>Staphylococcus aureus</i> , <i>Burkholderia cepacia</i> complex, <i>Aspergillus</i> spp., <i>Chromobacterium violaceum</i> ; often hard to culture organism; excessive inflammation with granulomas, frequent lymph node suppuration; granulomas can obstruct GI or GU tracts; gingivitis, aphthous ulcers, seborrheic dermatitis	No respiratory burst due to the lack of one of five NADPH oxidase subunits in neutrophils, monocytes, and eosinophils	DHR or NBT test; no superoxide and H ₂ O ₂ production by neutrophils; immunoblot for NADPH oxidase components; genetic detection
Chédiak-Higashi Syndrome (Autosomal Recessive)		
Recurrent pyogenic infections, especially with <i>S. aureus</i> ; many patients get lymphoma-like illness during adolescence; periodontal disease; partial oculocutaneous albinism, nystagmus, progressive peripheral neuropathy, mental retardation in some patients	Reduced chemotaxis and phagolysosome fusion, increased respiratory burst activity, defective egress from marrow, abnormal skin window; defect in CHS1	Giant primary granules in neutrophils and other granule-bearing cells (Wright's stain); genetic detection
Specific Granule Deficiency (Autosomal Recessive and Dominant)		
Recurrent infections of skin, ears, and sinopulmonary tract; delayed wound healing; decreased inflammation; bleeding diathesis	Abnormal chemotaxis, impaired respiratory burst and bacterial killing, failure to upregulate chemotactic and adhesion receptors with stimulation, defect in transcription of granule proteins; defect in CEBPE	Lack of secondary (specific) granules in neutrophils (Wright's stain), no neutrophil-specific granule contents (i.e., lactoferrin), no defensins, platelet α granule abnormality; genetic detection
Myeloperoxidase Deficiency (Autosomal Recessive)		
Clinically normal except in patients with underlying disease such as diabetes mellitus; then candidiasis or other fungal infections	No myeloperoxidase due to pre- and posttranslational defects in myeloperoxidase deficiency	No peroxidase in neutrophils; genetic detection
Leukocyte Adhesion Deficiency		
Type 1: Delayed separation of umbilical cord, sustained neutrophilia, recurrent infections of skin and mucosa, gingivitis, periodontal disease	Impaired phagocyte adherence, aggregation, spreading, chemotaxis, phagocytosis of C3bi-coated particles; defective production of CD18 subunit common to leukocyte integrins	Reduced phagocyte surface expression of the CD18-containing integrins with monoclonal antibodies against LFA-1 (CD18/CD11a), Mac-1 or CR3 (CD18/CD11b), p150,95 (CD18/CD11c); genetic detection
Type 2: Mental retardation, short stature, Bombay (hh) blood phenotype, recurrent infections, neutrophilia	Impaired phagocyte rolling along endothelium; due to defects in fucose transporter	Reduced phagocyte surface expression of Sialyl-Lewis ^x , with monoclonal antibodies against CD15s; genetic detection
Type 3: Petechial hemorrhage, recurrent infections	Impaired signaling for integrin activation resulting in impaired adhesion due to mutation in FERM73	Reduced signaling for adhesion through integrins; genetic detection
Phagocyte Activation Defects (X-Linked and Autosomal Recessive)		
NEMO deficiency: mild hypohidrotic ectodermal dysplasia; broad-based immune defect: pyogenic and encapsulated bacteria, viruses, <i>Pneumocystis</i> , mycobacteria; X-linked IRAK4 and MyD88 deficiency: susceptibility to pyogenic bacteria such as staphylococci, streptococci, clostridia; resistant to <i>Candida</i> ; autosomal recessive	Impaired phagocyte activation by IL-1, IL-18, TLR, CD40L, TNF-α leading to problems with inflammation and antibody production	Poor in vitro response to endotoxin; impaired NF-κB activation; genetic detection
DOCK8 deficiency (autosomal recessive), severe eczema, atopic dermatitis, cutaneous abscesses, HSV, HPV, and molluscum infections, severe allergies, cancer	Impaired phagocyte activation by endotoxin through TLR and other pathways; TNF-α signaling preserved	Poor in vitro response to endotoxin; lack of NF-κB activation by endotoxin; genetic detection
Hyper IgE-Recurrent Infection Syndrome (Autosomal Dominant) (Job's Syndrome)		
Eczematoid or pruritic dermatitis, "cold" skin abscesses, recurrent pneumonias with <i>S. aureus</i> with bronchopleural fistulae and cyst formation, mild eosinophilia, mucocutaneous candidiasis, characteristic facies, restrictive lung disease, scoliosis, delayed primary dental deciduation	Reduced chemotaxis in some patients, reduced memory T and B cells; mutation in STAT3	Somatic and immune features involving lungs, skeleton, and immune system; serum IgE >2000 IU/mL; genetic testing
DOCK8 deficiency (autosomal recessive), severe eczema, atopic dermatitis, cutaneous abscesses, HSV, HPV, and molluscum infections, severe allergies, cancer	Impaired T-cell proliferation to mitogens; mutation in DOCK8	Severe allergies, viral infections, high IgE, eosinophilia, low IgM, progressive lymphopenia, genetic detection
Mycobacteria Susceptibility (Autosomal Dominant and Recessive Forms)		
Severe extrapulmonary or disseminated infections with bacille Calmette-Guérin (BCG), nontuberculous mycobacteria, salmonella, histoplasmosis, coccidioidomycosis, poor granuloma formation	Inability to kill intracellular organisms due to low IFN-γ production or response; mutations in IFN-γ receptors, IL-12 receptors, IL-12 p40, STAT1, NEMO, ISG15, GATA2	Abnormally low or very high levels of IFN-γ receptor 1; functional assays of cytokine production and response; genetic detection
GATA2 Deficiency (Autosomal Dominant)		
Persistent or disseminated warts, disseminated mycobacterial disease, low monocytes, NK cells, B cells; hypoplastic myelodysplasia, leukemia, cytogenetic abnormalities, pulmonary alveolar proteinosis	Impaired macrophage activity, cytopenias; mutations in GATA2	Profound circulating monocytopenia, NK and B-cell cytopenias; genetic detection

Abbreviations: C/EBPs, CCAAT/enhancer binding protein-ε; DHR, dihydrorhodamine (oxidation test); DOCK8, dedicator of cytokinesis 8; GI, gastrointestinal; GU, genitourinary; HPV, human papilloma virus; HSV, herpes simplex virus; IFN, interferon; IL, interleukin; IRAK4, IL-1 receptor-associated kinase 4; LFA-1, leukocyte function-associated antigen 1; MyD88, myeloid differentiation primary response gene 88; NADPH, nicotinamide-adenine dinucleotide phosphate; NBT, nitroblue tetrazolium (dye test); NEMO, NF-κB essential modulator; NF-κB, nuclear factor-κB; NK, natural killer; STAT1–3, signal transducer and activator of transcription 1–3; TLR, Toll-like receptor; TNF, tumor necrosis factor.

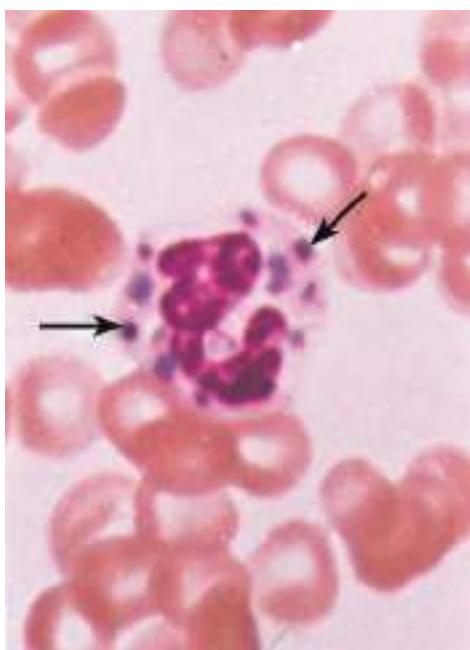


FIGURE 60-9 Chédiak-Higashi syndrome. The granulocytes contain huge cytoplasmic granules formed from aggregation and fusion of azurophilic and specific granules. Large abnormal granules are found in other granule-containing cells throughout the body.

NK cell function is also impaired. CHS patients may develop a severe disabling peripheral neuropathy in adulthood.

Specific granule deficiency is a rare autosomal recessive disease in which the production of secondary granules and their contents, as well as the primary granule component defensins, is defective. The defect in killing leads to severe bacterial infections. One type of specific granule deficiency is due to a mutation in the CCAAT/enhancer binding protein- ϵ , a regulator of expression of granule components. A dominant mutation in *C/EBP- ϵ* has also been described.

CHRONIC GRANULOMATOUS DISEASE Chronic granulomatous disease (CGD) is a group of disorders of granulocyte and monocyte oxidative metabolism. Although CGD is rare, with an incidence of ~1 in 200,000 individuals, it is an important model of defective neutrophil oxidative metabolism. In about two-thirds of patients, CGD is inherited as an X-linked recessive trait; the remainder patients inherit the disease in an autosomal recessive pattern. Mutations in the genes for the five proteins that assemble at the plasma membrane account for all patients with CGD. Two proteins (a 91-kDa protein, abnormal in X-linked CGD, and a 22-kDa protein, absent in one form of autosomal recessive CGD) form the heterodimer cytochrome b-558 in the plasma membrane. Three other proteins (40, 47, and 67 kDa, abnormal in the other autosomal recessive forms of CGD) are cytoplasmic in origin and interact with the cytochrome after cell activation to form the NADPH oxidase, required for hydrogen peroxide production. Leukocytes from patients with CGD have severely diminished hydrogen peroxide production. The genes involved in each of the defects have been cloned and sequenced and the chromosome locations identified. Patients with CGD characteristically have increased numbers of infections due to catalase-positive microorganisms (organisms that destroy their own hydrogen peroxide) such as *S. aureus*, *Burkholderia cepacia* complex, and *Aspergillus* species. When patients with CGD become infected, they often have extensive inflammatory reactions, and lymph node suppuration is common despite the administration of appropriate antibiotics. Aphthous ulcers and chronic inflammation of the nares are often present. Granulomas are frequent and can obstruct the gastrointestinal or genitourinary tracts. The excessive inflammation is due to failure to downregulate inflammation, reflecting a failure to inhibit the synthesis of, degradation of, or response to ILs or chemoattractants, leading to persistent myeloid reaction. Impaired killing of intracellular microorganisms by macrophages may lead to persistent cell-mediated immune

activation and granuloma formation. Autoimmune complications such as immune thrombocytopenic purpura and juvenile rheumatoid arthritis are also increased in CGD. In addition, for unexplained reasons, discoid lupus is more common in X-linked carriers. Late complications, including nodular regenerative hyperplasia and portal hypertension, are increasingly recognized in older patients with CGD.

DISORDERS OF PHAGOCYTE ACTIVATION Phagocytes depend on cell-surface stimulation to induce signals that evoke multiple levels of the inflammatory response, including cytokine synthesis, chemotaxis, and antigen presentation. Mutations affecting the major pathway that signals through NF- κ B have been noted in patients with a variety of infection susceptibility syndromes. If the defects are at a very late stage of signal transduction, in the protein critical for NF- κ B activation known as the NF- κ B essential modulator (NEMO), then affected males develop ectodermal dysplasia and severe immune deficiency with susceptibility to bacteria, fungi, mycobacteria, and viruses. If the defects in NF- κ B activation are closer to the cell-surface receptors, in the proteins transducing Toll-like receptor signals, IL-1 receptor-associated kinase 4 (IRAK4), and myeloid differentiation primary response gene 88 (MyD88), then children have a marked susceptibility to pyogenic infections early in life but develop resistance to infection later.

MONONUCLEAR PHAGOCYTES

The mononuclear phagocyte system is composed of monoblasts, promonocytes, and monocytes, in addition to the structurally diverse tissue macrophages that make up what was previously referred to as the reticuloendothelial system. Macrophages are long-lived phagocytic cells capable of many of the functions of neutrophils. They are also secretory cells that participate in many immunologic and inflammatory processes distinct from neutrophils. Monocytes leave the circulation by diapedesis more slowly than neutrophils and have a half-life in the blood of 12–24 h.

After blood monocytes arrive in the tissues, they differentiate into macrophages ("big eaters") with specialized functions suited for specific anatomic locations. Macrophages are particularly abundant in capillary walls of the lung, spleen, liver, and bone marrow, where they function to remove microorganisms and other noxious elements from the blood. Alveolar macrophages, liver Kupffer cells, splenic macrophages, peritoneal macrophages, bone marrow macrophages, lymphatic macrophages, brain microglial cells, and dendritic macrophages all have specialized functions. Macrophage-secreted products include lysozyme, neutral proteases, acid hydrolases, arginase, complement components, enzyme inhibitors (plasmin, α_2 -macroglobulin), binding proteins (transferrin, fibronectin, transcobalamin II), nucleosides, and cytokines (TNF- α ; IL-1, 8, 12, 18). IL-1 (**Chaps. 15 and 342**) has many functions, including initiating fever in the hypothalamus, mobilizing leukocytes from the bone marrow, and activating lymphocytes and neutrophils. TNF- α is a pyrogen that duplicates many of the actions of IL-1 and plays an important role in the pathogenesis of gram-negative shock (**Chap. 297**). TNF- α stimulates production of hydrogen peroxide and related toxic oxygen species by macrophages and neutrophils. In addition, TNF- α induces catabolic changes that contribute to the profound wasting (cachexia) associated with many chronic diseases.

Other macrophage-secreted products include reactive oxygen and nitrogen metabolites, bioactive lipids (arachidonic acid metabolites and platelet-activating factors), chemokines, CSFs, and factors stimulating fibroblast and vessel proliferation. Macrophages help regulate the replication of lymphocytes and participate in the killing of tumors, viruses, and certain bacteria (*Mycobacterium tuberculosis* and *Listeria monocytogenes*). Macrophages are key effector cells in the elimination of intracellular microorganisms. Their ability to fuse to form giant cells that coalesce into granulomas in response to some inflammatory stimuli is important in the elimination of intracellular microbes and is under the control of IFN- γ . Nitric oxide induced by IFN- γ is an important effector against intracellular parasites, including tuberculosis and *Leishmania*.

Macrophages play an important role in the immune response (**Chap. 342**). They process and present antigen to lymphocytes and secrete cytokines that modulate and direct lymphocyte development

and function. Macrophages participate in autoimmune phenomena by removing immune complexes and other substances from the circulation. Polymorphisms in macrophage receptors for immunoglobulin (Fc γ RII) determine susceptibility to some infections and autoimmune diseases. In wound healing, they dispose of senescent cells, and they contribute to atheroma development. Macrophage elastase mediates development of emphysema from cigarette smoking.

DISORDERS OF THE MONONUCLEAR PHAGOCYTE SYSTEM

Many disorders of neutrophils extend to mononuclear phagocytes. Monocytosis is associated with tuberculosis, brucellosis, subacute bacterial endocarditis, Rocky Mountain spotted fever, malaria, and visceral leishmaniasis (kala azar). Monocytosis also occurs with malignancies, leukemias, myeloproliferative syndromes, hemolytic anemias, chronic idiopathic neutropenia, and granulomatous diseases such as sarcoidosis, regional enteritis, and some collagen vascular diseases. Patients with LAD, hyperimmunoglobulin E-recurrent infection (Job's) syndrome, CHS, and CGD all have defects in the mononuclear phagocyte system.

Monocyte cytokine production or response is impaired in some patients with disseminated nontuberculous mycobacterial infection who are not infected with HIV. Genetic defects in the pathways regulated by IFN- γ and IL-12 lead to impaired killing of intracellular bacteria, mycobacteria, salmonellae, and certain viruses (Fig. 60-10).

Certain viral infections impair mononuclear phagocyte function. For example, influenza virus infection causes abnormal monocyte chemotaxis. Mononuclear phagocytes can be infected by HIV using CCR5, the chemokine receptor that acts as a co-receptor with CD4 for HIV. T lymphocytes produce IFN- γ , which induces FcR expression and phagocytosis and stimulates hydrogen peroxide production by mononuclear phagocytes and neutrophils. In certain diseases, such as AIDS, IFN- γ production may be deficient, whereas in other diseases, such as T-cell lymphomas, excessive release of IFN- γ may be associated with erythrophagocytosis by splenic macrophages.

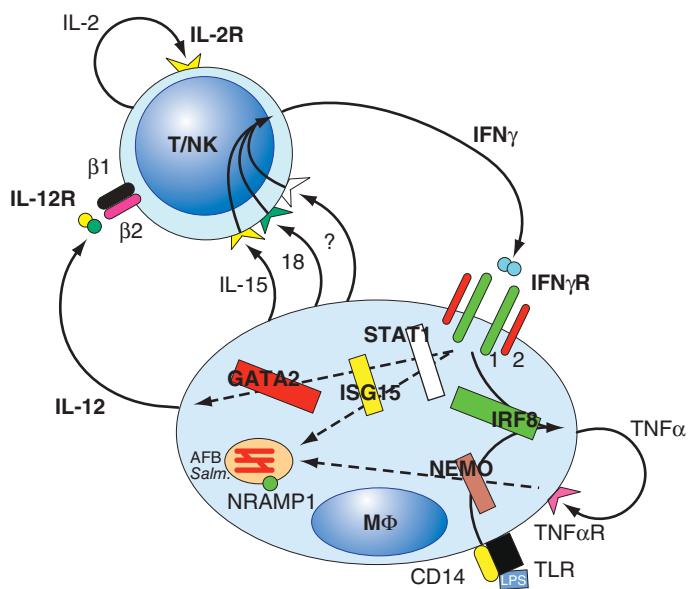


FIGURE 60-10 Lymphocyte-macrophage interactions underlying resistance to mycobacteria and other intracellular pathogens such as *Salmonella*, *Histoplasma*, and *Coccidioides*. Mycobacteria (and others) infect macrophages, leading to the production of IL-12, which activates T or NK cells through its receptor, leading to production of IL-2 and IFN- γ . IFN- γ acts through its receptor on macrophages to upregulate TNF- γ and IL-12 and kill intracellular pathogens. Other critical interacting molecules include signal transducer and activator of transcription 1 (STAT1), interferon regulatory factor 8 (IRF8), GATA2, and ISG15. Mutant forms of the cytokines and receptors shown in bold type have been found in severe cases of nontuberculous mycobacterial infection, salmonellosis and other intracellular pathogens. AFB, acid-fast bacilli; IFN, interferon; IL, interleukin; NEMO, nuclear factor- κ B essential modulator; NK, natural killer; TLR, Toll-like receptor; TNF, tumor necrosis factor.

Autoinflammatory diseases are characterized by abnormal cytokine regulation, leading to excess inflammation in the absence of infection. These diseases can mimic infectious or immunodeficient syndromes. Gain-of-function mutations in the TNF- α receptor cause TNF- α receptor-associated periodic syndrome (TRAPS), which is characterized by recurrent fever in the absence of infection, due to persistent stimulation of the TNF- α receptor (Chap. 362). Diseases with abnormal IL-1 regulation leading to fever include familial Mediterranean fever due to mutations in PYRIN. Mutations in *cold-induced autoinflammatory syndrome 1* (CIAS1) lead to neonatal-onset multisystem autoinflammatory disease, familial cold urticaria, and Muckle-Wells syndrome. The syndrome of pyoderma gangrenosum, acne, and sterile pyogenic arthritis (PAPA syndrome) is caused by mutations in PSTPIP1. In contrast to these syndromes of overexpression of proinflammatory cytokines, blockade of TNF- α by the antagonists infliximab, adalimumab, certolizumab, golimumab, or etanercept has been associated with severe infections due to tuberculosis, nontuberculous mycobacteria, and fungi (Chap. 362).

Monocytopenia occurs with acute infections, with stress, and after treatment with glucocorticoids. Drugs that suppress neutrophil production in the bone marrow can cause monocytopenia. Persistent severe circulating monocytopenia is seen in GATA2 deficiency, even though macrophages are found at the sites of inflammation. Monocytopenia also occurs in aplastic anemia, hairy cell leukemia, acute myeloid leukemia, and as a direct result of myelotoxic drugs.

EOSINOPHILS

Eosinophils and neutrophils share similar morphology, many lysosomal constituents, phagocytic capacity, and oxidative metabolism. Eosinophils express a specific chemoattractant receptor and respond to a specific chemokine, eotaxin, but little is known about their required role. Eosinophils are much longer lived than neutrophils, and unlike neutrophils, tissue eosinophils can recirculate. During most infections, eosinophils appear unimportant. However, in invasive helminthic infections, such as hookworm, schistosomiasis, strongyloidiasis, toxocariasis, trichinosis, filariasis, echinococcosis, and cysticercosis, the eosinophil plays a central role in host defense. Eosinophils are associated with bronchial asthma, cutaneous allergic reactions, and other hypersensitivity states.

The distinctive feature of the red-staining (Wright's stain) eosinophil granule is its crystalline core consisting of an arginine-rich protein (major basic protein) with histaminase activity, important in host defense against parasites. Eosinophil granules also contain a unique eosinophil peroxidase that catalyzes the oxidation of many substances by hydrogen peroxide and may facilitate killing of microorganisms.

Eosinophil peroxidase, in the presence of hydrogen peroxide and halide, initiates mast cell secretion in vitro and thereby promotes inflammation. Eosinophils contain cationic proteins, some of which bind to heparin and reduce its anticoagulant activity. Eosinophil-derived neurotoxin and eosinophil cationic protein are ribonucleases that can kill respiratory syncytial virus. Eosinophil cytoplasm contains Charcot-Leyden crystal protein, a hexagonal bipyramidal crystal first observed in a patient with leukemia and then in sputum of patients with asthma; this protein is lysophospholipase and may function to detoxify certain lysophospholipids.

Several factors enhance the eosinophil's function in host defense. T cell-derived factors enhance the ability of eosinophils to kill parasites. Mast cell-derived eosinophil chemotactic factor of anaphylaxis (ECFa) increases the number of eosinophil complement receptors and enhances eosinophil killing of parasites. Eosinophil CSFs (e.g., IL-5) produced by macrophages increase eosinophil production in the bone marrow and activate eosinophils to kill parasites.

EOSINOPHILIA

Eosinophilia is the presence of >500 eosinophils per μ L of blood and is common in many settings besides parasite infection. Significant tissue eosinophilia can occur without an elevated blood count. A common cause of eosinophilia is allergic reaction to drugs (iodides, aspirin, sulfonamides, nitrofurantoin, penicillins, and cephalosporins). Allergies such as hay fever, asthma, eczema, serum sickness,

allergic vasculitis, and pemphigus are associated with eosinophilia. Eosinophilia also occurs in collagen vascular diseases (e.g., rheumatoid arthritis, eosinophilic fasciitis, allergic angiitis, and periarteritis nodosa) and malignancies (e.g., Hodgkin's disease; mycosis fungoides; chronic myeloid leukemia; and cancer of the lung, stomach, pancreas, ovary, or uterus), as well as in STAT3 deficient Job's syndrome, DOCK8 deficiency (see below), and CGD. Eosinophilia is commonly present in helminthic infections. IL-5 is the dominant eosinophil growth factor. Therapeutic administration of the cytokines IL-2 or GM-CSF frequently leads to transient eosinophilia. The most dramatic hypereosinophilic syndromes are Loeffler's syndrome, tropical pulmonary eosinophilia, Loeffler's endocarditis, eosinophilic leukemia, and idiopathic hyper-eosinophilic syndrome (50,000–100,000/ μ L). IL-5 is the dominant eosinophil growth factor and can be specifically inhibited with the monoclonal antibody mepolizumab.

The idiopathic hypereosinophilic syndrome represents a heterogeneous group of disorders with the common feature of prolonged eosinophilia of unknown cause and organ system dysfunction, including the heart, central nervous system, kidneys, lungs, gastrointestinal tract, and skin. The bone marrow is involved in all affected individuals, but the most severe complications involve the heart and central nervous system. Clinical manifestations and organ dysfunction are highly variable. Eosinophils are found in the involved tissues and likely cause tissue damage by local deposition of toxic eosinophil proteins such as eosinophil cationic protein and major basic protein. In the heart, the pathologic changes lead to thrombosis, endocardial fibrosis, and restrictive endomyocardopathy. The damage to tissues in other organ systems is similar. Some cases are due to mutations involving the platelet-derived growth factor receptor, and these are extremely sensitive to the tyrosine kinase inhibitor imatinib. Glucocorticoids, hydroxyurea, and IFN- α each have been used successfully, as have therapeutic antibodies against IL-5. Cardiovascular complications are managed aggressively.

The *eosinophilia-myalgia syndrome* is a multisystem disease, with prominent cutaneous, hematologic, and visceral manifestations, that frequently evolves into a chronic course and can occasionally be fatal. The syndrome is characterized by eosinophilia (eosinophil count >1000/ μ L) and generalized disabling myalgias without other recognized causes. Eosinophilic fasciitis, pneumonitis, and myocarditis; neuropathy culminating in respiratory failure; and encephalopathy may occur. The disease is caused by ingesting contaminants in L-tryptophan-containing products. Eosinophils, lymphocytes, macrophages, and fibroblasts accumulate in the affected tissues, but their role in pathogenesis is unclear. Activation of eosinophils and fibroblasts and the deposition of eosinophil-derived toxic proteins in affected tissues may contribute. IL-5 and transforming growth factor β have been implicated as potential mediators. Treatment is withdrawal of products containing L-tryptophan and the administration of glucocorticoids. Most patients recover fully, remain stable, or show slow recovery, but the disease can be fatal in up to 5% of patients.

Eosinophilic neoplasms are discussed in Chap. 106.

■ EOSINOPENIA

Eosinopenia occurs with stress, such as acute bacterial infection, and after treatment with glucocorticoids. The mechanism of eosinopenia of acute bacterial infection is unknown but is independent of endogenous glucocorticoids, because it occurs in animals after total adrenalectomy. There is no known adverse effect of eosinopenia.

HYPERIMMUNOGLOBULIN E-RECURRENT INFECTION SYNDROME

The hyperimmunoglobulin E-recurrent infection syndrome, or Job's syndrome, is a rare multisystem disease in which the immune and somatic systems are affected, including neutrophils, monocytes, T cells, B cells, and osteoclasts. Autosomal dominant inhibitory mutations in signal transducer and activator of transcription 3 (STAT3) lead to inhibition of normal STAT signaling with broad and profound effects. Patients have characteristic facies with broad nose, kyphoscoliosis, and eczema. The primary teeth erupt normally but do not deciduate,

often requiring extraction. Patients develop recurrent sinopulmonary and cutaneous infections that tend to be much less inflamed than appropriate for the degree of infection and have been referred to as "cold abscesses." Characteristically, pneumonias cavitate, leading to pneumatoceles. Coronary artery aneurysms are common, as are cerebral demyelinated plaques that accumulate with age. Importantly, IL-17-producing T cells, which are thought responsible for protection against extracellular and mucosal infections, are profoundly reduced in Job's syndrome. Despite very high IgE levels, these patients have only mildly elevated levels of allergy. An important syndrome with clinical overlap with the dominant negative STAT3 deficiency is due to autosomal recessive defects in dedicator of cytokinesis 8 (DOCK8). In DOCK8 deficiency, IgE elevation is joined to severe allergy, viral susceptibility, and increased rates of cancer. Autosomal dominant *gain-of-function* mutations in STAT3 lead to a disease characterized by onset in childhood of lymphadenopathy, autoimmune cytopenias, multiorgan autoimmunity, infections, and interstitial lung disease.

LABORATORY DIAGNOSIS AND MANAGEMENT

Initial studies of WBC and differential and often a bone marrow examination may be followed by assessment of bone marrow reserves (steroid challenge test), marginated circulating pool of cells (epinephrine challenge test), and marginating ability (endotoxin challenge test) (Fig. 60-7). In vivo assessment of inflammation is possible with a Re buck skin window test or an in vivo skin blister assay, which measures the ability of leukocytes and inflammatory mediators to accumulate locally in the skin. In vitro tests of phagocyte aggregation, adherence, chemotaxis, phagocytosis, degranulation, and microbial activity (for *S. aureus*) may help pinpoint cellular or humoral lesions. Deficiencies of oxidative metabolism are detected with either the nitroblue tetrazolium (NBT) dye test or the dihydrorhodamine (DHR) oxidation test. These tests are based on the ability of products of oxidative metabolism to alter the oxidation states of reporter molecules so that they can be detected microscopically (NBT) or by flow cytometry (DHR). Qualitative studies of superoxide and hydrogen peroxide production may further define neutrophil oxidative function.

Patients with leukopenias or leukocyte dysfunction often have delayed inflammatory responses. Therefore, clinical manifestations may be minimal despite overwhelming infection, and unusual infections must always be suspected. Early signs of infection demand prompt, aggressive culturing for microorganisms, use of antibiotics, and drainage of abscesses. Prolonged courses of antibiotics are often required. In patients with CGD, prophylactic antibiotics (trimethoprim-sulfamethoxazole) and antifungals (itraconazole) markedly diminish the frequency of life-threatening infections. Glucocorticoids may relieve gastrointestinal or genitourinary tract obstruction by granulomas in patients with CGD. Although TNF- α -blocking agents may markedly relieve inflammatory bowel symptoms, extreme caution must be exercised in their use in CGD inflammatory bowel disease, because it profoundly increases these patients' already heightened susceptibility to infection. Recombinant human IFN- γ , which nonspecifically stimulates phagocytic cell function, reduces the frequency of infections in patients with CGD by 70% and reduces the severity of infection. This effect of IFN- γ in CGD is additive to the effect of prophylactic antibiotics. The recommended dose is 50 μ g/m² subcutaneously three times weekly. IFN- γ has also been used successfully in the treatment of leprosy, non-tuberculous mycobacteria, and visceral leishmaniasis.

Rigorous oral hygiene reduces but does not eliminate the discomfort of gingivitis, periodontal disease, and aphthous ulcers; chlorhexidine mouthwash and tooth brushing with a hydrogen peroxide-sodium bicarbonate paste help many patients. Oral anti-fungal agents (fluconazole, itraconazole, voriconazole, posaconazole) have reduced mucocutaneous candidiasis in patients with Job's syndrome. Androgens, glucocorticoids, lithium, and immunosuppressive therapy have been used to restore myelopoiesis in patients with neutropenia due to impaired production. Recombinant G-CSF is useful in the management of certain forms of neutropenia due to depressed neutrophil production, including those related to cancer

chemotherapy. Patients with chronic neutropenia with evidence of a good bone marrow reserve need not receive prophylactic antibiotics. Patients with chronic or cyclic neutrophil counts <500/ μ L may benefit from prophylactic antibiotics and G-CSF during periods of neutropenia. Oral trimethoprim-sulfamethoxazole (160/800 mg) twice daily can prevent infection. Increased numbers of fungal infections are not seen in patients with CGD on this regimen. Oral quinolones such as levofloxacin and ciprofloxacin are alternatives.

In the setting of cytotoxic chemotherapy with severe, persistent lymphocyte dysfunction, trimethoprim-sulfamethoxazole prevents *Pneumocystis jiroveci* pneumonia. These patients, and patients with phagocytic cell dysfunction, should avoid heavy exposure to airborne soil, dust, or decaying matter (mulch, manure), which are often rich in *Nocardia* and the spores of *Aspergillus* and other fungi. Restriction of activities or social contact has no proven role in reducing risk of infection for phagocyte defects.

Although aggressive medical care for many patients with phagocytic disorders can allow them to go for years without a life-threatening infection, there may still be delayed effects of prolonged antimicrobials and other inflammatory complications. Cure of most congenital phagocyte defects is possible by bone marrow transplantation, and rates of success are improving (**Chap. 110**). The identification of specific gene defects in patients with LAD 1, CGD, and other immunodeficiencies has led to gene therapy trials in a number of genetic white cell disorders.

FURTHER READING

- CASANOVA JL: Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proc Natl Acad Sci USA* 112:E7128, 2015.
- KOLACZKOWSKA E, KUBES P: Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol* 13:159, 2013.
- LEIDING JW et al (eds): *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2017. 2012 August 9 [updated 2016 February 11].
- LIONAKIS MS et al: Immunity against fungi. *JCI Insight* 2: pii: 93156, 2017.
- MOUTSOPoulos NM et al: Interleukin-12 and interleukin-23 blockade in leukocyte adhesion deficiency type 1. *N Engl J Med* 376:1141, 2017.
- SOEHNLEIN O et al: Neutrophils as protagonists and targets in chronic inflammation. *Nat Rev Immunol* 17:248, 2017.
- WILLIAMS KW et al: Eosinophilia associated with disorders of immune deficiency or immune dysregulation. *Immunol Allergy Clin North Am* 35:523, 2015.
- WU UI, HOLLAND SM: Host susceptibility to non-tuberculous mycobacterial infections. *Lancet Infect Dis* 15:968, 2015.

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. Because this receptor is the key mediator of platelet aggregation, it has become an effective target for antiplatelet therapy.

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. 61-1**. 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).

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 primarily dependent on its activation by factor XIIa but rather on its positive feedback activation by thrombin. Thus, factor Xla functions in the propagation and amplification, rather than in the initiation, of the coagulation cascade.

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

61

Bleeding and Thrombosis

Barbara A. Konkle

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.

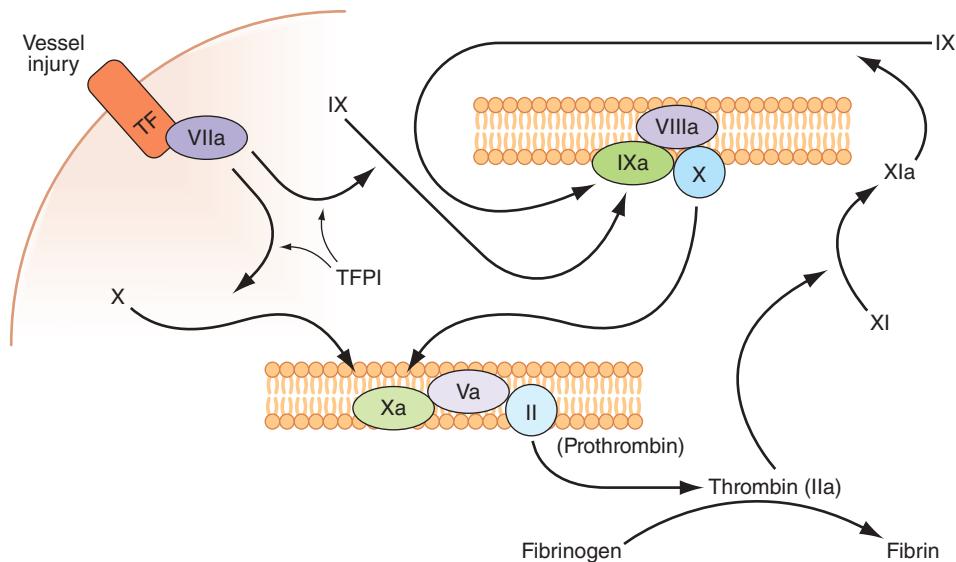


FIGURE 61-1 Coagulation is initiated by tissue factor (TF) exposure, which, with factor (F) VIIa, activates FIX and FX, which in turn, with FVIII and FV as cofactors, respectively, results in thrombin formation and subsequent conversion of fibrinogen to fibrin. Thrombin activates FXI, FVIII, and FV, 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.

associations (Fig. 61-2). Thrombin also activates factor XIII (fibrin-stabilizing factor) to factor XIIIa, which covalently cross-links and thereby stabilizes the fibrin clot.

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, antithrombin, TF pathway inhibitor, and thrombomodulin. They also activate fibrinolytic mechanisms through the production of tissue plasminogen activator 1, urokinase, plasminogen activator inhibitor, 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. 61-3.) 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 mutation at its target cleavage site in factor V (factor V Leiden), lead to hypercoagulable states.

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 (LMWH).

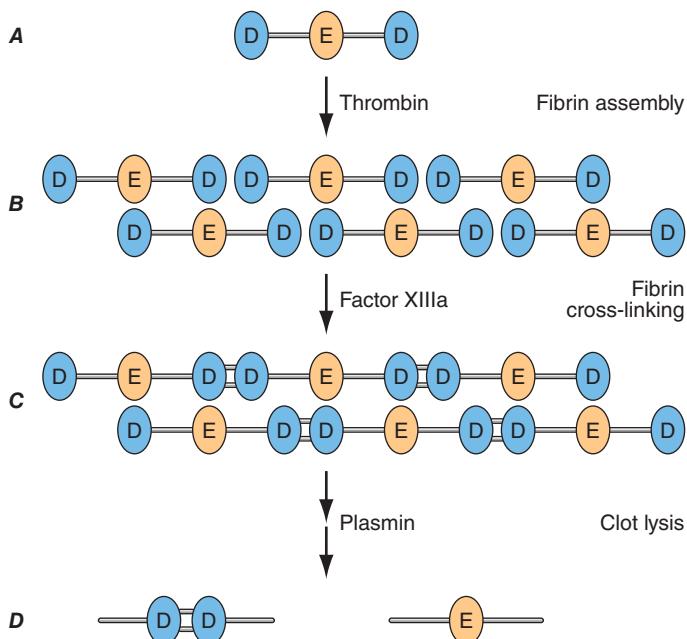


FIGURE 61-2 Fibrin formation and dissolution. (A) Fibrinogen is a trinodular structure consisting of two D domains and one E domain. Thrombin activation results in an ordered lateral assembly of protofibrils (B) with noncovalent associations. Factor XIIIa cross-links the D domains on adjacent molecules (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.

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

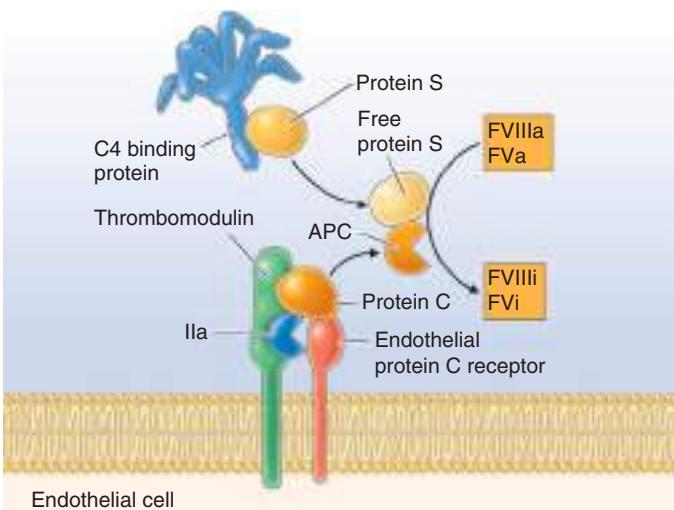


FIGURE 61-3 The activated protein C pathway in regulation of thrombosis. Thrombin generation results in protein C activation through interaction with thrombomodulin and protein C bound to the endothelial protein C receptor (EPCR). Activated protein C (APC) with free protein S converts activated factors (F) VIII and V to inactivate forms, thus in turn decreasing thrombin generation. APC, activated protein C; C4BP, C4 binding protein; EC, endothelial cell; EPCR, endothelial protein C receptor; F, factor; IIa, thrombin; PC, protein C; PS, protein S; TM, thrombomodulin.

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. 61-4.

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

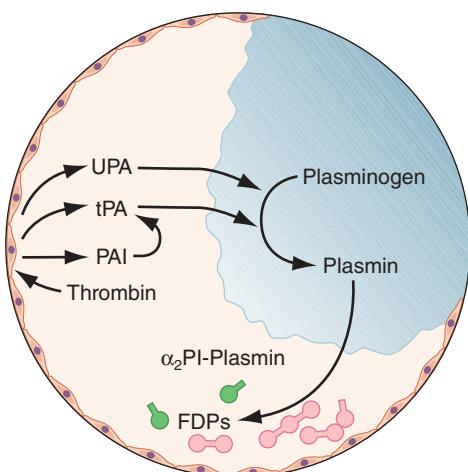


FIGURE 61-4 A schematic diagram of the fibrinolytic system. Tissue plasminogen activator (tPA) is released from endothelial cells, binds the fibrin clot, and activates plasminogen to plasmin. Excess fibrin is degraded by plasmin to distinct degradation products (FDPs). Any free plasmin is complexed with α₂-antiplasmin (α₂PI). PAI, plasminogen activator inhibitor; UPA, urokinase-type plasminogen activator.

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. 61-2). 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. In addition, d-dimer measurement can be used to stratify patients, particularly women, for risk of recurrent venous thromboembolism (VTE) when measured 1 month after discontinuation of anticoagulation given for treatment of an initial idiopathic event. D-Dimer levels increase with age. Whether a higher cut-off should be used in the elderly 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) α₂-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. α₂-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, 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 61-1.

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]). This is the most useful tool in excluding the diagnosis of a bleeding disorder, and thus avoiding unnecessary testing. One study found that a low bleeding score (≤ 3) and a normal activated partial thromboplastin time (aPTT) had 99.6% negative predictive value for the diagnosis of VWD. Bleeding symptoms that appear to be more common in patients with bleeding disorders include prolonged bleeding with surgery, dental procedures and extractions, and/or trauma, heavy menstrual bleeding (HMB), or postpartum hemorrhage (PPH), and large bruises (often described with lumps).

TABLE 61-1 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

Easy bruising and HMB 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 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 HMB include bleeding resulting in iron-deficiency anemia or a need for blood transfusion, passage of clots >1 in. in diameter, and changing a pad or tampon more than hourly. HMB is a common symptom in women with underlying bleeding disorders and is reported in the majority of women with VWD, women with factor XI deficiency, and symptomatic carriers of hemophilia. Women with underlying bleeding disorders are more likely to have other bleeding symptoms, including bleeding after dental extractions, postoperative bleeding, and postpartum bleeding, and are much more likely to have HMB beginning at menarche than women with HMB due to other causes.

PPH is a common symptom in women with underlying bleeding disorders. In women with type 1 VWD and symptomatic carriers of hemophilia A in whom levels of VWF and factor VIII usually

normalize during pregnancy, PPH may be delayed. Women with a history of PPH may have a higher risk of recurrence with subsequent pregnancies. Rupture of ovarian cysts with intraabdominal hemorrhage has also been reported in women with underlying bleeding disorders.

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 ~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, has been associated with angiodyplasia 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 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.

Prehemorrhagic 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. This aspirin effect lasts for the life of the platelet, though 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 P2Y₁₂ 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 61-2**). 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 PGI₃, a more potent platelet inhibitor than prostacyclin (PGI₂), and more thromboxane A₂, a less potent platelet activator than thromboxane A₂. 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

TABLE 61-2 Herbal Supplements Associated with Increased Bleeding**Herbs with Potential Antiplatelet Activity**

Ginkgo (<i>Ginkgo biloba</i> L.)
Garlic (<i>Allium sativum</i>)
Bilberry (<i>Vaccinium myrtillus</i>)
Ginger (<i>Zingiber officinale</i>)
Dong quai (<i>Angelica sinensis</i>)
Feverfew (<i>Tanacetum parthenium</i>)
Asian ginseng (<i>Panax ginseng</i>)
American ginseng (<i>Panax quinquefolius</i>)
Siberian ginseng/eleuthero (<i>Eleutherococcus senticosus</i>)
Turmeric (<i>Circuma longa</i>)
Meadowsweet (<i>Filipendula ulmaria</i>)
Willow (<i>Salix</i> spp.)
Coumarin-Containing Herbs
Motherwort (<i>Leonurus cardiaca</i>)
Chamomile (<i>Matricaria recutita</i> , <i>Chamaemelum nobilis</i>)
Horse chestnut (<i>Aesculus hippocastanum</i>)
Red clover (<i>Trifolium pratense</i>)
Fenugreek (<i>Trigonella foenum-graecum</i>)

results in combined factor deficiencies. This is often compounded by thrombocytopenia associated with liver failure 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 . The level needed for major surgery will depend on the type of surgery and the patient's underlying medical state, although a count of ~80,000/ μL is likely sufficient.

HISTORY OF THROMBOSIS

The risk of thrombosis, like that of bleeding, is influenced by both genetic and environmental influences. 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 61-3**.

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 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,

TABLE 61-3 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
Elevated factor VIII	Hormonal therapy
Acquired	Polycythemia vera
Age	Essential thrombocythemia
Previous thrombosis	Paroxysmal nocturnal hemoglobinuria
Immobilization	Thrombotic thrombocytopenic purpura
Major surgery	Heparin-induced thrombocytopenia
Pregnancy and puerperium	Disseminated intravascular coagulation
Hospitalization	Unknown^a
Obesity	Elevated factor II, IX, XI
Infection	Elevated TAFI levels
APC resistance, nongenetic	Low levels of TFPI
Smoking	

^aUnknown whether risk is inherited or acquired.

Abbreviations: APC, activated protein C; TAFI, thrombin-activatable fibrinolysis inhibitor; TFPI, tissue factor pathway inhibitor.

is a minor determinant of risk in an elderly individual undergoing a high-risk surgical procedure. As illustrated in **Fig. 61-5**, 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.

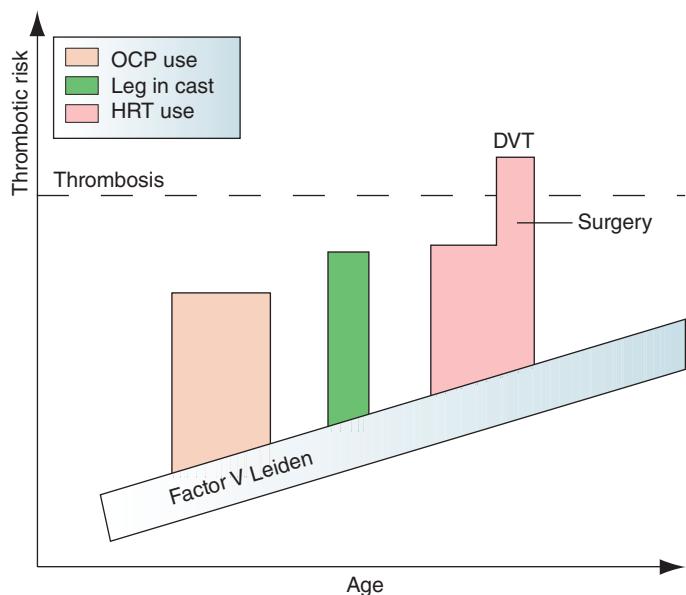


FIGURE 61-5 Thrombotic risk over time. Shown schematically is an individual's thrombotic risk over time. An underlying factor V Leiden mutation provides a "theoretically" constant increased risk. The thrombotic risk increases with age and, intermittently, with oral contraceptive (OCP) or hormone replacement therapy (HRT) use; other events may 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. (From BA Konkle, A Schafer, in DP Zipes et al [eds]: Braunwald's Heart Disease, 7th ed. Philadelphia, Saunders, 2005; modified with permission from FR Rosendaal: Venous thrombosis: A multicausal disease. Lancet 353:1167, 1999.)

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 complement, 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 but is not broadly applicable for the diagnosis of disorders of hemostasis and thrombosis.

For routine preoperative and pre-procedure 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 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 result 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. 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. 61-6). 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 61-4. 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;

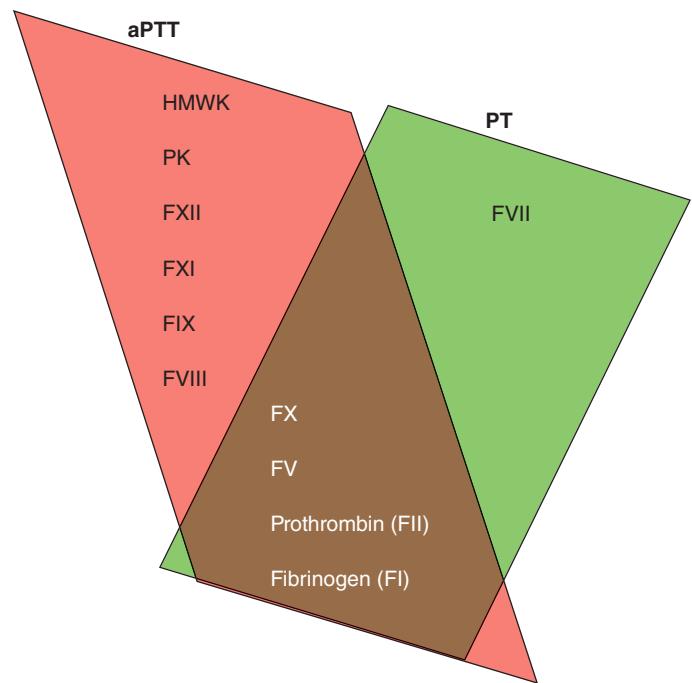


FIGURE 61-6 Coagulation factor activity tested in the activated partial thromboplastin time (aPTT) in red and prothrombin time (PT) in green, or both. F, factor; HMWK, high-molecular-weight kininogen; PK, prekallikrein.

TABLE 61-4 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

Factors 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

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

high-molecular-weight kininogen; and factor XII (Fig. 61-6). 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 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) and the tissue thromboplastin inhibition (TTI) test are modifications of standard tests with the phospholipid reagent decreased, thus increasing the sensitivity to antibodies that interfere with the phospholipid component. The 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 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. Anti-phospholipid 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. In most instances, warfarin anticoagulation can be stopped after the initial 3–6 months of treatment, and testing can be performed at least 3 weeks later. As a sensitive marker of coagulation activation, the quantitative D-dimer assay, drawn 4 weeks after stopping anticoagulation, can be used to stratify risk of recurrent thrombosis in patients, particularly women, who have an idiopathic event.

Measures of Platelet Function The bleeding time has been used 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, as with the bleeding time, 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 tests are affected by many factors, including numerous medications, and the association between minor defects in aggregation or secretion in these assays and bleeding risk is not clearly established.

FURTHER READING

- GIANNAKOPOULOS B, KRILIS SA: The pathogenesis of the antiphospholipid syndrome. *N Engl J Med* 368:11, 2013.
- HICKS LK et al: The ASH choosing wisely® campaign: Five hematologic tests and treatments to question. *Blood* 122:3879, 2013.

KONKLE BA: Direct oral anticoagulants: Monitoring anticoagulant effect, in *Direct Oral Anticoagulants in Clinical Practice*, Connors JM, ed., Hematol Oncol Clin North Am 30:995, 2016.

MACKIE I et al: Guidelines on the laboratory aspect of assays used in haemostasis and thrombosis. Int Jnl Lab Hem 35:1, 2013.

MIDDEL DORP S: Evidence-based approach to thrombophilia testing. J Thromb Haemost 31:275, 2011.

RYDZ N, JAMES PD: The evolution and value of bleeding assessment tools. J Thromb Haemost 10:2223, 2012.

WAGENMAN BL et al: The laboratory approach to inherited and acquired coagulation factor deficiencies. Clin Lab Med 29:229, 2009.

YAU JW et al: Endothelial cell control of thrombosis. BMC Cardiovasc Disord 15:130, 2015.

62

Enlargement of Lymph Nodes and Spleen

Dan L. Longo



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 62-1**. Many of these disorders are infrequent causes of lymphadenopathy. In primary care practice, more than two-thirds of patients with lymphadenopathy have non-specific 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,

TABLE 62-1 Diseases Associated with Lymphadenopathy

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.

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 62-1) 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 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 $<1.0\text{ cm}^2$ in area ($1.0\text{ cm} \times 1.0\text{ cm}$ 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 $>2\text{ cm}$ 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 ($1.5\text{ cm} \times 1.5\text{ cm}$) was the best size limit for distinguishing malignant or granulomatous lymphadenopathy from other causes of lymphadenopathy. Patients with node(s) $\leq 1.0\text{ cm}^2$ 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 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 two 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 cm², hard texture, and lack of pain or tenderness. Negative predictive value was evident for age <40 years, node size <1.0 cm², 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.

SPLENOMEGLY

STRUCTURE AND FUNCTION OF THE SPLEEN

The spleen is a reticuloendothelial organ that has its embryologic origin in the dorsal mesogastrium at about five 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:

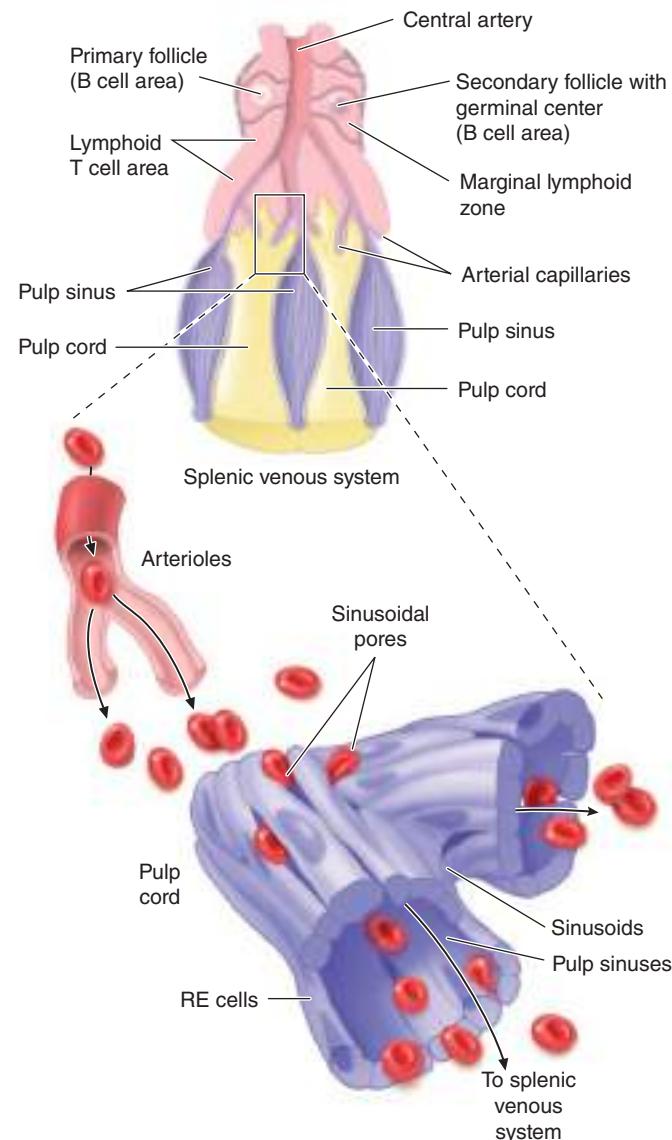


FIGURE 62-1 Schematic spleen structure. The spleen comprises many units of red and white pulp centered around small branches of the splenic artery, called *central arteries*. White pulp is lymphoid in nature and contains B-cell follicles, a marginal zone around the follicles, and T-cell-rich areas sheathing 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 from RS Hillman, KA Ault: *Hematology in Clinical Practice*, 4th ed. New York, McGraw-Hill, 2005.)

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. 62-1).
2. Synthesis of antibodies in the white pulp.
3. The removal of antibody-coated bacteria and antibody-coated blood cells from the circulation.

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. 219, 220, and A6**), nuclear residua (Howell-Jolly bodies, see Fig. 59-6), 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 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, ballottement, 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 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. 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 62-2**. 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," palpable >8 cm below the left costal margin or its drained weight is ≥ 1000 g (**Table 62-3**). The vast majority of such patients will have non-Hodgkin's lymphoma, chronic

TABLE 62-2 Diseases Associated with Splenomegaly Grouped by Pathogenic Mechanism

Enlargement Due to Increased Demand for Splenic Function

Reticuloendothelial system hyperplasia (for removal of defective erythrocytes)	Malaria
Spherocytosis	Leishmaniasis
Early sickle cell anemia	Trypanosomiasis
Ovalocytosis	Ehrlichiosis
Thalassemia major	Disordered immunoregulation
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
Bacterial septicemia	Thyrotoxicosis (benign lymphoid hypertrophy)
Congenital syphilis	Interleukin 2 therapy
Splenic abscess	Extramedullary hematopoiesis
Tuberculosis	Myelofibrosis
Histoplasmosis	Marrow damage by toxins, radiation, strontium
	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 thrombocythosis)
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	

TABLE 62-3 Diseases Associated with Massive Splenomegaly*

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

*The spleen extends >8 cm below left costal margin and/or weighs >1000 g.

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 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.

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/ μ L) and thrombocytosis (up to $1 \times 10^6/\mu$ L) 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 three 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 post-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 (Prevenar, 7-valent). 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.
- FACCHETTI F: Tumors of the spleen. Int J Surg Pathol 18:136S, 2010.
- GIRARD E et al: Management of splenic and pancreatic trauma. J Visc Surg 153(suppl 4):45, 2016.
- GRAVES SA et al: Does this patient have splenomegaly? JAMA 270:2218, 1993.
- KIM DK et al: Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older—United States, 2017. MMWR 66:136, 2017.
- 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.
- MCINTYRE OR, EBAUGH FG Jr: Palpable spleens: Ten year follow-up. Ann Intern Med 90:130, 1979.
- PANGALIS GA et al: Clinical approach to lymphadenopathy. Semin Oncol 20:570, 1993.
- WILLIAMSON HA Jr: Lymphadenopathy in a family practice: A descriptive study of 240 cases. J Fam Pract 20:449, 1985.



Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among healthcare 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. The goal of this chapter is to describe the principles of clinical pharmacology that can be used for the safe and optimal use of available and new drugs.

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 despite equivalent 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.

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 effects. 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. 64*.

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 two broad categories: those designed to alleviate a symptom and those designed to prolong useful life. An increasing emphasis on the principles of evidence-based medicine and techniques such as large clinical trials and meta-analyses has defined benefits of drug therapy in broad patient populations. However, establishing the balance between risk and benefit is not always simple. An increasing body of evidence supports the idea, with which practitioners are very familiar, that individual patients may display responses that are not expected from large population studies and often have comorbidities that typically exclude them from large clinical trials. 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 adverse drug reactions 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 adverse effects (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 post-marketing 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. 63-1). 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

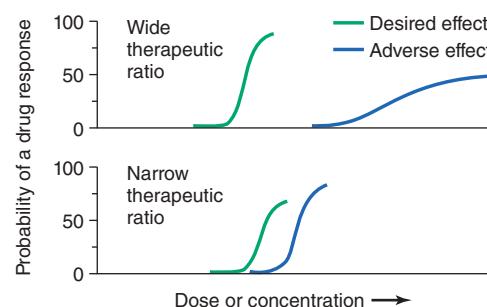


FIGURE 63-1 The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. **Top.** A drug with a wide therapeutic ratio, that is, a wide separation of the two curves. **Bottom.** A drug with a narrow 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.

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.

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

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug actually entering the systemic circulation may be less than with the intravenous route (Fig. 63-2A). The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) absorption is reduced, or (2) the drug undergoes metabolism or elimination prior to entering the systemic circulation. Occasionally, the administered drug formulation is inconsistent or has degraded with time; for example, the anticoagulant dabigatran degrades rapidly (over weeks) once exposed to air, so the amount administered may be less than prescribed.

When a drug is administered by a non-intravenous route, the peak concentration occurs later and is lower than after the same dose given by rapid intravenous injection, reflecting absorption from the site of administration (Fig. 63-2). The extent of absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at its 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.

“First-Pass” Effect When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the

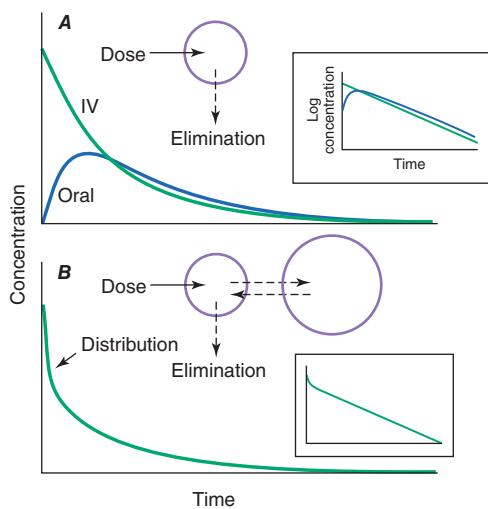


FIGURE 63-2 Idealized time-plasma concentration curves after a single dose of drug. **A.** The time course of drug concentration after an instantaneous IV bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV, 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.

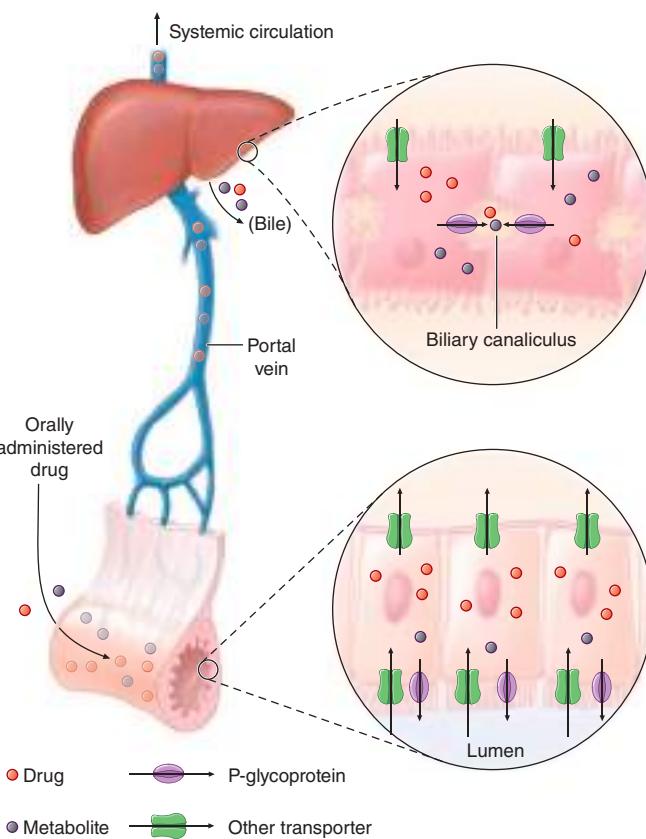


FIGURE 63-3 Mechanism of presystemic clearance. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to 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.)

liver prior to entering the systemic circulation (Fig. 63-3). 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 systemic 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*.

■ 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. 63-3). 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

TABLE 63-1 Molecular Pathways Mediating Drug Disposition

MOLECULE	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	
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	
	Some sulfonamides	
UGT1A1 ^b	Irinotecan	
Pseudocholinesterase ^b	Succinylcholine	
P-glycoprotein	Digoxin	Quinidine
	HIV protease inhibitors	Amiodarone
	Many CYP3A substrates	Verapamil
		Cyclosporine
		Itraconazole
		Erythromycin
SLCO1B1 ^b	Simvastatin and some other statins	

^aInhibitors affect the molecular pathway, and thus may affect substrate. ^bClinically important genetic variants described; see Chap. 64.

Note: A listing of CYP substrates, inhibitors, and inducers is maintained at <http://medicine.iupui.edu/clinpharm/ddis/main-table>.

by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs and other molecules that are especially important for drug metabolism are presented in Table 63-1, 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 further below.

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 deacetylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

■ 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"). In the simplest pharmacokinetic model (Fig. 63-2A), 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 the logarithm of concentration versus time is a straight line (Fig. 63-2A, 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–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. 63-2B). 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–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. 63-4). 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.

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. 63-4).

■ DRUG DISTRIBUTION

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

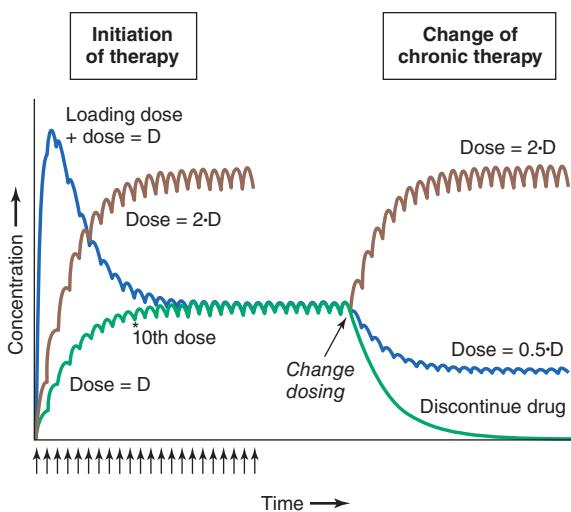


FIGURE 63-4 Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state 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.)

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. 63-4). Nevertheless, the time required for 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. 63-2 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. 241) 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:

$$C_{ss} = \text{dosing rate}/Cl \quad \text{or} \quad \text{dosing rate} = Cl \cdot C_{ss}$$

When 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%:

$$\text{Dose}/\text{time} = Cl \cdot C_{avg,ss}/F$$

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 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 des pointes ventricular tachycardia (Chap. 247) 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 anti-estrogen 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 generation of reactive metabolites that mediate certain adverse drug effects (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 *second* 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

The Onset of Drug Action For drugs used in the urgent treatment of acute symptoms, little or no delay is anticipated (or desired) between 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 between the interaction of a drug with its pharmacologic target(s) and a clinical effect is clinically acceptable. Common pharmacokinetic mechanisms that can contribute to such a delay include slow elimination (resulting in slow accumulation to steady state), uptake into peripheral compartments, or accumulation of active metabolites. A common pharmacodynamic explanation for such a delay is that the clinical effect develops as a downstream consequence of the initial molecular effect the drug produces. Thus, administration of a proton pump inhibitor or an H₂-receptor blocker produces an immediate increase in gastric pH but ulcer healing that is delayed. Cancer chemotherapy similarly produces delayed therapeutic effects.

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 adverse effects. 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 cardio-pulmonary problem. As a result, alternate antiarrhythmic therapies are 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 down-regulated 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.

■ 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 adverse effects. Previous experience with the drug, in controlled clinical trials or in post-marketing use, defines the relationships between dose or plasma concentration and these dual effects (Fig. 63-1) 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

diseases 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 maximum-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 adverse drug effects (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.*

Failure of Efficacy Assuming the diagnosis is correct and the correct drug is prescribed, 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 post-dose 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 (pre-dose). 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. 63-4), 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. 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. 63-5).

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

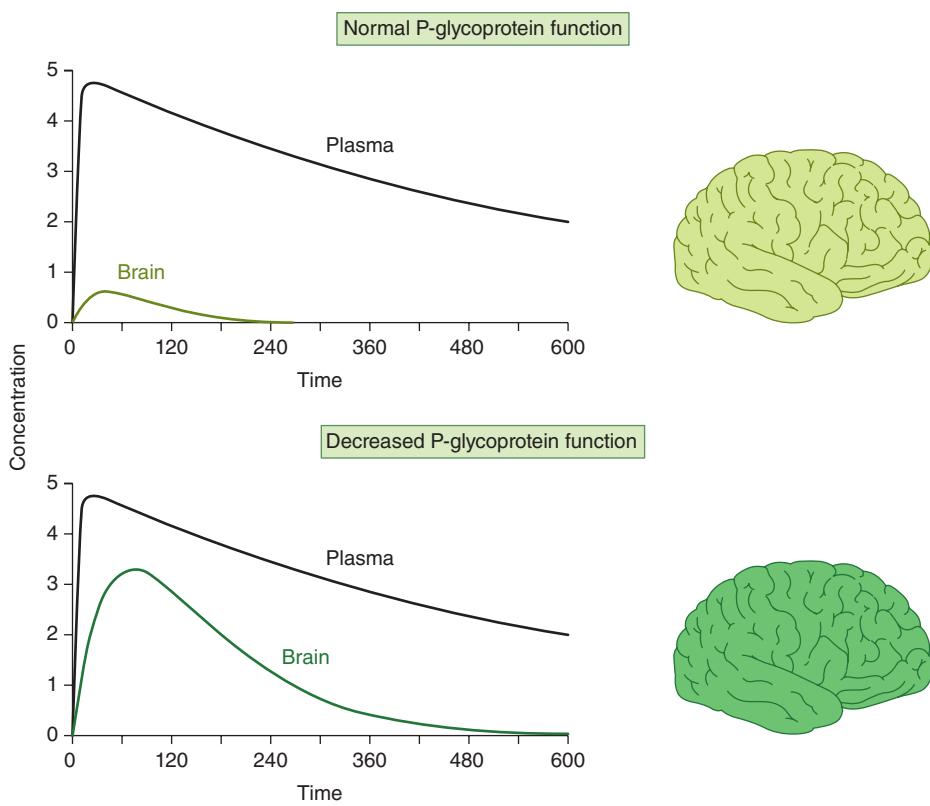


FIGURE 63-5 The efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain and so constitutes a key element of the blood-brain barrier. Thus, reduced P-glycoprotein function (e.g., due to drug interactions) increases penetration of substrate drugs into the brain, even when plasma concentrations are unchanged.

disproportionate increases in plasma concentration; examples include phenytoin and theophylline.

An increase in dosage is usually best achieved by changing the drug dose but not the dosing interval (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 adverse effects (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 non-renally, 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. 252). 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. As well, 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 adverse effects. 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, 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.

Adverse drug reactions 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 adverse drug reaction in the previous year.

■ DRUG USE IN CHILDREN

While most drugs used to treat disease in children are the same as those in adults, there are few studies that provide solid data to guide dosing. Drug metabolism pathways mature at different rates after birth, and disease mechanisms may be different in children. In practice, doses are adjusted for size (weight or body surface area) as a first approximation unless age-specific data are available.

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. *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 include examination of the patient's medications and, if necessary, calls to the pharmacist to identify prescriptions. It should also address the use of 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. Examples are presented below and in Table 63-2. Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

■ PHARMACOKINETIC INTERACTIONS CAUSING DECREASED DRUG EFFECTS

Gastrointestinal absorption can be reduced if a drug interaction results in drug binding in the gut, as with aluminum-containing antacids, kaolin-pectin suspensions, or bile acid sequestrants. Drugs such as histamine H₂-receptor antagonists or proton pump inhibitors that alter gastric pH may decrease the solubility and hence absorption of weak bases such as ketoconazole.

Expression of some genes responsible for drug elimination, notably CYP3A and ABCB1, can be markedly increased by inducing drugs, such as rifampin, carbamazepine, phenytoin, St. John's wort, and

TABLE 63-2 Drugs with a High Risk of Generating Pharmacokinetic Interactions

DRUG	MECHANISM	EXAMPLES
Antacids	Reduced absorption	Antacids/tetracyclines
Bile acid sequestrants		Cholestyramine/digoxin
Proton pump inhibitors	Altered gastric pH	Ketoconazole absorption decreased
H ₂ -receptor blockers		
Rifampin	Induction of CYPs and/or P-glycoprotein	Decreased concentration and effects of warfarin
Carbamazepine		quinidine
Barbiturates		cyclosporine
Phenytoin		losartan
St. John's wort		oral contraceptives
Glutethimide		
Nevirapine (CYP3A; CYP2B6)		methadone, dabigatran
Tricyclic antidepressants	Inhibitors of CYP2D6	Increased effect of many β blockers
Fluoxetine		Decreased codeine effect; possible decreased tamoxifen effect
Quinidine		
Cimetidine	Inhibitor of multiple CYPs	Increased concentration and effects of warfarin
		theophylline
		phenytoin
Ketoconazole, itraconazole	Inhibitor of CYP3A	Increased concentration and toxicity of some HMG-CoA reductase inhibitors, colchicine
Erythromycin, clarithromycin		Cyclosporine, cisapride, terfenadine (now withdrawn)
Calcium channel blockers		Increased concentration and effects of indinavir (with ritonavir)
Ritonavir		Decreased clearance and dose requirement for cyclosporine (with calcium channel blockers)
Allopurinol	Xanthine oxidase inhibitor	Azathioprine and 6-mercaptopurine toxicity
Amiodarone	Inhibitor of many CYPs and of P-glycoprotein	Decreased clearance (risk of toxicity) for warfarin
		digoxin
		quinidine
Gemfibrozil (and other fibrates)	CYP3A inhibition	Rhabdomyolysis when co-prescribed with some HMG-CoA reductase inhibitors
Quinidine	P-glycoprotein inhibition	Risk of toxicity with P-glycoprotein substrates (e.g., digoxin, dabigatran)
Amiodarone		
Verapamil		
Cyclosporine		
Itraconazole		
Erythromycin		
Phenylbutazone	Inhibition of renal tubular transport	Increased risk of methotrexate toxicity with salicylates
Probenecid		
Salicylates		

glutethimide, and by smoking, exposure to chlorinated insecticides, and chronic alcohol ingestion. Administration of inducing agents lowers plasma levels, and thus effects, over 2–3 weeks as gene expression is increased. If a drug dose is stabilized in the presence of an inducer that is subsequently stopped, major toxicity can occur as clearance returns to preinduction levels and drug concentrations rise. Individuals vary in the extent to which drug metabolism can be induced, likely through genetic mechanisms.

Interactions that inhibit the bioactivation of prodrugs will decrease drug effects (Table 63-1).

Interactions that decrease drug delivery to intracellular sites of action can decrease drug effects: tricyclic antidepressants can blunt the antihypertensive effect of clonidine by decreasing its uptake into

adrenergic neurons. Reduced CNS penetration of multiple human immunodeficiency virus (HIV) protease inhibitors (with the attendant risk of facilitating viral replication in a sanctuary site) appears attributable to P-glycoprotein-mediated exclusion of the drug from the CNS; indeed, inhibition of P-glycoprotein has been proposed as a therapeutic approach to enhance drug entry to the CNS (Fig. 63-5).

■ PHARMACOKINETIC INTERACTIONS CAUSING INCREASED DRUG EFFECTS

The most common mechanism here is inhibition of drug elimination. In contrast to induction, new protein synthesis is not involved, and the effect develops as drug and any metabolites accumulate (a function of their elimination half-lives). Since shared substrates of a single enzyme can compete for access to the active site of the protein, many CYP substrates are also inhibitors. However, some drugs are especially potent as inhibitors (and occasionally may not even be substrates) of specific drug elimination pathways, and so it is in the use of these agents that clinicians must be most alert to the potential for interactions (Table 63-2). Commonly implicated interacting drugs of this type include amiodarone, cimetidine, erythromycin and some other macrolide antibiotics (clarithromycin but not azithromycin), ketoconazole and other azole antifungals, the antiretroviral agent ritonavir, and high concentrations of grapefruit juice. The consequences of such interactions will depend on the drug whose elimination is being inhibited (see "The Concept of High-Risk Pharmacokinetics," above). Examples include CYP3A inhibitors increasing the risk of cyclosporine toxicity or of rhabdomyolysis with some 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors (lovastatin, simvastatin, atorvastatin, but not pravastatin), and P-glycoprotein inhibitors increasing the risk of toxicity with digoxin therapy or of bleeding with the thrombin inhibitor dabigatran.

These interactions can occasionally be exploited to therapeutic benefit. The antiviral ritonavir is a very potent CYP3A4 inhibitor that has been added to anti-HIV regimens, not because of its antiviral effects but because it decreases clearance, and hence increases efficacy, of other anti-HIV agents. Similarly, calcium channel blockers have been deliberately coadministered with cyclosporine to reduce its clearance and thus its maintenance dosage and cost.

Phenytoin, an inducer of many systems, including CYP3A, inhibits CYP2C9 and thus can reduce the bioactivation of losartan, with potential loss of antihypertensive effect, or the elimination of S-warfarin, with attendant increased bleeding risk.

Grapefruit (but not orange) juice inhibits CYP3A, especially at high doses; patients receiving drugs where even modest CYP3A inhibition may increase the risk of adverse effects (e.g., cyclosporine, some HMG-CoA reductase inhibitors) should therefore avoid grapefruit juice.

CYP2D6 is markedly inhibited by quinidine, a number of neuroleptic drugs (chlorpromazine and haloperidol), and the selective serotonin reuptake inhibitors (SSRIs) fluoxetine and paroxetine. The clinical consequences of fluoxetine's interaction with CYP2D6 substrates may not be apparent for weeks after the drug is started, because of its very long half-life and slow generation of a CYP2D6-inhibiting metabolite.

Azathioprine is metabolized to 6-mercaptopurine, which is then metabolized by thiopurine methyltransferase and by xanthine oxidase. When allopurinol, an inhibitor of xanthine oxidase, is administered with standard doses of azathioprine or 6-mercaptopurine, life-threatening toxicity (bone marrow suppression) can result.

A number of drugs are secreted by the renal tubular transport systems for organic anions. Inhibition of these systems can cause excessive drug accumulation. Salicylate, for example, reduces the renal clearance of methotrexate, an interaction that may lead to methotrexate toxicity. Renal tubular secretion contributes substantially to the elimination of penicillin, which can be inhibited (to increase its therapeutic effect) by probenecid. Similarly, inhibition of tubular cation transport by cimetidine decreases the renal clearance of dofetilide.

■ DRUG INTERACTIONS NOT MEDIATED BY CHANGES IN DRUG DISPOSITION

Drugs may act on separate components of a common process to generate effects greater than either has alone. While antithrombotic

therapy with combinations of antiplatelet agents (glycoprotein IIb/IIIa inhibitors, aspirin, clopidogrel) and anticoagulants (e.g., warfarin, heparins, dabigatran, apixaban, rivaraxaban, edoxaban) is often used in the treatment of vascular disease, such combinations do carry an increased risk of bleeding.

Nonsteroidal anti-inflammatory drugs (NSAIDs) cause gastric ulcers, and in patients treated with oral anticoagulants, the risk of upper gastrointestinal bleeding is increased almost threefold by concomitant use of an NSAID.

Indomethacin, piroxicam, and probably other NSAIDs antagonize the antihypertensive effects of β -adrenergic receptor blockers, diuretics, ACE inhibitors, and other drugs. The resulting elevation in blood pressure ranges from trivial to severe. This effect is not seen with aspirin and sulindac but has been found with the cyclooxygenase 2 (COX-2) inhibitor celecoxib.

Torsades de pointes ventricular tachycardia during administration of QT-prolonging antiarrhythmics (quinidine, sotalol, dofetilide) occurs much more frequently in patients receiving diuretics, probably reflecting hypokalemia. Low potassium not only prolongs the QT interval in the absence of drug but also potentiates drug block of ion channels that results in QT prolongation. Also, some diuretics have direct electrophysiologic actions that prolong QT.

The administration of supplemental potassium leads to more frequent and more severe hyperkalemia when potassium elimination is reduced by concurrent treatment with ACE inhibitors, spironolactone, eplerenone, amiloride, or triamterene.

The pharmacologic effects of sildenafil result from inhibition of the phosphodiesterase type 5 isoform that inactivates cyclic guanosine monophosphate (GMP) in the vasculature. Nitroglycerin and related nitrates used to treat angina produce vasodilation by elevating cyclic GMP. Thus, coadministration of these nitrates with sildenafil can cause profound hypotension, which can be catastrophic in patients with coronary disease.

Sometimes, combining drugs can increase overall efficacy and/or reduce drug-specific toxicity. Such therapeutically useful interactions are described in chapters dealing with specific disease entities.

ADVERSE DRUG REACTIONS

The beneficial effects of drugs are coupled with the inescapable risk of untoward effects. The morbidity and mortality from these adverse effects often present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. As well, 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.

Adverse reactions 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. 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.

Drugs may increase the frequency of an event that is common in a general population, and this may be especially difficult to recognize; an excellent example is the increase in myocardial infarctions with the COX-2 inhibitor rofecoxib. Drugs can also cause rare and serious adverse effects, such as hematologic abnormalities, arrhythmias, severe skin reactions, or hepatic or renal dysfunction. Prior to regulatory approval and marketing (see below), 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 adverse effects 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 adverse events.

Elucidating mechanisms underlying adverse drug effects 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 adverse reactions 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 reporting of a newly recognized adverse reaction 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 adverse drug effects. Similarly, patients commit errors in taking OTC drugs by not reading or following the 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 adverse drug reactions, and that 2.3% of these patients (0.15%) died as a result. The most common culprit drugs were aspirin, other NSAIDs, diuretics, warfarin, ACE inhibitors, antidepressants, opiates, digoxin, steroids, and clopidogrel. One study in the late 1990s suggested that adverse drug reactions were responsible for >100,000 in-hospital deaths in the United States, making them the 4th to 6th commonest 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 adverse drug reactions. When <6 different drugs are given to hospitalized patients, the probability of an adverse reaction is ~5%, but if >15 drugs are given, the probability is >40%. Serious adverse reactions 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. 333). 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 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. 98) is due to an immune-based drug reaction.

Serum sickness (Chap. 345) 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 Steven-Johnson Syndrome/Toxic Epidermal Necrolysis (SJS/TEN), which can result in death due to T-cell-mediated massive skin sloughing. As described in Chap. 64, specific genetic variants appear necessary but not sufficient to elicit SJS/TEN. The mechanism is thought to be T cell activation by hapten-“self-peptide” interactions or direct binding of drug to HLA or T cell receptors.

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 adverse reactions 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 adverse drug reaction 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 adverse effect 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 as 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 adverse drug reactions 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 adverse drug effects, 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. 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. Every physician should determine what drugs a patient has been taking, for the previous month or two ideally, before prescribing any medications. 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 adverse drug effects 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 adverse effect has occurred in the patient or a family member ([see Chap. 64](#)).

Once an adverse reaction 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 and if the attempt would not entail undue risk. With concentration-dependent adverse reactions, lowering the dosage may cause the reaction to disappear, and raising it may cause the reaction to reappear. When the reaction is thought to be immunologic, however, readministration of the drug may be hazardous, since anaphylaxis may develop.

If the patient is receiving many drugs when an adverse reaction 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 adverse effect to disappear depends on the time required for the concentration to fall below the range associated with the adverse effect; 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), dose finding (phase 2), and efficacy (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. 64](#), 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.

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.
- Electronic medical record 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 reactions.
- Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READING

- LANDRIGAN CP et al: Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 363:2124, 2010.
- LAZAROU J et al: Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* 279:1200, 1998.
- MACRAE CA et al: The future of cardiovascular therapeutics. *Circulation* 133:2610, 2016.
- PIRMOHAMED M et al: Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients. *Br Med J* 329:15, 2004.
- WHEATLEY LM et al: Report from the National Institute of Allergy and Infectious Diseases workshop on drug allergy. *J Allergy Clin Immunol* 136:262, 2015.

Types of Genetic Variants Influencing Drug Response

Table 64-1 The commonest 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*. 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 64-1 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 human leukocyte antigens, HLA-B) are currently best evaluated by sequencing.

Table 64-1 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. 64-1, 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.

Genome-Wide Association Studies The field has also had some success with “unbiased” approaches such as genome-wide association (GWA) (Chap. 456), particularly in identifying single variants associated with high risk for certain forms of drug toxicity. GWA studies have identified variants in the HLA-B 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. GWA approaches have also implicated interferon variants in antileukemic responses and in response to therapy in hepatitis C. African-American subjects are known to have higher dose requirements to achieve stable anticoagulation with warfarin, due in part to variation in *CYP2C9* and *VKORC1*, discussed below. In addition, a GWA study identified novel SNPs near *CYP2C9* that contribute to this effect in African Americans.

64

Pharmacogenomics

Dan M. Roden



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 outline areas of controversy and future work.

PRINCIPLES OF GENETIC VARIATION AND DRUG RESPONSE (SEE ALSO CHAPS. 456 AND 457)

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members (Chap. 457). 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 generally used to identify and validate DNA variants contributing to variable drug actions.

GENETIC VARIANTS AFFECTING PHARMACOKINETICS

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition (Table 64-2). A distinct multimodal distribution of drug disposition (as shown in Fig. 64-1) 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 *diplopotype*) can complicate the use of genotyping in clinical practice. Furthermore, some variants produce only partial loss of

TABLE 64-1 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 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*4 ^c	rs3892097	Splicing defect: Loss of function	23.1	11.9	0.4
	CYP2D6*10 ^c	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
Insertion/deletion	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
	CYP2D6*4xN	Duplication of loss of function allele		Up to 3% in North Africa and the Middle East		
				0.3	1.4	b

Note: Allele frequencies from <http://exac.broadinstitute.org/> and <https://cpicpgx.org/>. ^aIncludes heterozygotes and homozygotes. ^bAllele frequency <0.05%.

^cCYP2D6 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 <http://www.cypalleles.ki.se>.

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 (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (normal metabolizers, often termed *extensive metabolizers*, EMs). *Ultra-rapid metabolizers* (UMs) with especially high enzymatic activity (occasionally due to gene duplication; Table 64-1 and Fig. 64-1) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways (see Chap. 63, Table 63-1), 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.

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 non-synonymous 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, with 5–10% of European- and African-derived populations (but very few Asians) displaying the PM phenotype (Fig. 64-1). Dozens of loss-of-function variants in CYP2D6 have been described; the PM phenotype arises in individuals with two such alleles. In addition, ultra-rapid metabolizers 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. 64-1B), steady state concentrations are higher and the time taken to achieve steady state is longer than in EMs (see Chap. 63). 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. In the case of drugs with beta-blocking properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bronchospasm, bradycardia) are seen 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. Ultra-rapid metabolizers may require very high dosages of nortriptyline and other tricyclic antidepressants

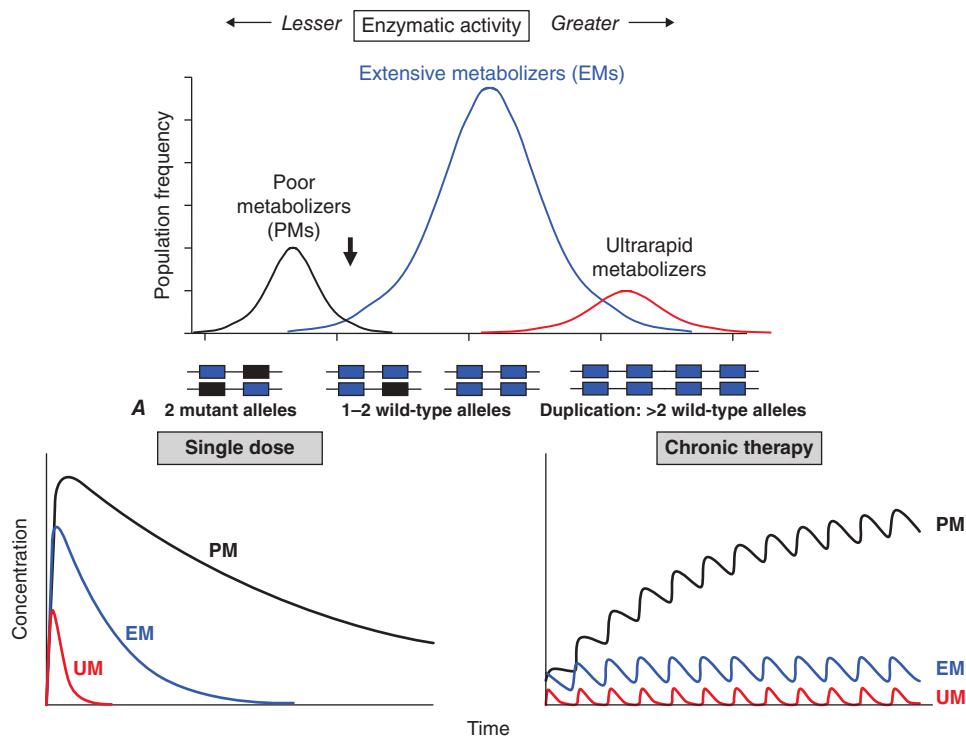


FIGURE 64-1 **A.** Distribution of CYP2D6 metabolic activity across a population. The heavy arrow indicates an antimode, separating poor metabolizer subjects (PMs, black), with two loss-of-function CYP2D6 alleles (black), indicated by the intron-exon structures below the chart. Individuals with one or 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).

TABLE 64-2 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
	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.

(Continued)

TABLE 64-2 Genetic Variants and Drug Responses (Continued)

GENE	DRUGS	EFFECT OF GENETIC VARIANTS ^a
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
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
<i>IFNL3</i> (IL28B)	interferon	Variable response in hepatitis C therapy
SLCO1B1	simvastatin	Encodes a drug uptake transporter; variant non-synonymous 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 non-synonymous 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 detectable
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

^aDrug effect in homozygotes unless otherwise specified. ^bMany tricyclic antidepressants and selective serotonin uptake inhibitors are metabolized by either CYP2D6, CYP2C19, or both, and some metabolites have pharmacologic activity. See <https://www.pharmgkb.org/view/dosing-guidelines.do>.

Note: 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); UM, ultra-rapid metabolizer (enzymatic activity much greater than normal, e.g., with gene duplication, Fig. 64-1). SJS/TEN: Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis.

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>

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.

CYP2C19 The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (2–3%) in other populations. 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 should improve therapy. CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and several large retrospective 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 reduction 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 controlling vitamin K metabolism such as CYP4F2. 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.

DYPD Individuals homozygous for loss of function alleles in dihydropyrimidine dehydrogenase, encoded by DYPD, are at high 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) bioactivates 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.

N-acetylation is catalyzed 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. 330**). 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 64-1), reduces transcriptional activity and warfarin dose requirement. Multiple polymorphisms identified in the β_2 -adrenergic receptor appear to be linked to specific phenotypes in asthma and congestive heart failure, diseases in which β_2 -receptor function might be expected to determine prognosis. 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. 409**). 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. 96**) 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. 241**), and in some patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

Immunologically Mediated Drug Reactions The Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are potentially fatal skin reactions now increasingly recognized to be linked to specific HLA alleles (see Table 64-2). Some cases of hepatotoxicity have also been linked to variants in this region. The frequency of risk alleles often varies by ancestry (Table 64-1). 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 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. 67**). 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). BCR-Abl1 is not only active but may be central to the pathogenesis of CML; use of imatinib and other BCR-Abl1 inhibitors has resulted in remarkable efficacy not only in CML but also in other BCR-Abl1-positive tumors such as gastrointestinal stromal tumors (see **Chap. 67**). Similarly, the anti-epidermal growth factor receptor (EGFR) antibodies cetuximab and panitumumab appear especially effective in colon cancers in which K-ras, a G protein in the EGFR pathway, is not mutated. 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.

Reactive versus Preemptive Approaches Two approaches to pharmacogenetic implementation have been put in place at both "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 then used to guide therapy with that specific drug. The alternative to this "reactive" 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 healthcare 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 64-1 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) has 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. CPIC does not, however, address the question of when or how such genetic testing should be undertaken.

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 non-genetic 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 was that while the use of carbamazepine dropped, it was often substituted by 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. These RCTs focused on time in therapeutic range in the first 4–12 weeks of treatment, and were not powered to examine outcomes such as recurrent thrombosis or bleeding. Retrospective analyses of bleeding cases vs non-bleeding controls in EHRs and administrative databases have suggested a role for CYP2C9*3 or the variants in V433M variant in CYP4F2 in mediating this risk.

While large retrospective analyses indicate that CYP2C19 loss of function variants decrease clopidogrel efficacy, RCTs are difficult to design: many argue that it is unethical to randomize individuals known to be homozygous for loss of function alleles, since administering clopidogrel is then tantamount to administering placebo. However, trials examining outcomes in only heterozygotes might require very large numbers of subjects.

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 reserve alternate more expensive therapies for subjects likely to have variant responses to warfarin or clopidogrel. As price drops and as experience grows with newer agents, it is likely that clopidogrel and warfarin will be largely supplanted.

GENETICS AND DRUG DEVELOPMENT

Genetic tools are now being increasingly used to identify or validate new drug targets. Initial studies in this field 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.

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, decreased incidence of coronary artery disease, and no deleterious consequences in other organ systems. These data triggered the development of PCSK9 antagonists which were marketed less than 10 years after the

initial population studies. Other targets implicated by similar population genetic studies include SLC30A8 for the prevention of type 2 diabetes and APOC3 for hypertriglyceridemia. In the latter examples, the identification of an apparently protective effect of rare loss-of-function alleles required very large datasets (>100,000) coupling DNA to longitudinal clinical information; long-term epidemiologic studies like the Framingham Heart Study or EHR systems are now being harnessed to address this opportunity.

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 are drugs that reverse immune system inhibition (**Chap. 69**). In some patients the release of this “break” 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 datasets is another approach that is being explored in drug development. For example, a GWA 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. 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 datasets. 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 electronic health record systems.
- Incorporating genetic approaches into drug development projects hold the promise of more rapid development of targeted, safe, and effective therapies.

FURTHER READING

- MALLAL S et al: HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 358:568, 2008.
- NELSON MR et al: The support of human genetic evidence for approved drug indications. *Nat Genet* 47:856, 2015.
- RELLING MV, EVANS WE: Pharmacogenomics in the clinic. *Nature* 526:343, 2015.
- WANG L et al: Genomics and drug response. *N Engl J Med* 364:1144, 2011.
- WEEKE P, RODEN DM: Applied pharmacogenomics in cardiovascular medicine. *Annu Rev Med* 65:81, 2014.

Section 1 Neoplastic Disorders

65

Approach to the Patient with Cancer

Dan L. Longo



The application of current treatment techniques (surgery, radiation therapy, chemotherapy, and biologic therapy) results in the cure of nearly two of three patients diagnosed with cancer. Nevertheless, patients experience the diagnosis of cancer as one of the most traumatic and revolutionary events that has ever happened to them. Independent of prognosis, the diagnosis brings with it a change in a person's self-image and in his or her role in the home and workplace. The prognosis of a person who has just been found to have pancreatic cancer is the same as the prognosis of the person with aortic stenosis who develops the first symptoms of congestive heart failure (median survival, ~8 months). However, the patient with heart disease may remain functional and maintain a self-image as a fully intact person with just a malfunctioning part, a diseased organ ("a bum ticker"). By contrast, the patient with pancreatic cancer has a completely altered self-image and is viewed differently by family and anyone who knows the diagnosis. He or she is being attacked and invaded by a disease that could be anywhere in the body. Every ache or pain takes on desperate significance. Cancer is an exception to the coordinated interaction among cells and organs. In general, the cells of a multicellular organism

are programmed for collaboration. Many diseases occur because the specialized cells fail to perform their assigned task. Cancer takes this malfunction one step further. Not only is there a failure of the cancer cell to maintain its specialized function, but it also strikes out on its own; the cancer cell competes to survive using natural mutability and natural selection to seek advantage over normal cells in a recapitulation of evolution. One consequence of the traitorous behavior of cancer cells is that the patient feels betrayed by his or her body. The cancer patient feels that he or she, and not just a body part, is diseased.

THE MAGNITUDE OF THE PROBLEM

No nationwide cancer registry exists; therefore, the incidence of cancer is estimated on the basis of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database, which tabulates cancer incidence and death figures from 13 sites, accounting for about 10% of the U.S. population, and from population data from the U.S. Census Bureau. In 2017, 1,688 million new cases of invasive cancer (836,150 men and 852,630 women) were diagnosed, and 600,920 persons (318,420 men and 282,500 women) died from cancer. The percent distribution of new cancer cases and cancer deaths by site for men and women is shown in **Table 65-1**. Cancer incidence has been declining by about 2% each year since 1992. Cancer is the cause of one in four deaths in the United States.

The most significant risk factor for cancer overall is age; two-thirds of all cases were in those aged >65 years. Cancer incidence increases as the third, fourth, or fifth power of age in different sites. For the interval between birth and age 49 years, 1 in 29 men and 1 in 19 women will develop cancer; for the interval between ages 50 and 59 years, 1 in 15 men and 1 in 17 women will develop cancer; for the interval between ages 60 and 69 years, 1 in 6 men and 1 in 10 women will develop cancer;

TABLE 65-1 Distribution of Cancer Incidence and Deaths for 2017

MALE			FEMALE		
SITES	%	NUMBER	SITES	%	NUMBER
Cancer Incidence					
Prostate	19	161,360	Breast	30	252,710
Lung	14	116,990	Lung	12	105,510
Colorectal	9	71,420	Colorectal	8	64,010
Bladder	7	60,490	Endometrial	6	61,380
Melanoma	6	52,170	Thyroid	5	42,470
Kidney	5	40,610	Melanoma	4	34,940
Lymphoma	5	40,080	Lymphoma	4	32,160
Leukemia	4	36,290	Leukemia	3	25,840
Oral Cavity	4	35,720	Pancreas	3	25,700
Liver	3	29,200	Kidney	3	23,380
All others	23	191,820	All others	22	184,530
All sites	100	836,150	All sites	100	852,630
Cancer Deaths					
Lung	27	84,590	Lung	25	71,280
Colorectal	9	27,150	Breast	14	40,610
Prostate	8	26,730	Colorectal	9	23,110
Pancreas	7	22,300	Pancreas	7	20,790
Liver	6	19,610	Ovary	5	14,080
Leukemia	4	14,300	Endometrial	4	10,920
Esophagus	4	12,720	Leukemia	4	10,200
Bladder	4	12,240	Liver	3	9310
Lymphoma	4	11,450	Lymphoma	3	8690
CNS	3	9620	CNS	3	7080
All others	24	77,710	All others	24	66,880
All sites	100	318,420	All sites	100	282,500

Source: From RL Siegel et al: Cancer statistics, 2017. CA Cancer J Clin 67:7, 2017.

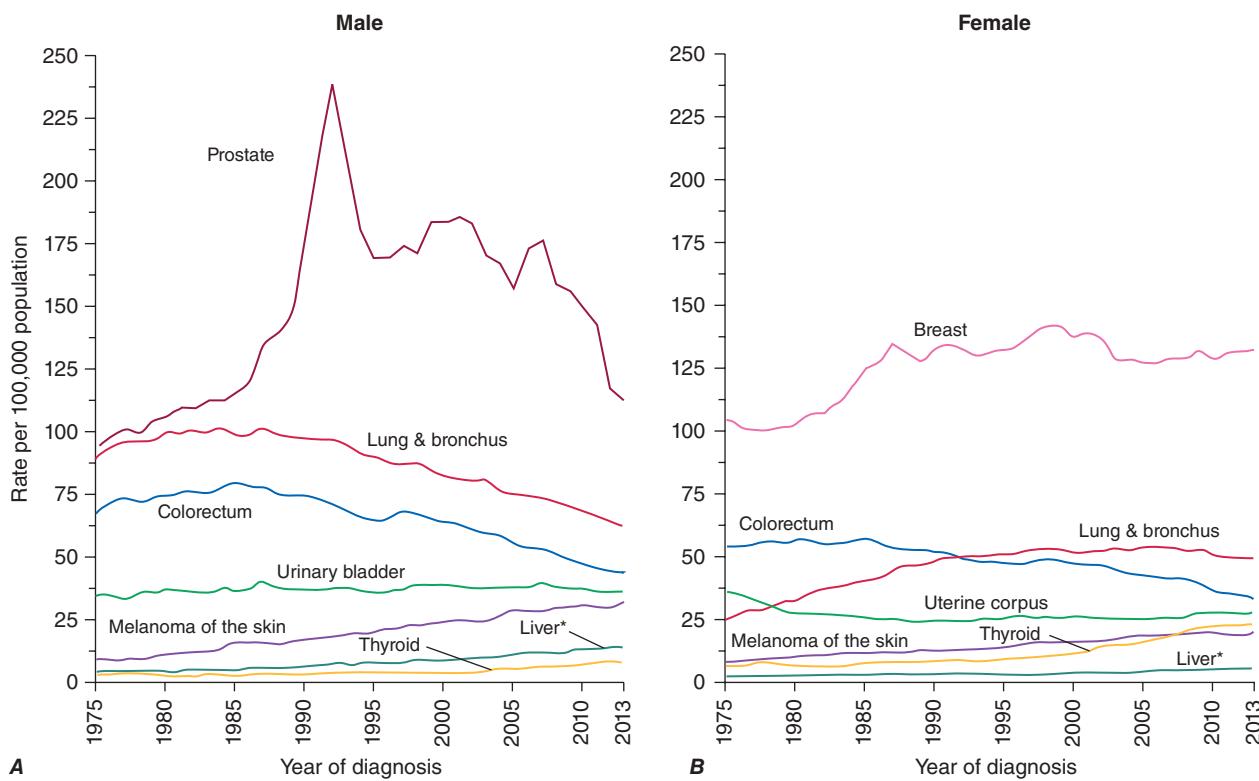


FIGURE 65-1 Incidence rates for particular types of cancer over the last 38 years in men (A) and women (B). (From RL Siegel et al: CA Cancer J Clin 67:7, 2017.)

and for people aged ≥ 70 , 1 in 3 men and 1 in 4 women will develop cancer. Overall, men have a 44% risk of developing cancer at some time during their lives; women have a 38% lifetime risk.

Cancer is the second leading cause of death behind heart disease. Deaths from heart disease have declined 45% in the United States since 1950 and continue to decline. Cancer has overtaken heart disease as the number one cause of death in persons aged <85 years. Incidence trends over time are shown in Fig. 65-1. After a 70-year period of increase, cancer deaths began to decline in 1990–1991 (Fig. 65-2). Between 1990 and 2010, cancer deaths decreased by 21% among men and 12.3% among women. The magnitude of the decline is illustrated in Fig. 65-3. The five leading causes of cancer deaths are shown for various populations in Table 65-2. The 5-year survival for white patients was 39% in 1960–1963 and 69% in 2003–2009. Cancers are more often deadly in blacks; the 5-year survival was 61% for the 2003–2009 interval; however, the racial differences are narrowing over time. Incidence and mortality vary among racial and ethnic groups (Table 65-3). The basis for these differences is unclear.

CANCER AROUND THE WORLD

 In 2008, 12.7 million new cancer cases and 7.6 million cancer deaths were estimated worldwide, according to estimates of GLOBOCAN 2008, developed by the International Agency for Research on Cancer (IARC). When broken down by region of the world, ~45% of cases were in Asia, 26% in Europe, 14.5% in North America, 7.1% in Central/South America, 6% in Africa, and 1% in Australia/New Zealand (Fig. 65-4). Lung cancer is the most common cancer and the most common cause of cancer death in the world. Its incidence is highly variable, affecting only 2 per 100,000 African women but as many as 61 per 100,000 North American men. Breast cancer is the second most common cancer worldwide; however, it ranks fifth as a cause of death behind lung, stomach, liver, and colorectal cancer. Among the eight most common forms of cancer, lung (2-fold), breast (3-fold), prostate (2.5-fold), and colorectal (3-fold) cancers are more common in more developed countries than in less developed countries. By contrast, liver (twofold), cervical (twofold), and esophageal (two- to threefold) cancers are more common in less developed countries. Stomach cancer incidence is similar in more and less developed countries but is much more common in Asia than North

America or Africa. The most common cancers in Africa are cervical, breast, and liver cancers. It has been estimated that nine modifiable risk factors are responsible for more than one-third of cancers worldwide. These include smoking, alcohol consumption, obesity, physical inactivity, low fruit and vegetable consumption, unsafe sex, air pollution, indoor smoke from household fuels, and contaminated injections.

PATIENT MANAGEMENT

Important information is obtained from every portion of the routine history and physical examination. The duration of symptoms may reveal the chronicity of disease. The past medical history may alert the physician to the presence of underlying diseases that may affect the choice of therapy or the side effects of treatment. The social history may reveal occupational exposure to carcinogens or habits, such as smoking or alcohol consumption, that may influence the course of disease and its treatment. The family history may suggest an underlying familial cancer predisposition and point out the need to begin surveillance or other preventive therapy for unaffected siblings of the patient. The review of systems may suggest early symptoms of metastatic disease or a paraneoplastic syndrome.

DIAGNOSIS

The diagnosis of cancer relies most heavily on invasive tissue biopsy and should never be made without obtaining tissue; no noninvasive diagnostic test is sufficient to define a disease process as cancer. Although in rare clinical settings (e.g., thyroid nodules), fine-needle aspiration is an acceptable diagnostic procedure, the diagnosis generally depends on obtaining adequate tissue to permit careful evaluation of the histology of the tumor, its grade, and its invasiveness and to yield further molecular diagnostic information, such as the expression of cell-surface markers or intracellular proteins that typify a particular cancer, or the presence of a molecular marker, such as the t(8;14) translocation of Burkitt's lymphoma. Increasing evidence links the expression of certain genes with the prognosis and response to therapy (Chaps. 67 and 68).

Occasionally, a patient will present with a metastatic disease process that is defined as cancer on biopsy but has no apparent primary site of disease. Efforts should be made to define the primary site based on age,

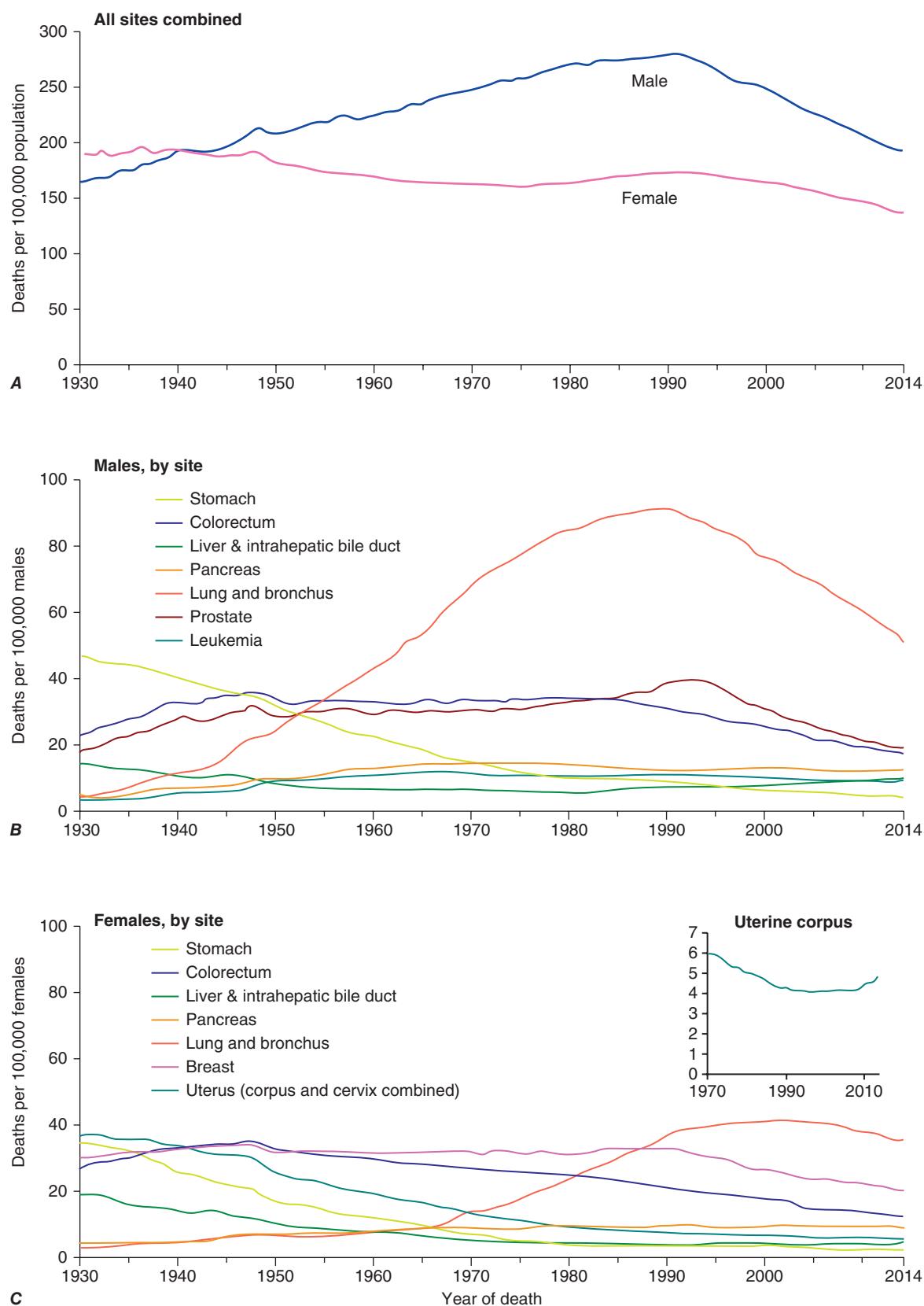


FIGURE 65-2 Eighty-five-year trend in cancer death rates for (A) women and (B) men by site in the United States, 1930–2014. Rates are per 100,000 age-adjusted to the 2000 U.S. standard population. All sites combined (A), individual sites in men (B) and individual sites in women (C) are shown. (From RL Siegel et al: CA Cancer J Clin 67:7, 2017.)

sex, sites of involvement, histology and tumor markers, and personal and family history. Particular attention should be focused on ruling out the most treatable causes (Chap. 88).

Once the diagnosis of cancer is made, the management of the patient is best undertaken as a multidisciplinary collaboration among

the primary care physician, medical oncologists, surgical oncologists, radiation oncologists, oncology nurse specialists, pharmacists, social workers, rehabilitation medicine specialists, and a number of other consulting professionals working closely with each other and with the patient and family.

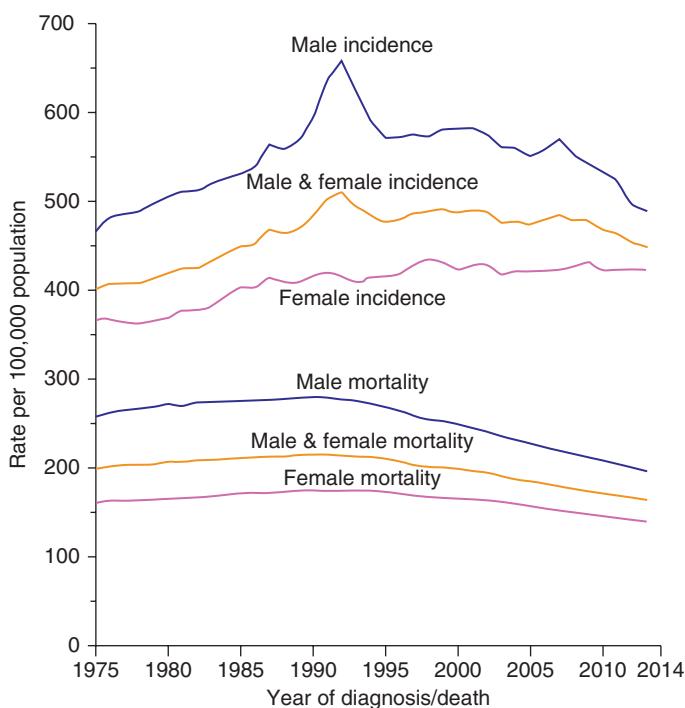


FIGURE 65-3 Trends in cancer incidence and death rates for men and women from 1975 to 2014. (From RL Siegel et al: CA Cancer J Clin 67:7, 2017.)

DEFINING THE EXTENT OF DISEASE AND THE PROGNOSIS

The first priority in patient management after the diagnosis of cancer is established and shared with the patient is to determine the extent of disease. The curability of a tumor usually is inversely proportional to the tumor burden. Ideally, the tumor will be diagnosed before symptoms develop or as a consequence of screening efforts (**Chap. 66**). A very high proportion of such patients can be cured. However, most patients with cancer present with symptoms related to the cancer, caused either by mass effects of the tumor or by alterations associated with the production of cytokines or hormones by the tumor.

For most cancers, the extent of disease is evaluated by a variety of noninvasive and invasive diagnostic tests and procedures. This process is called *staging*. There are two types. *Clinical staging* is based on physical examination, radiographs, isotopic scans, computed tomography (CT) scans, and other imaging procedures; *pathologic staging* takes into account information obtained during a surgical procedure, which might include intraoperative palpation, resection of regional lymph nodes and/or tissue adjacent to the tumor, and inspection and biopsy of organs commonly involved in disease spread. Pathologic staging includes histologic examination of all tissues removed during

the surgical procedure. Surgical procedures performed may include a simple lymph node biopsy or more extensive procedures such as thoracotomy, mediastinoscopy, or laparotomy. Surgical staging may occur in a separate procedure or may be done at the time of definitive surgical resection of the primary tumor.

Knowledge of the predilection of particular tumors for spreading to adjacent or distant organs helps direct the staging evaluation.

Information obtained from staging is used to define the extent of disease as localized, as exhibiting spread outside of the organ of origin to regional but not distant sites, or as metastatic to distant sites. The most widely used system of staging is the tumor, node, metastasis (TNM) system codified by the International Union Against Cancer and the American Joint Committee on Cancer. The TNM classification is an anatomically based system that categorizes the tumor on the basis of the size of the primary tumor lesion (T1–4, where a higher number indicates a tumor of larger size), the presence of nodal involvement (usually N0 and N1 for the absence and presence, respectively, of involved nodes, although some tumors have more elaborate systems of nodal grading), and the presence of metastatic disease (M0 and M1 for the absence and presence, respectively, of metastases). The various permutations of T, N, and M scores (sometimes including tumor histologic grade [G]) are then broken into stages, usually designated by the Roman numerals I through IV. Tumor burden increases and curability decreases with increasing stage. Other anatomic staging systems are used for some tumors, e.g., the Dukes classification for colorectal cancers, the International Federation of Gynecologists and Obstetricians classification for gynecologic cancers, and the Ann Arbor classification for Hodgkin's disease.

Certain tumors cannot be grouped on the basis of anatomic considerations. For example, hematopoietic tumors such as leukemia, myeloma, and lymphoma are often disseminated at presentation and do not spread like solid tumors. For these tumors, other prognostic factors have been identified (**Chaps. 101–107**).

In addition to tumor burden, a second major determinant of treatment outcome is the physiologic reserve of the patient. Patients who are bedridden before developing cancer are likely to fare worse, stage for stage, than fully active patients. Physiologic reserve is a determinant of how a patient is likely to cope with the physiologic stresses imposed by the cancer and its treatment. This factor is difficult to assess directly. Instead, surrogate markers for physiologic reserve are used, such as the patient's age or Karnofsky performance status (**Table 65-4**) or Eastern Cooperative Oncology Group (ECOG) performance status (**Table 65-5**). Older patients and those with a Karnofsky performance status <70 or ECOG performance status ≥3 have a poor prognosis unless the poor performance is a reversible consequence of the tumor.

Increasingly, biologic features of the tumor are being related to prognosis. The expression of particular oncogenes, drug-resistance genes, apoptosis-related genes, and genes involved in metastasis is being found to influence response to therapy and prognosis. The presence of selected cytogenetic abnormalities may influence survival. Tumors

TABLE 65-2 The Five Leading Primary Tumor Sites for Patients Dying of Cancer Based on Age and Sex in 2017

RANK	SEX	ALL AGES	AGE, YEARS				
			UNDER 20	20–39	40–59	60–79	>80
1	M	Lung	CNS	CNS	Lung	Lung	Lung
	F	Lung	CNS	Breast	Lung	Lung	Lung
2	M	Prostate	Leukemia	Leukemia	Colorectal	Colorectal	Prostate
	F	Breast	Leukemia	Cervix	Breast	Breast	Breast
3	M	Colorectal	Bone sarcoma	Colorectal	Liver	Prostate	Colorectal
	F	Colorectal	Bone sarcoma	Colorectal	Colorectal	Colorectal	Colorectal
4	M	Pancreas	Soft tissue sarcoma	Lymphoma	Pancreas	Pancreas	Bladder
	F	Pancreas	Soft tissue sarcoma	Leukemial	Ovary	Pancreas	Pancreas
5	M	Liver	Lymphoma	Lung	Esophagus	Liver	Pancreas
	F	Ovary	Lymphoma	CNS	Pancreas	Ovary	Leukemia

Abbreviations: CNS, central nervous system; F, female; M, male.

Source: From RL Siegel et al: Cancer statistics, 2017. CA Cancer J Clin 67:7, 2017.

TABLE 65-3 Cancer Incidence and Mortality in Racial and Ethnic Groups, United States, 2009–2013

SITE	SEX	WHITE	BLACK	ASIAN/PACIFIC ISLANDER	AMERICAN INDIAN ^a	HISPANIC
Incidence per 100,000 Population						
All	M	519.3	577.3	310.2	426.7	498.1
	F	436.0	408.5	287.1	387.3	329.6
Breast		128.3	125.1	89.3	98.1	91.7
Colorectal	M	46.1	58.3	37.8	51.4	42.8
	F	35.2	42.7	27.8	41.2	29.8
Kidney	M	21.9	24.4	10.8	29.9	20.7
	F	11.3	13.0	4.8	17.6	11.9
Liver	M	9.7	16.9	20.4	18.5	19.4
	F	3.3	5.0	7.6	8.9	7.5
Lung	M	77.7	90.8	46.6	71.3	42.2
	F	58.2	51.0	28.3	56.2	25.6
Prostate		114.8	198.4	63.5	85.1	104.9
Cervix		7.0	9.8	6.1	9.7	9.9
Deaths per 100,000 Population						
All	M	204.0	253.4	122.7	183.6	142.5
	F	145.5	165.9	88.8	129.1	97.7
Breast		21.1	30.0	11.3	14.1	14.4
Colorectal	M	17.3	25.9	12.4	19.5	15.0
	F	12.3	16.9	8.8	14.0	9.2
Kidney	M	5.8	5.7	2.7	8.9	4.9
	F	2.5	2.5	1.1	4.2	2.3
Liver	M	8.0	13.3	14.3	14.9	13.1
	F	3.3	4.6	6.1	6.8	5.8
Lung	M	58.3	69.8	31.7	46.2	27.3
	F	39.8	35.5	18.0	30.8	13.4
Prostate		20.0	42.8	8.8	19.4	16.5
Cervix		2.3	3.9	1.7	2.8	2.6

^aBased on Indian Health Service delivery areas.

Abbreviations: F, female; M, male.

Source: From RL Siegel R et al: Cancer statistics, 2017. CA Cancer J Clin 67:7, 2017.

with higher growth fractions, as assessed by expression of proliferation-related markers such as proliferating cell nuclear antigen, behave more aggressively than tumors with lower growth fractions. Information obtained from studying the tumor itself will increasingly be used to influence treatment decisions. Host genes involved in drug metabolism can influence the safety and efficacy of particular treatments.

Enormous heterogeneity has been noted by studying tumors; we have learned that morphology is not capable of discerning certain distinct subsets of patients whose tumors have different sets of abnormalities. Tumors that look the same by light microscopy can be very different. Similarly, tumors that look quite different from one another histologically can share genetic lesions that predict responses to treatments. Furthermore, tumor cells vary enormously within a single patient even though the cells share a common origin.

MAKING A TREATMENT PLAN

From information on the extent of disease and the prognosis and in conjunction with the patient's wishes, it is determined whether the treatment approach should be curative or palliative in intent. Cooperation among the various professionals involved in cancer treatment is of the utmost importance in treatment planning. For some cancers, chemotherapy or chemotherapy plus radiation therapy delivered before the use of definitive surgical treatment (so-called neoadjuvant therapy) may improve the outcome, as seems to be the case for locally advanced breast cancer and head and neck cancers. In certain settings in which combined-modality therapy is intended, coordination among the medical oncologist, radiation oncologist, and surgeon is crucial to achieving optimal results. Sometimes the chemotherapy and radiation therapy need to be delivered sequentially, and other times concurrently.

Surgical procedures may precede or follow other treatment approaches. It is best for the treatment plan either to follow a standard protocol precisely or else to be part of an ongoing clinical research protocol evaluating new treatments. Ad hoc modifications of standard protocols are likely to compromise treatment results.

The choice of treatment approaches was formerly dominated by the local culture in both the university and the practice settings. However, it is now possible to gain access electronically to standard treatment protocols and to every approved clinical research study in North America through a personal computer interface with the Internet.¹

The skilled physician also has much to offer the patient for whom curative therapy is no longer an option. Often a combination of guilt and frustration over the inability to cure the patient and the pressure of a busy schedule greatly limit the time a physician spends with a patient who is receiving only palliative care. Resist these forces. In addition to the medicines administered to alleviate symptoms (see below), it is important to remember the comfort that is provided by holding the patient's hand, continuing regular examinations, and taking time to talk.

¹The National Cancer Institute maintains a database called PDQ (Physician Data Query) that is accessible on the Internet under the name CancerNet at www.cancer.gov/cancertopics/pdq/cancerdatabase. Information can be obtained through a facsimile machine using CancerFax by dialing 301-402-5874. Patient information is also provided by the National Cancer Institute in at least three formats: on the Internet via CancerNet at www.cancer.gov, through the CancerFax number listed above, or by calling 1-800-4-CANCER. The quality control for the information provided through these services is rigorous.

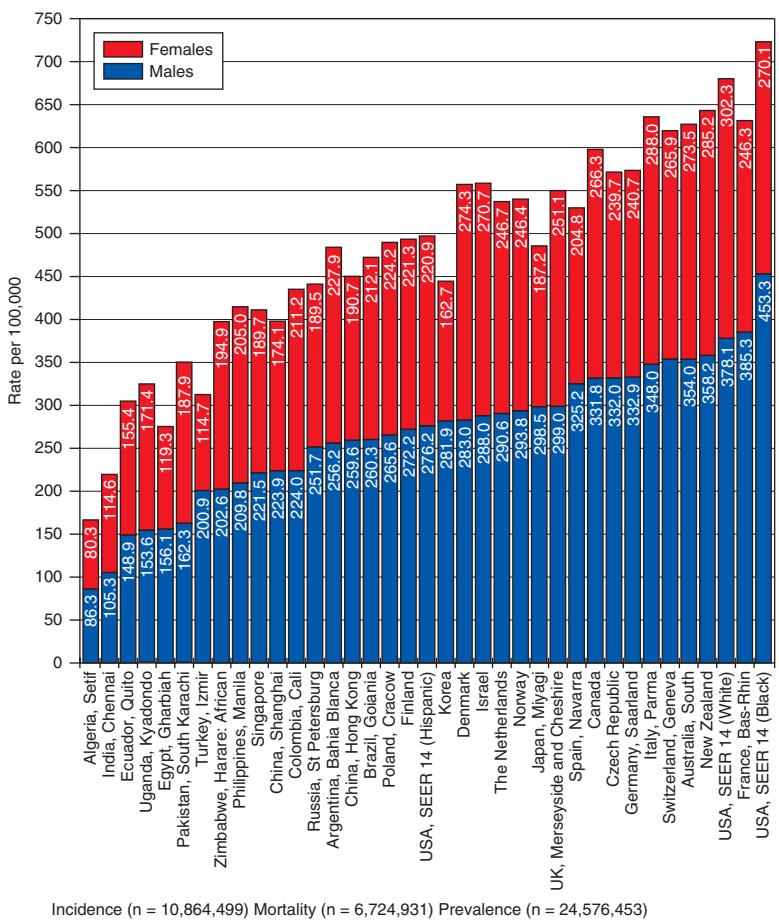


FIGURE 65-4 Worldwide overall annual cancer incidence, mortality, and 5-year prevalence for the period of 1993–2001. (Adapted from A Jemal et al: *Cancer Epidemiol Biomarkers Prev* 19:1893, 2010.)

■ MANAGEMENT OF DISEASE AND TREATMENT COMPLICATIONS

Because cancer therapies are toxic (Chap. 69), patient management involves addressing complications of both the disease and its treatment as well as the complex psychosocial problems associated with cancer. In the short term during a course of curative therapy, the patient's functional status may decline. Treatment-induced toxicity is less acceptable if the goal of therapy is palliation. The most common

side effects of treatment are nausea and vomiting (see below), febrile neutropenia (Chap. 70), and myelosuppression (Chap. 69). Tools are now available to minimize the acute toxicity of cancer treatment.

New symptoms developing in the course of cancer treatment should always be assumed to be reversible until proven otherwise. The fatalistic attribution of anorexia, weight loss, and jaundice to recurrent or progressive tumor could result in a patient dying from a reversible intercurrent cholecystitis. Intestinal obstruction may be due to reversible adhesions rather than progressive tumor. Systemic infections, sometimes with unusual pathogens, may be a consequence of the immunosuppression associated with cancer therapy. Some drugs used to treat cancer or its complications (e.g., nausea) may produce central nervous system symptoms that look like metastatic disease or may mimic paraneoplastic syndromes such as the syndrome of inappropriate antidiuretic hormone. A definitive diagnosis should be pursued and may even require a repeat biopsy.

A critical component of cancer management is assessing the response to treatment. In addition to a careful physical examination in which all sites of disease are physically measured and recorded in a flow chart by date, response assessment usually requires periodic repeating of imaging tests that were abnormal at the time of staging. If imaging tests have become normal, repeat biopsy of previously involved tissue is performed to document complete response by pathologic criteria. Biopsies are not usually required if there is macroscopic residual disease. A *complete response* is defined as disappearance of all evidence of disease, and a *partial response* as >50% reduction in the sum of the products of the perpendicular diameters of all measurable lesions. The determination of partial response may also be based on a 30% decrease in the sums of the longest diameters of lesions (Response Evaluation Criteria in Solid Tumors [RECIST]). *Progressive disease* is defined as the appearance of any new lesion or an increase of >25% in the sum of the products of the perpendicular diameters of all measurable lesions (or an increase of 20% in the sums of the longest diameters by RECIST). Tumor shrinkage or growth that does not meet any of these criteria is considered *stable disease*. Some sites of involvement (e.g., bone) or patterns of involvement (e.g., lymphangitic lung or diffuse pulmonary infiltrates) are considered unmeasurable. No response is complete without biopsy documentation of their resolution, but partial responses may exclude their assessment unless clear objective progression has occurred.

For some hematologic neoplasms, flow cytometric and genetic assays may determine the presence of residual tumor cells that escape microscopic detection. In general, these techniques can reliably detect as few as 1 tumor cell among 10,000 cells. If such tests do not detect tumor cells, the patient is said to have minimal residual disease negativity, a finding generally associated with more durable remissions. Accumulating data are defining interventions in patients with minimal

TABLE 65-4 Karnofsky Performance Index

PERFORMANCE STATUS	FUNCTIONAL CAPABILITY OF THE PATIENT
100	Normal; no complaints; no evidence of disease
90	Able to carry on normal activity; minor signs or symptoms of disease
80	Normal activity with effort; some signs or symptoms of disease
70	Cares for self; unable to carry on normal activity or do active work
60	Requires occasional assistance but is able to care for most needs
50	Requires considerable assistance and frequent medical care
40	Disabled; requires special care and assistance
30	Severely disabled; hospitalization is indicated, although death is not imminent
20	Very sick; hospitalization is necessary; active supportive treatment is necessary
10	Moribund, fatal processes progressing rapidly
0	Dead

TABLE 65-5 The Eastern Cooperative Oncology Group (ECOG) Performance Scale

ECOG Grade 0: Fully active, able to carry on all predisease performance without restriction
ECOG Grade 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work
ECOG Grade 2: Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about >50% of waking hours
ECOG Grade 3: Capable of only limited self-care, confined to bed or chair >50% of waking hours
ECOG Grade 4: Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
ECOG Grade 5: Dead

Source: From MM Oken et al: *Am J Clin Oncol* 5:649, 1982.

residual disease positivity that can extend remission duration and survival.

Tumor markers may be useful in patient management in certain tumors. Response to therapy may be difficult to gauge with certainty. However, some tumors produce or elicit the production of markers that can be measured in the serum or urine, and in a particular patient, rising and falling levels of the marker are usually associated with increasing or decreasing tumor burden, respectively. Some clinically useful tumor markers are shown in **Table 65-6**. Tumor markers are not in themselves specific enough to permit a diagnosis of malignancy to be made, but once a malignancy has been diagnosed and shown to be associated with elevated levels of a tumor marker, the marker can be used to assess response to treatment.

The recognition and treatment of depression are important components of management. The incidence of depression in cancer patients is ~25% overall and may be greater in patients with greater debility. This diagnosis is likely in a patient with a depressed mood (dysphoria) and/or a loss of interest in pleasure (anhedonia) for at least 2 weeks. In addition, three or more of the following symptoms are usually present: appetite change, sleep problems, psychomotor retardation or agitation, fatigue, feelings of guilt or worthlessness, inability to concentrate, and suicidal ideation. Patients with these symptoms should receive therapy. Medical therapy with a serotonin reuptake inhibitor such as fluoxetine (10–20 mg/d), sertraline (50–150 mg/d), or paroxetine (10–20 mg/d) or a tricyclic antidepressant such as amitriptyline (50–100 mg/d) or desipramine (75–150 mg/d) should be tried, allowing 4–6 weeks for response. Effective therapy should be continued at least 6 months after

TABLE 65-6 Tumor Markers

TUMOR MARKERS	CANCER	NONNEOPLASTIC CONDITIONS
Hormones		
Human chorionic gonadotropin	Gestational trophoblastic disease, gonadal germ cell tumor	Pregnancy
Calcitonin	Medullary cancer of the thyroid	
Catecholamines	Pheochromocytoma	
Oncofetal Antigens		
α Fetoprotein	Hepatocellular carcinoma, gonadal germ cell tumor	Cirrhosis, hepatitis
Carcinoembryonic antigen	Adenocarcinomas of the colon, pancreas, lung, breast, ovary	Pancreatitis, hepatitis, inflammatory bowel disease, smoking
Enzymes		
Prostatic acid phosphatase	Prostate cancer	Prostatitis, prostatic hypertrophy
Neuron-specific enolase	Small-cell cancer of the lung, neuroblastoma	
Lactate dehydrogenase	Lymphoma, Ewing's sarcoma	Hepatitis, hemolytic anemia, many others
Tumor-Associated Proteins		
Prostate-specific antigen	Prostate cancer	Prostatitis, prostatic hypertrophy
Monoclonal immunoglobulin	Myeloma	Infection, MGUS
CA-125	Ovarian cancer, some lymphomas	Menstruation, peritonitis, pregnancy
CA 19-9	Colon, pancreatic, breast cancer	Pancreatitis, ulcerative colitis
CD30	Hodgkin's disease, anaplastic large-cell lymphoma	—
CD25	Hairy cell leukemia, adult T-cell leukemia/lymphoma	—

Abbreviation: MGUS, monoclonal gammopathy of uncertain significance.

resolution of symptoms. If therapy is unsuccessful, other classes of antidepressants may be used. In addition to medication, psychosocial interventions such as support groups, psychotherapy, and guided imagery may be of benefit.

Many patients opt for unproven or unsound approaches to treatment when it appears that conventional medicine is unlikely to be curative. Those seeking such alternatives are often well educated and may be early in the course of their disease. Unsound approaches are usually hawked on the basis of unsubstantiated anecdotes and not only cannot help the patient but may be harmful. Physicians should strive to keep communications open and nonjudgmental, so that patients are more likely to discuss with the physician what they are actually doing. The appearance of unexpected toxicity may be an indication that a supplemental therapy is being taken.²

■ LONG-TERM FOLLOW-UP/LATE COMPLICATIONS

At the completion of treatment, sites originally involved with tumor are reassessed, usually by radiography or imaging techniques, and any persistent abnormality is biopsied. If disease persists, the multidisciplinary team discusses a new salvage treatment plan. If the patient has been rendered disease-free by the original treatment, the patient is followed regularly for disease recurrence. The optimal guidelines for follow-up care are not known. For many years, a routine practice has been to follow the patient monthly for 6–12 months, then every other month for a year, every 3 months for a year, every 4 months for a year, every 6 months for a year, and then annually. At each visit, a battery of laboratory and radiographic and imaging tests were obtained on the assumption that it is best to detect recurrent disease before it becomes symptomatic. However, where follow-up procedures have been examined, this assumption has been found to be untrue. Studies of breast cancer, melanoma, lung cancer, colon cancer, and lymphoma have all failed to support the notion that asymptomatic relapses are more readily cured by salvage therapy than symptomatic relapses. In view of the enormous cost of a full battery of diagnostic tests and their manifest lack of impact on survival, new guidelines are emerging for less frequent follow-up visits, during which the history and physical examination are the major investigations performed.

As time passes, the likelihood of recurrence of the primary cancer diminishes. For many types of cancer, survival for 5 years without recurrence is tantamount to cure. However, important medical problems can occur in patients treated for cancer and must be examined (**Chap. 91**). Some problems emerge as a consequence of the disease and some as a consequence of the treatment. An understanding of these disease- and treatment-related problems may help in their detection and management.

Despite these concerns, most patients who are cured of cancer return to normal lives.

■ SUPPORTIVE CARE

In many ways, the success of cancer therapy depends on the success of the supportive care. Failure to control the symptoms of cancer and its treatment may lead patients to abandon curative therapy. Of equal importance, supportive care is a major determinant of quality of life. Even when life cannot be prolonged, the physician must strive to preserve its quality. Quality-of-life measurements have become common endpoints of clinical research studies. Furthermore, palliative care has been shown to be cost-effective when approached in an organized fashion. A credo for oncology could be to cure sometimes, to extend life often, and to comfort always.

Pain Pain occurs with variable frequency in the cancer patient: 25–50% of patients present with pain at diagnosis, 33% have pain associated with treatment, and 75% have pain with progressive disease. The pain may have several causes. In ~70% of cases, pain is caused by the tumor itself—by invasion of bone, nerves, blood vessels, or

²Information about unsound methods may be obtained from the National Council Against Health Fraud, Box 1276, Loma Linda, CA 92354, or from the Center for Medical Consumers and Health Care Information, 237 Thompson Street, New York, NY 10012.

mucous membranes or obstruction of a hollow viscus or duct. In ~20% of cases, pain is related to a surgical or invasive medical procedure, to radiation injury (mucositis, enteritis, or plexus, or spinal cord injury), or to chemotherapy injury (mucositis, peripheral neuropathy, phlebitis, steroid-induced aseptic necrosis of the femoral head). In 10% of cases, pain is unrelated to cancer or its treatment.

Assessment of pain requires the methodical investigation of the history of the pain, its location, character, temporal features, provocative and palliative factors, and intensity (**Chap. 10**); a review of the oncologic history and past medical history as well as personal and social history; and a thorough physical examination. The patient should be given a 10-division visual analogue scale on which to indicate the severity of the pain. The clinical condition is often dynamic, making it necessary to reassess the patient frequently. Pain therapy should not be withheld while the cause of pain is being sought.

A variety of tools are available with which to address cancer pain. About 85% of patients will have pain relief from pharmacologic intervention. However, other modalities, including antitumor therapy (such as surgical relief of obstruction, radiation therapy, and strontium-89 or samarium-153 treatment for bone pain), neurostimulatory techniques, regional analgesia, or neuroablative procedures, are effective in an additional 12% or so. Thus, very few patients will have inadequate pain relief if appropriate measures are taken. **A specific approach to pain relief is detailed in Chap. 9.**

Nausea Emesis in the cancer patient is usually caused by chemotherapy (**Chap. 69**). Its severity can be predicted from the drugs used to treat the cancer. Three forms of emesis are recognized on the basis of their timing with regard to the noxious insult. *Acute emesis*, the most common variety, occurs within 24 h of treatment. *Delayed emesis* occurs 1–7 days after treatment; it is rare, but, when present, usually follows cisplatin administration. *Anticipatory emesis* occurs before administration of chemotherapy and represents a conditioned response to visual and olfactory stimuli previously associated with chemotherapy delivery.

Acute emesis is the best understood form. Stimuli that activate signals in the chemoreceptor trigger zone in the medulla, the cerebral cortex, and peripherally in the intestinal tract lead to stimulation of the vomiting center in the medulla, the motor center responsible for coordinating the secretory and muscle contraction activity that leads to emesis. Diverse receptor types participate in the process, including dopamine, serotonin, histamine, opioid, and acetylcholine receptors. The serotonin receptor antagonists ondansetron and granisetron are effective drugs against highly emetogenic agents, as are neurokinin receptor antagonists like aprepitant and fosaprepitant (see **Chap. 69**).

As with the analgesia ladder, emesis therapy should be tailored to the situation. For mildly and moderately emetogenic agents, prochlorperazine, 5–10 mg PO or 25 mg PR, is effective. Its efficacy may be enhanced by administering the drug before the chemotherapy is delivered. Dexamethasone, 10–20 mg IV, is also effective and may enhance the efficacy of prochlorperazine. For highly emetogenic agents such as cisplatin, mechlorethamine, dacarbazine, and streptozocin, combinations of agents work best and administration should begin 6–24 h before treatment. Ondansetron, 8 mg PO every 6 h the day before therapy and IV on the day of therapy, plus dexamethasone, 20 mg IV before treatment, is an effective regimen. Addition of oral aprepitant (a substance P/neurokinin 1 receptor antagonist) to this regimen (125 mg on day 1, 80 mg on days 2 and 3) further decreases the risk of both acute and delayed vomiting. Like pain, emesis is easier to prevent than to alleviate.

Delayed emesis may be related to bowel inflammation from the therapy and can be controlled with oral dexamethasone and oral metoclopramide, a dopamine receptor antagonist that also blocks serotonin receptors at high dosages. The best strategy for preventing anticipatory emesis is to control emesis in the early cycles of therapy to prevent the conditioning from taking place. If this is unsuccessful, prophylactic antiemetics the day before treatment may help. Experimental studies are evaluating behavior modification.

Effusions Fluid may accumulate abnormally in the pleural cavity, pericardium, or peritoneum. Asymptomatic malignant effusions may not require treatment. Symptomatic effusions occurring in tumors responsive to systemic therapy usually do not require local treatment but respond to the treatment for the underlying tumor. Symptomatic effusions occurring in tumors unresponsive to systemic therapy may require local treatment in patients with a life expectancy of at least 6 months.

Pleural effusions due to tumors may or may not contain malignant cells. Lung cancer, breast cancer, and lymphomas account for ~75% of malignant pleural effusions. Their exudative nature is usually gauged by an effusion/serum protein ratio of ≥ 0.5 or an effusion/serum lactate dehydrogenase ratio of ≥ 0.6 . When the condition is symptomatic, thoracentesis is usually performed first. In most cases, symptomatic improvement occurs for <1 month. Chest tube drainage is required if symptoms recur within 2 weeks. Fluid is aspirated until the flow rate is <100 mL in 24 h. Then either 60 units of bleomycin or 1 g of doxycycline is infused into the chest tube in 50 mL of 5% dextrose in water; the tube is clamped; the patient is rotated on four sides, spending 15 min in each position; and, after 1–2 h, the tube is again attached to suction for another 24 h. The tube is then disconnected from suction and allowed to drain by gravity. If <100 mL drains over the next 24 h, the chest tube is pulled, and a radiograph is taken 24 h later. If the chest tube continues to drain fluid at an unacceptably high rate, sclerosis can be repeated. Bleomycin may be somewhat more effective than doxycycline but is very expensive. Doxycycline is usually the drug of first choice. If neither doxycycline nor bleomycin is effective, talc can be used.

Symptomatic pericardial effusions are usually treated by creating a pericardial window or by stripping the pericardium. If the patient's condition does not permit a surgical procedure, sclerosis can be attempted with doxycycline and/or bleomycin.

Malignant ascites is usually treated with repeated paracentesis of small volumes of fluid. If the underlying malignancy is unresponsive to systemic therapy, peritoneovenous shunts may be inserted. Despite the fear of disseminating tumor cells into the circulation, widespread metastases are an unusual complication. The major complications are occlusion, leakage, and fluid overload. Patients with severe liver disease may develop disseminated intravascular coagulation.

Nutrition Cancer and its treatment may lead to a decrease in nutrient intake of sufficient magnitude to cause weight loss and alteration of intermediary metabolism. The prevalence of this problem is difficult to estimate because of variations in the definition of cancer cachexia, but most patients with advanced cancer experience weight loss and decreased appetite. A variety of both tumor-derived factors (e.g., bombesin, adrenocorticotrophic hormone) and host-derived factors (e.g., tumor necrosis factor, interleukins 1 and 6, growth hormone) contribute to the altered metabolism, and a vicious cycle is established in which protein catabolism, glucose intolerance, and lipolysis cannot be reversed by the provision of calories.

It remains controversial how to assess nutritional status and when and how to intervene. Efforts to make the assessment objective have included the use of a prognostic nutritional index based on albumin levels, triceps skinfold thickness, transferrin levels, and delayed-type hypersensitivity skin testing. However, a simpler approach has been to define the threshold for nutritional intervention as <10% unexplained body weight loss, serum transferrin level <1500 mg/L (150 mg/dL), and serum albumin <34 g/L (3.4 g/dL).

The decision is important, because it appears that cancer therapy is substantially more toxic and less effective in the face of malnutrition. Nevertheless, it remains unclear whether nutritional intervention can alter the natural history. Unless some pathology is affecting the absorptive function of the gastrointestinal tract, enteral nutrition provided orally or by tube feeding is preferred over parenteral supplementation. However, the risks associated with the tube may outweigh the benefits. Megestrol acetate, a progestational agent, has been advocated as a pharmacologic intervention to improve nutritional status. Research

in this area may provide more tools in the future as cytokine-mediated mechanisms are further elucidated.

Psychosocial Support The psychosocial needs of patients vary with their situation. Patients undergoing treatment experience fear, anxiety, and depression. Self-image is often seriously compromised by deforming surgery and loss of hair. Women who receive cosmetic advice that enables them to look better also feel better. Loss of control over how one spends time can contribute to the sense of vulnerability. Juggling the demands of work and family with the demands of treatment may create enormous stresses. Sexual dysfunction is highly prevalent and needs to be discussed openly with the patient. An empathetic health care team is sensitive to the individual patient's needs and permits negotiation where such flexibility will not adversely affect the course of treatment.

Cancer survivors have other sets of difficulties. Patients may have fears associated with the termination of a treatment they associate with their continued survival. Adjustments are required to physical losses and handicaps, real and perceived. Patients may be preoccupied with minor physical problems. They perceive a decline in their job mobility and view themselves as less desirable workers. They may be victims of job and/or insurance discrimination. Patients may experience difficulty reentering their normal past life. They may feel guilty for having survived and may carry a sense of vulnerability to colds and other illnesses. Perhaps the most pervasive and threatening concern is the ever-present fear of relapse (the Damocles syndrome).

Patients in whom therapy has been unsuccessful have other problems related to the end of life.

Death and Dying The most common causes of death in patients with cancer are infection (leading to circulatory failure), respiratory failure, hepatic failure, and renal failure. Intestinal blockage may lead to inanition and starvation. Central nervous system disease may lead to seizures, coma, and central hypoventilation. About 70% of patients develop dyspnea preterminally. However, many months usually pass between the diagnosis of cancer and the occurrence of these complications, and during this period, the patient is severely affected by the possibility of death. The path of unsuccessful cancer treatment usually occurs in three phases. First, there is optimism at the hope of cure; when the tumor recurs, there is the acknowledgment of an incurable disease, and the goal of palliative therapy is embraced in the hope of being able to live with disease; finally, at the disclosure of imminent death, another adjustment in outlook takes place. The patient imagines the worst in preparation for the end of life and may go through stages of adjustment to the diagnosis. These stages include denial, isolation, anger, bargaining, depression, acceptance, and hope. Of course, patients do not all progress through all the stages or proceed through them in the same order or at the same rate. Nevertheless, developing an understanding of how the patient has been affected by the diagnosis and is coping with it is an important goal of patient management.

It is best to speak frankly with the patient and the family regarding the likely course of disease. These discussions can be difficult for the physician as well as for the patient and family. The critical features of the interaction are to reassure the patient and family that everything that can be done to provide comfort will be done. They will not be abandoned. Many patients prefer to be cared for in their homes or in a hospice setting rather than a hospital. The American College of Physicians has published a book called *Home Care Guide for Cancer: How to Care for Family and Friends at Home* that teaches an approach to successful problem-solving in home care. With appropriate planning, it should be possible to provide the patient with the necessary medical care as well as the psychological and spiritual support that will prevent the isolation and depersonalization that can attend in-hospital death.

The care of dying patients may take a toll on the physician. A "burnout" syndrome has been described that is characterized by fatigue, disengagement from patients and colleagues, and a loss of self-fulfillment. Efforts at stress reduction, maintenance of a balanced life, and setting realistic goals may combat this disorder.

End-of-Life Decisions Unfortunately, a smooth transition in treatment goals from curative to palliative may not be possible in all cases because of the occurrence of serious treatment-related complications or rapid disease progression. Vigorous and invasive medical support for a reversible disease or treatment complication is assumed to be justified. However, if the reversibility of the condition is in doubt, the patient's wishes determine the level of medical care. These wishes should be elicited before the terminal phase of illness and reviewed periodically. Information about advance directives can be obtained from the American Association of Retired Persons, 601 E Street, NW, Washington, DC 20049, 202-434-2277, or Choice in Dying, 250 West 57th Street, New York, NY 10107, 212-366-5540. Some states allow physicians to assist patients who choose to end their lives. This subject is challenging from an ethical and a medical point of view. Discussions of end-of-life decisions should be candid and involve clear informed consent, waiting periods, second opinions, and documentation. **A full discussion of end-of-life management is in Chap. 9.**

FURTHER READING

- BRANDT JM et al: Chronic and refractory pain: A systematic review of pharmacologic management in oncology. *Clin J Oncol Nurs* 21:31, 2017.
- KELLEY AS, MORRISON RS: Palliative care for the seriously ill. *N Engl J Med* 373:747, 2015.
- NAVARI RM, AAPRO M: Antiemetic prophylaxis for chemotherapy-induced nausea and vomiting. *N Engl J Med* 374:1356, 2016.
- SIEGEL RL et al: Cancer statistics, 2017. *CA Cancer J Clin* 67:7, 2017.

66

Prevention and Early Detection of Cancer



Jennifer M. Croswell, Otis W. Brawley,
Barnett S. Kramer

Improved understanding of carcinogenesis has allowed cancer prevention and early detection to expand beyond the identification and avoidance of carcinogens. Specific interventions to reduce cancer mortality by preventing cancer in those at risk, and effective screening for early detection of cancer, are the goals.

Carcinogenesis is a process that usually extends over years, a continuum of discrete tissue and cellular changes over time resulting in aberrant physiologic processes. Prevention concerns the identification and manipulation of the biologic, environmental, social, and genetic factors in the causal pathway of cancer.

EDUCATION AND HEALTHFUL HABITS

Public education on the avoidance of identified risk factors for cancer and encouraging healthy habits contributes to cancer prevention. The clinician is a powerful messenger in this process. The patient-provider encounter provides an opportunity to teach patients about the hazards of smoking, features of a healthy lifestyle, and use of proven cancer screening methods.

SMOKING CESSATION

Tobacco smoking is a strong, modifiable risk factor for cardiovascular disease, pulmonary disease, and cancer. Smokers have an ~1 in 3 lifetime risk of dying prematurely from a tobacco-related cancer, cardiovascular, or pulmonary disease. Tobacco use causes more deaths from cardiovascular disease than from cancer. Lung cancer and cancers of the larynx, oropharynx, esophagus, kidney, bladder, colon, pancreas, and stomach are all tobacco-related.

The number of cigarettes smoked per day and the level of inhalation of cigarette smoke are correlated with risk of lung cancer mortality. Light- and low-tar cigarettes are not safer, because smokers tend to inhale them more frequently and deeply.

Those who stop smoking have a 30–50% lower 10-year lung cancer mortality rate compared to those who continue smoking, despite the fact that some carcinogen-induced gene mutations persist for years after smoking cessation. Smoking cessation and avoidance would save more lives than any other public health activity.

The risk of tobacco smoke is not limited to the smoker. Environmental tobacco smoke, known as secondhand or passive smoke, causes lung cancer and other cardiopulmonary diseases in nonsmokers.

Tobacco use prevention is a pediatric issue. More than 80% of adult American smokers began smoking before the age of 18 years. Approximately 13% of Americans in grades 9 through 12 reported using two or more tobacco products in the past month. Electronic cigarettes have been advanced as a tool to achieve smoking cessation in adult smokers, but there is concern that they serve as a “gateway” to cigarette uptake in adolescents and are increasing in use. Counseling of adolescents and young adults is critical to prevent smoking. A clinician’s simple advice can be of benefit. Providers should query patients on tobacco use and offer smokers assistance in quitting.

Current approaches to smoking cessation recognize smoking as an addiction ([Chap. 448](#)). The smoker who is quitting goes through identifiable stages including: contemplation of quitting, an action phase in which the smoker quits, and a maintenance phase. Smokers who quit completely are more likely to be successful than those who gradually reduce the number of cigarettes smoked or change to lower-tar or lower-nicotine cigarettes. More than 90% of the Americans who have successfully quit smoking did so on their own, without participation in an organized cessation program, but cessation programs are helpful for some. The Community Intervention Trial for Smoking Cessation (COMMIT) was a 4-year program showing that light smokers (<25 cigarettes per day) were more likely to benefit from simple cessation messages and cessation programs than those who did not receive an intervention. Quit rates were 30.6% in the intervention group and 27.5% in the control group. The COMMIT interventions were unsuccessful in heavy smokers (>25 cigarettes per day). Heavy smokers may need an intensive broad-based cessation program that includes counseling, behavioral strategies, and pharmacologic adjuncts, such as nicotine replacement (gum, patches, sprays, lozenges, and inhalers), bupropion, and/or varenicline.

The health risks of cigars are similar to those of cigarettes. Smoking one or two cigars daily doubles the risk for oral and esophageal cancers; smoking three or four cigars daily increases the risk of oral cancers more than eightfold and esophageal cancer fourfold. The risks of occasional use are unknown.

Smokeless tobacco also represents a substantial health risk. Chewing tobacco is a carcinogen linked to dental caries, gingivitis, oral leukoplakia, and oral cancer. The systemic effects of smokeless tobacco (including snuff) may increase risks for other cancers. Esophageal cancer is linked to carcinogens in tobacco dissolved in saliva and swallowed. The net effects of e-cigarettes on health are poorly studied.

■ PHYSICAL ACTIVITY

Physical activity is associated with a decreased risk of colon and breast cancer. A variety of mechanisms have been proposed. However, such studies are prone to confounding factors such as recall bias, association of exercise with other health-related practices, and effects of preclinical cancers on exercise habits (reverse causality).

■ DIET MODIFICATION

International epidemiologic studies suggest that diets high in fat are associated with increased risk for cancers of the breast, colon, prostate, and endometrium. These cancers have their highest incidence and mortalities in Western cultures, where fat composes an average of one-third of the total calories consumed.

Despite correlations, dietary fat has not been proven to cause cancer. Case-control and cohort epidemiologic studies give conflicting results.

In addition, diet is a highly complex exposure to many nutrients and chemicals. Low-fat diets are associated with many dietary changes beyond simple subtraction of fat. Other lifestyle changes are also associated with adherence to a low-fat diet.

In observational studies, dietary fiber is associated with a reduced risk of colonic polyps and invasive cancer of the colon. However, cancer-protective effects of increasing fiber and lowering dietary fat have not been proven in the context of a prospective clinical trial. The putative protective mechanisms are complex and speculative. Fiber binds oxidized bile acids and generates soluble fiber products, such as butyrate, that may have differentiating properties. Fiber does not increase bowel transit times. Two large prospective cohort studies of >100,000 health professionals showed no association between fruit and vegetable intake and risk of cancer.

The Polyp Prevention Trial randomly assigned 2000 elderly persons, who had polyps removed, to a low-fat, high-fiber diet versus routine diet for 4 years. No differences were noted in polyp formation.

The U.S. National Institutes of Health Women’s Health Initiative, launched in 1994, was a long-term clinical trial enrolling >100,000 women age 45–69 years. It placed women into 22 intervention groups. Participants received calcium/vitamin D supplementation; hormone replacement therapy; and counseling to increase exercise, eat a low-fat diet with increased consumption of fruits, vegetables, and fiber, and cease smoking. The study showed that although dietary fat intake was lower in the diet intervention group, invasive breast cancers were not reduced over an 8-year follow-up period compared to the control group. No reduction was seen in the incidence of colorectal cancer in the dietary intervention arm. The difference in dietary fat averaged ~10% between the two groups. Evidence does not currently establish the anticarcinogenic value of vitamin, mineral, or nutritional supplements in amounts greater than those provided by a balanced diet.

■ ENERGY BALANCE

Risk of certain cancers appears to increase modestly (relative risks generally in the 1.0–2.0 range) as body mass index (BMI) increases beyond 25 kg/m². A cohort study of >5 million adults included in the U.K. Clinical Practice Research Datalink (a primary care database) found that each 5 kg/m² increase in BMI was linearly associated with cancers of the uterus, gallbladder, kidney, cervix, thyroid, and leukemia. Positive associations were also noted between BMI and colon, liver, ovarian, and postmenopausal breast cancers, but these associations were not linear and the effect varied by individual characteristics. High BMI appears to have an inverse association with prostate and premenopausal breast cancer.

■ SUN AVOIDANCE

Nonmelanoma skin cancers (basal cell and squamous cell) are induced by cumulative exposure to ultraviolet (UV) radiation. Intermittent acute sun exposure and sun damage have been linked to melanoma, but the evidence is inconsistent. Sunburns, especially in childhood and adolescence, may be associated with an increased risk of melanoma in adulthood. Reduction of sun exposure through use of protective clothing and changing patterns of outdoor activities can reduce skin cancer risk. Sunscreens decrease the risk of actinic keratoses, the precursor to squamous cell skin cancer, but melanoma risk may not be reduced. Sunscreens prevent burning, but they may encourage more prolonged exposure to the sun and may not filter out wavelengths of energy that cause melanoma.

Appearance-focused behavioral interventions in young women can decrease indoor tanning use and other UV exposures and may be more effective than messages about long-term cancer risks. Self-examination for skin pigment characteristics associated with skin cancer, such as freckling, may be useful in identifying people at high risk. Those who recognize themselves as being at risk tend to be more compliant with sun-avoidance recommendations. Risk factors for melanoma include a propensity to sunburn, a large number of benign melanocytic nevi, and atypical nevi.

CANCER CHEMOPREVENTION

Chemoprevention involves the use of specific natural or synthetic chemical agents to reverse, suppress, or prevent carcinogenesis before the development of invasive malignancy.

Cancer develops through an accumulation of tissue abnormalities associated with genetic and epigenetic changes, and growth regulatory pathways that are potential points of intervention to prevent cancer. The initial changes are termed *initiation*. The alteration can be inherited or acquired through the action of physical, infectious, or chemical carcinogens. Like most human diseases, cancer arises from an interaction between genetics and environmental exposures (Table 66-1). Influences that cause the initiated cell and its surrounding tissue microenvironment to progress through the carcinogenic process and change phenotypically are termed *promoters*. Promoters include hormones such as androgens, linked to prostate cancer, and estrogen, linked to breast and endometrial cancer. The distinction between an initiator and promoter is indistinct; some components of cigarette smoke are “complete carcinogens,” acting as both initiators and promoters.

TABLE 66-1 Suspected Carcinogens

CARCINOGENS ^a	ASSOCIATED CANCER OR NEOPLASM
Alkylating agents	Acute myeloid leukemia, bladder cancer
Androgens	Prostate cancer
Aromatic amines (dyes)	Bladder cancer
Arsenic	Cancer of the lung, skin
Asbestos	Cancer of the lung, pleura, peritoneum
Benzene	Acute myelocytic leukemia
Chromium	Lung cancer
Diethylstilbestrol (prenatal)	Vaginal cancer (clear cell)
Epstein-Barr virus	Burkitt's lymphoma, nasal T-cell lymphoma
Estrogens	Cancer of the endometrium, liver, breast
Ethyl alcohol	Cancer of the breast, liver, esophagus, head and neck
<i>Helicobacter pylori</i>	Gastric cancer, gastric MALT lymphoma
Hepatitis B or C virus	Liver cancer
Human immunodeficiency virus	Non-Hodgkin's lymphoma, Kaposi's sarcoma, squamous cell carcinomas (especially of the urogenital tract)
Human papilloma virus	Cancers of the cervix, anus, oropharynx
Human T-cell lymphotropic virus type 1 (HTLV-1)	Adult T-cell leukemia/lymphoma
Immunosuppressive agents (azathioprine, cyclosporine, glucocorticoids)	Non-Hodgkin's lymphoma
Ionizing radiation (therapeutic or diagnostic)	Breast, bladder, thyroid, soft tissue, bone, hematopoietic, and many more
Nitrogen mustard gas	Cancer of the lung, head and neck, nasal sinuses
Nickel dust	Cancer of the lung, nasal sinuses
Diesel exhaust	Lung cancer (miners)
Phenacetin	Cancer of the renal pelvis and bladder
Polycyclic hydrocarbons	Cancer of the lung, skin (especially squamous cell carcinoma of scrotal skin)
Radon gas	Lung cancer
Schistosomiasis	Bladder cancer (squamous cell)
Sunlight (ultraviolet)	Skin cancer (squamous cell and melanoma)
Tobacco (including smokeless)	Cancer of the upper aerodigestive tract, bladder
Vinyl chloride	Liver cancer (angiosarcoma)

^aAgents that are thought to act as cancer initiators and/or promoters.

Cancer can be prevented or controlled through interference with the factors that cause cancer initiation, promotion, or progression. Compounds of interest in chemoprevention often have antimutagenic, hormone modulation, anti-inflammatory, antiproliferative, or proapoptotic activity (or a combination).

■ CHEMOPREVENTION OF CANCERS OF THE UPPER AERODIGESTIVE TRACT

Smoking causes diffuse epithelial injury in the oral cavity, neck, esophagus, and lung. Patients cured of squamous cell cancers of the lung, esophagus, oral cavity, and neck are at risk (as high as 5% per year) of developing second cancers of the upper aerodigestive tract. Cessation of cigarette smoking does not markedly decrease the cured cancer patient's risk of second malignancy, even though it does lower the cancer risk in those who have never developed a malignancy. Smoking cessation may halt the early stages of the carcinogenic process (such as metaplasia), but it may have no effect on late stages of carcinogenesis. This “field carcinogenesis” hypothesis for upper aerodigestive tract cancer has made “cured” patients an important population for chemoprevention of second malignancies.

Persistent oral human papilloma virus (HPV) infection, particularly HPV-16, increases the risk for cancers of the oropharynx. This association exists even in the absence of other risk factors such as smoking or alcohol use (although the magnitude of increased risk appears greater than additive when HPV infection and smoking are both present). Oral HPV infection is believed to be largely sexually acquired. Although the evidence is not definitive, the introduction of the HPV vaccine may eventually reduce oropharyngeal cancer rates.

Oral leukoplakia, a premalignant lesion commonly found in smokers, has been used as an intermediate marker of chemopreventive activity in smaller shorter-duration, randomized, placebo-controlled trials. Response was associated with upregulation of retinoic acid receptor-β (RAR-β). Therapy with high, relatively toxic doses of isotretinoin (13-cis-retinoic acid) causes regression of oral leukoplakia. However, the lesions recur when the therapy is withdrawn, suggesting the need for long-term administration. More tolerable doses of isotretinoin have not shown benefit in the prevention of head and neck cancer. Isotretinoin did not prevent second malignancies in patients cured of early-stage non-small cell lung cancer; mortality rates were actually increased in current smokers.

Several large-scale trials have assessed agents in the chemoprevention of lung cancer in patients at high risk. In the α-tocopherol/β-carotene (ATBC) Lung Cancer Prevention Trial, participants were male smokers, age 50–69 years at entry. Participants had smoked an average of one pack of cigarettes per day for 35.9 years. Participants received α-tocopherol, β-carotene, and/or placebo in a randomized, two-by-two factorial design. After median follow-up of 6.1 years, lung cancer incidence and mortality were statistically significantly increased in those receiving β-carotene. α-Tocopherol had no effect on lung cancer mortality, with no apparent interaction between the two drugs. Patients receiving α-tocopherol had a higher incidence of hemorrhagic stroke.

The β-Carotene and Retinol Efficacy Trial (CARET) involved 17,000 American smokers and workers with asbestos exposure. Entrants were randomly assigned to one of four arms and received β-carotene, retinol, and/or placebo in a two-by-two factorial design. This trial also demonstrated harm from β-carotene: a lung cancer rate of 5 per 1000 subjects per year for those taking placebo versus 6 per 1000 subjects per year for those taking β-carotene.

The ATBC and CARET results demonstrate the importance of testing chemoprevention hypotheses thoroughly before widespread implementation because the results contradict a number of observational studies. The Physicians' Health Trial showed no change in the risk of lung cancer for those taking β-carotene; however, fewer of its participants were smokers than those in the ATBC and CARET studies.

■ CHEMOPREVENTION OF COLON CANCER

Many colon cancer prevention trials are based on the premise that most colorectal cancers develop from adenomatous polyps. These trials use adenoma recurrence or disappearance as a surrogate endpoint (not

yet validated) for colon cancer prevention. Early clinical trial results suggest that nonsteroidal anti-inflammatory drugs (NSAIDs), such as piroxicam, sulindac, and aspirin, may prevent adenoma formation or cause regression of adenomatous polyps. The mechanism of action of NSAIDs is unknown, but they are presumed to work through the cyclooxygenase pathway. Although two randomized controlled trials (the Physicians' Health Study and the Women's Health Study) did not show an effect of aspirin on colon cancer or adenoma incidence in persons with no previous history of colonic lesions after 10 years of therapy, these trials did show an approximately 18% relative risk reduction for colonic adenoma incidence in persons with a previous history of adenomas after 1 year. A meta-analysis of four randomized controlled trials (albeit primarily designed to examine aspirin's effects on cardiovascular events) found that aspirin at doses of at least 75 mg/d resulted in a 33% relative reduction in colorectal cancer incidence after 20 years, with no clear increase in efficacy at higher doses. Based on a systematic review of evidence from randomized trials for primary prevention of cardiovascular disease, the U.S. Preventive Services Task Force concluded that the balance of benefits and harms favored initiating low-dose aspirin for colorectal cancer prevention in adults age 50–59 if they have a 10% or greater 10-year risk of cardiovascular disease. Cyclooxygenase-2 (COX-2) inhibitors have also been considered for colorectal cancer and polyp prevention. Trials with COX-2 inhibitors were initiated, but an increased risk of cardiovascular events in those taking the COX-2 inhibitors was noted, suggesting that these agents are not suitable for chemoprevention in the general population.

Epidemiologic studies suggest that diets high in calcium lower colon cancer risk. Calcium binds bile and fatty acids, which cause proliferation of colonic epithelium. It is hypothesized that calcium reduces intraluminal exposure to these compounds. The randomized controlled Calcium Polyp Prevention Study found that calcium supplementation decreased the absolute risk of adenomatous polyp recurrence by 7% at 4 years; extended observational follow-up demonstrated a 12% absolute risk reduction 5 years after cessation of treatment. However, in the Women's Health Initiative, combined use of calcium carbonate and vitamin D twice daily did not reduce the incidence of invasive colorectal cancer compared with placebo after 7 years.

The Women's Health Initiative demonstrated that postmenopausal women taking estrogen plus progestin have a 44% lower relative risk of colorectal cancer compared to women taking placebo. Of >16,600 women randomized and followed for a median of 5.6 years, 43 invasive colorectal cancers occurred in the hormone group and 72 in the placebo group. The positive effect on colon cancer is mitigated by the modest increase in cardiovascular and breast cancer risks associated with combined estrogen plus progestin therapy.

Most case-control and cohort studies have not confirmed early reports of an association between regular statin use and a reduced risk of colorectal cancer. No randomized controlled trials have addressed this hypothesis. A meta-analysis of statin use showed no protective effect of statins on overall cancer incidence or death.

CHEMOPREVENTION OF BREAST CANCER

Tamoxifen is an antiestrogen with partial estrogen agonistic activity in some tissues, such as endometrium and bone. One of its actions is to upregulate transforming growth factor β , which decreases breast cell proliferation. In a randomized placebo-controlled prevention trial involving >13,000 pre- and postmenopausal women at high risk, tamoxifen decreased the risk of developing breast cancer by 49% (from 43.4 to 22 per 1000 women) after a median follow-up of nearly 6 years. Tamoxifen also reduced bone fractures; a small increase in risk of endometrial cancer, stroke, pulmonary emboli, and deep vein thrombosis was noted. The International Breast Cancer Intervention Study (IBIS-I) and the Italian Randomized Tamoxifen Prevention Trial also demonstrated a reduction in breast cancer incidence with tamoxifen use. A trial comparing tamoxifen with another selective estrogen receptor modulator, raloxifene, performed in postmenopausal women showed that raloxifene is comparable to tamoxifen in cancer prevention, but without the risk of endometrial cancer. Raloxifene was associated with more invasive breast cancers and a trend toward more

noninvasive breast cancers, but fewer thromboembolic events than tamoxifen; the drugs are similar in risks of other cancers, fractures, ischemic heart disease, and stroke. Both tamoxifen and raloxifene (the latter for postmenopausal women only) have been approved by the U.S. Food and Drug Administration (FDA) for reduction of breast cancer in women at high risk for the disease (1.66% risk at 5 years based on the Gail risk model: <http://www.cancer.gov/bcrisktool/>).

Because the aromatase inhibitors are even more effective than tamoxifen in adjuvant breast cancer therapy, it has been hypothesized that they would be more effective in breast cancer prevention. A randomized, placebo-controlled trial of exemestane reported a 65% relative reduction (from 5.5 to 1.9 per 1000 women) in the incidence of invasive breast cancer in women at elevated risk after a median follow-up of about 3 years. Common adverse effects included arthralgias, hot flashes, fatigue, and insomnia. No trial has directly compared aromatase inhibitors with selective estrogen receptor modulators for breast cancer chemoprevention.

CHEMOPREVENTION OF PROSTATE CANCER

Finasteride and dutasteride are 5- α -reductase inhibitors. They inhibit conversion of testosterone to dihydrotestosterone (DHT), a potent stimulator of prostate cell proliferation. The Prostate Cancer Prevention Trial (PCPT) randomly assigned men age 55 years or older at average risk of prostate cancer to finasteride or placebo. All men in the trial were being regularly screened with prostate-specific antigen (PSA) levels and digital rectal examination. After 7 years of therapy, the incidence of prostate cancer was 18.4% in the finasteride arm, compared with 24.4% in the placebo arm, a statistically significant difference. However, the finasteride group had more patients with tumors of Gleason score 7 and higher compared with the placebo arm (6.4 vs 5.1%). Long-term (10–15 years) follow-up did not reveal any statistically significant differences in overall mortality between all men in the finasteride and placebo arms or in men diagnosed with prostate cancer, but the power to detect a difference was limited.

Dutasteride has also been evaluated as a preventive agent for prostate cancer. The Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial was a randomized double-blind trial in which ~8200 men with an elevated PSA (2.5–10 ng/mL for men age 50–60 years and 3–10 ng/mL for men age 60 years or older) and negative prostate biopsy on enrollment received daily 0.5 mg of dutasteride or placebo. The trial found a statistically significant 23% relative risk reduction in the incidence of biopsy-detected prostate cancer in the dutasteride arm at 4 years of treatment (659 cases vs 858 cases, respectively). Overall, across years 1 through 4, there was no difference between the arms in the number of tumors with a Gleason score of 7 to 10; however, during years 3 and 4, there was a statistically significant difference in tumors with Gleason score of 8 to 10 in the dutasteride arm (12 tumors vs 1 tumor, respectively).

The clinical importance of the apparent increased incidence of higher-grade tumors in the 5- α -reductase inhibitor arms of these trials is controversial. It may represent an increased sensitivity of PSA and digital rectal exam for high-grade tumors in men receiving these agents. The FDA has analyzed both trials, and it determined that the use of a 5- α -reductase inhibitor for prostate cancer chemoprevention would result in one additional high-grade (Gleason score 8 to 10) prostate cancer for every three to four lower-grade (Gleason score <6) tumors averted. Although it acknowledged that detection bias may have accounted for the finding, a causative role for 5- α -reductase inhibitors could not be conclusively dismissed. These agents are therefore not FDA-approved for prostate cancer prevention.

Because all men in both the PCPT and REDUCE trials were being screened and because screening approximately doubles the rate of prostate cancer, it is not known if finasteride or dutasteride decreases the risk of prostate cancer in men who are not being screened or simply reduces the risk of non-life threatening cancers detectable by screening.

Several favorable laboratory and observational studies led to the formal evaluation of selenium and α -tocopherol (vitamin E) as potential prostate cancer preventives. The Selenium and Vitamin E Cancer Prevention Trial (SELECT) assigned 35,533 men to receive 200 μ g/d

selenium, 400 IU/d α -tocopherol, selenium plus vitamin E, or placebo. After a median follow-up of 7 years, a trend toward an increased risk of developing prostate cancer was observed for those men taking vitamin E alone as compared to the placebo arm (hazard ratio 1.17; 95% confidence interval, 1.004–1.36).

VACCINES AND CANCER PREVENTION

A number of infectious agents cause cancer. Hepatitis B and C are linked to liver cancer; some HPV strains are linked to cervical, anal, and head and neck cancer; and *Helicobacter pylori* is associated with gastric adenocarcinoma and gastric lymphoma. Vaccines to protect against these agents may therefore reduce the risk of their associated cancers.

The hepatitis B vaccine is effective in preventing hepatitis and hepatomas due to chronic hepatitis B infection.

A nonavalent vaccine (covering HPV strains 6, 11, 16, 18, 31, 33, 45, 52, and 58) is available for use in the United States. HPV types 6 and 11 cause genital papillomas. The remaining HPV types cause cervical and anal cancer; reduction in HPV types 16 and 18 alone could prevent >70% of cervical cancers worldwide. For individuals not previously infected with these HPV strains, the vaccine demonstrates high efficacy in preventing persistent strain-specific HPV infections. Studies also confirm the vaccine's ability to prevent preneoplastic lesions (cervical or anal intraepithelial neoplasia [CIN/AIN] I, II, and III). The durability of the immune response beyond 8–10 years is not currently known. The vaccine does not appear to impact preexisting infections and the efficacy appears to be lower for populations that had previously been exposed to vaccine-specific HPV strains. A two-dose schedule is currently recommended in the United States for females and males age 9–14 years; teens and young adults who start the series between 15 and 26 years are recommended to receive three doses of the vaccine.

SURGICAL PREVENTION OF CANCER

Some organs in some individuals are at such high risk of developing cancer that surgical removal of the organ at risk may be considered. Women with severe cervical dysplasia are treated with laser or loop electrosurgical excision or conization and occasionally even hysterectomy. Colectomy is used to prevent colon cancer in patients with familial polyposis or ulcerative colitis.

Prophylactic bilateral mastectomy may be chosen for breast cancer prevention among women with genetic predisposition to breast cancer. In a prospective series of 139 women with *BRCA1* and *BRCA2* mutations, 76 chose to undergo prophylactic mastectomy and 63 chose close surveillance. At 3 years, no cases of breast cancer had been diagnosed in those opting for surgery, but eight patients in the surveillance group had developed breast cancer. A larger ($n = 639$) retrospective cohort study reported that three patients developed breast cancer after prophylactic mastectomy compared with an expected incidence of 30–53 cases: a 90–94% reduction in breast cancer risk. Postmastectomy breast cancer-related deaths were reduced by 81–94% for high-risk women compared with sister controls and by 100% for moderate-risk women when compared with expected rates.

Prophylactic salpingo-oophorectomy may also be employed for the prevention of ovarian and breast cancers among high-risk women. A prospective cohort study evaluating the outcomes of *BRCA* mutation carriers demonstrated a statistically significant association between prophylactic salpingo-oophorectomy and a reduced incidence of ovarian or primary peritoneal cancer (36% relative risk reduction, or a 4.5% absolute difference). Studies of prophylactic oophorectomy for prevention of breast cancer in women with genetic mutations have shown relative risk reductions of approximately 50%; the risk reduction may be greatest for women having the procedure at younger (i.e., <50 years) ages. The observation that most high-grade serous "ovarian cancers" actually arise in the fallopian tube fimbria raises the possibility that this lethal subtype may be prevented by ovary-sparing salpingectomy.

All of the evidence concerning the use of prophylactic mastectomy and salpingo-oophorectomy for prevention of breast and ovarian cancer in high-risk women has been observational in nature; such studies are prone to a variety of biases, including case selection bias, family relationships between patients and controls, and inadequate

information about hormone use. Thus, they may give an overestimate of the magnitude of benefit.

CANCER SCREENING

Screening is a means of early detection in asymptomatic individuals, with the goal of decreasing morbidity and mortality. While screening can potentially reduce disease-specific deaths and has been shown to do so in cervical, colon, lung, and breast cancer, it is also subject to a number of biases that can suggest a benefit when actually there is none. Biases can even mask net harm. Early detection does not in itself confer benefit. Cause-specific mortality, rather than survival after diagnosis, is the preferred endpoint (see below).

Because screening is done on asymptomatic, healthy persons, it should offer substantial likelihood of benefit that outweighs harm. Screening tests and their appropriate use should be carefully evaluated before their use is widely encouraged in screening programs.

A large and increasing number of genetic mutations and nucleotide polymorphisms have been associated with an increased risk of cancer. Testing for these genetic mutations could in theory define a high-risk population. However, most of the identified mutations have very low penetrance and individually provide limited predictive accuracy. The ability to predict the development of a particular cancer may someday present therapeutic options as well as ethical dilemmas. It may eventually allow for early intervention to prevent a cancer or limit its severity. People at high risk may be ideal candidates for chemoprevention and screening; however, efficacy of these interventions in the high-risk population should be investigated. Currently, persons at high risk for a particular cancer can engage in intensive screening. While this course is clinically reasonable, it is not known if it reduces mortality in these populations.

The Accuracy of Screening A screening test's accuracy or ability to discriminate disease is described by four indices: sensitivity, specificity, positive predictive value, and negative predictive value (Table 66-2). *Sensitivity*, also called the true-positive rate, is the proportion of persons with the disease who test positive in the screen (i.e., the ability of the test to detect disease when it is present). *Specificity*, or 1 minus the false-positive rate, is the proportion of persons who do not have the disease that test negative in the screening test (i.e., the ability of a test to correctly indicate that the disease is not present). The *positive predictive value* is the proportion of persons who test positive that actually have the disease. Similarly, *negative predictive value* is the proportion testing negative that do not have the disease. The sensitivity and specificity of a test are independent of the underlying prevalence (or risk) of the disease in the population screened, but the predictive values depend strongly on the prevalence of the disease.

TABLE 66-2 Assessment of the Value of a Diagnostic Test^a

	CONDITION PRESENT	CONDITION ABSENT
Positive test	<i>a</i>	<i>b</i>
Negative test	<i>c</i>	<i>d</i>
<i>a</i> = true positive		
<i>b</i> = false positive		
<i>c</i> = false negative		
<i>d</i> = true negative		
Sensitivity	The proportion of persons with the condition who test positive: $a / (a + c)$	
Specificity	The proportion of persons without the condition who test negative: $d / (b + d)$	
Positive predictive value (PPV)	The proportion of persons with a positive test who have the condition: $a / (a + b)$	
Negative predictive value	The proportion of persons with a negative test who do not have the condition: $d / (c + d)$	

Prevalence, sensitivity, and specificity determine PPV

$$\text{PPV} = \frac{\text{prevalence} \times \text{sensitivity}}{(\text{prevalence} \times \text{sensitivity}) + (1 - \text{prevalence})(1 - \text{specificity})}$$

^aFor diseases of low prevalence, such as cancer, poor specificity has a dramatic adverse effect on PPV such that only a small fraction of positive tests are true positives.

Screening is most beneficial, efficient, and economical when the target disease is common in the population being screened. Specificity is at least as important to the ultimate feasibility and success of a screening test as sensitivity.

Potential Biases of Screening Tests Common biases of screening are lead time, length-biased sampling, and selection. These biases can make a screening test seem beneficial when actually it is not (or even causes net harm). Whether beneficial or not, screening can create the false impression of an epidemic by increasing the number of cancers diagnosed. It can also produce a shift in the proportion of patients diagnosed at an early stage (even without a reduction in absolute incidence of late-stage disease) and inflate survival statistics without reducing mortality (i.e., the number of deaths from a given cancer relative to the number of those at risk for the cancer). In such a case, the apparent duration of survival (measured from date of diagnosis) increases without lives being saved or life expectancy changed.

Lead-time bias occurs whether or not a test influences the natural history of the disease; the patient is merely diagnosed at an earlier date. Survival appears increased even if life is not prolonged. The screening test only prolongs the time the subject is aware of the disease and spends as a patient.

Length-biased sampling occurs because screening tests generally can more easily detect slow-growing, less aggressive cancers than fast-growing cancers. Cancers diagnosed due to the onset of symptoms between scheduled screenings are on average more aggressive, and treatment outcomes are not as favorable. An extreme form of length bias sampling is termed *overdiagnosis*, the detection of “pseudo disease.” The reservoir of some undetected slow-growing tumors is large. Many of these tumors fulfill the histologic criteria of cancer but will never become clinically significant or cause death during the patient’s remaining lifespan. This problem is compounded by the fact that the most common cancers appear most frequently at ages when competing causes of death are more frequent.

Selection bias occurs because the population most likely to seek screening often differs from the general population to which the screening test might be applied. In general, volunteers for studies are more health conscious and likely to have a better prognosis or lower mortality rate, irrespective of the screening result. This is termed the *healthy volunteer effect*.

Potential Drawbacks of Screening Risks associated with screening include harm caused by the screening intervention itself, harm due to the further investigation of persons with positive tests (both true and false positives), and harm from the treatment of persons with a true-positive result, whether or not life is extended by treatment (e.g., even if a screening test reduces relative cause-specific mortality by 20–30%, 70–80% of those diagnosed still go on to die of the target cancer). The diagnosis and treatment of cancers that would never have caused medical problems can lead to the harm of unnecessary treatment and give patients the anxiety of a cancer diagnosis. The psychosocial impact of cancer screening can be substantial when applied to the entire population.

Assessment of Screening Tests Good clinical trial design can offset some biases of screening and demonstrate the relative risks and benefits of a screening test. A randomized controlled screening trial with cause-specific mortality as the endpoint provides the strongest support for a screening intervention. Overall mortality should also be reported to detect an adverse effect of screening and treatment on other disease outcomes (e.g., cardiovascular disease). In a randomized trial, two like populations are randomly established. One is given the usual standard of care (which may be no screening at all) and the other receives the screening intervention being assessed. Efficacy for the population studied is established when the group receiving the screening test has a better cause-specific mortality rate than the control group. Studies showing a reduction in the incidence of advanced-stage disease, improved survival, or a stage shift are weaker (and possibly misleading) evidence of benefit. These latter criteria are early indicators but not sufficient to establish the value of a screening test.

Although a randomized, controlled screening trial provides the strongest evidence to support a screening test, it is not perfect. Unless the trial is population-based, it does not remove the question of generalizability to the target population. Screening trials generally involve thousands of persons and last for years. Less definitive study designs are therefore often used to estimate the effectiveness of screening practices. However, every nonrandomized study design is subject to strong confounders. In descending order of strength, evidence may also be derived from the findings of internally controlled trials using intervention allocation methods other than randomization (e.g., allocation by birth date, date of clinic visit); the findings of analytic observational studies; or the results of multiple time series studies with or without the intervention.

Screening for Specific Cancers Screening for cervical, colon, and breast cancer has the potential to be beneficial for certain age groups. Depending on age and smoking history, lung cancer screening can also be beneficial in specific settings. Special surveillance of those at high risk for a specific cancer because of a family history or a genetic risk factor may be prudent, but few studies have assessed the effect on mortality. A number of organizations have considered whether or not to endorse routine use of certain screening tests. Because criteria have varied, they have arrived at different recommendations. The American Cancer Society (ACS) and the U.S. Preventive Services Task Force (USPSTF) publish screening guidelines (Table 66-3); the American Academy of Family Practitioners (AAFP) often follow/endorse the USPSTF recommendations; and the American College of Physicians (ACP) develops recommendations based on structured reviews of other organizations’ guidelines.

BREAST CANCER Breast self-examination, clinical breast examination by a caregiver, mammography, and magnetic resonance imaging (MRI) have all been variably advocated as useful screening tools.

A number of trials have suggested that annual or biennial screening with mammography or mammography plus clinical breast examination in normal-risk women older than age 50 years decreases breast cancer mortality. Each trial has been criticized for design flaws. In most trials, breast cancer-related mortality rates were decreased by 15–30%. Experts disagree on whether average-risk women age 40–49 years should receive regular screening (Table 66-3). The U.K. Age Trial, the only randomized trial of breast cancer screening to specifically evaluate the impact of mammography in women age 40–49 years, found no statistically significant difference in breast cancer mortality for screened women versus controls after about 11 years of follow-up (relative risk 0.83; 95% confidence interval 0.66–1.04); however, <70% of women received screening in the intervention arm, potentially diluting the observed effect. A meta-analysis of nine large randomized trials showed an 8% relative reduction in mortality (relative risk 0.92; 95% confidence interval 0.75–1.02) from mammography screening for women age 39–49 years after 11–20 years of follow-up. This is equivalent to 3 breast cancer deaths prevented per 10,000 women >10 years (although the result is not statistically significant). At the same time, nearly half of women age 40–49 years screened annually will have false-positive mammograms necessitating further evaluation, often including biopsy. Estimates of overdiagnosis range from 10 to 40% of diagnosed invasive cancers. In the United States, widespread screening over the last several decades has not been accompanied by a reduction in incidence of metastatic breast cancer despite a large increase in early-stage disease, suggesting a substantial amount of overdiagnosis at the population level.

Digital breast tomosynthesis is an emerging method of breast cancer screening that reconstructs multiple x-ray images of the breast into superimposed “three-dimensional” slices. Although some evidence is available concerning the test characteristics of this modality, there are currently no data on its effects on health outcomes such as breast cancer-related morbidity, mortality, or overdiagnosis rates.

No study of breast self-examination has shown it to decrease mortality. A randomized controlled trial of approximately 266,000 women in China demonstrated no difference in breast cancer mortality between a group that received intensive breast self-exam instruction and reinforcement/reminders and controls at 10 years of follow-up.

TABLE 66-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition^a

Cancer Type	Test or Procedure	USPSTF	ACS
Breast	Self-examination	"D" ^b (Not in current recommendations; from 2009)	Women, all ages: No specific recommendation
	Clinical examination	Women ≥40 years: "I" (as a stand-alone without mammography) (Not in current recommendations; from 2009)	Women, all ages: Do not recommend
	Mammography	Women 40–49 years: The decision to start screening mammography in women prior to age 50 years should be an individual one. Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years. ("C") Women 50–74 years: Every 2 years ("B") Women ≥75 years: "I"	Women 40–44 years: Provide the opportunity to begin annual screening Women 45–54 years: Screen annually Women ≥55 years: Transition to biennial screening or have the opportunity to continue annual screening Women ≥40 should continue screening mammography as long as their overall health is good and they have a life expectancy of 10 years or longer
	Magnetic resonance imaging (MRI)	"I" (Not in current recommendations; from 2009)	Women with >20% lifetime risk of breast cancer: Screen with MRI plus mammography annually Women with 15–20% lifetime risk of breast cancer: Discuss option of MRI plus mammography annually Women with <15% lifetime risk of breast cancer: Do not screen annually with MRI
	Tomosynthesis	Women, all ages: "I"	No specific recommendation
Cervical	Pap test (cytology)	Women 21–65 years: Screen every 3 years ("A") Women <21 years: "D" Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women 21–29 years: Screen every 3 years Women 30–65 years: Acceptable approach to screen with cytology every 3 years (see HPV test below) Women <21 years: No screening Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
	HPV test	Women 30–65 years: Screen in combination with cytology every 5 years if woman desires to lengthen the screening interval (see Pap test above) ("A") Women <30 years: "D" Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women 30–65 years: Preferred approach to screen with HPV and cytology co-testing every 5 years (see Pap test above) Women <30 years: Do not use HPV testing Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
	Sigmoidoscopy	Adults, 50–75 years: "A" Screen for colorectal cancer; the risks and benefits of the different screening methods vary Adults, 76 to 85 years: "C" The decision to screen should be an individual one, taking into account the patient's overall health and prior screening history Every 5 years; modeling suggests improved benefit if performed every 10 years in combination with annual FIT	Adults ≥50 years: Screen every 5 years
	Fecal occult blood testing (FOBT)	Every year	Adults ≥50 years: Screen every year
Colorectal	Colonoscopy	Every 10 years	Adults ≥50 years: Screen every 10 years
	Fecal DNA testing	Every 1 or 3 years	Adults ≥50 years: Screen, but interval uncertain
	Fecal immuno-chemical testing (FIT)	Every year	Adults ≥50 years: Screen every year
	CT colonography	Every 5 years	Adults ≥50 years: Screen every 5 years
Lung	Low-dose computed tomography (CT) scan	Adults 55–80 years, with a ≥30 pack-year smoking history, still smoking or have quit within past 15 years: "B" Discontinue once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy or the ability to have curative lung surgery	Men and women, 55–74 years, with ≥30 pack-year smoking history, still smoking or have quit within past 15 years: Discuss benefits, limitations, and potential harms of screening; only perform screening in facilities with the right type of CT scanner and with high expertise/specialists
Ovarian	CA-125 Transvaginal ultrasound	Women, all ages: "D" Women, all ages: "D"	There is no sufficiently accurate test proven effective in the early detection of ovarian cancer. For women at high risk of ovarian cancer and/or who have unexplained, persistent symptoms, the combination of CA-125 and transvaginal ultrasound with pelvic exam may be offered.

(Continued)

TABLE 66-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition^a (Continued)

CANCER TYPE	TEST OR PROCEDURE	USPSTF	ACS
Prostate	Prostate-specific antigen (PSA)	Men, all ages: "D"	Starting at age 50, men should talk to a doctor about the pros and cons of testing so they can decide if testing is the right choice for them. If African American or have a father or brother who had prostate cancer before age 65, men should have this talk starting at age 45. How often they are tested will depend on their PSA level
	Digital rectal examination (DRE)	No individual recommendation	As for PSA; if men decide to be tested, they should have the PSA blood test with or without a rectal exam
Skin	Complete skin examination by clinician or patient	Adults, all ages: "I"	Self-examination monthly; clinical exam as part of routine cancer-related checkup

^aSummary of the screening procedures recommended for the general population by the USPSTF and the ACS. These recommendations refer to asymptomatic persons who are not known to have risk factors, other than age or gender, for the targeted condition. ^bUSPSTF lettered recommendations are defined as follows: "A": The USPSTF recommends the service, because there is high certainty that the net benefit is substantial; "B": The USPSTF recommends the service, because there is high certainty that the net benefit is moderate or moderate certainty that the net benefit is moderate to substantial; "C": The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences; there is at least moderate certainty that the net benefit is small; "D": The USPSTF recommends against the service because there is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits; "I": The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service.

Abbreviations: ACS, American Cancer Society; USPSTF, U.S. Preventive Services Task Force.

However, more benign breast lesions were discovered and more breast biopsies were performed in the self-examination arm.

Genetic screening for *BRCA1* and *BRCA2* mutations and other markers of breast cancer risk has identified a group of women at high risk for breast cancer. Unfortunately, when to begin and the optimal frequency of screening have not been defined. Mammography is less sensitive at detecting breast cancers in women carrying *BRCA1* and *BRCA2* mutations, possibly because such cancers occur in younger women, in whom mammography is known to be less sensitive. MRI screening may be more sensitive than mammography in women at high risk due to genetic predisposition or in women with very dense breast tissue, but specificity may be lower. An increase in overdiagnosis may accompany the higher sensitivity. The impact of MRI on breast cancer mortality with or without concomitant use of mammography has not been evaluated in a randomized controlled trial.

CERVICAL CANCER Screening with Papanicolaou (Pap) smears decreases cervical cancer mortality. The cervical cancer mortality rate has fallen substantially since the widespread use of the Pap smear. With the onset of sexual activity comes the risk of sexual transmission of HPV, the fundamental etiologic factor for cervical cancer. Screening guidelines recommend regular Pap testing for all women who have reached the age of 21 (before this age, even in individuals that have begun sexual activity, screening may cause more harm than benefit). The recommended interval for Pap screening is 3 years. Screening more frequently adds little benefit but leads to important harms, including unnecessary procedures and overtreatment of transient lesions. Beginning at age 30, guidelines also offer the alternative of combined Pap smear and HPV testing for women. The screening interval for women who test normal using this approach may be lengthened to 5 years.

An upper age limit at which screening ceases to be effective is not known, but women age 65 years with no abnormal results in the previous 10 years may choose to stop screening. Screening should be discontinued in women who have undergone a hysterectomy with cervical excision for noncancerous reasons.

Although the efficacy of the Pap smear in reducing cervical cancer mortality has never been directly confirmed in a randomized, controlled setting, a clustered randomized trial in India evaluated the impact of one-time cervical visual inspection and immediate colposcopy, biopsy, and/or cryotherapy (where indicated) versus counseling on cervical cancer deaths in women age 30–59 years. After 7 years of follow-up, the age-standardized rate of death due to cervical cancer was 39.6 per 100,000 person-years in the intervention group versus 56.7 per 100,000 person-years in controls.

COLORECTAL CANCER Fecal occult blood testing (FOBT), digital rectal examination (DRE), rigid and flexible sigmoidoscopy, colonoscopy, and

computed tomography (CT) colonography have been considered for colorectal cancer screening. A meta-analysis of five randomized controlled trials demonstrated a 22% relative reduction in colorectal cancer mortality after 2 to 9 rounds of biennial FOBT at 30 years of follow-up; annual screening was shown to result in a greater colorectal cancer mortality reduction in a single trial (a 32% relative reduction). The sensitivity for FOBT is increased if specimens are rehydrated before testing, but at the cost of lower specificity. The false-positive rate for rehydrated FOBT is high; 1–5% of persons tested have a positive test. Only 2–10% of those with occult blood in the stool have cancer. The high false-positive rate of FOBT substantially increases the number of colonoscopies performed.

Fecal immunochemical tests (FIT) have higher sensitivity for colorectal cancer than nonrehydrated FOBT tests. Multi-targeted stool DNA testing is an emerging screening modality that combines FIT with testing for altered DNA biomarkers in cells that are shed into the stool. Although limited evidence demonstrates that it has a higher single-test sensitivity for colorectal cancer than fecal immunochemical testing alone, its specificity is much lower, resulting in a higher number of false-positive tests and follow-up colonoscopies. There are no studies evaluating its effects on colorectal cancer incidence, morbidity, or mortality.

A blood test for the methylated *SEPT9* gene associated with colorectal cancer is available. However, its sensitivity is low, no longitudinal data have been collected on its performance or efficacy, and it is not recommended as a first-line screening test.

Two meta-analyses of five randomized controlled trials of sigmoidoscopy (i.e., the NORCCAP, SCORE, PLCO, Telemark, and U.K. trials) found an 18% relative reduction in colorectal cancer incidence and a 28% relative reduction in colorectal cancer mortality. Participant ages ranged from 50 to 74 years, with follow-up ranging from 6 to 13 years. Diagnosis of adenomatous polyps by sigmoidoscopy should lead to evaluation of the entire colon with colonoscopy. The most efficient interval for screening sigmoidoscopy is unknown, but an interval of 5 years is often recommended. Case-control studies suggest that intervals of up to 15 years may confer benefit; the randomized U.K. trial demonstrated benefit with one-time screening.

One-time colonoscopy detects ~25% more advanced lesions (polyps >10 mm, villous adenomas, adenomatous polyps with high-grade dysplasia, invasive cancer) than one-time FOBT with sigmoidoscopy; comparative *programmatic* performance of the two modalities over time is not known. Perforation rates are about 4/10,000 for colonoscopy and 1/10,000 for sigmoidoscopy. Debate continues on whether colonoscopy is too expensive and invasive and whether sufficient provider capacity exists to be recommended as the preferred screening tool in standard-risk populations. Some observational studies suggest

that efficacy of colonoscopy to decrease colorectal cancer mortality is primarily limited to the left side of the colon.

CT colonography, if done at expert centers, appears to have a sensitivity for polyps ≥ 6 mm comparable to colonoscopy. However, the rate of extracolonic findings of abnormalities of uncertain significance that must nevertheless be worked up is high (~5–37%); the long-term cumulative radiation risk of repeated colonography screenings is also a concern.

LUNG CANCER Chest x-ray and sputum cytology have been evaluated in several randomized lung cancer screening trials. The most recent and largest ($n = 154,901$) of these, a component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, found that, compared with usual care, annual chest x-ray did not reduce the risk of dying from lung cancer (relative risk 0.99; 95% confidence interval 0.87–1.22) after 13 years. Low-dose CT has also been evaluated in several randomized trials. The largest and longest of these, the National Lung Screening Trial (NLST), was a randomized controlled trial of screening for lung cancer in ~53,000 persons age 55–74 years with a 30+ pack-year smoking history. It demonstrated a statistically significant relative reduction of about 15–20% in lung cancer mortality in the CT arm compared to the chest x-ray arm (or about 3 fewer deaths per 1000 people screened with CT). However, the harms include the potential radiation risks associated with multiple scans, the discovery of incidental findings of unclear significance, and a high rate of false-positive test results. Both incidental findings and false-positive tests can lead to invasive diagnostic procedures associated with anxiety, expense, and complications (e.g., pneumo- or hemothorax after lung biopsy). The NLST was performed at experienced screening centers, and the balance of benefits and harms may differ in the community setting at less experienced centers.

OVARIAN CANCER Adnexal palpation, transvaginal ultrasound (TVUS), and serum CA-125 assay have been considered for ovarian cancer screening. A large randomized controlled trial has shown that an annual screening program of TVUS and CA-125 in average-risk women does not reduce deaths from ovarian cancer (relative risk 1.21; 95% confidence interval 0.99–1.48). Adnexal palpation was dropped early in the study because it did not detect any ovarian cancers that were not detected by either TVUS or CA-125. A second large randomized trial that used a two-stage screening approach incorporating a risk of ovarian cancer algorithm which determined whether additional testing with CA-125 or TVUS was required. At 14 years of follow-up, there was no statistically significant reduction in ovarian cancer deaths. The risks and costs associated with the high number of false-positive results are impediments to routine use of these modalities for screening. In the PLCO trial, 10% of participants had a false-positive result from TVUS or CA-125, and one-third of these women underwent a major surgical procedure; the ratio of surgeries to screen-detected ovarian cancer was approximately 20:1. In September 2016, the FDA issued a safety communication recommending against using any screening test, including the risk of ovarian cancer algorithm, for ovarian cancer.

PROSTATE CANCER The most common prostate cancer screening modalities are digital rectal exam (DRE) and serum PSA assay. An emphasis on PSA screening has caused prostate cancer to become the most common nonskin cancer diagnosed in American males. This disease is prone to lead-time bias, length bias, and overdiagnosis, and substantial debate continues among experts as to whether screening should be offered unless the patient specifically asks to be screened. Virtually all organizations stress the importance of informing men about the uncertainty regarding screening efficacy and the associated harms. Prostate cancer screening clearly detects many asymptomatic cancers, but the ability to distinguish tumors that are lethal but still curable from those that pose little or no threat to health is limited, and randomized trials indicate that the effect of PSA screening on prostate cancer mortality across a population is, at best, small. Men older than age 50 years have a high prevalence of indolent, clinically insignificant prostate cancers (about 30–50% of men, increasing further as men age).

Two major randomized controlled trials of the impact of PSA screening on prostate cancer mortality have been published. The PLCO

Cancer Screening Trial was a multicenter U.S. trial that randomized almost 77,000 men age 55–74 years to receive either annual PSA testing for 6 years or usual care. At 13 years of follow-up, no statistically significant difference in the number of prostate cancer deaths were noted between the arms (rate ratio 1.09; 95% confidence interval 0.87–1.36). More than half of men in the control arm received at least one PSA test during the trial, which may have potentially diluted a small effect.

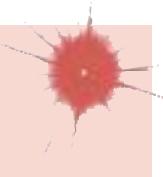
The European Randomized Study of Screening for Prostate Cancer (ERSPC) was a multinational study that randomized ~182,000 men between age 50 and 74 years (with a predefined “core” screening group of men age 55–69 years) to receive PSA testing or no screening. Recruitment and randomization procedures, as well as actual frequency of PSA testing, varied by country. After a median follow-up of 13 years, a 21% relative reduction in the risk of prostate cancer death in the screened arm was noted in the “core” screening group. The trial found that 781 (95% CI 490–1,929) men would need to be invited to screening, and 27 (95% CI 17–66) cases of prostate cancer detected, to avert 1 death from prostate cancer. Of the seven countries included in the mortality analysis, two demonstrated statistically significant reductions in prostate cancer deaths, whereas five did not. There was also an imbalance in treatment between the two study arms, with a higher proportion of men with clinically localized cancer receiving radical prostatectomy in the screening arm and receiving it at experienced referral centers.

Screening must be linked to effective therapy in order to have any benefit. In a trial conducted in the United States after the initiation of widespread PSA testing, random assignment to radical prostatectomy compared with “watchful waiting” did not result in a statistically significant decrease in prostate cancer deaths (absolute risk reduction 2.7%; 95% confidence interval—1.3 to 6.2%). Likewise, in a randomized trial conducted in the U.K. comparing monitoring (no curative treatment) to radical prostatectomy and to radiotherapy in men diagnosed in a screening program, prostate-cancer specific survival was very good (about 99%), and nearly identical, in all three study arms at a median of 10 years follow-up. Treatments for low-stage prostate cancer, such as surgery and radiation therapy, can cause substantial morbidity, including impotence and urinary incontinence.

SKIN CANCER Visual examination of all skin surfaces by the patient or by a health care provider is used in screening for basal and squamous cell cancers and melanoma. No prospective randomized study has been performed to look for a mortality decrease. Unfortunately, screening is associated with a substantial rate of overdiagnosis.

FURTHER READING

- CARTER JL, COLETTI RJ, HARRIS RP: Quantifying and monitoring overdiagnosis in cancer screening: A systematic review of methods. *BMJ* 350:g7773, 2015.
- CHUBAK J et al: Aspirin for the prevention of cancer incidence and mortality: Systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med* 164:814, 2016.
- FUTURE II STUDY GROUP: Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med* 356(19):1915, 2007.
- HAMDY FC et al: 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med* 375:1415, 2016.
- HUMPHREY LL et al: Screening for lung cancer with low-dose computed tomography: A systematic review to update the U.S. Preventive Services Task Force recommendation. *Ann Intern Med* 159:411, 2013.
- ILIC D et al: Screening for prostate cancer. *Cochrane Database of Systematic Reviews* 0.1002/14651858.CD004720.pub3, 2013.
- KRAMER BS, CROSWELL JM: Cancer screening: The clash of science and intuition. *Annu Rev Med* 60:125, 2009.
- LIN JS et al: Screening for colorectal cancer: Updated evidence report and systematic review for the U.S. Preventive Services Task Force. *JAMA* 315:2576, 2016.
- PACE LE, KEATING NL: A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA* 311:1327, 2014.
- PEIRSON L et al: Screening for cervical cancer: A systematic review and meta-analysis. *Syst Rev* 2:35, 2013.



CANCER IS A GENETIC DISEASE

Cancer arises through a series of somatic alterations in DNA that result in unrestrained cellular proliferation. Most of these alterations involve subtle sequence changes in DNA (i.e., mutations). The somatic mutations may originate as a consequence of random replication errors or exposure to carcinogens (e.g., radiation) and can be exacerbated by faulty DNA repair processes. While most cancers arise sporadically, clustering of cancers occurs in families that carry a germline mutation in a cancer gene.

HISTORICAL PERSPECTIVE

The idea that cancer progression is driven by sequential somatic mutations in specific genes has only gained general acceptance in the past 30 years. Before the advent of the microscope, cancer was believed to be composed of aggregates of mucus or other noncellular matter. By the middle of the nineteenth century, it became clear that tumors were masses of cells and that these cells arose from the normal cells of the tissue from which the cancer originated. The molecular basis for the uncontrolled proliferation of cancer cells was to remain a mystery for another century. During that time, a number of theories for the origin of cancer were postulated. The great biochemist Otto Warburg proposed the combustion theory of cancer, which stipulated that cancer was due to abnormal oxygen metabolism. Others believed that all cancers were caused by viruses, and that cancer was in fact a contagious disease.

In the end, observations of cancer occurring in chimney sweeps, studies of x-rays, and the overwhelming data demonstrating cigarette smoke as a causative agent in lung cancer, together with Ames's work on chemical mutagenesis, were consistent with the idea that cancer originated through changes in DNA. However, it was not until the somatic mutations responsible for cancer were identified at the molecular level that the genetic basis of cancer was definitively established. Although the viral theory of cancer did not prove to be generally accurate (with the exception of human papillomaviruses, which can cause cervical and other cancers), the study of retroviruses led to the discovery of the first human *oncogenes* in the late 1970s. Oncogenes are one of the two major classes of cancer driver genes. The study of families with genetic predisposition to cancer was instrumental to the discovery of the other major class of cancer driver genes, called *tumor-suppressor genes*. Current technologies permit the sequence analysis of entire cancer genomes, and provide a comprehensive view of the genetic changes that cause tumors to arise and become malignant. The field that studies the various types of mutations, as well as the consequences of these mutations in tumor cells, is now known as *cancer genetics*.

THE CLONAL ORIGIN AND MULTISTEP NATURE OF CANCER

Nearly all cancers originate from a single cell; this clonal origin is a critical discriminating feature between neoplasia and hyperplasia. Multiple cumulative mutational events are invariably required for the progression of a tumor from normal to fully malignant phenotype. The process can be seen as Darwinian microevolution in which, at each successive step, the mutated cells gain a growth advantage resulting in the expansion of a neoplastic clone (Fig. 67-1). Based on observations of cancer frequency increases during aging, the epidemiologists Armitage and Doll and Nordling independently proposed that cancer is a result of three discrete cellular changes. Remarkably, this early model has been validated by extensive sequencing of cancer genomes. These studies revealed that just three causal mutations are required for the development of several of the most common cancers. Overall, it is currently believed that most common solid tumors require a minimum of three mutated cancer driver genes (either oncogenes or tumor suppressor genes) for their development. One or two mutations is

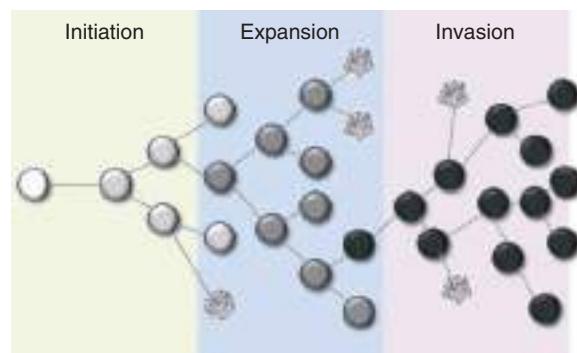


FIGURE 67-1 Multistep clonal development of malignancy. In this diagram a series of three cumulative mutations, each with a modest growth advantage acting alone, eventually results in a malignant tumor. Note that not all such alterations result in progression. The actual number of cumulative mutations necessary to transform from the normal to the malignant state has been estimated to be three for several of the most common types of cancer. (After P Nowell: Science 194:23, 1976, with permission.)

sufficient for benign tumorigenesis, but not for the invasive capacity that distinguishes cancers from benign tumors. Less common tumors, such as liquid tumors (leukemias or lymphomas), sarcomas, and childhood tumors, require two driver gene alterations for malignancy. Note that a cancer driver gene is best defined as one containing a mutation that increases the selective growth advantage of the cell containing it. Normally, cell birth and cell death are in perfect equilibrium; every time a cell is born, another in the same lineage dies. Cancer driver gene mutations alter this equilibrium, so that more cells are born than die. The imbalance is often slight, so that the difference between cell birth and cell death is <1%. This explains why tumorigenesis—the journey from a normal cell to a typical malignant, solid tumor—often takes decades.

We now know the precise nature of the genetic alterations responsible for nearly all malignancies and are beginning to understand how these alterations promote the distinct stages of tumor growth. The prototypical example is colon cancer, in which analyses of genomes from the entire spectrum of neoplastic growths—from normal colon epithelium through adenoma to carcinoma—have identified mutations that are highly characteristic of each type of lesion (Fig. 67-2).

TWO TYPES OF CANCER GENES: ONCOGENES AND TUMOR-SUPPRESSOR GENES

As briefly mentioned above, there are two major types of cancer genes. The first type comprises genes that positively influence growth and are known as *tumor-suppressor genes*. Both oncogenes and tumor-suppressor genes exert their effects on tumor growth through their ability to determine cell fates, influence cell survival and contribute to genome maintenance. The underlying molecular mechanisms can be extremely complex. While tightly regulated in normal cells, oncogenes acquire mutations that typically relieve this control and lead to increased activity of the gene products. This activating mutational event occurs in a single allele and acts in a dominant fashion. In contrast, the normal function of tumor-suppressor genes is usually to restrain cell growth, and this function is lost in cancer. Because of the diploid nature of mammalian cells, both alleles must be inactivated for a cell to completely lose the function of a tumor-suppressor gene. Thus, it requires two genetic events to inactivate a tumor-suppressor gene mutation, while only one genetic event is required to activate an oncogene.

A subset of tumor-suppressor genes controls the ability of the cell to maintain the integrity of its genome. Cells with a deficiency in these genes acquire an increased number of mutations throughout their genomes, including those in oncogenes and tumor-suppressor genes. This “mutator” phenotype was first hypothesized by Loeb to explain how the multiple rare mutational events required for tumorigenesis can occur in the lifetime of an individual. A mutator phenotype underlies several forms of cancer, such as those associated with deficiencies

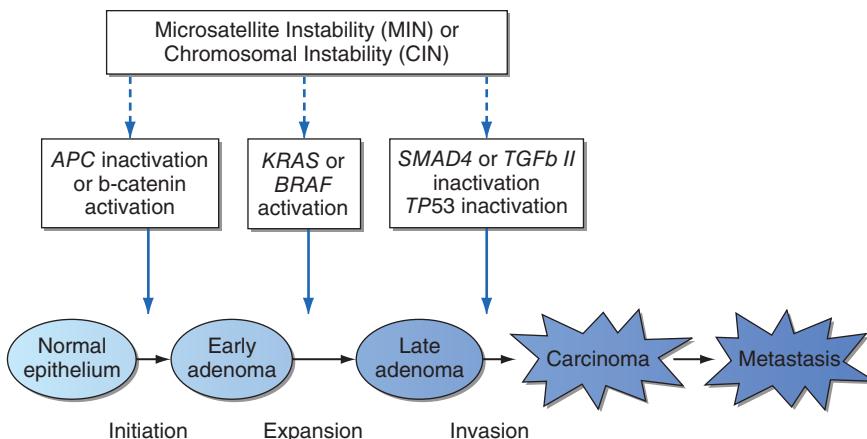


FIGURE 67-2 Progressive somatic mutational steps in the development of colon carcinoma. The accumulation of alterations in a number of different genes results in the progression from normal epithelium through adenoma to full-blown carcinoma. Genetic instability (microsatellite or chromosomal) accelerates the progression by increasing the likelihood of mutation at each step. Patients with familial polyposis are already one step into this pathway, because they inherit a germline alteration of the APC gene. TGF, transforming growth factor.

in DNA mismatch repair. The great majority of cancers do not harbor repair deficiencies, and their rate of mutation is similar to that observed in normal cells. Many of these cancers, however, appear to harbor a different kind of genetic instability, affecting the loss or gains of whole chromosomes or large parts thereof (as explained in more detail below).

ONCOGENES IN HUMAN CANCER

Work by Peyton Rous in the early 1900s revealed that a chicken sarcoma could be transmitted from animal to animal in cell-free extracts, suggesting that cancer could be induced by an agent acting positively to promote tumor formation. The agent responsible for the transmission of the cancer was a retrovirus (Rous sarcoma virus, RSV) and the oncogene responsible was identified 75 years later as *V-SRC*. Other oncogenes were also discovered through their presence in the genomes of retroviruses that are capable of causing cancers in chickens, mice, and rats. The non-mutated cellular homologues of these viral genes are called proto-oncogenes and are often targets of mutation or aberrant regulation in human cancer. Whereas many oncogenes were discovered on the basis of their presence in retroviruses, other oncogenes, particularly those involved in translocations characteristic of particular leukemias and lymphomas, were identified through genomic approaches. Investigators cloned the sequences surrounding the chromosomal translocations observed cytogenetically and identified the genes activated at the breakpoints (see below). Some of these were oncogenes previously found in retroviruses (like *ABL*, involved in chronic myeloid leukemia [CML]), whereas others were new (like *BCL2*, involved in B-cell lymphoma). In the normal cellular environment, proto-oncogenes have crucial roles in cell proliferation and differentiation. **Table 67-1** is a partial list of oncogenes known to be involved in human cancer.

The normal growth and differentiation of cells is controlled by growth factors that bind to receptors on the surface of the cell. The signals generated by the membrane receptors are transmitted inside the cells through signaling cascades involving kinases, G proteins, and other regulatory proteins. Ultimately, these signals affect the activity of transcription factors in the nucleus, which regulate the expression of genes crucial in cell proliferation, cell differentiation, and cell death. Oncogene products have been found to function at critical steps in these pathways (**Chap. 68**). Inappropriate activation of these pathways can lead to tumorigenesis.

MECHANISMS OF ONCOGENE ACTIVATION

POINT MUTATION

Point mutation (alternatively known as single nucleotide substitution) is a common mechanism of oncogene activation. For example, mutations in KRAS are present in > 95% of pancreatic cancers and 40% of colon cancers but are less common in other cancer types, although they can occur at significant frequencies in leukemia, lung, and thyroid cancers. Remarkably—and in contrast to the diversity of mutations found in tumor-suppressor genes—most of the activated KRAS alleles contain point mutations in codons 12, 13, or 61. These mutations reduce RAS GTPase activity, leading to constitutive activation of the mutant RAS protein. The restricted pattern of mutations observed in oncogenes compared to that of tumor-suppressor genes reflects the fact that gain-of-function mutations must occur at specific sites, while a broad variety of mutations can lead to loss of activity. Indeed, inactivation of a gene can in theory be

accomplished through the introduction of a stop codon anywhere in the coding sequence, whereas activations require precise substitutions at residues that can somehow lead to an increase in the activity of the encoded protein under particular circumstances within the cell.

DNA AMPLIFICATION

The second mechanism for activation of oncogenes is DNA sequence amplification, leading to overexpression of the gene product. This increase in DNA copy number may cause cytologically recognizable chromosome alterations referred to as *homogeneously staining regions* (HSRs) if integrated within chromosomes, or *double minutes* (dmins) if extrachromosomal. The recognition of DNA amplification is accomplished through various DNA sequence-based methods for copy number analysis. With both microarray and sequencing technologies, the entire genome can be surveyed for gains and losses of DNA sequences, thus pinpointing chromosomal regions likely to contain genes important in the development or progression of cancer.

Numerous genes have been reported to be amplified in cancer. Several of these genes, including *NMYC* and *LMYC*, were identified through their presence within the amplified DNA sequences of a tumor and had homology to known oncogenes. Because the region amplified often includes hundreds of thousands of base pairs, multiple oncogenes may be amplified in a single amplicon in some cancers.

TABLE 67-1 Oncogenes Commonly Altered in Human Cancers

ONCOGENE	FUNCTION	ALTERATION IN CANCER	NEOPLASM
<i>AKT1</i>	Serine/threonine kinase	Point mutation	Skin
<i>BRAF</i>	Serine/threonine kinase	Point mutation	Melanoma, thyroid, colorectal
<i>CCND1</i>	Cell cycle progression	Amplification	Esophageal, head and neck
<i>CTNNB1</i>	Signal transduction	Point mutation	Colon, liver, uterine, melanoma
<i>EGFR</i>	Signal transduction	Point mutation	Lung
<i>FLT3</i>	Signal transduction	Point mutation	AML
<i>IDH1</i>	Chromatin modification	Point mutation	Glioma
<i>MDM2</i>	Inhibitor of p53	Amplification	Sarcoma, glioma
<i>MDM4</i>	Inhibitor of p53	Amplification	Breast
<i>MYC</i>	Transcription factor	Amplification	Prostate, ovarian, breast, liver, pancreatic
<i>MYCL1</i>	Transcription factor	Amplification	Ovarian, bladder
<i>MYCN</i>	Transcription factor	Amplification	Neuroblastoma
<i>PIK3CA</i>	Phosphoinositol-3-kinase	Point Mutation	Multiple cancers
<i>KRAS</i>	GTPase	Point mutation	Pancreatic, colorectal, lung
<i>NRAS</i>	GTPase	Point mutation	Melanoma

Abbreviation: AML, acute myeloid leukemia.

TABLE 67-2 Representative Oncogenes at Chromosomal Translocations

GENE (CHROMOSOME)	TRANSLOCATION	MALIGNANCY
BCR-ABL	(9;22)(q34;q11)	Chronic myeloid leukemia
BCL1 (11q13.3)-IgH (14q32)	(11;14)(q13;q32)	Mantle cell lymphoma
BCL2 (18q21.3)-IgH (14q32)	(14;18)(q32;q21)	Follicular lymphoma
FLI-EWSR1	(11;22)(q24;q12)	Ewing's sarcoma
LCK-TCRB	(1;7)(p34;q35)	T-cell acute lymphocytic leukemia
PAX3-FOXO1	(2;13)(q35;q14)	Rhabdomyosarcoma
PAX8-PPARG	(2;3)(q13;p25)	Thyroid
IL21R-BCL6	(3;16)(q27;p11)	Non-Hodgkin's lymphoma
TAL1-TCTA	(1;3)(p34;p21)	Acute T cell leukemia
TMPRSS2-ERG	Rearrangement on Chr21q22	Prostate

(particularly in sarcomas). Indeed, *MDM2*, *GLI*, *CDK4*, and *TPSPAN31* at chromosomal location 12q13-15 have been shown to be co-amplified in several types of sarcomas and other tumors. Amplification of a cellular gene is often a predictor of poor prognosis; for example, *ERBB2/HER2* and *NMYC* are often amplified in aggressive breast cancers and neuroblastoma, respectively.

CHROMOSOMAL REARRANGEMENT

Chromosomal alterations provide important clues to the genetic changes in cancer. The chromosomal alterations in human solid tumors such as carcinomas are heterogeneous and complex and occur as a result of the frequent chromosomal instability observed in these tumors (see below). In contrast, the chromosome alterations in myeloid and lymphoid tumors are often simple translocations, that is, reciprocal transfers of chromosome arms from one chromosome to another. The breakpoints of recurring chromosome abnormalities usually occur at the site of cellular oncogenes. **Table 67-2** lists representative examples of recurring chromosome alterations in malignancy and the associated gene(s) rearranged or deregulated by the chromosomal rearrangement. Translocations are often observed in liquid tumors in general and are particularly common in lymphoid tumors, probably because these cell types have the capability to rearrange their DNA to generate antigen

receptors. Indeed, antigen receptor genes are commonly involved in the translocations, implying that an imperfect regulation of receptor gene rearrangement may be involved in their pathogenesis. In addition to transcription factors and signal transduction molecules, translocation may result in the overexpression of cell cycle regulatory proteins or proteins such as cyclins and of proteins that regulate cell death. Recurrent translocations have more recently been identified in solid tumors such as prostate cancers. Fusions between *TMPRSS2* and *ERG*, which are normally located in tandem on chromosome 21, contribute to about one-third of all prostate cancers and correlate with more aggressive disease.

The first reproducible chromosome abnormality detected in human malignancy was the Philadelphia chromosome detected in CML. This cytogenetic abnormality is generated by reciprocal translocation involving the *ABL* oncogene on chromosome 9, encoding a tyrosine kinase, being placed in proximity to the breakpoint cluster region (BCR) gene on chromosome 22. **Figure 67-3** illustrates the generation of the translocation and its protein product. The consequence of expression of the *BCR-ABL* gene product is the activation of signal transduction pathways leading to cell growth independent of normal external signals. Imatinib (marketed as Gleevec), a drug that specifically blocks the activity of Abl tyrosine kinase, has shown remarkable efficacy with little toxicity in patients with CML. The successful targeting of *BCR-ABL* by imatinib is the paradigm for molecularly targeted anti-cancer therapies.

CHROMOSOMAL INSTABILITY IN SOLID TUMORS

Solid tumors generally contain an abnormal number of chromosomes, a state known as aneuploidy; chromosomes from aneuploid tumors exhibit structural alterations such as translocations, deletions, and amplifications. These abnormalities reflect an underlying defect in cancer cells known as chromosomal instability. While aneuploidy is a striking cellular phenotype, chromosomal instability is manifest as only a small increase in the tendency of cells to gain, lose, or rearrange chromosomes during any given cell cycle. This intrinsically low rate of chromosome aberration implies that cancer cells become aneuploid only after many generations of clonal expansion. The molecular basis of aneuploidy remains incompletely understood. It is widely believed that defects in checkpoints, the quality-control mechanisms that halt the cell cycle if chromosomes are damaged or misaligned, contribute to chromosomal instability. This hypothesis emerged from experimental

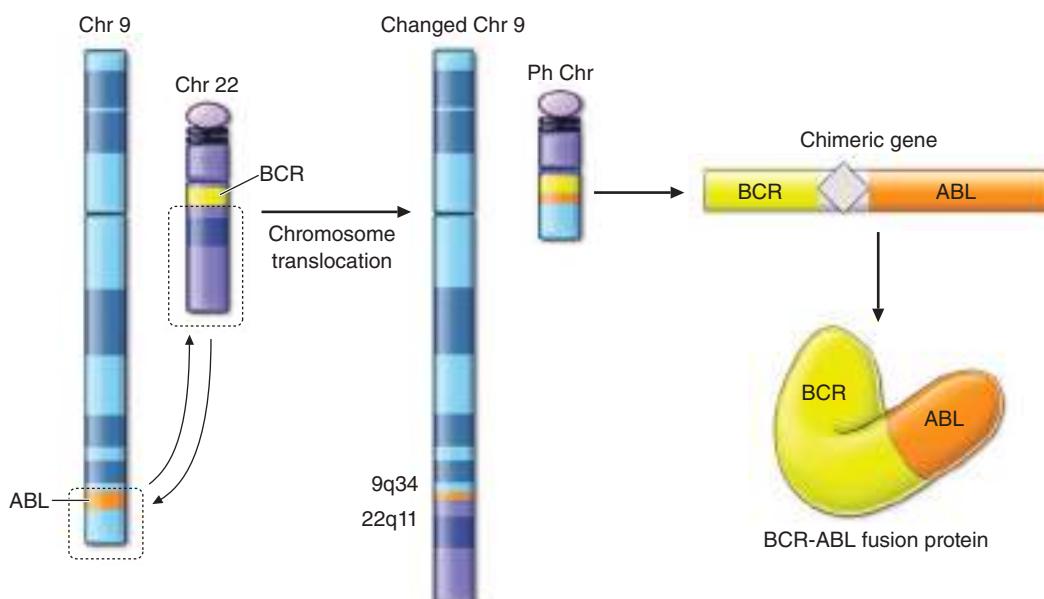


FIGURE 67-3 Specific translocation seen in chronic myeloid leukemia (CML). The Philadelphia chromosome (Ph) is derived from a reciprocal translocation between chromosomes 9 and 22 with the breakpoint joining the sequences of the *ABL* oncogene with the *BCR* gene. The fusion of these DNA sequences allows the generation of an entirely novel fusion protein with modified function.

observations that the tumor suppressor p53 controls checkpoints that regulate the initiation of DNA replication and the onset of mitosis. These processes are therefore defective in many cancer cells. The mitotic spindle checkpoint, which ensures proper chromosome attachment to the mitotic spindle before allowing the sister chromatids to separate, is also altered in some cancers, irrespective of p53 status. The precise relationship between checkpoint deficiency and chromosomal instability remains unclear, but it is believed that even a subtle perturbation of the highly orchestrated process of cell division can impact the ability of a cell to faithfully replicate and segregate its complement of chromosomes. From a therapeutic standpoint, the checkpoint defects that are prevalent in cancers have been proposed as vulnerabilities that may be exploited by novel agents and combinatorial strategies.

In contrast to the genome-wide cytogenetic changes that are typical indications of an underlying chromosomal instability, more focal patterns of chromosomal rearrangement have been recurrently detected in several cancer types. A curious phenomenon known as *chromothripsis* causes dozens of distinct breakpoints that are localized on one or several chromosomes. These striking structural alterations are thought to reflect a single event in which a chromosome is fragmented and then imprecisely reassembled. While the exact process that underlies chromothripsis remains obscure, and its effects on driver genes is not yet clear, a transient period of extreme instability stands in contrast to the gradual loss, gain and rearrangement of chromosomes that is typically observed in serially cultured cancer cells.

TUMOR-SUPPRESSOR GENE INACTIVATION IN CANCER

The first indication of the functional existence of tumor-suppressor genes came from experiments showing that fusion of mouse cancer cells with normal mouse fibroblasts led to a nonmalignant phenotype in the fused cells. The normal role of tumor-suppressor genes is to

restrain cell growth, and the function of these genes is inactivated in cancer. The three major types of somatic lesions observed in tumor-suppressor genes during tumor development are *point mutations*, small insertions and/or deletions known as *indels*, and *large deletions*. Point mutations or indels in the coding region of tumor-suppressor genes will frequently lead to truncated protein products or allele-specific loss of RNA expression by the process of *nonsense-mediated decay*. Unlike the highly recurrent point mutations that are found in critical positions of activated oncogenes, known as mutational *hotspots*, the point mutations that cause tumor-suppressor gene inactivation tend to be distributed throughout the open reading frame. Large deletions lead to the loss of a functional product and sometimes encompass the entire gene or even the entire chromosome arm, leading to loss of heterozygosity (LOH) in the tumor DNA compared to the corresponding normal tissue DNA (Fig. 67-4). LOH in tumor DNA often indicates the presence of a tumor-suppressor gene at a particular chromosomal location, and LOH studies have been useful in the positional cloning of many tumor-suppressor genes. The rate of LOH is increased in the presence of chromosomal instability, a relationship that would account for the high prevalence of aneuploidy in late-stage cancers.

Gene silencing, an epigenetic change that leads to the loss of gene expression, occurs in conjunction with hypermethylation of the promoter and histone deacetylation, and is another mechanism of tumor-suppressor gene inactivation. An *epigenetic modification* refers to a covalent modification of chromatin, heritable by cell progeny that may involve DNA but does not involve a change in the DNA sequence. The inactivation of the second X chromosome in female cells is an example of an epigenetic silencing that prevents gene expression from the inactivated chromosome. Genomic regions of hypermethylated and hypomethylated DNA can be detected by specialized techniques, and a subset of these regional modifications has consequences on the cell's behavior.

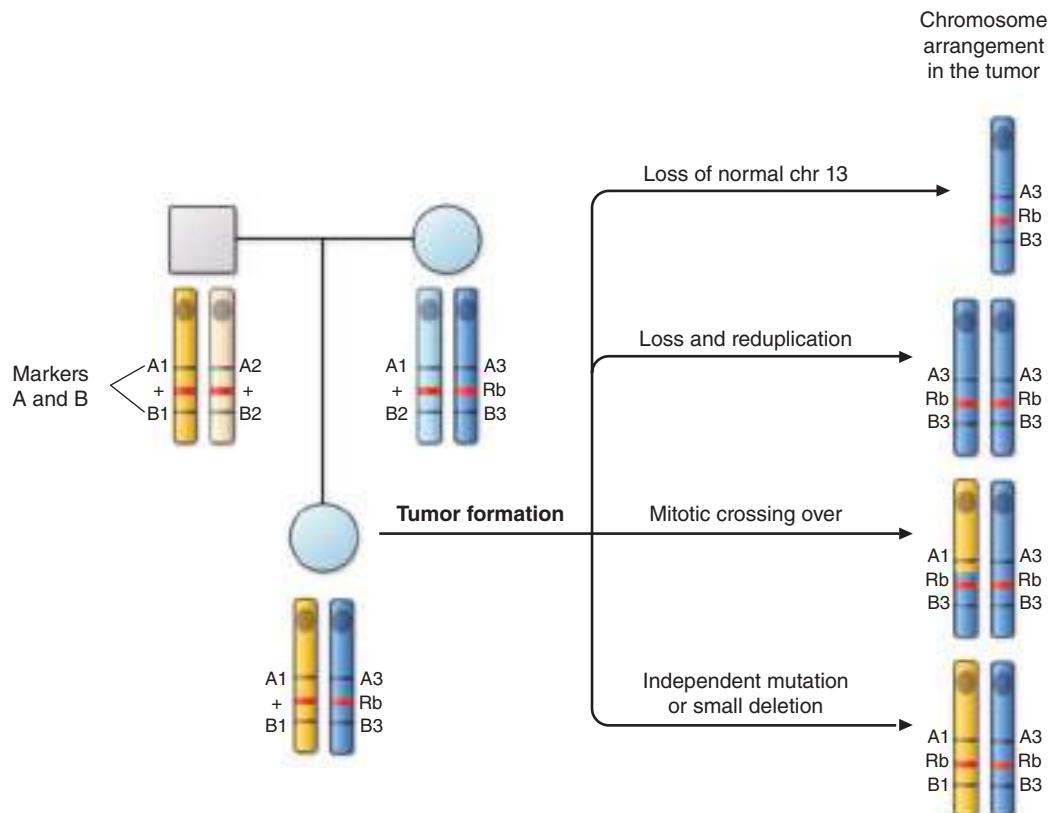


FIGURE 67-4 Diagram of possible mechanisms for tumor formation in an individual with hereditary (familial) retinoblastoma. On the left is shown the pedigree of an affected individual who has inherited the abnormal (Rb) allele from her affected mother. The normal allele is shown as a (+). The four chromosomes of her two parents are drawn to indicate their origin. Flanking the retinoblastoma locus are genetic markers (A and B) also analyzed in this family. Markers A3 and B3 are on the chromosome carrying the retinoblastoma disease gene. Tumor formation results when the normal allele, which this patient inherited from her father, is inactivated. On the right are shown four possible ways in which this could occur. In each case, the resulting chromosome 13 arrangement is shown. Note that in the first three situations, the normal allele (B1) has been lost in the tumor tissue, which is referred to as loss of heterozygosity (LOH) at this locus.

FAMILIAL CANCER SYNDROMES

A small fraction of cancers occurs in patients with a genetic predisposition. Based on studies of inherited and sporadic forms of retinoblastoma, Knudson and others formulated a hypothesis that explains the differences between sporadic and inherited forms of the same tumor type. In inherited forms of cancer, called *cancer predisposition syndromes*, one allele of a particular tumor suppressor gene is inherited in mutant form. This germline mutation is not sufficient to initiate a tumor, however; the other allele, inherited from the unaffected parent, must become somatically mutated in a normal stem cell for tumorigenesis to be initiated. In sporadic (non-inherited) forms of the same disease, all cells in the body start out with two normal copies of the tumor suppressor gene. A single cell must then sequentially acquire mutations in both alleles of the tumor suppressor gene to initiate a tumor. Thus bi-allelic mutations of the same tumor suppressor gene are required for both inherited and non-inherited forms of the disease; the only difference is that individuals with the inherited form have a “head-start”: they already have one allele mutated, from conception, and only need one additional mutation to initiate the process (Fig. 67-4). This distinction explains why those with inherited forms of the disease develop more cancers, at an earlier age, than the general population. It also explains why, even though every cell in an individual with a cancer predisposition syndrome has a mutant gene, only a relatively small number of tumors arise during his/her lifetime. The reason is that the vast majority of cells within such individuals are functionally normal because one of the two alleles of the tumor suppressor gene is normal. Mutations are uncommon events, and only the rare cells that develop a mutation in the remaining normal allele will exhibit uncontrolled proliferation. The

same principle applies to virtually all types of cancer predisposition syndromes, though the particular genes differ. For example, inherited mutations in *RB1*, *WT1*, *VHL*, *APC*, and *BRCA1* lead to predispositions to retinoblastomas, Wilms’ tumors, renal cell carcinomas, colorectal carcinomas, and breast carcinomas, respectively (Table 67-3). Also note that the biallelic inactivation of any of these genes is not sufficient to develop cancer; it requires other, additional somatic mutations for the initiating cells to evolve to malignancy, as noted above.

Roughly 100 familial cancer syndromes have been reported; the great majority are very rare. Most of these syndromes exhibit an autosomal dominant pattern of inheritance, although some of those associated with DNA repair abnormalities (xeroderma pigmentosum, Fanconi’s anemia, ataxia telangiectasia) are inherited in an autosomal recessive fashion. Table 67-3 shows a number of cancer predisposition syndromes and the responsible genes.

The next section examines inherited colon cancer predispositions in detail because several lessons of general importance have been derived from the study of these syndromes.

Familial adenomatous polyposis (FAP) is a dominantly inherited colon cancer syndrome caused by germline mutations in the adenomatous polyposis coli (*APC*) tumor-suppressor gene on chromosome 5. Affected individuals develop hundreds to thousands of adenomas in the colon. In each of these adenomas, the *APC* allele inherited from the affected parent has been inactivated by virtue of a somatic mutation (Fig. 67-2). This inactivation usually occurs through a gross chromosomal event resulting in loss of all or a large part of the long arm of chromosome 5, where *APC* resides. In other cases, the remaining allele is inactivated by a subtle intragenic mutation of *APC*, which as a single

TABLE 67-3 Cancer Predisposition Syndromes and Associated Genes

SYNDROME	GENE	CHROMOSOME	INHERITANCE	TUMORS
Ataxia telangiectasia	ATM	11q22-q23	AR	Breast
Autoimmune lymphoproliferative syndrome	FAS	10q24 1q23	AD	Lymphomas
	FASL			
Bloom’s syndrome	BLM	15q26.1	AR	Various
Cowden’s syndrome	PTEN	10q23	AD	Breast, thyroid
Familial adenomatous polyposis	APC	5q21	AD	Colorectal (early onset)
	MUTYH	1p34.1	AR	
Familial melanoma	CDKN2A	9p21	AD	Melanoma, pancreatic
Familial Wilms’ tumor	WT1	11p13	AD	Kidney (pediatric)
Hereditary breast/ovarian cancer	BRCA1	17q21	AD	Breast, ovarian, prostate
	BRCA2	13q12.3		
Hereditary diffuse gastric cancer	CDH1	16q22	AD	Stomach
Hereditary multiple exostoses	EXT1	8q24	AD	Exostoses, chondrosarcoma
	EXT2	11p11-12		
Hereditary retinoblastoma	RB1	13q14.2	AD	Retinoblastoma, osteosarcoma
Hereditary nonpolyposis colon cancer (HNPCC)	MSH2	2p16	AD	Colon, endometrial, ovarian, stomach, small bowel, ureter carcinoma
	MLH1	3p21.3		
	MSH6	2p16		
	PMS2	7p22		
Hereditary papillary renal carcinoma	MET	7q31	AD	Papillary kidney
Juvenile polyposis syndrome	SMAD4	18q21	AD	Gastrointestinal, pancreatic
	BMPR1A			
Li-Fraumeni syndrome	TP53	17p13.1	AD	Sarcoma, breast
Multiple endocrine neoplasia type 1	MEN1	11q13	AD	Parathyroid, endocrine, pancreas, and pituitary
Multiple endocrine neoplasia type 2a	RET	10q11.2	AD	Medullary thyroid carcinoma, pheochromocytoma
Neurofibromatosis type 1	NF1	17q11.2	AD	Neurofibroma, neurofibrosarcoma, brain
Neurofibromatosis type 2	NF2	22q12.2	AD	Vestibular schwannoma, meningioma, spine
Nevus basal cell carcinoma syndrome (Gorlin’s syndrome)	PTCH1	9q22.3	AD	Basal cell carcinoma, medulloblastoma, jaw cysts
Tuberous sclerosis	TSC1	9q34	AD	Angiofibroma, renal angiomyolipoma
	TSC2	16p13.3		
von Hippel-Lindau disease	VHL	3p25-26	AD	Kidney, cerebellum, pheochromocytoma

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

base substitution resulting in a nonsense codon. Gross chromosomal losses occur more commonly than point mutations in normal cells, explaining why these are the predominant mechanism underlying the inactivation of the normal allele of *APC*. The same is true for other cancer predisposition syndromes caused by other inherited tumor suppressor gene mutations; gross chromosomal events are generally responsible for inactivation of the tumor suppressor gene allele inherited from the non-affected parent. Several thousand adenomas form in FAP patients, and a small subset of the billions of cells within these adenomas will acquire a second mutation, leading to tumor progression, that is, a larger adenoma. A third mutation in such a larger adenoma may convert it to a carcinoma. If untreated (by colectomy), at least one of the adenomas will progress to cancer by the time patients are in their mid-40s. *APC* can be considered to be a gatekeeper for colon tumorigenesis in that the absence of mutation of this gatekeeper (or a gene acting within the same pathway), a colorectal tumor simply cannot be initiated. **Figure 67-5** shows the germline and somatic mutations found in the *APC* gene. A negative regulator of a signaling pathway that determines cell fate during development, the *APC* protein provides differentiation and apoptotic cues to colonic epithelial cells as they migrate up the crypts. Defects in this process can lead to abnormal accumulation of cells that would otherwise differentiate and eventually undergo apoptosis.

In contrast to patients with FAP, patients with hereditary nonpolyposis colon cancer (HNPCC, or Lynch's syndrome) do not develop polyposis, but instead develop only one or a small number of adenomas that rapidly progress to cancer. HNPCC is due to inherited mutations in one of four DNA mismatch repair genes (Table 67-3) that are components of a repair system responsible for correcting errors in newly replicated DNA. Germline mutations in *MSH2* and *MLH1* account for more than 90% of HNPCC cases, and mutations in *MSH6* and *PMS2* account for the remainder. When a somatic mutation inactivates the remaining wild-type allele of a mismatch repair gene, the cell develops a hypermutable phenotype characterized by profound genomic instability that is most readily apparent in short repeated sequences called *microsatellites* and is sometimes called microsatellite instability (MSI). The high rate of mutation in such cells impacts all genes, including oncogenes and tumor suppressor genes, and thereby accelerates the activation of the former and the inactivation of the latter

(Fig. 67-2). HNPCC can be considered a disease of tumor progression; once tumors are initiated (by an inactivating mutation of *APC* or by some other gene in the *APC* pathway), tumors rapidly progress because of the accelerated mutation rate. Progression from a tiny adenoma to carcinoma takes only a few years in HNPCC patients instead of the two or three decades this progression takes in patients with FAP (or in patients with sporadic colorectal tumors). Approximately half of HNPCC patients develop colorectal cancers by the time they are in their mid-40s—similar to that of FAP patients. This coincidence in age of onset emphasizes that both tumor initiation (abnormal in FAP patients) and tumor progression (abnormal in HNPCC patients) are the two pillars of cancer development and are equally important for cancer development.

Another general principle is apparent from the comparison between FAP and HNPCC patients. The tumors in FAP patients, like those in patients without hereditary predisposition to cancers, are chromosomal instability. MSI and chromosomal instability appear to be mutually exclusive in colon cancers, suggesting that they represent alternative mechanisms for the generation of genomic instability (Fig. 67-2). Other cancer types rarely exhibit MSI. Chromosomal instability is far more prevalent than MSI among all cancer types, perhaps explaining why nearly all cancers are aneuploid.

Although most autosomal dominant inherited cancer syndromes are due to mutations in tumor-suppressor genes (Table 67-3), there are a few interesting exceptions. Multiple endocrine neoplasia type 2, a dominant disorder characterized by pituitary adenomas, medullary carcinoma of the thyroid, and (in some pedigrees) pheochromocytoma, is due to gain-of-function mutations in the proto-oncogene *RET* on chromosome 10. Similarly, gain-of-function mutations in the tyrosine kinase domain of the *MET* oncogene lead to hereditary papillary renal carcinoma. Interestingly, loss-of-function mutations in the *RET* gene cause a completely different disease, Hirschsprung's disease (aganglionic megacolon [Chaps. 321 and 381]).

Although the heritable forms of cancer have taught us much about the mechanisms of growth control, most forms of cancer do not follow simple Mendelian patterns of inheritance. The majority of human cancers arise in a sporadic fashion, solely as a result of somatic mutation, and in the absence of any mutations in cancer-predisposing genes in their germlines.

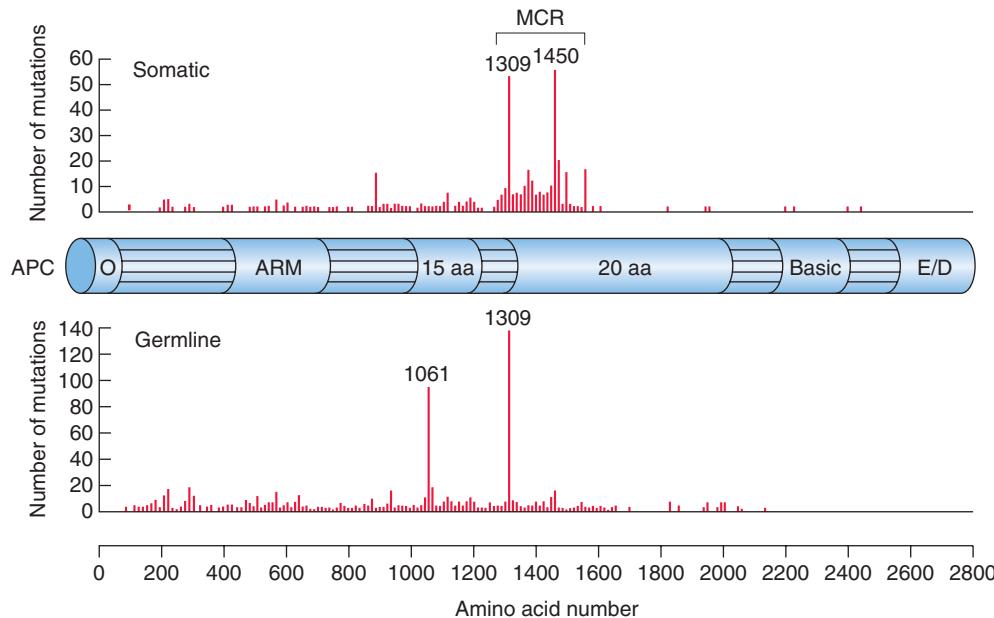


FIGURE 67-5 Germline and somatic mutations in the tumor-suppressor gene *adenomatous polyposis coli* (*APC*). *APC* encodes a 2843-amino-acid protein with six major domains: an oligomerization region (O), armadillo repeats (ARM), 15-amino-acid repeats (15 aa), 20-amino-acid repeats (20 aa), a basic region, and a domain involved in binding EB1 and the *Drosophila* discs large homologue (E/D). Shown are 650 somatic and 826 germline mutations representative of the mutations that occur within the *APC* gene (from the *APC* database at www.umd.be/APC). All known pathogenic mutations of *APC* result in the truncation of the *APC* protein. Germline mutations are found to be relatively evenly distributed up to codon 1600 except for two mutation hotspots surrounding amino acids 1061 and 1309, which together account for one-third of the mutations found in familial adenomatous polyposis (FAP) families.

458 GENETIC TESTING FOR FAMILIAL CANCER

The discovery of cancer susceptibility genes raises the possibility of DNA testing to predict the risk of cancer in individuals of affected families. An algorithm for cancer risk assessment and decision making in high-risk families using genetic testing is shown in Fig. 67-6. Once a mutation is discovered in a family, subsequent testing of asymptomatic family members can be crucial in patient management. A negative gene test in these individuals can prevent years of anxiety in the knowledge that their cancer risk is no higher than that of the general population. On the other hand, a positive test may lead to alteration of clinical management, such as increased frequency of cancer screening and, when feasible and appropriate, prophylactic surgery. Potential negative consequences of a positive test result include psychological distress (anxiety, depression) and discrimination, although the Genetic Information Nondiscrimination Act (GINA) makes it illegal for predictive genetic information to be used to discriminate in health insurance or employment. Testing should therefore not be conducted without counseling before and after disclosure of the test result.

Recent technological developments have made it feasible to obtain high-quality sequence of all of the protein-coding DNA sequences, and even of the entire genome, in any given individual. The redundant nature of modern DNA sequencing provides an extremely high level of sensitivity, such that mutations and polymorphisms will inevitably be identified in every subject. In patients lacking a clear family history, the significance of these DNA sequence findings will not be apparent. Even mutations in tumor suppressor genes are difficult to interpret unless there is an obvious functional implication, such as the truncation of the open reading frame, or that particular mutation has previously

been associated with cancer. Such germline mutations are very rare in the general population. Vastly more common are *variants of unknown significance* (VUS). VUS that are found during genetic testing cannot be used to evaluate the relative risk of cancer, but may nonetheless cause anxiety because they represent a deviation from the reference allele that is established as "normal." Because of the low yield of informative mutations that modify cancer risk and the frequent identification of VUS, it is generally not appropriate to use DNA sequencing to assess cancer risk in individuals unless the family history is suggestive of a germline mutation. Conversely, testing may be appropriate in some subpopulations with a known increased risk, even without a defined family history. For example, two mutations in the breast cancer susceptibility gene *BRCA1*, 185delAG and 5382insC, exhibit a sufficiently high frequency in the Ashkenazi Jewish population that genetic testing based on ethnicity alone may be warranted.

It is important that genetic test results be communicated to families by trained genetic counselors, especially for high-risk high-penetrance conditions such as the hereditary breast and ovarian cancer syndrome (*BRCA1/BRCA2*). To ensure that the families clearly understand its advantages and disadvantages and the impact it may have on disease management and psyche, genetic testing should never be done before counseling. Significant expertise is needed to communicate the results of genetic testing to families.

VIRUSES IN HUMAN CANCER

Several human malignancies are associated with viruses. Examples include Burkitt's lymphoma (Epstein-Barr virus; Chap. 189), hepatocellular carcinoma (hepatitis viruses), cervical cancer (human papillomavirus [HPV]; Chap. 193), and T cell leukemia (retroviruses; Chap. 196). There are several types of HPV, including the high-risk types 16 and 18 that are strongly associated with the development of cervical, vulvar, vaginal, penile, anal, and oropharyngeal cancer. The mechanisms of action of all these viruses involve inactivation of tumor suppressor genes. For example, HPV proteins E6 and E7 bind to and inactivate cellular tumor suppressors p53 and pRB, respectively. This is the reason that HPV is such a potent initiator of cancer: infection with a virus is tantamount to having two of the three mutant driver genes required for cancer, that is, one viral oncogene inactivates p53 and the other inactivates Rb. Though these two inactivated gene products are not sufficient for tumorigenesis, only one additional mutant gene is required to develop a malignancy.

CANCER GENOMES

The advent of relatively inexpensive technologies for rapid and high-throughput DNA sequencing has facilitated the comprehensive analysis of numerous genomes from many types of tumors. This unprecedented view into the genetic nature of cancer has provided remarkable insights. Most cancers do not arise in the context of a mutator phenotype, and accordingly the number of mutations in even the most advanced cancers is relatively modest. Common solid tumors harbor 30–70 subtle mutations that are non-synonymous (i.e., result in an amino acid change in the encoded protein). Liquid tumors such as lymphomas and leukemias, as well as pediatric tumors, typically have fewer than 20 mutations. The vast majority of the mutations detected in tumors are not functionally significant, they simply arose by chance in a single cell that gave rise to an expanding clone. Such mutations, which provide no selective advantage to the cell in which they occur, are known as *passenger* mutations. As noted above, only a small number of the mutations confer a selective growth advantage and thereby promote tumorigenesis. These functional mutations are known as *driver* mutations, and the genes in which they occur are called driver genes.

The frequency and distribution of driver mutations within a single tumor type can be represented as a topographical landscape (Fig. 67-7). The picture that emerges from these studies reveals that most genes that are mutated in tumors are actually mutated at relatively low frequencies, as would be expected of passenger genes, whereas a small number of genes (the driver genes) are mutated in a large proportion of tumors. There are a total of ~200 driver genes that

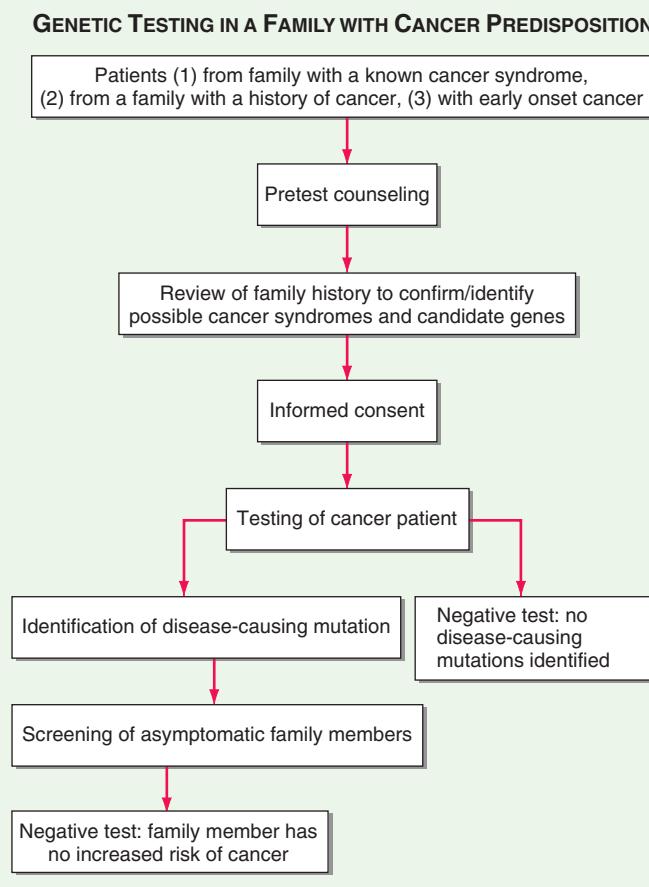


FIGURE 67-6 Algorithm for genetic testing in a family with cancer predisposition.

The key step is the identification of a disease-mutation in a cancer patient, which is an indication for the testing of asymptomatic family members. Asymptomatic family members who test positive may require increased screening or surgery, whereas those who test negative are at no greater risk for cancer than the general population. It should be emphasized that no molecular assay used for this sort of testing is 100% sensitive; negative results must be interpreted with this caveat in mind.

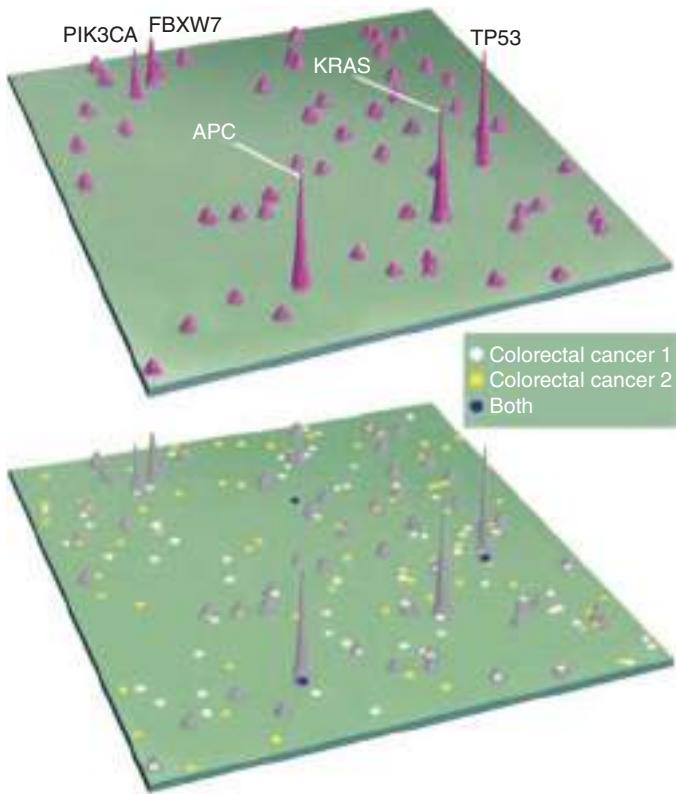


FIGURE 67-7 The mutational topography of colorectal cancer. The two-dimensional landscape represents the positions of the individual genes along the chromosomes. The height of each peak represents the mutation frequency at that locus. The top map is a representation of many sequenced colorectal cancers. The taller peaks represent the genes that are commonly mutated in colon cancer, while the smaller hills indicate the genes that are mutated at lower frequency. On the lower map, the mutations of two individual tumors are indicated. Note that there is little overlap between the mutated genes of the two colorectal tumors shown. These differences represent Type I heterogeneity, as noted in the text, which is the foundation for personalized medicine in cancer. (From LD Wood et al: Science 318:1108, 2007, with permission.)

are responsible for the development of all solid tumors, representing only ~1% of the total number of human protein-encoding genes. The majority of the mutations in these driver genes provide a direct selective growth advantage by altering the signaling pathways that mediate cell survival or the determination of cell fate. The remaining driver gene mutations indirectly provide a selective growth advantage by accelerating the mutation rate of proto-oncogenes and tumor suppressor genes. The functions of all these driver genes can be organized into a dozen signaling pathways, as shown in **Table 67-4**.

TUMOR HETEROGENEITY

The mutant cells that compose a single tumor are not genetically identical. Rather, cells obtained from different sites on a tumor will harbor common mutations as well as mutations that are unique to each sample. Genetic heterogeneity results from the ongoing acquisition of mutations during tumor growth. Each time a genome is replicated, there is a small but quantifiable probability that a mutation will spontaneously arise as a result of a replication error and be passed on to the cellular progeny. This is true in normal cells or in tumor cells. Any randomly chosen cell from the skin of one individual will harbor hundreds of genetic alterations that distinguish it from a different randomly chosen skin cell, and the same is true for all organs of self-renewing tissues. Tumors are actually *less* genetically heterogeneous than normal cells; any two randomly chosen cells from a tumor of an individual will have fewer differences than any two randomly chosen cells from that individual's normal tissues. The reason for this decrease in heterogeneity is clonal expansion, the fundamental feature of tumorigenesis. Every time a clonal expansion occurs, a genetic bottleneck wipes out heterogeneity among the cells that didn't expand; these unexpanded cells either die or form only a minute proportion of the total cells in the expanding tumor.

The mutations that vary between cells of a given tumor are invariably passenger mutations that arose since the last evolutionary bottleneck, that is, those mutations that arose during the expansion of the founder cell that gave rise to the final clonal expansion. In contrast, the passenger mutations that were present in the founder cell will be uniformly present in every cell in the tumor. In that respect, these passenger mutations that are not heterogeneously distributed, that is, those that are present in every cancer cell, are like the driver gene mutations, which are also present in virtually all cancer cells. The total number of mutations and their distribution within tumor cells therefore represents a complex interplay between the age of the patient (the older the patient, the more passenger mutations will have accumulated in the founding cell of the first clonal expansion) and the evolutionary history of the cancer (its age and number of clonal expansions it experienced).

Tumor heterogeneity has been recognized for decades at the cytogenetic, biochemical, and histopathologic levels. However, it is only recently, with the advent of a deep understanding of cancer genetics that genetic heterogeneity can be interpreted in a medically relevant fashion. The first important point to recognize about tumor heterogeneity is that it is only the variation in driver gene alterations that is important; the cellular distribution of passenger gene mutations is completely irrelevant. In this discussion of heterogeneity, we can expand the definition of "driver genes" to include those that provide a selective growth advantage in the face of therapy in addition to those that provide a selective growth advantage during tumor evolution, prior to treatment.

Type I heterogeneity refers to that among tumors of the same type from different patients (Fig. 67-8). Though adenocarcinomas of the lung generally harbor mutations in three or more driver genes, the genes differ among the patients and the precise mutations within the same gene can vary considerably. Type I heterogeneity is the basis for precision medicine, where the goal is to treat patients with drugs that target the proteins encoded by genetic alterations within their specific tumors. Type 2 heterogeneity refers to the genetic heterogeneity among different cells from the same primary tumor. Tumors continue to evolve as they grow, and different cells of the same cancer, in its original site (e.g., the colon), may acquire another driver gene mutations that are not shared among the other cells of the tumor. Such a mutation can result in a small clonal expansion that may or may not be important biologically. In cases in which the primary tumor can be surgically excised, such mutations are unimportant unless they give rise to Type III heterogeneity (described below). The reason they are important is because all primary tumor cells, whether homogeneous or not, are removed by the surgical procedure. In primary tumors that cannot be completely excised (such as most advanced brain tumors and many pancreatic ductal adenocarcinomas), heterogeneity is biomedically important because it can give rise to drug resistance, analogously to that described for Type IV heterogeneity (see below). Type III heterogeneity refers

TABLE 67-4 Signaling Pathways Altered in Cancer

PROCESS	PATHWAY	REPRESENTATIVE DRIVER GENES
Cell survival	Cell cycle regulation/apoptosis	<i>RB1, BCL2</i>
	RAS	<i>KRAS, BRAF</i>
	PIK3CA	<i>PTEN, PIK3CA</i>
	JAK/STAT	<i>JAK2, FLT3</i>
	MAPK	<i>MAP3K, ERK</i>
	TGF- β	<i>BMPR1A, SMAD4</i>
Cell fate	Notch	<i>NOTCH1, FBXW7</i>
	Hedgehog	<i>PTCH1, SMO</i>
	WNT/APC	<i>APC, CTNNB1</i>
	Chromatin modification	<i>DNMT1, IDH1</i>
	Transcriptional regulation	<i>AR, KLF4</i>
Genome maintenance	DNA damage signaling and repair	<i>ATM, BRCA1</i>

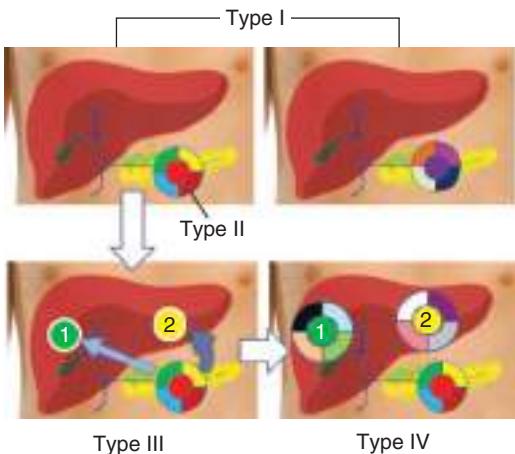


FIGURE 67-8 The four types of tumor heterogeneity. Tumor heterogeneity is the inevitable result of cell proliferation, as new mutations are introduced during clonal expansion. This concept is illustrated by a primary tumor in the pancreas and two metastatic tumors in the liver. The tumors of the founding populations are shown in the middle of each circle, while the distinct subclones are shown around the periphery. Type I: the heterogeneity of tumors that occur among different patients. Type II: the heterogeneity among the cells of a primary tumor, also known as intratumoral heterogeneity. Type III: the heterogeneity among the founding cells of distinct metastatic lesions (marked as 1 and 2) that arise in the same patient, also known as intermetastatic heterogeneity. Type IV: heterogeneity among the cells of each metastasis that develops as each tumor grows, also known as intrametastatic heterogeneity.

to the genetic differences among the founder cells of the metastatic lesions from the same patient. For example, a patient with melanoma may have 100 different metastases distributed throughout various organs. Only if a mutant *BRAF* is present in every founder cell of every metastasis, then the patient has a chance at a complete response to a *BRAF* inhibitor. There have been several recent detailed studies of the metastases from various tumor types. Fortunately, these studies suggest there is very little, if any, Type III heterogeneity among driver genes, a necessary prerequisite for the successful implementation of future targeted therapies. Finally, Type IV heterogeneity refers to that among cells of individual metastatic lesions. As the founder cell of each metastasis expands to become detectable, it acquires mutations, a small number of which can act as “drivers” if the patient is exposed to therapeutics. This type of heterogeneity is of major clinical importance, as it has been shown to be responsible for the development of resistance in virtually all targeted therapies. The development of such resistance is a fait accompli based simply on known mutation rates and known genetic resistance mechanisms. The only way to circumvent acquired resistance is to treat metastatic tumors earlier (i.e., in adjuvant setting, before much tumor expansion has occurred) or to treat with combinations of drugs for which cross-resistance is genetically impossible.

PERSONALIZED CANCER DETECTION AND TREATMENT

High-throughput DNA sequencing has led to an unprecedented understanding of cancer at the molecular level. A comprehensive mutation profile provides a molecular history of a given tumor and insights into how it arose. Because tumor cells and tumor DNA are shed into the blood and other bodily fluids, common driver mutations can be used as highly specific biomarkers for early detection. For diagnosed tumors, tumor-specific mutations can be used to estimate tumor burden, to assess treatment responses and to detect recurrence.

In some cases, information regarding specific genes and pathways that are altered provides patients and physicians with options for personalized therapy. This general approach is sometimes referred to

as *precision medicine*. Because tumor behavior is highly variable, even within a tumor type, personalized information-based medicine can supplement and perhaps eventually supplant histology-based tumor assessment, especially in the case of tumors that are resistant to conventional therapeutic approaches. Conversely, molecular nosology has revealed similarities in tumors of diverse histotype. The success of the precision medicine approach in any given patient depends on the presence of tumor-associated genetic alterations that are actionable (i.e., can be targeted with a specific drug). Examples of currently actionable changes include mutations in *BRAF* (targeted by the drug vemurafenib) and *RET* (targeted by sunitinib and sorafenib), and *ALK* rearrangements (targeted by crizotinib). At present, the proportion of tumors that can be treated with such precision medicine approaches is small, but future therapeutic development will hopefully change this situation. The development of new targeted agents is at present hindered by the fact that such agents can only target activated oncogenes, while the great majority of genetic alterations in common solid tumors are those that inactivate tumor suppressor genes. Because all drugs, whether for use in oncology or any other purpose, can only inhibit protein actions, drugs cannot be used to directly target the proteins encoded by inactivated tumor suppressor genes; these proteins are already inactive. More information about the pathways through which tumor suppressor genes act may provide a way around this obstacle. For example, when a tumor suppressor gene is inactivated, some downstream component of the pathway is likely to be activated, thereby presenting a realistic target. An example of this is provided by PARP-1 inhibitors, which have been successfully used to treat patients whose tumors have inactivating mutations of genes involved in DNA repair processes, such as *BRCA1*. Patterns of global gene expression can be used to help unravel such pathways and are already being used to predict drug sensitivities and provide prognostic information in addition to that provided by DNA sequence analysis. Evaluation of proteomic and metabolomics patterns may also prove useful.

THE FUTURE

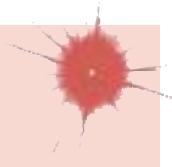
A revolution in cancer genetics has occurred in the past 30 years. Most types of cancer are now understood at the DNA sequence level and this accomplishment has led us to an increasingly refined understanding of tumorigenesis. Cancer gene mutations have proven to be reliable biomarkers for cancer detection and monitoring as well as for informing therapeutics through precision medicine approaches. Gene-based tests are already standard of care for certain tumor types, such as melanoma, colorectal and pancreatic cancers, and the utility of these tests will undoubtedly be expanding greatly in the coming years as new therapies and ways of predicting responses to therapies are developed. While effective treatment of advanced cancers remains difficult, it is expected that breakthroughs in these areas will continue to emerge and be applicable to an ever-increasing number of cancers. Moreover, with the hoped-for advances in diagnostics, particularly in the earlier detection of cancers, the new and old therapies for cancer can be expected to have a much greater impact on reducing cancer deaths.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the past contributions of Pat J. Morin, Jeff Trent, and Francis Collins to earlier versions of this chapter.

FURTHER READING

- BUNZ F: *Principles of Cancer Genetics*, 2nd ed. Dordrecht, Springer, 2016.
- SIMON R, ROYCHOWDHURY S: Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Disc* 12:358, 2013.
- VOGELSTEIN B et al: Cancer genome landscapes. *Science* 339:1546, 2013.
- VOGELSTEIN B, KINZLER KW: The path to cancer—three strikes and you’re out. *N Engl J Med* 373:1895, 2015.



CANCER CELL BIOLOGY

Cancers are characterized by unregulated cell division, avoidance of cell death, tissue invasion, and the ability to metastasize. A neoplasm is *benign* when it grows in an unregulated fashion without tissue invasion. The presence of unregulated growth and tissue invasion is characteristic of *malignant* neoplasms. Cancers are named based on their origin: those derived from epithelial tissue are called *carcinomas*, those derived from mesenchymal tissues are *sarcomas*, and those derived from hematopoietic tissue are *leukemias*, *lymphomas*, and *plasma cell dyscrasias* (including *multiple myeloma*).

Cancers nearly always arise as a consequence of genetic alterations, the vast majority of which begin in a single cell and therefore are monoclonal in origin. However, because a wide variety of genetic and epigenetic changes can occur in different cells within malignant tumors over time, most cancers are characterized by marked heterogeneity in the populations of cells. This heterogeneity significantly complicates the treatment of most cancers because it is likely that there are subsets of cells that will be resistant to therapy and will therefore survive and proliferate even if the majority of cells are killed.

A few cancers appear to, at least initially, be primarily driven by an alteration in a dominant gene that produces uncontrolled cell proliferation. Examples include chronic myeloid leukemia (*abl*), about half of melanomas (*braf*), Burkitt's lymphoma (*c-myc*), and subsets of lung adenocarcinomas (*egfr*, *alk*, *ros1*, *met*, and *ret*). The genes that can promote cell growth when altered are often called *oncogenes*. They were first identified as critical elements of viruses that cause animal tumors; it was subsequently found that the viral genes had normal counterparts with important functions in the cell and had been captured and mutated by viruses as they passed from host to host.

However, the vast majority of human cancers are characterized by a multiple step process involving many genetic abnormalities, each of which contributes to the loss of control of cell proliferation and differentiation and the acquisition of capabilities, such as tissue invasion, the ability to metastasize, and angiogenesis (development of new blood vessels required for tumor growth). These properties are not found in the normal adult cell from which the tumor is derived. Indeed, normal cells have a large number of safeguards against DNA damage (including multiple DNA repair and extensive DNA damage response mechanisms), uncontrolled proliferation, and invasion. Many cancers go through recognizable steps of progressively more abnormal phenotypes: hyperplasia, to adenoma, to dysplasia, to carcinoma *in situ*, to invasive cancer with the ability to metastasize (Table 68-1). For most cancers, these changes occur over a prolonged period of time, usually many years.

In most organs, only primitive undifferentiated cells are capable of proliferating and the cells lose the capacity to proliferate as they differentiate and acquire functional capability. The expansion of the primitive cells (stem cells) is linked to some functional need in the host through receptors that receive signals from the local environment or through hormonal and other influences delivered by the vascular supply. In the absence of such signals, the cells are at rest. The signals that keep the primitive cells at rest remain incompletely understood. These signals must be environmental, based on the observations that a regenerating liver stops growing when it has replaced the portion that has been surgically removed post partial hepatectomy and regenerating bone marrow stops growing when the peripheral blood counts return to normal. Cancer cells clearly have lost responsiveness to such controls and do not recognize when they have overgrown the niche normally occupied by the organ from which they are derived. A better understanding of the mechanisms of growth regulation is evolving.

TABLE 68-1 Phenotypic Characteristics of Malignant Cells

Deregulated cell proliferation: Loss of function of negative growth regulators (tumor suppressor genes, i.e., *Rb*, *p53*), and increased action of positive growth regulators (oncogenes, i.e., *Ras*, *Myc*). Leads to aberrant cell cycle control and includes loss of normal checkpoint responses.

Failure to differentiate: Arrest at a stage before terminal differentiation. May retain stem cell properties. (Frequently observed in leukemias due to transcriptional repression of developmental programs by the gene products of chromosomal translocations.)

Loss of normal apoptosis pathways: Inactivation of *p53*, increases in *Bcl-2* (anti-apoptotic) family members. This defect enhances the survival of cells with oncogenic mutations and genetic instability and allows clonal expansion and diversification within the tumor without activation of physiologic cell death pathways.

Genetic instability: Defects in DNA repair pathways leading to either single or oligo-nucleotide mutations (as in microsatellite instability, *MIN*) or more commonly chromosomal instability (CIN) leading to aneuploidy (abnormal number of chromosomes in a cell). Caused by loss of function of a number of proteins including *p53*, *BRCA1/2*, mismatch repair genes, DNA repair enzymes, and the spindle checkpoint. Leads to accumulation of a variety of mutations in different cells within the tumor and heterogeneity.

Loss of replicative senescence: Normal cells stop dividing *in vitro* after 25–50 population doublings. Arrest is mediated by the *Rb*, *p16^{INK4a}*, and *p53* pathways. While most cells remain arrested, genetic and epigenetic changes in a subset of cells allows further replication leading to telomere loss, with crisis leading to death of many cells. Cells that survive often harbor gross chromosomal abnormalities and the ability to continue to proliferate. These cells express telomerase which maintains telomeres and is important for ongoing growth of these cells. Relevance to human *in vivo* cancer remains uncertain. Many human cancers express telomerase.

Non-responsiveness to external growth-inhibiting signals: Cancer cells have lost responsiveness to signals normally present to stop proliferating when they have overgrown the niche normally occupied by the organ from which they are derived. Our understanding about this mechanism of growth regulation remains limited.

Increased angiogenesis: Due to increased gene expression of proangiogenic factors (VEGF, FGF, IL-8, ANGIOPOEITIN) by tumor or stromal cells, or loss of negative regulators (endostatin, tumstatin, thrombospondin).

Invasion: Cell mobility and ability to move through extracellular matrix and into other tissues or organs. Loss of cell-cell contacts (gap junctions, cadherins) and increased production of matrix metalloproteinases (MMPs). Can take the form of epithelial-to-mesenchymal transition (EMT), with anchored epithelial cells becoming more like motile fibroblasts.

Metastasis: Spread of tumor cells to lymph nodes or distant tissue sites. Limited by the ability of tumor cells to migrate out of initial site and to survive in a foreign environment, including evading the immune system (see below).

Evasion of the immune system: Downregulation of MHC class I and II molecules; induction of T-cell tolerance; inhibition of normal dendritic cell and/or T-cell function; antigenic loss variants and clonal heterogeneity; increase in regulatory T cells.

Shift in cell metabolism: Complex changes including alterations due to tumor stress such as hypoxia, energy generation shifts from oxidative phosphorylation to aerobic glycolysis, generate building blocks for malignant cell production and proliferation.

Abbreviations: FGF, fibroblast growth factor; IL, interleukin; MHC, major histocompatibility complex; VEGF, vascular endothelial growth factor.

CELL CYCLE CHECKPOINTS

The cell division cycle consists of four phases—G1 (growth and preparation for DNA synthesis), S (DNA synthesis), G2 (preparation to divide), and M (mitosis, cell division). Cells can also exit the cell cycle and be quiescent (G0). Progression of a cell through the cell cycle is tightly regulated at a number of checkpoints (especially at the G1/S boundary, the G2/M boundary, and during M [spindle checkpoint]) by an array of genes that are targeted by specific genetic alterations in cancer. Critical proteins in these control processes that are frequently mutated or otherwise inactivated in cancers are called tumor-suppressor genes. Examples include *p53* and *Rb* (discussed below). In the first phase, G1, preparations are made to replicate the genetic material. The cell stops before entering the DNA synthesis phase, or S phase, to take inventory. Are we ready to replicate our DNA? Is the DNA repair machinery in

place to fix any mutations that are detected? Are the DNA replicating enzymes available? Is there an adequate supply of nucleotides? Is there sufficient energy to proceed? The main brake on the process is the retinoblastoma protein, Rb. When the cell determines that it is prepared to move ahead, sequential activation of cyclin-dependent kinases (CDKs) results in the inactivation of the brake, Rb, by phosphorylation. Phosphorylated Rb releases the S phase-regulating transcription factor, E2F/DP1, and genes required for S phase progression are expressed. If the cell determines that it is unready to move ahead with DNA replication, a number of inhibitors are capable of blocking the action of the CDKs, including p21^{Cip1/Waf1}, p16^{Ink4a}, and p27^{Kip1}. *Nearly every cancer has one or more defects in the G₁ checkpoint that permit progression to S phase despite abnormalities in DNA repair machinery or other deficiencies that would affect normal DNA synthesis.*

At the end of the G₂ phase and prior to the M phase, after the cell has exactly duplicated its DNA content, a second inventory is taken at the G₂ checkpoint. Have all of the chromosomes been fully duplicated? Were all segments of DNA copied only once? Has all damaged DNA been repaired? Do we have the right number of chromosomes and the right amount of DNA? If so, the cell proceeds to G₂, in which the cell prepares for division by synthesizing mitotic spindle and other proteins needed to produce two daughter cells. When DNA damage is detected, the p53 pathway is normally activated. Called the guardian of the genome, p53 is a transcription factor that is normally present in the cell in very low levels. Its level is generally regulated through its rapid turnover. Normally, p53 is bound to mdm2, an ubiquitin ligase that both inhibits p53 transcriptional activation and also targets p53 for degradation in the proteasome. When damage is sensed, the ATM (ataxia-telangiectasia mutated) pathway is activated; ATM phosphorylates mdm2, which no longer binds to p53, and p53 then stops cell cycle progression, directs the synthesis of repair enzymes, or if the damage is too great, initiates apoptosis (programmed cell death) of the cell to prevent the propagation of a damaged cell (Fig. 68-1).

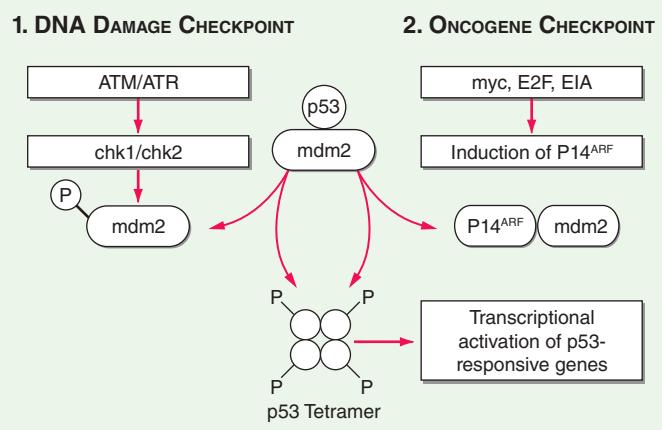


FIGURE 68-1 Induction of p53 by the DNA damage and oncogene checkpoints. In response to noxious stimuli, p53 and mdm2 are phosphorylated by the ataxiatelangiectasia mutated (ATM) and related ATR serine/threonine kinases, as well as the immediate downstream checkpoint kinases, Chk1 and Chk2. This causes dissociation of p53 from mdm2, leading to increased p53 protein levels and transcription of genes leading to cell cycle arrest (p21^{Cip1/Waf1}) or apoptosis (e.g., the proapoptotic Bcl-2 family members Noxa and Puma). Inducers of p53 include hypoxemia, DNA damage (caused by ultraviolet radiation, gamma irradiation, or chemotherapy), ribonucleotide depletion, and telomere shortening. A second mechanism of p53 induction is activated by oncogenes such as Myc, which promote aberrant G₁/S transition. This pathway is regulated by a second product of the Ink4a locus, p14^{ARF} (p19 in mice), which is encoded by an alternative reading frame (ARF) of the same stretch of DNA that codes for p16^{Ink4a}. Levels of ARF are upregulated by Myc and E2F, and ARF binds to mdm2 and rescues p53 from its inhibitory effect. This oncogene checkpoint leads to the death or senescence (an irreversible arrest in G₁ of the cell cycle) of renegade cells that attempt to enter S phase without appropriate physiologic signals. Senescent cells have been identified in patients whose premalignant lesions harbor activated oncogenes, for instance, dysplastic nevi that encode an activated form of BRAF (see below), demonstrating that induction of senescence is a protective mechanism that operates in humans to prevent the outgrowth of neoplastic cells.

A second method of activating p53 involves the induction of p14^{ARF} by hyperproliferative signals from oncogenes. p14^{ARF} competes with p53 for binding to mdm2, allowing p53 to escape the effects of mdm2 and accumulate in the cell. Then p53 stops cell cycle progression by activating CDK inhibitors such as p21 and/or initiating the apoptosis pathway. Not surprisingly given its critical role in controlling cell cycle progression, mutations in the gene for p53 on chromosome 17p are among the most frequent mutations in human cancers, although percentages vary between different cancers. Most commonly these mutations are acquired in the malignant tissue in one allele and the second allele is inactivated (such as by deletion), leaving the cell unprotected from DNA-damaging agents or activated oncogenes. Some environmental exposures produce signature mutations in p53; for example, aflatoxin exposure leads to mutation of arginine to serine at codon 249 and leads to hepatocellular carcinoma. In rare instances, p53 mutations are in the germline (Li-Fraumeni syndrome) and produce a familial cancer syndrome. The absence of p53 leads to chromosome instability and the accumulation of DNA damage including the acquisition of properties that give the abnormal cell a proliferative and survival advantage. *Like Rb dysfunction, most cancers have mutations that disable the p53 pathway.* Indeed, the importance of p53 and Rb in the development of cancer is underscored by the neoplastic transformation mechanism of human papillomavirus. This virus has two main oncogenes, E6 and E7. E6 acts to increase the rapid turnover of p53, and E7 acts to inhibit Rb function; inhibition of these two targets is required for transformation of epithelial cells.

Another cell cycle checkpoint exists when the cell is undergoing division (M phase), the spindle checkpoint which acts to ensure that there is proper attachment of chromosomes to the mitotic spindle before progression through the cell cycle can occur. If the spindle apparatus does not properly align the chromosomes for division, if the chromosome number is abnormal (i.e., greater or less than 4n), or if the centromeres are not properly paired with their duplicated partners, then the cell initiates a cell death pathway to prevent the production of aneuploid progeny (having an altered number of chromosomes). Abnormalities in the spindle checkpoint facilitate the development of aneuploidy which is frequently found in cancers. In some tumors, aneuploidy is a predominant genetic feature. In others, a defect in the cells' ability to repair errors in the DNA, such as due to mutations in genes coding for the proteins critical for mismatched DNA repair, is the primary genetic lesion. Mismatch repair is usually detected by finding alterations in repeat sequences of DNA (called microsatellites), or microsatellite instability, in malignant cells. In general, tumors either have defects in chromosome number or defective DNA repair pathways such as microsatellite instability, but not both. Defects that lead to cancer include abnormal cell cycle checkpoints, inadequate DNA repair, and failure to preserve genome integrity leading to DNA damage. These defects and the stress of the resultant increased DNA damage make cancer cells more vulnerable to additional DNA damage which can be exploited by chemotherapy, radiation therapy, and immunotherapy which are the major systemic therapeutic approaches effective against cancer.

Efforts are also under way to therapeutically restore the defects in cell cycle regulation that characterize cancer, although this remains a challenging problem because it is much more difficult to restore normal biologic function than to inhibit abnormal function of proteins driving cell proliferation, such as occurs with oncogenes. Newer approaches to gene editing (e.g., Clustered Regularly Interspaced Short Palindromic Repeats [CRISPR]) should make this more feasible.

CANCER AS AN ORGAN THAT IGNORES ITS NICHE

The fundamental cellular defects that create a malignant neoplasm act at the cellular level and some of these are cell autonomous. However, that is not the entire story. Cancers consist of both malignant cells as well as other cells in the cancer microenvironment and behave as organs that have lost their specialized function and stopped responding to signals that would limit their growth in tightly regulated normal tissue homeostasis. Human cancers usually become clinically detectable when a primary mass is at least 1 cm in diameter—such a mass consists of about 10⁹ cells. More commonly patients present

with tumors that are at least 10^{10} cells. A lethal tumor burden is about 10^{12} – 10^{13} cells, although there is significant variability depending on the type and location of the cancer. If all malignant cells were dividing at the time of diagnosis, patients would reach a lethal tumor burden in a very short time. However, human tumors grow by Gompertzian kinetics—this means that not every daughter cell produced by a cell division is actively dividing. In addition, the overall growth rate of a tumor depends on differences between growth rates of different cells within the tumor and rate of cell loss. The growth fraction of a tumor declines with time, largely due to factors in the microenvironment. The growth fraction of the first malignant cell is 100%, and by the time a patient presents for medical care, the growth fraction is estimated to be <10%, although the fraction varies between different types of cancers and even different cancers of the same type in different individuals. This fraction is similar to the growth fraction of normal bone marrow and normal intestinal epithelium, the most highly proliferative normal tissues in the human body, a fact that may explain the dose-limiting toxicities of agents that target dividing cells.

The implication of these data is that the tumor is slowing its own growth over time. How does it do this? The tumor cells have multiple genetic lesions that tend to promote proliferation, yet by the time the tumor is clinically detectable, its capacity for proliferation has declined. Better understanding of how a tumor slows its own growth would provide important clues for better cancer treatment. A number of factors, including those in the tumor microenvironment, are known to contribute to the failure of tumor cells to proliferate *in vivo*. Some cells are hypoxic and have inadequate supply of nutrients and energy. Some have sustained too much genetic damage to complete the cell cycle but have lost the capacity to undergo apoptosis and therefore survive but do not proliferate. However, an important subset is not actively dividing but retains the capacity to divide and can start dividing again under certain conditions such as when the tumor mass is reduced by treatments leading to improved conditions in the tumor microenvironment favorable for cell proliferation. Just as the bone marrow increases its rate of proliferation in response to bone marrow-damaging agents, the tumor also seems to sense when tumor cell numbers have been reduced and can respond by increasing growth rate. However, the critical difference is that the marrow stops growing when it has reached its production goals whereas tumors do not.

Additional tumor cell vulnerabilities are likely to be detected when we learn more about how normal cells respond to “stop” signals from their environment, and why and how tumor cells fail to heed such signals.

■ IS IN VITRO SENESCENCE RELEVANT TO CARCINOGENESIS?

When normal cells are placed in culture *in vitro*, most are not capable of sustained growth. Fibroblasts are an exception to this rule. When they are cultured, fibroblasts may divide 30–50 times and then they undergo what has been termed a “crisis” during which the majority of cells stop dividing (usually due to an increase in p21 expression, a CDK inhibitor), many die, and a small fraction emerge that have acquired genetic and epigenetic changes that permit their uncontrolled growth. The cessation of growth of normal cells in culture has been termed “senescence” and whether this phenomenon is relevant to any physiologic event *in vivo* is still an area of investigation, including identifying biomarkers of senescence *in vivo*.

Among the cellular changes during *in vitro* propagation is telomere shortening. DNA polymerase is unable to replicate the tips of chromosomes, resulting in the loss of DNA at the specialized ends of chromosomes (called *telomeres*) with each replication cycle. At birth, human telomeres are 15- to 20-kb pairs long and are composed of tandem repeats of a six-nucleotide sequence (TTAGGG) that associates with specialized telomere-binding proteins to form a T-loop structure that protects the ends of chromosomes from being mistakenly recognized as damaged. The loss of telomeric repeats with each cell division cycle causes gradual telomere shortening, leading to growth arrest (called *senescence*) when one or more critically short telomeres trigger a p53-regulated DNA-damage checkpoint response. Cells can bypass

this growth arrest if pRb and p53 are nonfunctional, but cell death usually ensues when the unprotected ends of chromosomes lead to chromosome fusions or other catastrophic DNA rearrangements. *The ability to bypass telomere-based growth limitations is thought to be a critical step in the evolution of most malignancies.* This occurs by the reactivation of telomerase expression in cancer cells. Telomerase is an enzyme that adds TTAGGG repeats onto the 3' ends of chromosomes. It contains a catalytic subunit with reverse transcriptase activity (hTERT) and an RNA component that provides the template for telomere extension. Most normal somatic cells do not express sufficient telomerase to prevent telomere attrition with each cell division. Exceptions include stem cells (such as those found in hematopoietic tissues, gut and skin epithelium, and germ cells) that require extensive cell division to maintain tissue homeostasis. More than 90% of human cancers express high levels of telomerase that prevent telomere shortening to critical levels and allow indefinite cell proliferation. *In vitro* experiments indicate that inhibition of telomerase activity leads to tumor cell apoptosis. Major efforts are underway to develop methods to inhibit telomerase activity in cancer cells. For example, the protein component of telomerase (hTERT) may act as one of the most widely expressed tumor-associated antigens and can be targeted by vaccine approaches. However, a caveat to targeting telomerase for anticancer treatment is an inadequate understanding of how important its presence is in certain normal cells to maintaining the normal physiologic state.

Although most of the functions of telomerase relate to cell division, it also has several other effects including interfering with the differentiated functions of at least certain stem cells. However, the impact on differentiated function of normal non-stem cells is less clear. The picture is further complicated by the fact that rare genetic defects in the telomerase enzyme seem to cause pulmonary fibrosis, aplastic anemia, or dyskeratosis congenita (characterized by abnormalities in skin, nails, and oral mucosa with increased risk for certain malignancies) but not defects in nutrient absorption in the gut, a site that might be presumed to be highly sensitive to defective cell proliferation. Much remains to be learned about how telomere shortening and telomere maintenance are related to human illness in general and cancer in particular.

■ SIGNAL TRANSDUCTION PATHWAYS IN CANCER CELLS

Signals that affect cell behavior come from adjacent cells, the stroma in which the cells are located, hormonal signals that originate remotely, and from the cells themselves (autocrine signaling). These signals generally exert their influence on the receiving cell through activation of signal transduction pathways that have as their end result the induction of activated transcription factors that mediate a change in cell behavior or function or the acquisition of effector machinery to accomplish a new task. Although signal transduction pathways can lead to a wide variety of outcomes, many such pathways rely on cascades of signals that sequentially activate different proteins or glycoproteins and lipids or glycolipids, and the activation steps often involve the addition or removal of one or more phosphate groups on a downstream target. Other chemical changes can result from signal transduction pathways, but phosphorylation and dephosphorylation play a major role. The proteins that add phosphate groups to proteins are called kinases. There are two major distinct classes of kinases; one class acts on tyrosine residues and the other acts on serine/threonine residues. The tyrosine kinases often play critical roles in signal transduction pathways; they may be receptor tyrosine kinases (RTKs) or they may be linked to other cell-surface receptors through associated docking proteins and transmit the signal into the cell (Fig. 68-2).

Normally, tyrosine kinase activity is short-lived and reversed by protein tyrosine phosphatases (PTPs). However, in many human cancers, tyrosine kinases or components of their downstream pathways are activated by mutation, gene amplification, or chromosomal translocations. Because these pathways regulate proliferation, survival, migration, and angiogenesis, they have been identified as important targets for cancer therapeutics.

Inhibition of kinase activity is effective in the treatment of a number of neoplasms. Lung cancers with mutations in the epidermal

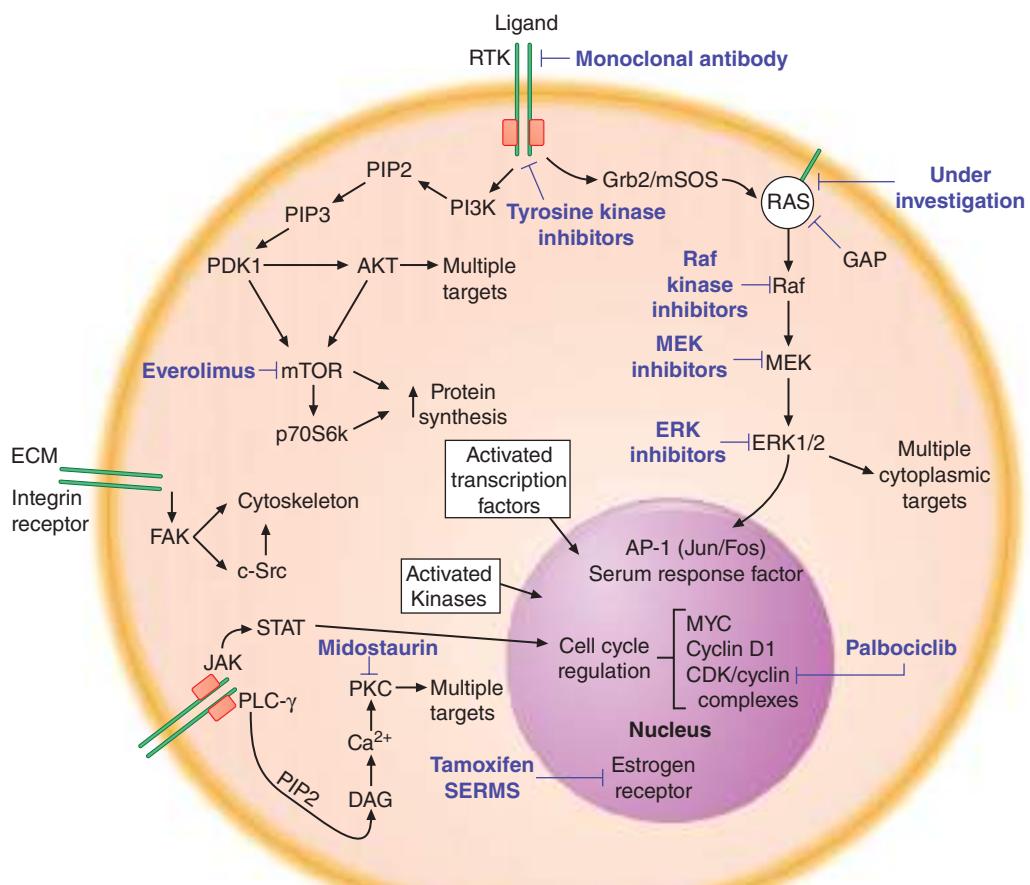


FIGURE 68-2 Therapeutic targeting of signal transduction pathways in cancer cells. Three major signal transduction pathways are activated by receptor tyrosine kinases (RTK). 1. The protooncogene Ras is activated by the Grb2/mSOS guanine nucleotide exchange factor, which induces an association with Raf and activation of downstream kinases (MEK and ERK1/2). 2. Activated PI3K phosphorylates the membrane lipid PIP₂ to generate PIP₃, which acts as a membrane-docking site for a number of cellular proteins including the serine/threonine kinases PDK1 and Akt. PDK1 has numerous cellular targets, including Akt and mTOR. Akt phosphorylates target proteins that promote resistance to apoptosis and enhance cell cycle progression, while mTOR and its target p70S6K upregulate protein synthesis to potentiate cell growth. 3. Activation of PLC-γ leads the formation of diacylglycerol (DAG) and increased intracellular calcium, with activation of multiple isoforms of PKC and other enzymes regulated by the calcium/calmodulin system. Other important signaling pathways involve non-RTKs that are activated by cytokine or integrin receptors. Janus kinases (JAK) phosphorylate STAT (signal transducer and activator of transcription) transcription factors, which translocate to the nucleus and activate target genes. Integrin receptors mediate cellular interactions with the extracellular matrix (ECM), inducing activation of FAK (focal adhesion kinase) and c-Src, which activate multiple downstream pathways, including modulation of the cell cytoskeleton. Many activated kinases and transcription factors migrate into the nucleus, where they regulate gene transcription, thus completing the path from extracellular signals, such as growth factors, to a change in cell phenotype, such as induction of differentiation or cell proliferation. The nuclear targets of these processes include transcription factors (e.g., Myc, AP-1, and serum response factor) and the cell cycle machinery (CDKs and cyclins). Inhibitors of many of these pathways have been developed for the treatment of human cancers. Examples of inhibitors that are either approved or are currently being evaluated in clinical trials are shown in purple type.

growth factor receptor are highly responsive to erlotinib and gefitinib (**Table 68-2**). Lung cancers with activation of anaplastic lymphoma kinase (ALK) or ROS1 by translocations respond to crizotinib, an ALK and ROS1 inhibitor and additional ALK inhibitors including ceritinib and alectinib are available for treating lung cancers with a number of additional inhibitors currently in trials. BRAF inhibitors are highly effective in melanomas and thyroid cancers in which BRAF is mutated. Targeting a protein (MEK) downstream of BRAF also has activity against BRAF mutant melanomas and combined inhibition of BRAF and MEK is more effective than either alone. Janus kinase (JAK) inhibitors are active in myeloproliferative syndromes in which JAK2 activation is a pathogenetic event. Imatinib (which targets a number of tyrosine kinases) is an effective agent in tumors that have translocations of the c-Abl and BCR gene (such as chronic myeloid leukemia), mutant c-Kit (gastrointestinal stromal cell tumors), or mutant platelet-derived growth factor receptor (PDGFR α ; gastrointestinal stromal tumors); second-generation inhibitors of BCR-Abl, dasatinib, and nilotinib are even more effective and the third generation agent bosutinib has activity in some patients who have progressed on other inhibitors, while the third generation ponatinib has activity against the T315I mutation, which is resistant to the other agents. Although almost all tyrosine kinase inhibitors are not entirely selective for one protein, certain inhibitors have significant activity against a broad number

of proteins. These include sorafenib, regorafenib, cabozantinib, sunitinib, and lenvatinib. These have shown antitumor activity in various malignancies, including renal cell cancer (RCC) (sorafenib, sunitinib, cabozantinib, lenvatinib), hepatocellular carcinoma (sorafenib, regorafenib, lenvatinib), gastrointestinal stromal tumor (GIST) (sunitinib, regorafenib), thyroid cancer (sorafenib, cabozantinib, lenvatinib), colorectal cancer (regorafenib), and pancreatic neuroendocrine tumors (sunitinib). Inhibitors of the mammalian target of rapamycin (mTOR) are active in RCC, pancreatic neuroendocrine tumors, and breast cancer. The list of active agents and treatment indications is growing rapidly (**Table 68-2**). These agents have ushered in a new era of personalized therapy. It is becoming more common for resected tumors to be assessed for specific molecular changes that predict response and to have clinical decision-making guided by those results. This is now an important component of standard therapy for metastatic lung, gastroesophageal, melanoma, breast, and colorectal cancers as well as in adjuvant therapy for breast cancer.

However, none of these therapies has yet been curative by themselves for any malignancy, although prolonged periods of disease control lasting many years frequently occur in CML, including a >80% survival rate at 10 years. The reasons for the failure to cure are not completely defined, although resistance to the treatment ultimately develops in most patients. In some tumors, resistance to kinase inhibitors is

TABLE 68-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
All-trans retinoic acid	PML-RAR α oncogene	Acute promyelocytic leukemia M3 AML; t(15;17)	Inhibits transcriptional repression by PML-RAR α
Imatinib	Bcr-Abl, c-Abl, c-Kit, PDGFR- α/β	Chronic myeloid leukemia; GIST	Blocks ATP binding to tyrosine kinase active site
Dasatinib, Nilotinib, Ponatinib, Bosutinib	Bcr-Abl (primarily)	Chronic myeloid leukemia	Blocks ATP binding to tyrosine kinase active site
Sunitinib	c-Kit, VEGFR-2, PDGFR- β , Flt-3	GIST; RCC; PNET	Inhibits activated c-Kit and PDGFR in GIST; inhibits VEGFR in RCC and probably in PNET
Sorafenib	RAF, VEGFR-2, PDGFR- α/β , Flt-3, c-Kit	RCC; hepatocellular carcinoma, differentiated thyroid cancer, desmoid	Targets VEGFR pathways in RCC and HCC. Possible activity against BRAF in thyroid cancer
Regorafenib	VEGFR1-3, TIE-2, FGFR1, KIT, RET, PDGFR	Colorectal cancer; GIST; HCC	Competitive inhibitor ATP binding site of tyrosine kinase domain multiple kinases including VEGFR
Axitinib	VEGFR 1-3	RCC	Competitive inhibitor ATP binding site of tyrosine kinase domain VEGF receptors
Erlotinib	EGFR	NSCLC; pancreatic cancer	Competitive inhibitor of the ATP-binding site of the EGFR
Afatinib	EGFR (and other HER family)	NSCLC	Irreversible inhibitor of ATP-binding site of HER family members
Osimertinib	EGFR(T790M)	NSCLC	Inhibits EGFR mutations including T790M mutant NSCLC
Lapatinib	HER2/neu	Breast Cancer	Competitive inhibitor of the ATP binding site of HER2
Crizotinib, Ceritinib, Alectinib	ALK, ROS1	NSCLC	Inhibitor of ALK and ROS1 tyrosine kinase
Palbociclib, Ribociclib, Abemaciclib	CDK4/6	Breast	Inhibitor of CDK4/6
Bortezomib, Carfilzomib, Ixazomib	Proteasome	Multiple myeloma	Inhibits proteolytic degradation of multiple cellular proteins
Vemurafenib, Dabrafenib	BRAF	Melanoma	Inhibitor of serine-threonine kinase domain of V600E mutant of BRAF
Trametinib, Cobimetinib	MEK	Melanoma	Inhibitor of serine-threonine kinase domain of MEK
Cabozantinib	RET, MET, VEGFR	MTC, RCC	Competitive inhibitor ATP binding site of tyrosine kinase domain multiple kinases, including VEGFR2 and RET
Vandetanib	RET, VEGFR, EGFR	MTC	Competitive inhibitor ATP-binding site of tyrosine kinase domain multiple kinases, including RET
Temsiroliumus	mTOR	RCC	Competitive inhibitor of mTOR serine-threonine kinase
Everolimus	mTOR	RCC; PNET	Binds to immunophilin FK binding protein-12 which forms complex that inhibits mTOR kinase
Vorinostat, Romidepsin, Belinostat	HDAC	CTCL/PTL	HDAC inhibitor, epigenetic modulation
Panobinostat	HDAC	MM	HDAC inhibitor, epigenetic modulation
Ruxolitinib	JAK-1, 2	Myelofibrosis	Competitive inhibitor of tyrosine kinase
Vismodegib	Hedgehog pathway	Basal cell cancer (skin)	Inhibits smoothened in Hedgehog Pathway
Lenvatinib	Multi-Kinase inhibitor (VEGFR, FGFR, PGFR-a, others)	RCC, Thyroid cancer	Competitive inhibitor ATP-binding site of tyrosine kinase domain multiple kinases
Olaparib, rucaparib	PARP	BRCA mutant ovarian (both) and breast (olaparib) cancers	Inhibits PARP and DNA repair
Venetoclax	BCL-2	CLL (with 17p deletion)	Inhibits BCL-2 and enhances apoptosis
Ibrutinib	Bruton Tyrosine Kinase (BTK)	CLL, MCL, MZL, SLL, WM	Inhibitor of BTK
Idealisib	PI3K-delta	CLL, SLL, FL	Inhibits PI3k-delta, preventing proliferation and inducing apoptosis
Monoclonal Antibodies Alone			
Trastuzumab	HER2/neu (ERBB2)	Breast cancer	Binds HER2 on tumor cell surface and induces receptor internalization
Pertuzumab	HER2/neu (ERBB2)	Breast cancer	Binds HER2 on tumor cell surface at distinct site from Trastuzumab and prevents binding to other receptors
Cetuximab	EGFR	Colon cancer, squamous cell carcinoma of the head and neck	Binds extracellular domain of EGFR and blocks binding of EGF and TGF- α ; induces receptor internalization. Potentiates the efficacy of chemotherapy and radiotherapy
Panitumumab	EGFR	Colon cancer	Similar to Cetuximab but fully humanized rather than chimeric
Necitumumab	EGFR	Squamous NSCLC	Binds EGFR
Rituximab	CD20	B-cell lymphomas and leukemias that express CD20	Multiple potential mechanisms, including direct induction of tumor cell apoptosis and immune mechanisms
Alemtuzumab	CD52	Chronic lymphocytic leukemia and CD52-expressing lymphoid tumors	Immune mechanisms
Bevacizumab	VEGF	Colorectal, lung cancers, RCC, glioblastoma	Inhibits angiogenesis by high-affinity binding to VEGF
Ziv-Aflibercept	VEGFA, VEGFB, PLGF	Colorectal cancers	Inhibits angiogenesis by high-affinity binding to VEGFA, B and PLGF
Ramucirumab	VEGFR	Gastric, colorectal, lung cancers	Inhibits angiogenesis by binding to VEGFR

(Continued)

TABLE 68-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer (Continued)

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
Ipilimumab	CTLA-4	Melanoma	Blocks CTLA-4 preventing interaction with CD80/86 and T-cell inhibition
Nivolumab Pembrolizumab	PD-1	Melanoma, Head and Neck Cancer, NSCLC, Hodgkin's Disease, Urothelial cancer, RCC, HCC, gastric cancer, MSI high cancers	Blocks PD-1 preventing interaction with PDL-1 and T-cell inhibition
Atezolizumab, Durvalumab	PDL1	NSCLC, Urothelial cancer	Blocks PDL1 preventing interaction with PD-1 and T-cell inhibition
Denosumab	Rank ligand	Breast, Prostate	Inhibits Rank ligand, primary signal for bone removal
Dinutuximab	Glycolipid GD2	Neuroblastoma (pediatric)	Immune mediated attack on GD2 expressing cells
Daratumumab	CD38	MM	Binds to CD38 on MM cells causing apoptosis by antibody dependent or complement mediated cytotoxicity
Elotuzumab	SLAMF7	MM	Activating NK cells to kill MM cells
Olarutumab	PDGFR α	Soft Tissue Sarcomas	Blocks PDGFR α activity
Blinotuzumab	CD19 and CD3	PH-relapsed precursor B-cell ALL	Binds CD19 on ALL cells and CD3 on T cells; Immune attack on CD19 expressing cells
Antibody-Chemotherapy Conjugates			
Brentuximab vedotin	CD30	HD, Anaplastic Lymphoma	Delivery of chemotherapeutic agent (MMAE) to CD30 expressing tumor cells
Ado-Trastuzumab emtansine	HER2	Breast Cancer	Delivery of chemotherapeutic agent emtansine to HER2 expressing breast cancer cells
CAR-T Cells			
Kymria, Yescarta	CD19	ALL (Kymria), DLBCL (Yescarta)	Targeted T-cells to protein on surface of malignant cells

Abbreviations: ALL, acute lymphocytic leukemia; AML, acute myeloid leukemia; DLBCL, diffuse large B-cell lymphoma; EGFR, epidermal growth factor receptor; Flt-3, fms-like tyrosine kinase-3; GIST, gastrointestinal stromal tumor; HDAC, histone deacetylases; MCL, mantle cell lymphoma; MSI, microsatellite instability; MZL, mantle zone lymphoma; NSCLC, nonsmall cell lung cancer; PARP, poly ADP ribose polymerase; PDGFR, platelet-derived growth factor receptor; PLGF, placenta growth factor; PML-RAR α , promyelocytic leukemia-retinoic acid receptor-alpha; PNET, pancreatic neuroendocrine tumors; RCC, renal cell cancer; t(15;17), translocation between chromosomes 15 and 17; SLL, small lymphocytic lymphoma; TGF- α , transforming growth factor-alpha; VEGFR, vascular endothelial growth factor receptor; WM, Waldenstrom's macroglobulinemia.

related to an acquired mutation in the target kinase that inhibits drug binding. Many of these kinase inhibitors act as competitive inhibitors of the ATP-binding pocket. ATP is the phosphate donor in these phosphorylation reactions. For example, mutation in the critical BCR-ABL kinase in the ATP-binding pocket (such as the threonine to isoleucine change at codon 315 [T315I]) can prevent imatinib binding. Other resistance mechanisms include alterations in other signal transduction pathways to bypass the inhibited pathway. As resistance mechanisms become better defined, rational strategies to overcome resistance will emerge. In addition, many kinase inhibitors are less specific for an oncogenic target than was hoped, and toxicities related to off-target inhibition of kinases limits the use of the agent at a dose that would optimally inhibit the cancer-relevant kinase.

Targeted agents can also be used to deliver highly toxic compounds. An important component of the technology for developing effective conjugates is the design of the linker between the two which needs to be stable. Examples of currently approved antibody drug conjugates include brentuximab vedotin, which links the microtubule toxin monomethyl auristatin E (MMAE) to an antibody targeting the cell surface antigen CD30, which is expressed on a number of malignant cells but especially in Hodgkin's lymphoma and anaplastic lymphoma. The linker in this case is cleavable which allows diffusion of the drug out of the cell after delivery. A second approved conjugate is ado-trastuzumab emtansine which links the microtubule formation inhibitor mertansine and the monoclonal antibody trastuzumab targeted against HER2 on breast cancer cells. In this case the linker is non-cleavable, thus trapping the chemotherapeutic agent within the cells. There are theoretical pluses and minuses to having either cleavable or non-cleavable linkers and it is likely that both will be used in future developments of antibody-drug conjugates.

Another strategy to enhance the antitumor effects of targeted agents is to use them in rational combinations with each other as well as with chemotherapy or immunotherapy agents that kill cells in ways distinct from agents targeting specific mutant or overexpressed proteins. Combinations of trastuzumab (a monoclonal antibody that targets the HER2 receptor [member of the EGFR family]) with chemotherapy have significant activity against breast and stomach cancers that have high levels of expression of the HER2 protein. The activity of trastuzumab and chemotherapy can be enhanced further by combinations with another targeted

monoclonal antibody (pertuzumab) which prevents dimerization of the HER2 receptor with other HER family members including HER3.

Although targeted therapies have not yet resulted in cures when used alone, their use in the adjuvant setting and when combined with other effective treatments has substantially increased the fraction of patients cured. For example, the addition of rituximab, an anti-CD20 antibody, to combination chemotherapy in patients with diffuse large B-cell lymphoma improves cure rates by ~15%. The addition of trastuzumab, antibody to HER2, to combination chemotherapy in the adjuvant treatment of HER2-positive breast cancer significantly improves overall survival.

A major effort is underway to develop targeted therapies for mutations in the *ras* family of genes which are the most common mutations in oncogenes in cancers (especially *kras*) but have proved to be very difficult targets for a number of reasons related to the structure of RAS proteins as well as mechanisms of activation and inactivation. Targeted therapies against a subset of proteins downstream of RAS in the signaling pathway (including BRAF and mitogen-activated protein [MAP] kinase) have proven to have significant antitumor activity against V600E BRAF mutant melanoma, with improved efficacy when they are used in combination. However, similar activity is not seen against *ras* mutant tumors. Additional targeted therapies against other proteins downstream of RAS (including ERK, or combinations of MAP kinase inhibitors and immunotherapy) are currently being studied, both individually and in combination. However, at this time, there is no established effective approach to inhibiting RAS mutant tumors. Inhibitors of phospholipid signaling pathways such as the phosphatidylinositol-3-kinase (PI3K) and phospholipase C-gamma pathways, which are involved in a large number of cellular processes that are important in cancer development and progression, are being evaluated. The targeting of a variety of other pathways that are activated in malignant cells, such as the MET pathway, hedgehog pathway, and various angiogenesis pathways are also being explored.

One of the strategies for new drug development is to take advantage of so-called oncogene addiction. This situation (**Fig. 68-3**) is created when a tumor cell develops an activating mutation in an oncogene that becomes a dominant pathway for survival and growth with reduced contributions from other pathways, even when there may be abnormalities in those pathways. This dependency on a single pathway creates

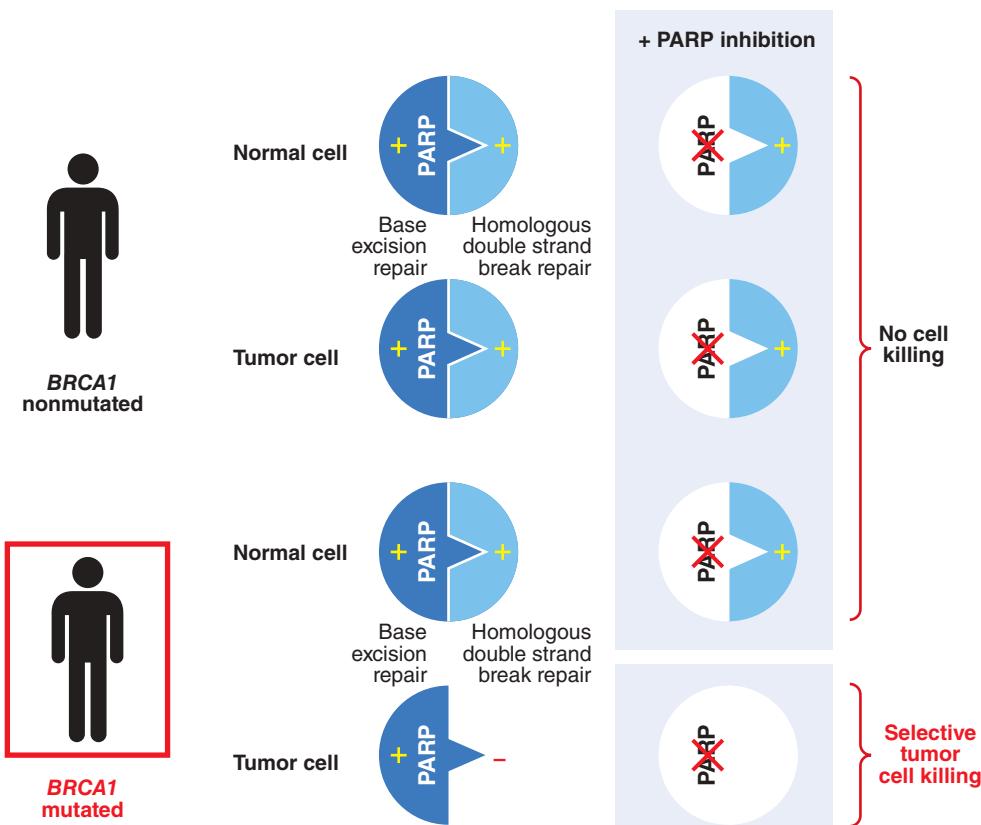


FIGURE 68-3 Synthetic lethality. Genes are said to have a synthetic lethal relationship when mutation of either gene alone is tolerated by the cell, but mutation of both genes leads to lethality, as originally noted by Bridges and later named by Dobzhansky. Thus, mutant gene *a* and gene *b* have a synthetic lethal relationship, implying that the loss of one gene makes the cell dependent on the function of the other gene. In cancer cells, loss of function of a DNA repair gene like *BRCA1*, which repairs double-strand breaks, makes the cell dependent on base excision repair mediated in part by *PARP*. If the *PARP* gene product is inhibited, the cell attempts to repair the break using the error-prone non-homologous end-joining method, which results in tumor cell death. High-throughput screens can now be performed using isogenic cell line pairs in which one cell line has a defined defect in a DNA repair pathway. Compounds can be identified that selectively kill the mutant cell line; targets of these compounds have a synthetic lethal relationship to the repair pathway, and are potentially important targets for future therapeutics.

a cell that is vulnerable to inhibitors of that oncogene pathway. For example, cells harboring mutations in *BRAF* are very sensitive to MEK inhibitors that inhibit downstream signaling in the *BRAF* pathway.

Proteins critical for transcription of other proteins essential for malignant cell survival or proliferation provide another potential target for treating cancers. The transcription factor NF- κ B is a heterodimer composed of p65 and p50 subunits that associate with an inhibitor, I κ B, in the cell cytoplasm. In response to growth factor or cytokine signaling, a multi-subunit kinase called IKK (I κ B-kinase) phosphorylates I κ B and directs its degradation by the ubiquitin/proteasome system. NF- κ B, free of its inhibitor, translocates to the nucleus and activates target genes, many of which promote the survival of tumor cells. One of the mechanisms by which novel drugs called *proteasome inhibitors* are thought to produce an anticancer effect is by blocking the proteolysis of I κ B, thereby preventing NF- κ B activation. For reasons that have not been fully elucidated, this has a differential toxicity affect on tumor, as compared to normal, cells. Although this mechanism appears to be an important aspect of the antitumor effects of proteasome inhibitors, there are other effects involving the inhibition of the degradation of multiple cellular proteins important in malignant cell survival or proliferation. Proteasome inhibitors (bortezomib, carfilzomib, ixazomib) have activity in patients with multiple myeloma, including partial and complete remissions. Inhibitors of IKK are also in development, with the hope of more selectively blocking the degradation of I κ B, thus “locking” NF- κ B in an inhibitory complex and rendering the cancer cell more susceptible to apoptosis-inducing agents. Many other transcription factors are activated by phosphorylation, which can be prevented by tyrosine- or serine/threonine kinase inhibitors, a number of which are currently in clinical trials.

Estrogen receptors (ERs) and androgen receptors (ARs), members of the steroid hormone family of nuclear receptors, are targets of inhibition by drugs used to treat breast and prostate cancers, respectively. Selective estrogen receptor modulators (SERMs) have been developed as a treatment approach for ER positive breast cancer. Tamoxifen, a partial agonist and antagonist of ER function, is frequently used in breast cancer and can mediate tumor regression in metastatic breast cancer and can prevent disease recurrence in the adjuvant setting. Tamoxifen binds to the ER and modulates its transcriptional activity, inhibiting activity in the breast but promoting activity in bone but unfortunately also in uterine epithelium leading to a small increased risk of uterine cancer. Attempts have been made to develop SERMs that would have anti-estrogenic effects in both breast and uterus while maintaining protective effects on bone. However, none of these to date has been an improvement over tamoxifen. Aromatase inhibitors, which block the conversion of androgens to estrogens in breast and subcutaneous fat tissues, have demonstrated improved clinical efficacy compared with tamoxifen in postmenopausal women and are often used as first-line therapy in post menopausal patients with ER-positive disease. They are occasionally utilized in premenopausal patients with ER positive disease in combination with ovarian suppression approaches such as leutinizing hormone receptor (LHRH) agonists. A number of approaches have been developed for blocking androgen

stimulation of prostate cancer, including decreasing production by the testicles (e.g., orchectomy, LHRH agonists or antagonists), directly blocking actions of androgen (a number of agents have been developed to do this), or blocking production by inhibiting the enzyme CYP17 which is central in production of androgens from cholesterol (Chap. 75).

CANCER-SPECIFIC GENETIC CHANGES AND SYNTHETIC LETHALITY

The concepts of oncogene addiction and synthetic lethality have spurred new drug development targeting oncogene- and tumor-suppressor pathways. As discussed earlier in this chapter and outlined in Fig. 68-3, cancer cells can become dependent upon signaling pathways containing activated oncogenes; this can effect proliferation (i.e., mutated KRAS, *BRAF*, overexpressed Myc, or activated tyrosine kinases). Additional genetic changes in malignant cells or unique features of tumors including defects in DNA repair (e.g., loss of *BRCA1* or *BRCA2* gene function), modifications in cell cycle control (e.g., changes in protein levels or mutations in cyclins and cyclin dependent kinases), enhanced survival mechanisms (overexpression of Bcl-2 or NF- κ B), altered cell metabolism (such as occurs when mutant KRAS enhances glucose uptake and aerobic glycolysis), tumor-stromal interactions, and angiogenesis (e.g., production of vascular endothelial growth factor [VEGF] in response to HIF-2 α in RCC) can also be successfully exploited to relatively specifically target cancers. However, resistance to inhibition of specific oncogenic pathways almost always eventually develops. In addition, targeting defects in tumor-suppressor genes has been much more difficult, both because the target of mutation is often deleted and because it is much more difficult to restore normal function than to inhibit abnormal function of a protein.

Synthetic lethality occurs when loss of function in either of two or more genes individually has limited effects on cell survival but loss of function in both (or more) genes leads to cell death. In the case of oncogene addicted pathways, identifying genes that have a synthetic lethal relationship with the activated pathway may allow enhanced cell killing and decreased resistance by targeting those genes or their proteins. In the case of mutant tumor-suppressor genes, identifying genes that have a synthetic lethal relationship to those mutated pathways may allow targeting by inhibiting proteins required uniquely by those cells for survival or proliferation (Fig. 68-3). This is a much more tractable approach than attempting to repair normal function of the mutant suppressor gene itself. Examples of synthetic lethality with potential clinical impact have been identified. For instance, cells with mutations in the *BRCA1* or *BRCA2* tumor-suppressor genes (e.g., a subset of breast and ovarian cancers) are unable to repair DNA damage by homologous recombination. Poly ADP ribose polymerase (PARP) are a family of proteins important for single-strand break (SSB) DNA repair. PARP inhibition results in selective killing of cancer cells which have lost *BRCA1* or *BRCA2* function. Trials have shown effectiveness of PARP inhibition in patients with BRCA mutant ovarian and breast cancers. Both olaparib (ovarian, breast) and rucaparib (ovarian) have been approved for this indication and others are in trials. The concept of synthetic lethality provides a framework for genetic screens to identify other synthetic lethal combinations involving known tumor-suppressor genes, and development of novel therapeutic agents to target dependent pathways. Other unique aspects of malignant tumors, including those outlined elsewhere in the chapter, may also be vulnerable to synthetic lethal interactions.

■ EPIGENETIC INFLUENCES ON CANCER GENE TRANSCRIPTION

Chromatin structure regulates the hierarchical order of sequential gene transcription that governs differentiation and tissue homeostasis. Disruption of chromatin remodeling (the process of modifying

chromatin structure to control exposure of specific genes to transcriptional proteins, thereby controlling the expression of those genes) leads to aberrant gene expression that can significantly alter the biology of cells including inducing proliferation or migration of cells. *Epigenetics* is defined as changes that alter the pattern of gene expression that persist across at least one cell division, but are not caused by changes in the DNA code.

Epigenetic changes include alterations of chromatin structure mediated by methylation of cytosine residues of DNA (primarily in context of CpG dinucleotides in somatic cells), modification of histones by altering acetylation or methylation, or changes in higher-order chromosome structure (Fig. 68-4). Appropriate control of DNA methylation is essential for normal cell function and development and both methylation and hypomethylation of histones occurs in cancers. Hypermethylation of DNA promoter regions is a common mechanism by which tumor-suppressor loci are epigenetically silenced in cancer cells. Thus one allele of a tumor suppressor gene may be inactivated by mutation or deletion (as occurs in loss of heterozygosity), while expression of the other allele is epigenetically silenced, usually by methylation leading to loss of gene function. Aberrant hypomethylation is also frequently found in a number of cancers consistent with the dysregulated pattern of gene transcription that is a hallmark of cancer cells with some genes being inappropriately turned off while others are inappropriately turned on.

Acetylation of the amino terminus of the core histones H3 and H4 induces an open chromatin conformation that promotes transcription initiation. Histone acetylases are components of coactivator complexes recruited to promoter/enhancer regions by sequence-specific transcription factors during the activation of genes (Fig. 68-4). Histone deacetylases (HDACs; multiple HDACs are encoded in the human genome) are recruited to genes by transcriptional repressors and prevent the initiation of gene transcription. Methylated cytosine residues in promoter regions become associated with methyl cytosine-binding proteins that recruit protein complexes with HDAC activity. The balance between permissive and inhibitory chromatin structure is therefore largely

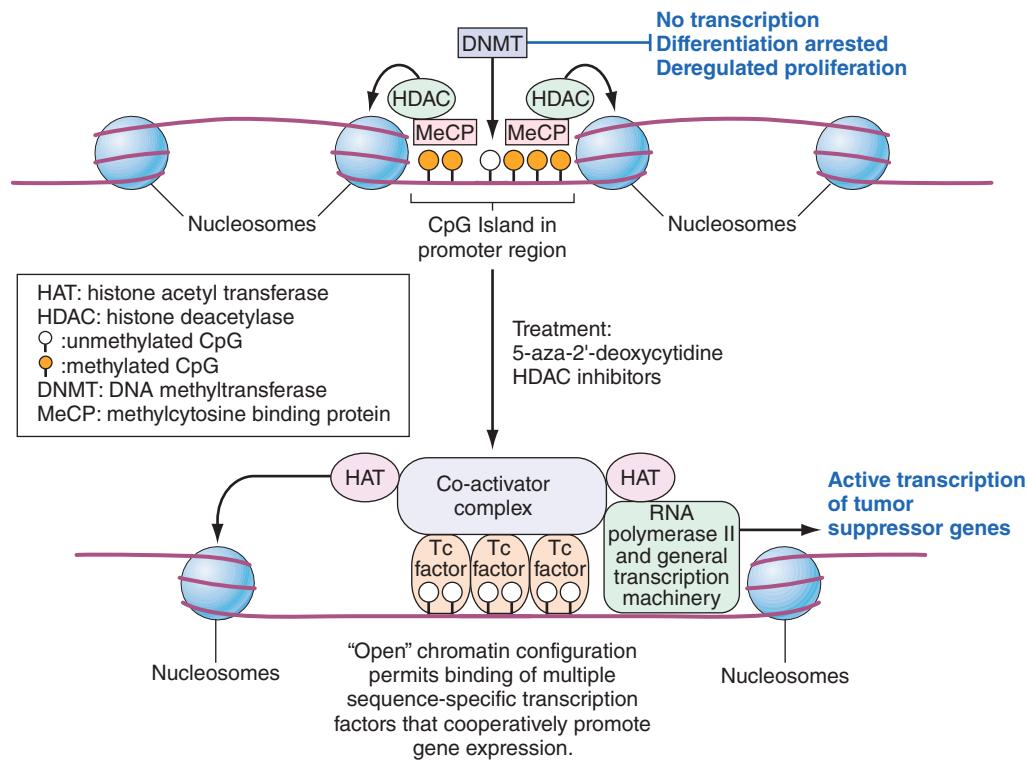


FIGURE 68-4 Epigenetic regulation of gene expression in cancer cells. Tumor-suppressor genes are often epigenetically silenced in cancer cells. In the upper portion, a CpG island within the promoter and enhancer regions of the gene has been methylated, resulting in the recruitment of methyl-cytosine binding proteins (MeCP) and complexes with histone deacetylase (HDAC) activity. Chromatin is in a condensed, nonpermissive conformation that inhibits transcription. Clinical trials are under way utilizing the combination of demethylating agents such as 5-aza-2'-deoxycytidine plus HDAC inhibitors, which together confer an open, permissive chromatin structure (lower portion). Transcription factors bind to specific DNA sequences in promoter regions and, through protein-protein interactions, recruit coactivator complexes containing histone acetyl transferase (HAT) activity. This enhances transcription initiation by RNA polymerase II and associated general transcription factors. The expression of the tumor-suppressor gene commences, with phenotypic changes that may include growth arrest, differentiation, or apoptosis.

determined by the activity of transcription factors in modulating the “histone code” and the methylation status of the genetic regulatory elements of genes.

The pattern of gene transcription is aberrant in all human cancers, and in many cases, epigenetic events are responsible. Epigenetic events play a critical role in carcinogenesis (e.g., long-lasting changes in methylation induced by smoking) and are found in pre-malignant lesions. Unlike genetic events that alter DNA primary structure (e.g., deletions), epigenetic changes are potentially reversible and appear amenable to therapeutic intervention. In certain human cancers, including a subset of pancreatic cancers and multiple myeloma, the p16^{Ink4a} promoter is inactivated by methylation, thus permitting the unchecked activity of CDK4/cyclin D and rendering pRb nonfunctional. In sporadic forms of renal, breast, and colon cancer, the von Hippel-Lindau (*VHL*), breast cancer 1 (*BRCA1*), and serine/threonine kinase 11 (*STK11*) genes, respectively, can be epigenetically silenced. Other targeted genes include the p15^{Ink4b} CDK inhibitor, glutathione-S-transferase (which detoxifies reactive oxygen species [ROS]), and the E-cadherin molecule (important for junction formation between epithelial cells). Epigenetic silencing can affect genes involved in DNA repair, thus predisposing to further genetic damage. Examples include MLH1 (mut L homologue in sporadic colon cancers that have microsatellite instability) and MSH2 in a subset of hereditary non-polyposis colon cancer patients who have a mutation in the 3' end of epithelial cell adhesion molecule (EPCAM). These are critical genes involved in repair of mismatched bases that occur during DNA synthesis and their silencing can lead to mutations in the DNA.

Human leukemias often have chromosomal translocations that code for novel fusion proteins with activities that alter chromatin structure by interacting with HDACs or histone acetyl transferases (HATs). For example, the promyelocytic leukemia-retinoic acid receptor α (PML-RAR α) fusion protein, generated by the t(15;17) translocation observed in most cases of acute promyelocytic leukemia (APL), binds to promoters containing retinoic acid response elements and recruits HDACs to these promoters, effectively inhibiting gene expression. This arrests differentiation at the promyelocyte stage and promotes tumor cell proliferation and survival. Treatment with pharmacologic doses of all-*trans* retinoic acid (ATRA), the ligand for RAR α , results in the release of HDAC activity and the recruitment of coactivators, which overcome the differentiation block. This induced differentiation of APL cells has improved treatment of these patients but also has led to a novel treatment toxicity when newly differentiated tumor cells infiltrate the lungs. ATRA represents a treatment paradigm for the reversal of epigenetic changes in cancer. Other leukemia-associated fusion proteins, such as Tel-acute myeloid leukemia (AML1), AML1-eight-twenty-one (ETO), and the MLL fusion proteins seen in AML and acute lymphocytic leukemia, also lead to repression through the HDAC complex. Therefore, efforts are ongoing to determine the structural basis for interactions between translocation fusion proteins and chromatin-remodeling proteins and to use this information to rationally design small molecules that will disrupt specific protein-protein associations, although this has proven to be technically difficult. Several drugs that block the enzymatic activity of HDACs are approved for cancer treatment and others are being tested. HDAC inhibitors have demonstrated sufficient antitumor activity against cutaneous T-cell lymphoma (vorinostat, romidepsin), peripheral T-cell lymphoma (romidepsin, belinostat), and multiple myeloma (panobinostat) to be approved by the FDA.

HDAC inhibitors (HDACi) have also demonstrated antitumor activity in clinical studies against some solid tumors and additional studies are ongoing. HDACi may target cancer cells via a number of mechanisms including both epigenetic modulation via histone acetylation as well as effects on other proteins which are acetylated. Some of HDACi's pleiotropic effects include: enhancement of apoptosis by upregulation of a number of proteins that enhance apoptosis including death receptors (DR4/5, FAS, and their ligands) and downregulation of proteins that inhibit apoptosis (e.g., X-linked inhibitor of apoptosis (XIAP); upregulation of proteins that inhibit cell cycle progression (e.g., p21Cip1/Waf1); inhibition of DNA repair and generation of ROS leading to increased DNA damage; and disruption of the chaperone protein HSP90.

Efforts are also under way to modulate other epigenetic processes such as reversing the hypermethylation of CpG islands that characterizes many malignancies. Drugs that induce DNA demethylation, such as 5-aza-2-deoxycytidine, can lead to reexpression of silenced genes in cancer cells with restoration of function, and 5-aza-2-deoxycytidine is approved for use in myelodysplastic syndrome (MDS). However, 5-aza-2-deoxycytidine has limited aqueous solubility and is myelo-suppressive limiting its usefulness. Other inhibitors of DNA methyltransferases are in development. In ongoing clinical trials, inhibitors of DNA methylation are being combined with HDAC inhibitors, with the idea that reversing coexisting epigenetic changes will reverse the deregulated patterns of gene transcription in cancer cells. Epigenetic gene regulation can also occur via microRNAs or long non-coding RNAs (lncRNA). MicroRNAs are short (average 22 nucleotides in length) RNA molecules that silence gene expression after transcription by binding and inhibiting the translation or promoting the degradation of mRNA transcripts. It is estimated that >1000 microRNAs are encoded in the human genome. Each tissue has a distinctive repertoire of microRNA expression and this pattern is altered in specific ways in cancers. Specific correlations between microRNA expression and tumor biology and clinical behavior are just now emerging. Therapies targeting microRNAs are not currently at hand but represent an ongoing area of treatment development. lncRNAs are longer than 200 nucleotides and comprise the largest group of noncoding RNAs. Some of them have been shown to play important roles in gene regulation. The potential for altering these RNAs for therapeutic benefit is an area of active investigation, although much more needs to be learned before this will be feasible.

APOPTOSIS AND OTHER MECHANISMS OF CELL DEATH

Tissue homeostasis requires a balance between the death of aged, terminally differentiated cells or severely damaged cells and their renewal by proliferation of committed progenitors. Genetic damage to growth-regulating genes of stem cells could lead to catastrophic results for the host as a whole. Thus, genetic events causing activation of oncogenes or loss of tumor suppressors, which would be predicted to lead to unregulated cell proliferation unless corrected, usually activate signal transduction pathways that block aberrant cell proliferation. These pathways can lead to a form of programmed cell death (*apoptosis*) or irreversible growth arrest (*senescence*). Much as a panoply of intra- and extracellular signals impinge upon the core cell cycle machinery to regulate cell division, so too these signals are transmitted to a core enzymatic machinery that regulates cell death and survival.

Apoptosis is a tightly regulated process induced by two main pathways (Fig. 68-5). The extrinsic pathway of apoptosis is activated by cross-linking members of the tumor necrosis factor (TNF) receptor superfamily, such as CD95 (Fas) and death receptors DR4 and DR5, by their ligands, Fas ligand or TRAIL (TNF-related apoptosis-inducing ligand), respectively. This induces the association of FADD (Fas-associated death domain) and procaspase-8 to death domain motifs of the receptors. Caspase-8 is activated and then cleaves and activates effector caspases-3 and -7, which then target cellular constituents (including caspase-activated DNase, cytoskeletal proteins, and a number of regulatory proteins), inducing the morphologic appearance characteristic of apoptosis, which pathologists term “karyorrhexis.” The intrinsic pathway of apoptosis is initiated by the release of cytochrome *c* and SMAC (second mitochondrial activator of caspases) from the mitochondrial intermembrane space in response to a variety of noxious stimuli, including DNA damage, loss of adherence to the extracellular matrix (ECM), oncogene-induced proliferation, and growth factor deprivation. Upon release into the cytoplasm, cytochrome *c* associates with dATP, procaspase-9, and the adaptor protein APAF-1, leading to the sequential activation of caspase-9 and effector caspases. SMAC binds to and blocks the function of inhibitor of apoptosis proteins (IAP), negative regulators of caspase activation.

The release of apoptosis-inducing proteins from the mitochondria is regulated by pro- and antiapoptotic members of the Bcl-2 family. Antiapoptotic members (e.g., Bcl-2, Bcl-XL, and Mcl-1) associate with

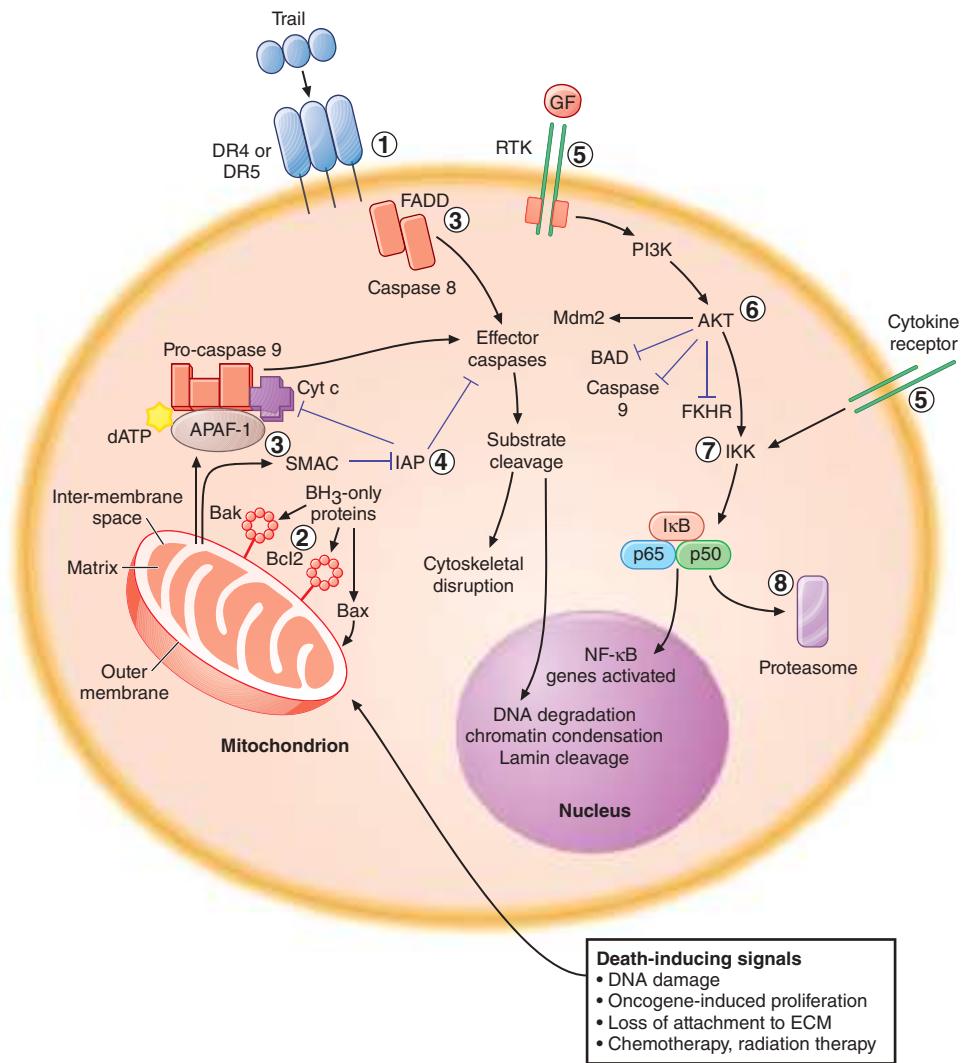


FIGURE 68-5 Therapeutic strategies to overcome aberrant survival pathways in cancer cells. 1. The extrinsic pathway of apoptosis can be selectively induced in cancer cells by TRAIL (the ligand for death receptors 4 and 5) or by agonistic monoclonal antibodies. 2. Inhibition of antiapoptotic Bcl-2 family members with antisense oligonucleotides or inhibitors of the BH₃-binding pocket will promote formation of Bak- or Bax-induced pores in the mitochondrial outer membrane. 3. Epigenetic silencing of APAF-1, caspase-8, and other proteins can be overcome using demethylating agents and inhibitors of histone deacetylases. 4. Inhibitor of apoptosis proteins (IAP) blocks activation of caspases; small-molecule inhibitors of IAP function (mimicking SMAC action) should lower the threshold for apoptosis. 5. Signal transduction pathways originating with activation of receptor tyrosine kinase receptors (RTKs) or cytokine receptors promote survival of cancer cells by a number of mechanisms. Inhibiting receptor function with monoclonal antibodies, such as trastuzumab or cetuximab, or inhibiting kinase activity with small-molecule inhibitors can block the pathway. 6. The Akt kinase phosphorylates many regulators of apoptosis to promote cell survival; inhibitors of Akt may render tumor cells more sensitive to apoptosis-inducing signals; however, the possibility of toxicity to normal cells may limit the therapeutic value of these agents. 7 and 8. Activation of the transcription factor NF-κB (composed of p65 and p50 subunits) occurs when its inhibitor, IκB, is phosphorylated by IκB-kinase (IKK), with subsequent degradation of IκB by the proteasome. Inhibition of IKK activity should selectively block the activation of NF-κB target genes, many of which promote cell survival. Inhibitors of proteasome function are FDA-approved and may work in part by preventing destruction of IκB, thus blocking NF-κB nuclear localization. NF-κB is unlikely to be the only target for proteasome inhibitors.

the mitochondrial outer membrane via their carboxyl termini, exposing to the cytoplasm a hydrophobic binding pocket composed of Bcl homology (BH) domains 1, 2, and 3 that is crucial for their activity. Perturbations of normal physiologic processes in specific cellular compartments lead to the activation of BH3-only proapoptotic family members (such as Bad, Bim, Bid, Puma, Noxa, and others) that can alter the conformation of the outer-membrane proteins Bax and Bak, which then oligomerize to form pores in the mitochondrial outer membrane resulting in cytochrome *c* release. If proteins comprised only by BH3 domains are sequestered by Bcl-2, Bcl-XL, or Mcl-1, pores do not form and apoptosis-inducing proteins are not released from the mitochondria. The ratio of levels of antiapoptotic Bcl-2 family members and the levels of proapoptotic BH3-only proteins at the mitochondrial membrane determines the activation state of the intrinsic pathway. The mitochondrion must therefore be recognized not only as an organelle with vital roles in intermediary metabolism and oxidative phosphorylation but also as a central regulatory structure of the apoptotic process.

The evolution of tumor cells to a more malignant phenotype requires the acquisition of genetic changes that subvert apoptosis pathways and promote cancer cell survival and resistance to anticancer therapies. However, cancer cells may be more vulnerable than normal cells to therapeutic interventions that target the apoptosis pathways that cancer cells depend upon. For instance, overexpression of Bcl-2 as a result of the t(14;18) translocation contributes to follicular lymphoma and it is highly expressed in many lymphoid malignancies including chronic lymphocytic leukemia (CLL). Upregulation of Bcl-2 expression is also observed in other cancers including prostate, breast, and lung cancers and melanoma. Targeting of antiapoptotic Bcl-2 family members has been accomplished by the identification of several low-molecular-weight compounds that bind to the hydrophobic pockets of either Bcl-2 or Bcl-XL and block their ability to associate with death-inducing BH3-only proteins. These compounds inhibit the antiapoptotic activities of Bcl-2 and Bcl-XL at nanomolar concentrations. An oral BH3 mimetic inhibitor of BCL-2, venetoclax, is approved for use in patients with refractory CLL with 17p deletion.

Preclinical studies targeting death receptors DR4 and -5 have demonstrated that recombinant, soluble, human TRAIL or humanized monoclonal antibodies with agonist activity against DR4 or -5 can induce apoptosis of tumor cells while sparing normal cells. The mechanisms for this selectivity may include expression of decoy receptors or elevated levels of intracellular inhibitors (such as FLIP, which competes with caspase-8 for FADD) by normal cells but not tumor cells. Synergy has been shown between TRAIL-induced apoptosis and chemotherapeutic agents in some preclinical studies. However, clinical studies have not yet shown significant activity of approaches targeting the TRAIL pathway.

Many of the signal transduction pathways perturbed in cancer promote tumor cell survival (Fig. 68-5). These include activation of the PI3K/Akt pathway, increased levels of the NF- κ B transcription factor, and epigenetic silencing of genes such as *APAF-1*(*apoptosis protease activating factor-1 involved in activating caspase-9 and essential for apoptosis*) and *caspase-8*. Each of these pathways is a target for therapeutic agents that, in addition to affecting cancer cell proliferation or gene expression, may render cancer cells more susceptible to apoptosis, thus promoting synergy when combined with other chemotherapeutic agents.

Some tumor cells resist drug-induced apoptosis indirectly by eliminating the noxious stimulus-inducing apoptosis through expression of one or more members of the ABC (ATP-binding cassette proteins) family of ATP-dependent efflux pumps that mediate the multidrug-resistance (MDR) phenotype. The prototype member of this family, P-glycoprotein (PGP), spans the plasma membrane 12 times and has two ATP-binding sites. Hydrophobic drugs (e.g., anthracyclines and vinca alkaloids) are recognized by PGP as they enter the cell and are pumped out. Numerous clinical studies have failed to demonstrate that drug resistance can be overcome using inhibitors of PGP. However, ABC transporters have different substrate specificities, and inhibition of a single family member may not be sufficient to overcome the MDR phenotype. Efforts to reverse PGP-mediated drug resistance continue.

Cells, including cancer cells, can also undergo other mechanisms of cell death including *autophagy* (degradation of proteins and organelles by lysosomal proteases) and *necrosis* (digestion of cellular components and rupturing of the cell membrane). Necrosis usually occurs in response to external forces resulting in release of cellular components, which leads to inflammation and damage to surrounding tissues. Although necrosis was thought to be unprogrammed, evidence now suggests that at least some aspects may also be programmed. The exact role of necrosis in cancer cell death in various settings is still being determined. In addition to its role in cell death, autophagy can also serve as a homeostatic mechanism to promote survival for the cell by recycling cellular components to provide necessary energy. The mechanisms that control the balance between enhancing survival versus leading to cell death are still not fully understood. Autophagy appears to play conflicting roles in the development and survival of cancer. Early in the carcinogenic process it can act as a tumor suppressor by preventing the cell from accumulating abnormal proteins and organelles. However, in established tumors, it may serve as a mechanism of survival for cancer cells when they are stressed by damage such as from chemotherapy. Preclinical studies have indicated that inhibition of this process can enhance the sensitivity of cancer cells to chemotherapy and ongoing trials are evaluating inhibitors of autophagy in combination with chemotherapy. Better understanding of the factors that control the survival-promoting versus death-inducing aspects of autophagy is required in order to know how to best manipulate it for therapeutic benefit.

■ METASTASIS

The metastatic process accounts for the vast majority of deaths from solid tumors and therefore an understanding of this process is critical for improvements in survival from cancer. The biology of metastasis is complex and requires multiple steps. The initial step involves cell migration and invasion through the ECM. The three major features of tissue invasion are cell adhesion to the basement membrane, local proteolysis of the membrane, and movement of the cell through the rent in the membrane and the ECM. Cells that lose contact with the

ECM normally undergo programmed cell death (anoikis-apoptosis induced by the loss of contact) and this process has to be suppressed in cells that metastasize. Another process important for many, but not necessarily all, metastasizing epithelial cancer cells is epithelial-mesenchymal transition (EMT). This is a process by which cells lose their epithelial properties and gain mesenchymal properties. This normally occurs during the developmental process in embryos, allowing cells to migrate to their appropriate destinations in the embryo. It also occurs in wound healing, tissue regeneration, and fibrotic reactions, but in all of these processes, cells stop proliferating when the process is complete. Malignant cells that metastasize often undergo EMT as an important step in that process but retain the capacity for unregulated proliferation. However, there is evidence that not all metastasizing cancer cells require EMT, and the exact role of EMT in different metastasizing cancer cells continues to be elucidated. Malignant cells that gain access to the circulation must then repeat those steps at a remote site, find a hospitable niche in a foreign tissue, avoid detection and elimination by host defenses including the immune system, and induce the growth of new blood vessels. Some metastatic cells occur as oligoclonal clusters which appear to be more potent in establishing metastasis than single cells, perhaps through differential and cooperative effects in evading host defenses. The rate-limiting step for metastasis is the ability for tumor cells to survive and expand in the novel microenvironment of the metastatic site, and multiple host-tumor interactions determine the ultimate outcome (Fig. 68-6). Few drugs have been developed to attempt to directly target the process of metastasis, in part because the specifics of the critical steps in the process that would be potentially good targets for drugs are still being identified. However, a number of potential targets are known. HER2 can enhance the metastatic potential of breast cancer cells and as discussed above, the monoclonal antibody trastuzumab which targets HER2, improves survival in the adjuvant setting for HER2+ breast cancer patients. A number of other potential targets that increase metastatic potential of cells in preclinical studies include: HIF-1 and 2, transcription factors induced by hypoxia within tumors; growth factors (e.g., cMET and VEGFR); oncogenes (e.g., SRC); adhesion molecules (e.g., focal adhesion kinase, FAK); ECM proteins (e.g., matrix metalloproteinases 1 and 2); and inflammatory molecules (e.g., COX-2).

The metastatic phenotype is likely restricted to a small fraction of tumor cells (Fig. 68-6). A number of genetic and epigenetic changes are required for tumor cells to be able to metastasize, including activation of metastatic-promoting genes and inhibition of genes that suppress the metastatic ability. Given the role of microRNAs in controlling gene expression (see epigenetic section) including those critical to the metastatic process, efforts are under way to modulate these to try to inhibit metastasis. Cells with metastatic capability frequently express chemokine receptors that are likely important in the metastatic process. A number of candidate metastasis-suppressor genes have been identified, including genes coding for proteins that enhance apoptosis, suppress cell division, are involved in the interactions of cells with each other or the ECM, or suppress cell migration. The loss of function of these genes enhances metastasis. Gene expression profiling is being used to study the metastatic process and other properties of tumor cells that may predict susceptibilities.

An example of the ability of malignant cells to survive and grow in a novel microenvironment is bone metastases. Bone metastases can be extremely painful, cause fractures of weight-bearing bones, can lead to hypercalcemia, and are a major cause of morbidity for cancer patients. Osteoclasts and their monocyte-derived precursors express the surface receptor RANK (receptor activator of NF- κ B), which is required for terminal differentiation and activation of osteoclasts. Osteoblasts and other stromal cells express RANK ligand (RANKL), as both a membrane-bound and soluble cytokine. Osteoprotegerin (OPG), a soluble receptor for RANKL produced by stromal cells, acts as a decoy receptor to inhibit RANK activation. The relative balance of RANKL and OPG determines the activation state of RANK on osteoclasts. Many tumors increase osteoclast activity by secretion of substances such as parathyroid hormone (PTH), PTH-related peptide, interleukin (IL)-1, or Mip1 that perturb the homeostatic balance of bone remodeling by

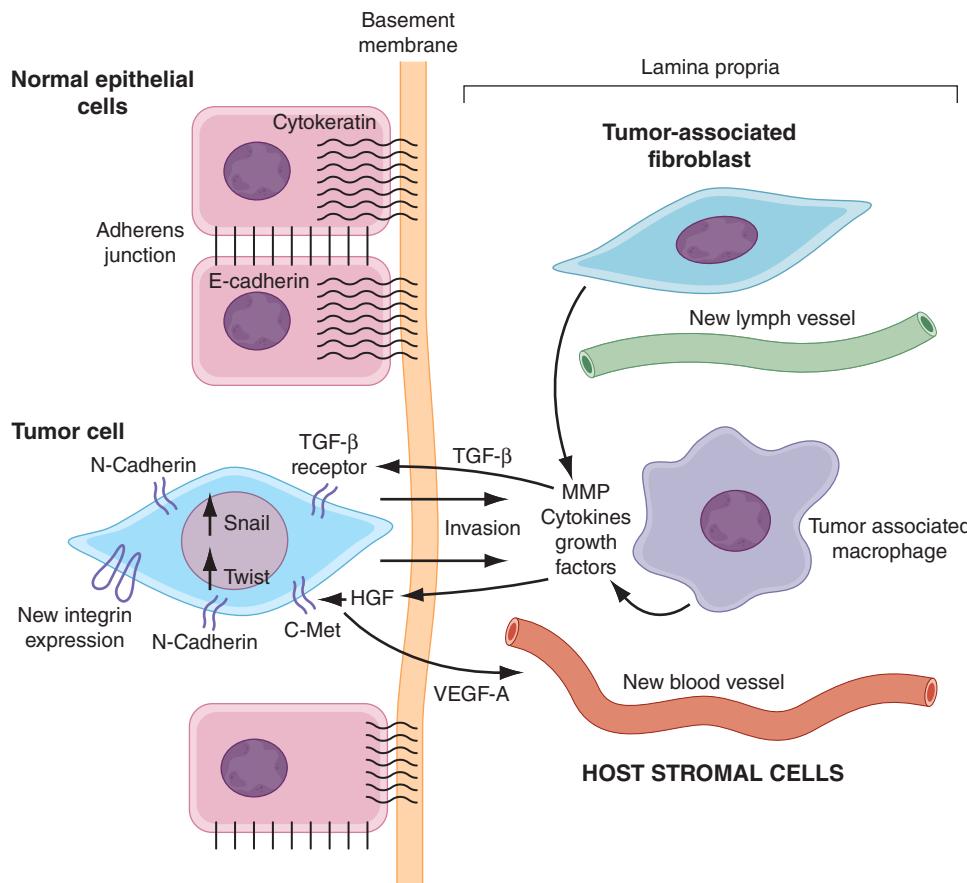


FIGURE 68-6 Oncogene signaling pathways are activated during tumor progression and promote metastatic potential. This figure shows a cancer cell that has undergone epithelial to mesenchymal transition (EMT) under the influence of several environmental signals. Critical components include activated transforming growth factor beta (TGF- β) and the hepatocyte growth factor (HGF)/c-Met pathways, as well as changes in the expression of adhesion molecules that mediate cell-cell and cell-extracellular matrix interactions. Important changes in gene expression are mediated by the Snail and Twist family of transcriptional repressors (whose expression is induced by the oncogenic pathways), leading to reduced expression of E-cadherin, a key component of adherens junctions between epithelial cells. This, in conjunction with upregulation of N-cadherin, a change in the pattern of expression of integrins (which mediate cell–extracellular matrix associations that are important for cell motility), and a switch in intermediate filament expression from cytokeratin to vimentin, results in the phenotypic change from adherent highly organized epithelial cells to motile and invasive cells with a fibroblast or mesenchymal morphology. EMT is thought to be an important step leading to metastasis in some human cancers. Host stromal cells, including tumor-associated fibroblasts and macrophages, play an important role in modulating tumor cell behavior through secretion of growth factors and proangiogenic cytokines, and matrix metalloproteinases that degrade the basement membrane. VEGF-A, -C, and -D are produced by tumor cells and stromal cells in response to hypoxemia or oncogenic signals, and induce production of new blood vessels and lymphatic channels through which tumor cells metastasize to lymph nodes or tissues.

increasing RANK signaling. One example is multiple myeloma, where tumor cell-stromal cell interactions activate osteoclasts and inhibit osteoblasts, leading to the development of multiple lytic bone lesions. Inhibition of RANK ligand by an antibody (denosumab) can prevent further bone destruction. Bisphosphonates are also effective inhibitors of osteoclast function that are used in the treatment of cancer patients with bone metastases.

CANCER STEM CELLS

Normal tissues have stem cells capable of self-renewal and repairing damaged tissue whereas the majority of cells in normal tissues do not have this capacity. Similarly, only a small proportion of the cells within a tumor are capable of initiating colonies *in vitro* or forming tumors at high efficiency when injected into immunocompromised NOD/SCID mice. For example, acute and chronic myeloid leukemias (AML and CML) have a small population of cells (estimated to be <1%) that have properties of stem cells, such as unlimited self-renewal and the capacity to cause leukemia when serially transplanted in mice. These cells have an undifferentiated phenotype (Thy1-CD34+CD38- and do not express other differentiation markers) and resemble normal stem cells in many ways, but are no longer under homeostatic control (Fig. 68-7). Solid tumors may also contain a population of stem cells. It is not yet known how often cancers may originate within a stem cell population. Cancer stem cells, like their normal counterparts, have unlimited proliferative capacity and paradoxically traverse the cell cycle at a slow rate; cancer growth occurs largely due to expansion of the stem cell pool, the

unregulated proliferation of an amplifying population, and failure of apoptosis pathways (Fig. 68-7). Slow cell cycle progression and high levels of expression of antiapoptotic Bcl-2 family members and drug efflux pumps of the MDR family render cancer stem cells less vulnerable to cancer chemotherapy or radiation therapy. Implicit in the cancer stem cell hypothesis is the idea that failure to cure most human cancers is due to the fact that current therapeutic agents do not kill the stem cells. Identification and isolation of cancer stem cells will allow determination of the aberrant signaling pathways that distinguish these cells from normal tissue stem cells. These would serve as potential therapeutic targets. Evidence that cells with stem cell properties can arise from other epithelial cells within the cancer by processes such as epithelial-mesenchymal transition also implies that it is essential to treat all of the cancer cells, and not just those with current stem cell-like properties, in order to eliminate the self-renewing cancer cell population. The exact nature of cancer stem cells remains an area of investigation. One of the unanswered questions is the exact origin of cancer stem cells for the different cancers.

PLASTICITY AND RESISTANCE

Cancer cells, and especially stem cells, have the capacity for significant plasticity allowing them to alter multiple aspects of cell biology in response to external factors (e.g., chemotherapy, radiation therapy, inflammation, immune response). In addition, heterogeneity between the different clones of cells within the tumor population and their interactions with each other and the tumor microenvironment provides the

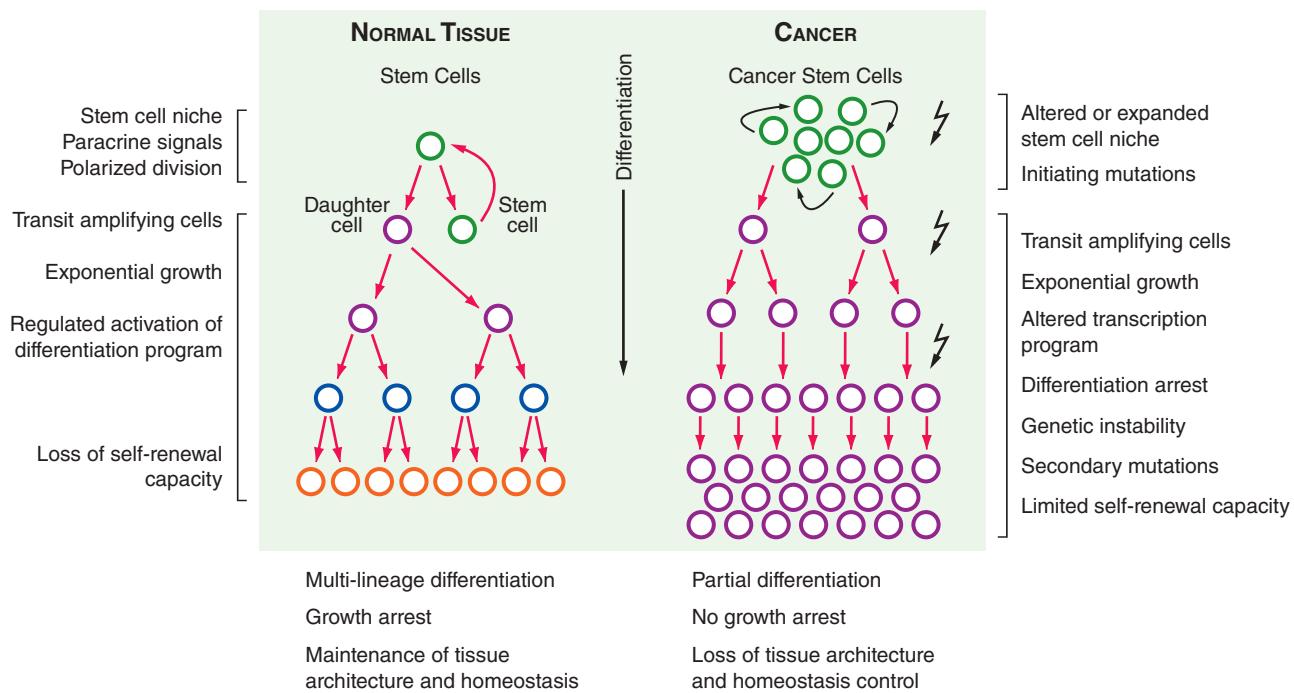


FIGURE 68-7 Cancer stem cells play a critical role in the initiation, progression, and resistance to therapy of malignant neoplasms. In normal tissues (left), homeostasis is maintained by asymmetric division of stem cells leading to one progeny cell that will differentiate and one cell that will maintain the stem cell pool. This occurs within highly specific niches unique to each tissue, such as in close apposition to osteoblasts in bone marrow, or at the base of crypts in the colon. Here, paracrine signals from stromal cells, such as sonic hedgehog or Notch-ligands, as well as upregulation of β -catenin and telomerase, help to maintain stem cell features of unlimited self-renewal while preventing differentiation or cell death. This occurs in part through upregulation of the transcriptional repressor Bmi-1 and inhibition of the p16^{Ink4a}/Arf and p53 pathways. Daughter cells leave the stem cells niche and enter a proliferative phase (referred to as *transit-amplifying*) for a specified number of cell divisions, during which time a developmental program is activated, eventually giving rise to fully differentiated cells that have lost proliferative potential. Cell renewal equals cell death, and homeostasis is maintained. In this hierarchical system, only stem cells are long-lived. The hypothesis is that cancers harbor stem cells that make up a small fraction (i.e., 0.001–1%) of all cancer cells. These cells share several features with normal stem cells, including an undifferentiated phenotype, unlimited self-renewal potential, a capacity for some degree of differentiation; however, due to initiating mutations (mutations are indicated by lightning bolts), they are no longer regulated by environmental cues. The cancer stem cell pool is expanded, and rapidly proliferating progeny, through additional mutations, may attain stem cell properties, although most of this population is thought to have a limited proliferative capacity. Differentiation programs are dysfunctional due to reprogramming of the pattern of gene transcription by oncogenic signaling pathways. Within the cancer transit-amplifying population, genomic instability generates aneuploidy and clonal heterogeneity as cells attain a fully malignant phenotype with metastatic potential. The cancer stem cell hypothesis has led to the idea that current cancer therapies may be effective at killing the bulk of tumor cells but do not kill tumor stem cells, leading to a regrowth of tumors that is manifested as tumor recurrence or disease progression. Research is in progress to identify unique molecular features of cancer stem cells that can lead to their direct targeting by novel therapeutic agents.

tumor with the capacity for significant plasticity in dealing with both internal and external stresses. Thus, a major problem in cancer therapy is that malignancies have a wide spectrum of mechanisms for both initial and adaptive resistance to treatments. These include inhibiting drug delivery to the cancer cells, blocking drug uptake and retention, increasing drug metabolism, altering levels of target proteins making them less sensitive to drugs, acquiring mutations in target proteins making them no longer sensitive to the drug, modifying metabolism and cell signaling pathways, using alternate signaling pathways, adjusting the cell replication process including mechanisms by which the cell deals with DNA damage, inhibiting apoptosis, and evading the immune system. Thus, most metastatic cancers (except those curable with chemotherapy such as germ cell tumors) eventually become resistant to the therapy being utilized. Overcoming resistance is a major area of research.

CANCER METABOLISM

One of the distinguishing characteristics of cancer cells is that they have altered metabolism as compared with normal cells in supporting survival, their high rates of proliferation, and ability to metastasize. Complicating studies evaluating metabolic differences between normal and malignant cells is that there is heterogeneity in metabolism between different cells within a cancer. Malignant cells must focus a significant fraction of their energy resources into synthesis of proteins and other molecules (building blocks required for the production of new cells) while still maintaining sufficient ATP production to survive and grow. Although normal proliferating cells also have similar needs, there are differences in how cancer cells metabolize glucose and a number of

other compounds including the amino acid glutamine as compared to normal cells in part because of genetic and epigenetic changes within cancer cells but also likely due to differences in the environments of cancer and normal cells. Many cancer cells utilize aerobic glycolysis (the Warburg effect) (Fig. 68-8) to metabolize glucose leading to increased lactic acid production whereas normal cells utilize oxidative phosphorylation in mitochondria under aerobic conditions, a much more efficient process for generating ATP for energy utilization but one that does not produce the same level of building blocks needed for new cells. One consequence is increased glucose uptake by cancer cells, a fact utilized in fluorodeoxyglucose (FDG)-positron emission tomography (PET) scanning to detect tumors. A number of proteins in cancer cells, including cMYC, HIF1, RAS, p53, pRB, and AKT are all involved in modulating glycolytic processes and controlling the Warburg effect. Although these pathways remain difficult to target therapeutically, both the p13kinase pathway with signaling through mTOR and the AMP-activated kinase (AMPK) pathway that inhibits mTORC1 (a protein complex that includes mTOR) are important in controlling the glycolytic process and thus provide potential targets for inhibiting this process. An inhibitor of MTOR is approved for use against RCC (temsirolimus) and another inhibitor (everolimus) has activity against breast, neuroendocrine, and RCC. Other MTOR inhibitors are in trials and modulators of AMPK are being investigated. The inefficient utilization of glucose by malignant cells also leads to a need for alternative metabolic pathways for other compounds as well, one of which is glutamine. Similar to glucose, this provides both a source for structural molecules as well as energy production. Similarly to glucose, glutamine is also inefficiently utilized by cancer cells. Cancer cells

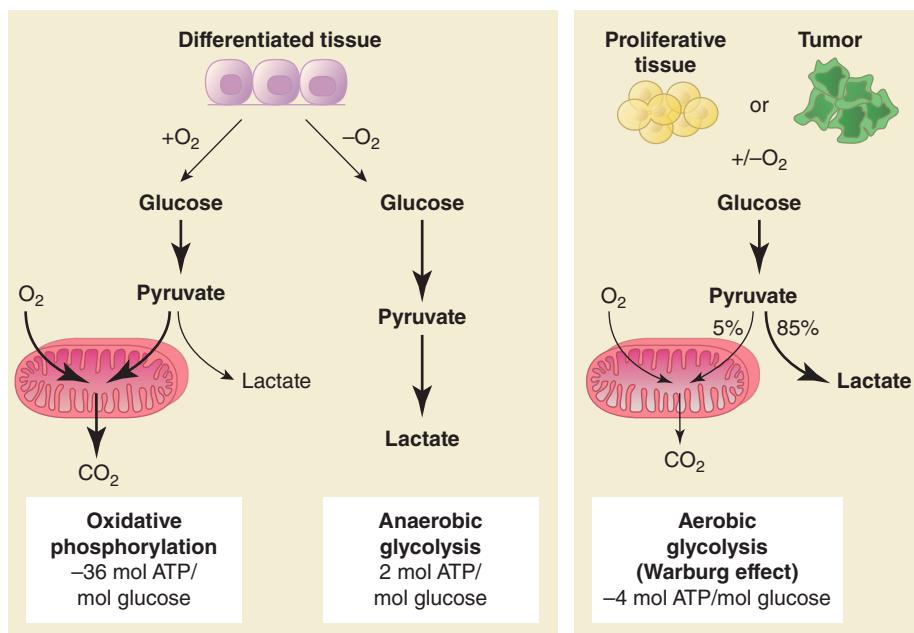


FIGURE 68-8 Warburg vs oxidative phosphorylation. In most normal tissues, the vast majority of cells are differentiated and dedicated to a particular function within the organ in which they reside. The metabolic needs are mainly for energy and not for building blocks for new cells. In these tissues, ATP is generated by oxidative phosphorylation that efficiently generates about 36 molecules of ATP for each molecule of glucose metabolized. By contrast, proliferative tumor tissues, especially in the setting of hypoxia, a typical condition within tumors, use aerobic glycolysis to generate energy for cell survival and generation of building blocks for new cells.

can also take up nutrients released by surrounding cells and tissues increasing the complexity of successfully therapeutically inhibiting metabolism in cancer.

Mutations in genes involved in the metabolic process occur in a number of cancers. Among the most frequently found to date are mutations in isocitrate dehydrogenases 1 and 2 (IDH1 and 2). These have been most commonly seen in gliomas, acute myeloid leukemias (AML), and intra-hepatic cholangiocarcinomas. These mutations lead to the production of an oncometabolite (2-hydroxyglutarate, 2HG) instead of the normal product α -ketoglutarate. Although the exact mechanisms of oncogenesis by 2HG are still being elucidated, α -ketoglutarate is a key cofactor for a number of dioxygenases involved in controlling DNA methylation. 2HG can act as a competitive inhibitor for α -ketoglutarate leading to alterations in methylation status (primarily hypermethylation) of genes (leading to epigenetic changes) that can have profound effects on a number of cellular processes including differentiation. Inhibitors of mutants IDH1 and IDH2 are being developed. To date, they have had some activity against IDH mutant AML but less activity against glioblastomas or cholangiocarcinomas.

Much needs to be learned about the specific differences in metabolism between cancer cells and normal cells; however, even with the currently limited state of knowledge, modulators of metabolism are being tested clinically. The first of these is the anti-diabetic agent metformin, both alone and in combination with chemotherapeutic agents. Metformin inhibits gluconeogenesis and may have direct effects on tumor cells by activating the 5'-adenosine monophosphate-activated kinase (AMPK), a serine/threonine protein kinase which is downstream of the LKB1 tumor suppressor, and thus inhibiting mammalian target of rapamycin complex 1 (mTORC1). This leads to decreased protein synthesis and proliferation. Studies to date have not yet established metformin to have a clear role as an anticancer agent. Additional approaches being evaluated include other modulators of glucose metabolism (e.g., pioglitazone) and inhibiting glutaminase (important for glutamine utilization).

TUMOR MICROENVIRONMENT, ANGIOGENESIS, AND IMMUNE EVASION

Tumors consist not only of malignant cells but also of a complex microenvironment including many other types of cells (including inflammatory cells), ECM, secreted factors (including growth factors), reactive oxygen and nitrogen species, mechanical factors, blood vessels, and

lymphatics. This microenvironment is not static but rather is dynamic and continually evolving. Both the complexity and dynamic nature of the microenvironment enhance the difficulty of treating tumors. There are also a number of mechanisms by which the microenvironment can contribute to resistance to anti-cancer therapies.

One of the critical elements of tumor cell proliferation is delivery of oxygen, nutrients, and circulating factors important for growth and survival. The diffusion limit for oxygen in tissues is ~100–200 μ m and thus a critical aspect in the growth of tumors is the development of new blood vessels, or angiogenesis. The growth of primary and metastatic tumors to larger than a few millimeters requires the recruitment of blood vessels and vascular endothelial cells (ECs) to support their metabolic requirements. Thus, a critical element in growth of primary tumors and formation of metastatic sites is the *angiogenic switch*: the ability of the tumor to promote the formation of new capillaries from preexisting host vessels. The angiogenic switch is a phase in tumor development when the dynamic balance of pro- and antiangiogenic factors is tipped in favor of vessel formation by the effects of the tumor on its immediate environment. Stimuli for tumor angiogenesis include hypoxemia, inflammation, and genetic lesions in oncogenes or tumor suppressors that alter tumor cell gene expression. Angiogenesis consists of several steps, including the stimulation of ECs by growth factors, degradation of the ECM by proteases, proliferation and migration of ECs into the tumor, and the eventual formation of new capillary tubes.

Tumor blood vessels are not normal; they have chaotic architecture and blood flow. Due to an imbalance of angiogenic regulators such as VEGF and angiopoietins (see below), tumor vessels are tortuous and dilated with an uneven diameter, excessive branching, and shunting. Tumor blood flow is variable, with areas of hypoxemia and acidosis leading to the selection of variants that are resistant to hypoxemia-induced apoptosis (often due to the loss of p53 expression). Tumor vessel walls have numerous openings, widened interendothelial junctions, and discontinuous or absent basement membrane; this contributes to the high vascular permeability of these vessels and, together with lack of functional intratumoral lymphatics, causes increased interstitial pressure within the tumor (which also interferes with the delivery of therapeutics to the tumor; Figs. 68-9, 68-10, and 68-11). Tumor blood vessels lack perivascular cells such as pericytes and smooth-muscle cells that normally regulate flow in response to tissue metabolic needs.

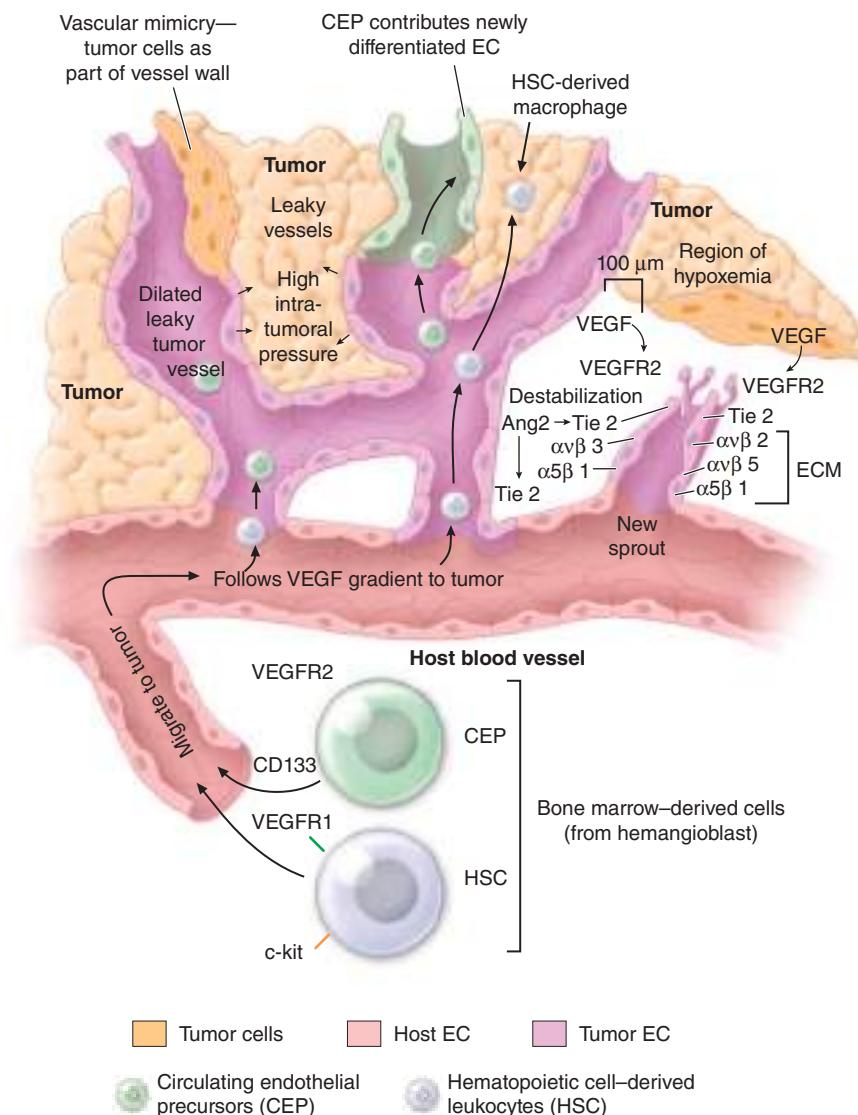


FIGURE 68-9 Tumor angiogenesis is a complex process involving many different cell types that must proliferate, migrate, invade, and differentiate in response to signals from the tumor microenvironment. Endothelial cells (ECs) sprout from host vessels in response to VEGF, bFGF, Ang2, and other proangiogenic stimuli. Sprouting is stimulated by VEGF/VEGFR2, Ang2/Tie-2, and integrin/extracellular matrix (ECM) interactions. Bone marrow-derived circulating endothelial precursors (CEPs) migrate to the tumor in response to VEGF and differentiate into ECs, while hematopoietic stem cells differentiate into leukocytes, including tumor-associated macrophages that secrete angiogenic growth factors and produce MMPs that remodel the ECM and release bound growth factors. Tumor cells themselves may directly form parts of vascular channels within tumors. The pattern of vessel formation is haphazard: vessels are tortuous, dilated, leaky, and branch in random ways. This leads to uneven blood flow within the tumor, with areas of acidosis and hypoxemia (which stimulate release of angiogenic factors) and high intratumoral pressures that inhibit delivery of therapeutic agents.

Unlike normal blood vessels, the vascular lining of tumor vessels is not a homogeneous layer of ECs but often consists of a mosaic of ECs and tumor cells with upregulated genes seen in ECs and vessel formation that can occur in hypoxic conditions because of their plasticity; the concept of cancer cell–derived vascular channels, which may be lined by ECM secreted by the tumor cells, is referred to as *vascular mimicry*. During tumor angiogenesis, ECs are highly proliferative and express a number of plasma membrane proteins that are characteristic of activated endothelium, including growth factor receptors and adhesion molecules such as integrins.

MECHANISMS OF TUMOR VESSEL FORMATION

Tumors use a number of mechanisms to promote vascularization, subverting normal angiogenic processes for this purpose (Fig. 68-9). Primary or metastatic tumor cells sometimes arise in proximity to host blood vessels and grow around these vessels, parasitizing nutrients by co-opting the local blood supply. However, most tumor blood vessels arise by the process of *sprouting*, in which tumors secrete trophic angiogenic molecules, the most potent being VEGFs, that induce the proliferation and migration of host ECs into the tumor. Sprouting in

normal and pathogenic angiogenesis is regulated by three families of transmembrane RTKs expressed on ECs and their ligands (VEGFs, angiopoietins, ephrins; Fig. 68-10), which are produced by tumor cells, inflammatory cells, or stromal cells in the tumor microenvironment.

When tumor cells arise in or metastasize to an avascular area, they grow to a size limited by hypoxemia and nutrient deprivation. Hypoxemia, a key regulator of tumor angiogenesis, causes the transcriptional induction of the genes encoding VEGF family members. VEGFs and their receptors are required for embryonic *vasculogenesis* (development of new blood vessels when none pre-exist) and normal (wound healing, corpus luteum formation) and pathologic angiogenesis (tumor angiogenesis, inflammatory conditions such as rheumatoid arthritis). VEGF-A is a heparin-binding glycoprotein with at least four isoforms (splice variants) that regulates blood vessel formation by binding to the RTKs VEGFR1 and VEGFR2, which are expressed on all ECs in addition to a subset of hematopoietic cells (Fig. 68-9). VEGFR2 regulates EC proliferation, migration, and survival, while VEGFR1 may act as an antagonist of R2 in ECs but is probably also important for angioblast differentiation during embryogenesis. Tumor vessels may be more dependent on VEGFR signaling for growth and survival than

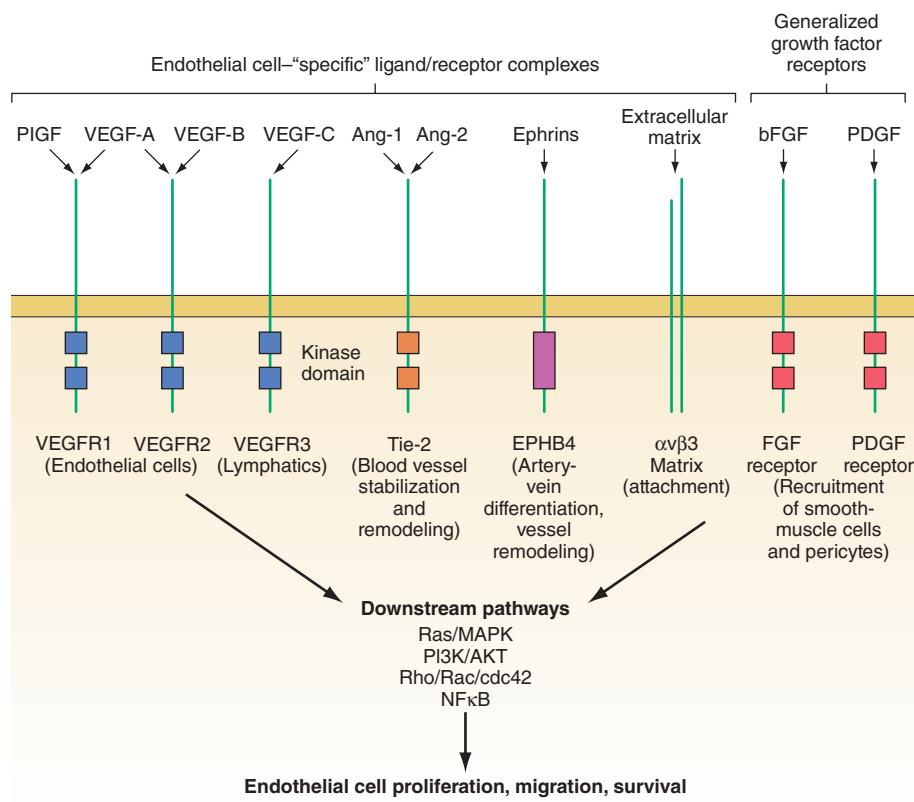


FIGURE 68-10 Critical molecular determinants of endothelial cell biology. Angiogenic endothelium expresses a number of receptors not found on resting endothelium. These include receptor tyrosine kinases (RTKs) and integrins that bind to the extracellular matrix (EC) and mediate endothelial cell (EC) adhesion, migration, and invasion. ECs also express RTK (i.e., the fibroblast growth factor [FGF] and platelet-derived growth factor [PDGF] receptors) that are found on many other cell types. Critical functions mediated by activated RTK include proliferation, migration, and enhanced survival of endothelial cells, as well as regulation of the recruitment of perivascular cells and bloodborne circulating endothelial precursors and hematopoietic stem cells to the tumor. Intracellular signaling via EC-specific RTK utilizes molecular pathways that may be targets for future antiangiogenic therapies.

normal ECs. While VEGF signaling is a critical initiator of angiogenesis, this is a complex process regulated by additional signaling pathways (Fig. 68-10). The angiopoietin, Ang1, produced by stromal cells, binds to the EC RTK Tie-2 and promotes the interaction of ECs with the ECM and perivascular cells, such as pericytes and smooth-muscle cells, to form tight, nonleaky vessels. PDGF and basic fibroblast growth factor (bFGF) help to recruit these perivascular cells. Ang1 is required for maintaining the quiescence and stability of mature blood vessels and prevents the vascular permeability normally induced by VEGF and inflammatory cytokines.

For tumor cell-derived VEGF to initiate sprouting from host vessels, the stability conferred by the Ang1/Tie2 pathway must be perturbed; this occurs by the secretion of Ang2 by ECs that are undergoing active remodeling. Ang2 binds to Tie2 and is a competitive inhibitor of Ang1 action: under the influence of Ang2, preexisting blood vessels become more responsive to remodeling signals, with less adherence of ECs to stroma and associated perivascular cells and more responsiveness to VEGF. Therefore, Ang2 is required at early stages of tumor angiogenesis for destabilizing the vasculature by making host ECs more sensitive to angiogenic signals. In the presence of Ang2, there is no stabilization by the Ang1/Tie2 interaction, and tumor blood vessels are leaky, hemorrhagic, and have poor association of ECs with underlying stroma. Sprouting tumor ECs express high levels of the transmembrane protein ephrin-B2 and its receptor, the RTK EPH, whose signaling appears to work with the angiopoietins during vessel remodeling. During embryogenesis, EPH receptors are expressed on the endothelium of primordial venous vessels while the transmembrane ligand ephrin-B2 is expressed by cells of primordial arteries; the reciprocal expression may regulate differentiation and patterning of the vasculature.

A number of ubiquitously expressed host molecules play critical roles in normal and pathologic angiogenesis. Proangiogenic cytokines, chemokines, and growth factors secreted by stromal cells or inflammatory cells make important contributions to neovascularization,

including bFGF, transforming growth factor- α (TGF- α), TNF- α , and IL-8. In contrast to normal endothelium, angiogenic endothelium over-expresses specific members of the integrin family of ECM-binding proteins that mediate EC adhesion, migration, and survival. Specifically, expression of integrins $\alpha_v\beta_3$, $\alpha_v\beta_5$, and $\alpha_5\beta_1$ mediates spreading and migration of ECs and is required for angiogenesis induced by VEGF and bFGF, which in turn can upregulate EC integrin expression. The $\alpha_v\beta_3$ integrin physically associates with VEGFR2 in the plasma membrane and promotes signal transduction from each receptor to promote EC proliferation (via focal adhesion kinase, src, PI3K, and other pathways) and survival (by inhibition of p53 and increasing the Bcl-2/Bax expression ratio). In addition, $\alpha_v\beta_3$ forms cell-surface complexes with matrix metalloproteinases (MMPs), zinc-requiring proteases that cleave ECM proteins, leading to enhanced EC migration and the release of heparin-binding growth factors, including VEGF and bFGF. EC adhesion molecules can be upregulated (i.e., by VEGF, TNF- α) or downregulated (by TGF- β); this, together with chaotic blood flow, explains poor leukocyte-endothelial interactions in tumor blood vessels and may help tumor cells avoid immune surveillance.

Lymphatic vessels also exist within tumors. Development of tumor lymphatics is associated with expression of VEGFR3 and its ligands VEGF-C and VEGF-D. The role of these vessels in tumor cell metastasis to regional lymph nodes remains to be determined. However, VEGF-C levels correlate significantly with metastasis to regional lymph nodes in lung, prostate, and colorectal cancers.

■ ANTIANGIOGENIC THERAPY

Angiogenesis inhibitors function by targeting the critical molecular pathways involved in EC proliferation, migration, and/or survival, many of which are highly expressed in the activated endothelium in tumors. Inhibition of growth factor and adhesion-dependent signaling pathways can induce EC apoptosis with concomitant inhibition of tumor growth. Different types of tumors can use distinct combinations

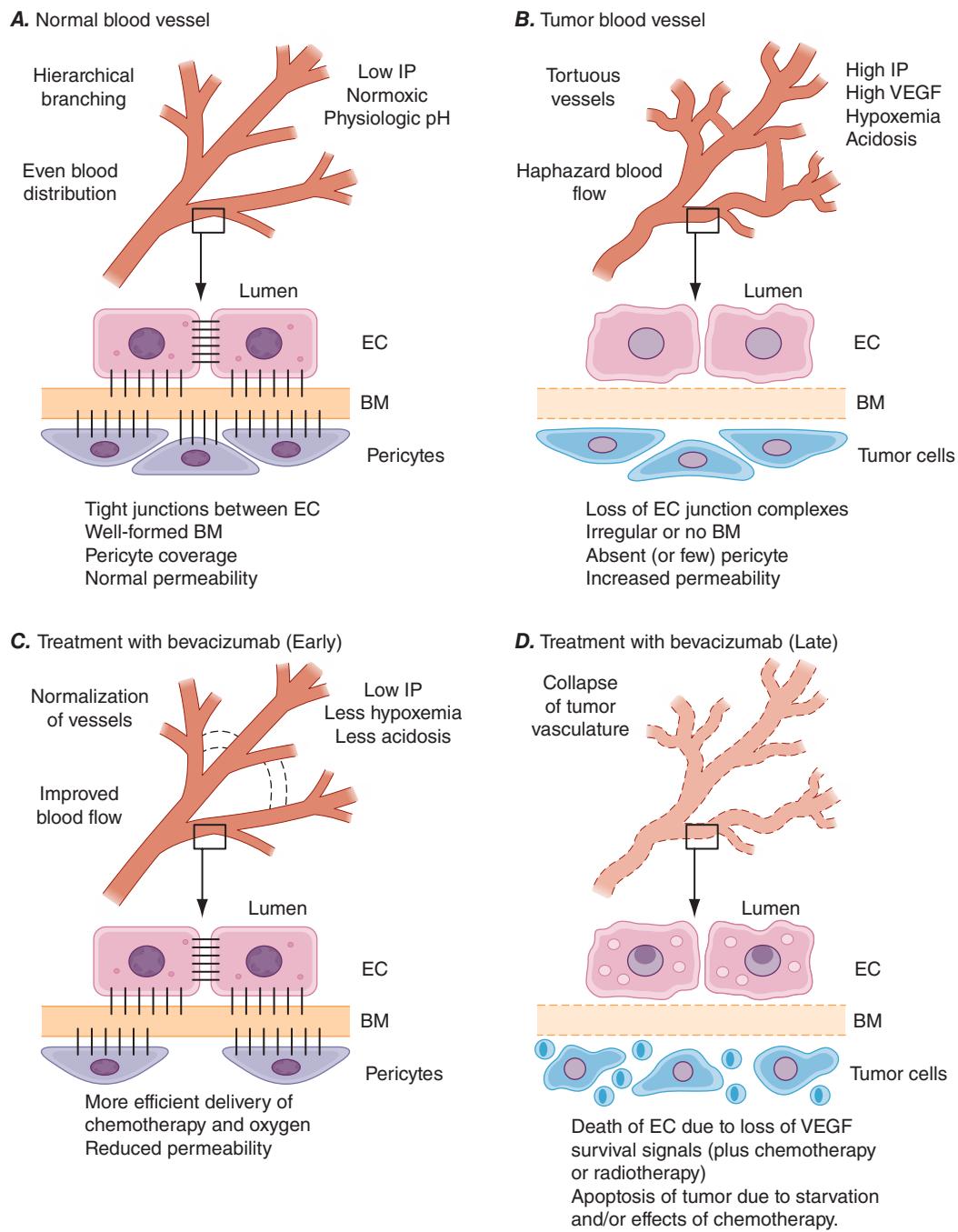


FIGURE 68-11 Normalization of tumor blood vessels due to inhibition of VEGF signaling. **A.** Blood vessels in normal tissues exhibit a regular hierarchical branching pattern that delivers blood to tissues in a spatially and temporally efficient manner to meet the metabolic needs of the tissue (top). At the microscopic level, tight junctions are maintained between endothelial cells (ECs), which are adherent to a thick and evenly distributed basement membrane (BM). Pericytes form a surrounding layer that provides trophic signals to the EC and helps maintain proper vessel tone. Vascular permeability is regulated, interstitial fluid pressure is low, and oxygen tension and pH are physiologic. **B.** Tumors have abnormal vessels with tortuous branching and dilated, irregular interconnecting branches, causing uneven blood flow with areas of hypoxemia and acidosis. This harsh environment selects genetic events that result in resistant tumor variants, such as the loss of p53. High levels of VEGF (secreted by tumor cells) disrupt gap junction communication, tight junctions, and adherens junctions between EC via src-mediated phosphorylation of proteins such as connexin 43, zonula occludens-1, VE-cadherin, and α/β -catenins. Tumor vessels have thin, irregular BM, and pericytes are sparse or absent. Together, these molecular abnormalities result in a vasculature that is permeable to serum macromolecules, leading to high tumor interstitial pressure, which can prevent the delivery of drugs to the tumor cells. This is made worse by the binding and activation of platelets at sites of exposed BM, with release of stored VEGF and microvessel clot formation, creating more abnormal blood flow and regions of hypoxemia. **C.** In experimental systems, treatment with bevacizumab or blocking antibodies to VEGFR2 leads to changes in the tumor vasculature that has been termed *vessel normalization*. During the first week of treatment, abnormal vessels are eliminated or pruned (dotted lines), leaving a more normal branching pattern. ECs partially regain features such as cell-cell junctions, adherence to a more normal BM, and pericyte coverage. These changes lead to a decrease in vascular permeability, reduced interstitial pressure, and a transient increase in blood flow within the tumor. Note that in murine models, this normalization period lasts only for ~5–6 days. **D.** After continued anti-VEGF/VEGFR therapy (which is often combined with chemo- or radiotherapy), ECs die, leading to tumor cell death (either due to direct effects of the chemotherapy or lack of blood flow).

of molecular mechanisms to activate the angiogenic switch. Therefore, it is doubtful that a single antiangiogenic strategy will suffice for all human cancers; rather, a number of agents or combinations of agents will be needed, depending on distinct programs of angiogenesis used by different human cancers. Despite this, experimental data indicate

that for some tumor types, blockade of a single growth factor (e.g., VEGF) may inhibit tumor-induced vascular growth.

Bevacizumab, an antibody which binds VEGF, potentiates the effects of a number of different types of active chemotherapeutic regimens used to treat a variety of different tumor types including colon, lung,

ovarian, and cervical cancers. It also has activity in combination with interferon against RCCs and alone for glioblastomas. Other protein inhibitors of the VEGF signaling pathway approved for anticancer therapy include ramucirumab (a monoclonal antibody directed against VEGFR2, approved for use against gastric/gastroesophageal, colon and lung cancers) and ziv-aflibercept (a recombinant protein inhibitor of VEGF, approved for colorectal cancer). Hypertension is the most common side effect of inhibitors of VEGF (or its receptors), but can be treated with antihypertensive agents and uncommonly requires discontinuation of therapy. Rare but serious potential risks include arterial thromboembolic events, including stroke and myocardial infarction, hemorrhage, bowel perforation, and inhibition of wound healing.

Several small-molecule inhibitors (SMI) that target VEGF RTK activity but are also inhibitory to other kinases have also been approved to treat certain cancers. Sunitinib (see above and Table 68-2) has activity directed against mutant c-Kit receptors (approved for GIST), but also targets VEGFR and PDGFR, and has antitumor activity against pancreatic neuroendocrine and metastatic renal cell carcinomas (RCC), presumably on the basis of its antiangiogenic activity. Similarly, sorafenib, originally developed as a Raf kinase inhibitor but with potent activity against VEGFR and PDGFR, has activity against RCC, differentiated thyroid and hepatocellular cancers as well as desmoid tumors. A closely related molecule to sorafenib, regorafenib, has activity against colorectal cancer, GIST, and hepatocellular cancer. Other inhibitors of the VEGF pathway approved for the treatment of various cancers include axitinib, pazopanib, lenvatinib, and cabozantinib.

The success in targeting tumor angiogenesis has led to enhanced enthusiasm for the development of drugs that target other aspects of the angiogenic process; some of these therapeutic approaches are outlined in Fig. 68-12. There is also evidence suggesting potential enhanced activity when anti-VEGF agents are used in combination with immunomodulators including immune check point inhibitors.

However, it is not yet known whether this will produce a clinically meaningful enhancement of anti-tumor activity.

EVASION OF THE IMMUNE SYSTEM BY CANCERS

There is a complex interaction between tumors and the host from the initiation of the cancer until the establishment of a clinical cancer. Cancers have a number of mechanisms that allow them to evade detection and elimination by the immune system. These include downregulation of cell surface proteins involved in immune recognition (including MHC proteins and tumor-specific antigens), expression of other cell surface proteins that inhibit immune function (including members of the B7 family of proteins such as PD-L1), secretion of proteins and other molecules that are immunosuppressive, recruitment and expansion of immunosuppressive cells such as regulatory T cells, induction of T-cell tolerance, and down regulation of death receptors. Due to the marked heterogeneity of cells within a cancer, a variety of immune suppressive mechanisms are continuously occurring and changing. In addition, the inflammatory effects of some of the immune mediator cells in the tumor microenvironment (especially tissue-associated macrophages and myeloid-derived suppressor cells) can suppress T-cell responses to the tumor as well as stimulate inflammation that can enhance tumor growth. Immunotherapy approaches to treat cancer aimed at activating the immune response against tumors using immunostimulatory molecules such as interferons, IL-2, and monoclonal antibodies have had some successes. A more direct approach to enhance the activity of T cells directed against specific tumors involves isolating T cells from patients and re-engineering the cells to express chimeric antigen receptors (CAR-T cells) that recognize antigens present on the cells of that individual's tumor. The most commonly studied approach to date has been to engineer the cells to express receptors targeting the CD19 antigen on ALL and DLBCL cells. These have been shown to have significant antitumor activity in the treatment of patients with

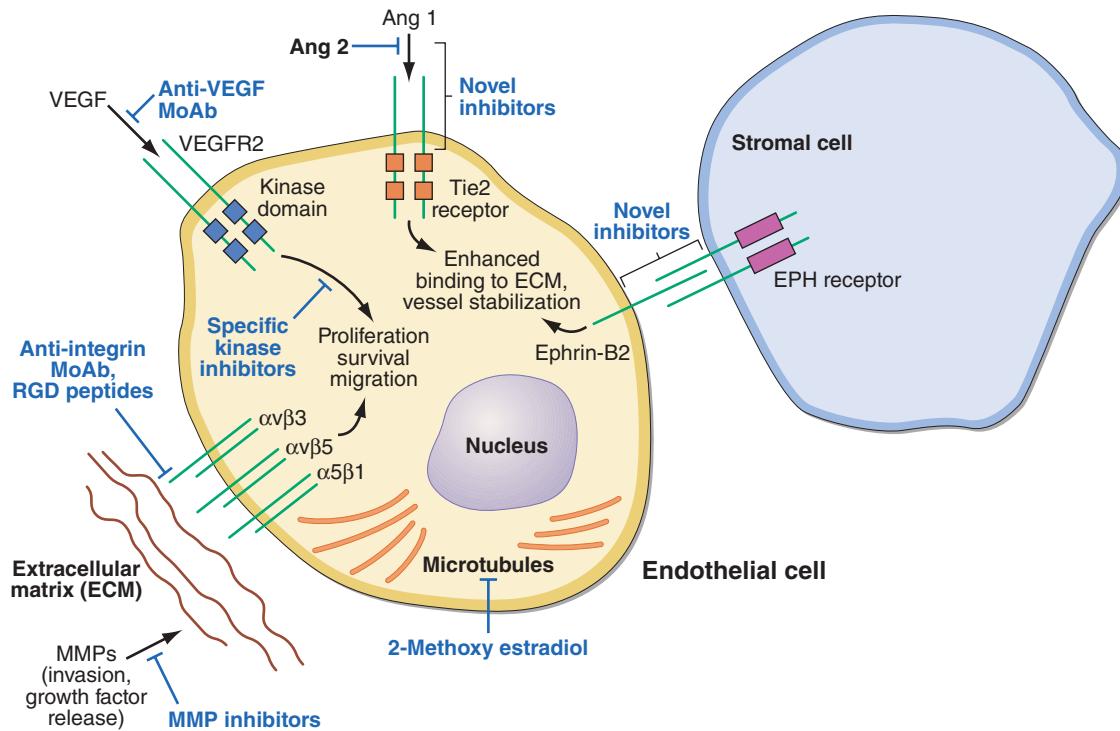


FIGURE 68-12 Knowledge of the molecular events governing tumor angiogenesis has led to a number of therapeutic strategies to block tumor blood vessel formation. The successful therapeutic targeting of VEGF and its receptors VEGFR is described in the text. Other endothelial cell-specific receptor tyrosine kinase pathways (e.g., angiopoietin/Tie2 and ephrin/EPH) are likely targets for the future. Ligation of the $\alpha_5\beta_3$ integrin is required for EC survival. Integrins are also required for EC migration and are important regulators of matrix metalloproteinase (MMP) activity, which modulates EC movement through the ECM as well as release of bound growth factors. Targeting of integrins includes development of blocking antibodies, small peptide inhibitors of integrin signaling, and arg-gly-asp-containing peptides that prevent integrin:ECM binding. Peptides derived from normal proteins by proteolytic cleavage, including endostatin and tumstatin, inhibit angiogenesis by mechanisms that include interfering with integrin function. Signal transduction pathways that are dysregulated in tumor cells indirectly regulate EC function. Inhibition of EGF-family receptors, whose signaling activity is upregulated in a number of human cancers (e.g., breast, colon, and lung cancers), results in downregulation of VEGF and IL-8, while increasing expression of the antiangiogenic protein thrombospondin-1. The Ras/MAPK, PI3K/Akt, and Src kinase pathways constitute important antitumor targets that also regulate the proliferation and survival of tumor-derived EC. The discovery that ECs from normal tissues express tissue-specific "vascular addressins" on their cell surface suggests that targeting specific EC subsets may be possible.

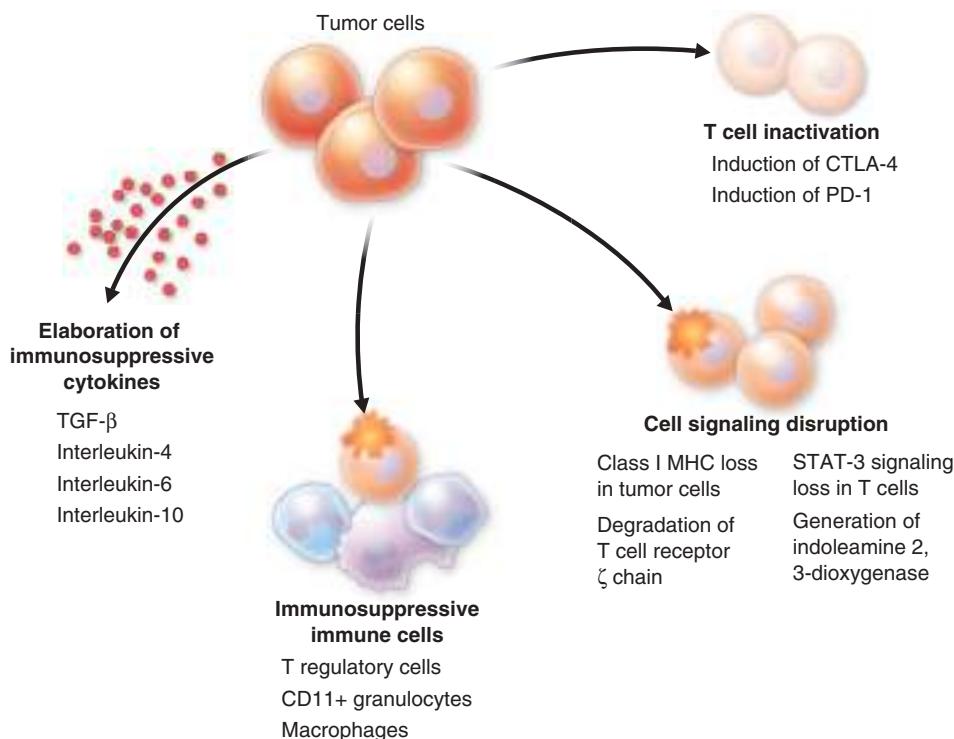


FIGURE 68-13 Tumor-host interactions that suppress the immune response to the tumor.

ALL and DLBCL including durable remissions in patients refractory to standard therapies and are approved for these malignancies. However, there have also been significant issues with toxicity including cytokine release syndrome, organ toxicity felt to be due to inadvertent targeting of antigens present in the organ, and neurotoxicity. These patients often require aggressive supportive care by individuals experienced in the delivery of CAR-T cells. In addition, as is true for most anticancer therapies, mechanisms of resistance have developed, most commonly the outgrowth of tumor cells no longer expressing the antigen. Mechanisms for preventing the development of resistant cells are being explored.

Another approach that has shown particular clinical promise is the targeting of proteins or cells (such as regulatory T cells) involved in normal homeostatic control to prevent autoimmune damage to the host but which malignant cells and their stroma can also utilize to inhibit the immune response directed against them. The approach that is furthest along clinically has involved targeting CTLA-4, PD-1, and PDL-1, co-inhibitory molecules that are expressed on the surface of cancer cells, cells of the immune system, and/or stromal cells and are involved in inhibiting the immune response against cancer (Fig. 68-13). A monoclonal antibody directed against CTLA-4 is approved for the treatment of melanoma and antibodies targeting PD-1 or PDL-1 are approved for use against melanoma, RCC, lung cancer, head and neck cancer, urothelial cancer, HCC, gastric cancer, MSI high cancers, and Hodgkin's lymphoma. There is evidence of activity against other cancers including gastroesophageal and hepatocellular cancers and they continue to be evaluated against other malignancies as well. Combination approaches targeting more than one protein or with other anticancer approaches (targeted agents, chemotherapy, radiation therapy) are also being explored and have shown promise in early studies. An important aspect of these approaches is balancing sufficient release of the negative control of the immune response to allow immune mediated attack on the tumors while not allowing too much release and inducing severe autoimmune effects (such as against lung, liver, skin, thyroid, pituitary gland, or the GI tract).

SUMMARY

Although each of the biological aspects of cancers and examples of targeting them has been addressed individually, clearly there is

complicated cross-talk between these that occurs in all cancers which needs to be understood to optimally treat different cancers. The explosion of information on tumor cell biology, metastasis, and tumor-host interactions (including angiogenesis, other tumor-stromal interactions, and immune evasion by tumors) has ushered in a new era of rational targeted therapy for cancer. Furthermore, it has become clear that specific molecular factors detected in individual tumors (specific gene mutations, gene-expression profiles, microRNA expression, overexpression of specific proteins) can be used to tailor therapy and maximize antitumor effects.

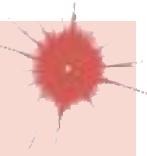
ACKNOWLEDGMENT

Robert G. Fenton contributed to this chapter in prior editions and important material from those prior chapters has been included here.

FURTHER READING

- BOUSSIOTIS VA: Molecular and biochemical aspects of the PD-1 checkpoint pathway. *N Engl J Med* 375:1767, 2016.
- DE PALMA M, BIZIATO D, PETROVA TV: Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer* 17:457, 2017.
- DU W, ELEMENTO O: Cancer systems biology: Embracing complexity to develop better anticancer therapeutic strategies. *Oncogen* 34:3215, 2015.
- FLEUREN ED et al: The kinome "at large" in cancer. *Nat Rev Cancer* 16:83, 2016.
- HE S, SHARPLESS NE: Senescence in health and disease. *Cell* 169:1000, 2017.
- LAMBERT AW et al: Emerging biological principles of metastasis. *Cell* 168:670, 2017.
- OTTO T, SICINSKI P: Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer* 17:93, 2017.
- TOMASETTI C et al: Stem cell divisions, somatic mutations, cancer etiology and cancer prevention. *Science* 355:1330, 2017.
- VANDER HEIDEN MG, DEBERARDINIS RJ: Understanding the intersections between metabolism and cancer biology. *Cell* 168:657, 2017.
- VOGELSTEIN B et al: Cancer genome landscapes. *Science* 339:1546, 2013.

Edward A. Sausville, Dan L. Longo



CANCER PRESENTATION

Localized or systemic cancer is frequent in the differential diagnosis of a variety of common complaints. Although not all forms of cancer are curable at initial diagnosis, affording patients the greatest opportunity for cure or meaningful prolongation of life is greatly aided by diagnosing cancer early in its natural history, and defining treatments that prevent or retard its systemic spread. Indeed, certain forms of cancer, notably breast, colon, and possibly lung cancers in certain patients, can be prevented by screening appropriately selected asymptomatic patients; screening is arguably the earliest point in the spectrum of possible cancer-related interventions where cure is possible (**Table 69-1**).

DETECTION OF A CANCER

The term *cancer*, as used here, is synonymous with the term *tumor*, whose original derivation from Latin simply meant “swelling,” not otherwise specified. We now understand that swelling as a common physical manifestation of a tumor reflects increased interstitial fluid pressure and increased cellular and stromal mass per volume, compared to normal tissue. Leukemias are a special case of a cancer of the blood-forming tissues presenting in a disseminated form frequently without definable tumor masses. In addition to localized swelling, tumors present by altered function of the organ they afflict, such as dyspnea on exertion from the anemia caused by leukemia replacing normal hematopoietic cells, cough from lung cancers, jaundice from tumors disrupting the hepatobiliary tree, or seizures and neurologic signs from brain tumors. Hemorrhage is also a frequent presenting sign of tumors involving hollow viscera, but also may reflect decreases in the number of platelets or altered blood coagulation. Tumors may also present owing to the effects of substances they secrete called a “paraneoplastic” syndrome. Thus, although statistically the fraction of patients with cancer underlying a particular presenting sign or symptom may be low, the implications for a patient with cancer of missing an early-stage tumor call for vigilance; therefore, persistent signs or symptoms should be evaluated as possibly coming from an early-stage tumor.

Evidence of a tumor’s existence can objectively be established by careful physical examination, detecting enlarged lymph nodes in lymphomas or a palpable mass in a breast or soft tissue site. A mass may also be detected or confirmed by an imaging modality, such as

TABLE 69-1 Spectrum of Cancer-Related Interventions

Screening for cancer in an asymptomatic patient

Consideration of cancer in a differential diagnosis

Physical examination, imaging, or endoscopy to define a possible tumor

Diagnosis of cancer by biopsy or removal:

Routine histology

Specialized histology: immunohistochemistry

Molecular studies

Cytogenetic studies

Staging the cancer: Where has it spread?

Treatment

Localized

Systemic

Supportive care

During treatment: related to tumor effects on patient

During treatment to counteract side effects of treatment

Palliative and end of life

When useful treatments are not feasible or desired

plain x-ray, computed tomography (CT) scan, ultrasound, positron emission tomography (PET) imaging, or nuclear magnetic resonance approaches. Another way of initially establishing the existence of a possible tumor is through direct visualization of an afflicted organ by endoscopy.

ESTABLISHING A CANCER DIAGNOSIS

Once the existence of a likely tumor is defined, unequivocally establishing the diagnosis is the next step in the intervention spectrum. This is usually accomplished by a biopsy procedure and the emergence after pathologic examination of an unequivocal statement that cancer is present, or a non-cancer diagnosis explains the abnormality. Due to tumor heterogeneity, pathologists are better able to make the diagnosis when they have more tissue to examine. In addition to light microscopic inspection of a tumor, sufficient tissue also allows definition of genetic abnormalities and protein expression patterns, such as hormone receptor expression in breast cancers, that may aid in the differential diagnosis or provide information about prognosis or likely response to treatment. Efforts to define “personalized” information from the biology of each patient’s tumor and pertinent to each patient’s treatment plan are becoming increasingly important in selecting treatment options. The general internist should make sure that a patient’s cancer biopsy is appropriately referred from the surgical suite for important molecular studies that can advise the best treatment (**Table 69-2**).

Coordination among the surgeon, pathologist, and primary care physician is essential to ensure that the amount of information learned from the biopsy material is maximized. These goals are best met by an *excisional biopsy* in which the entire tumor mass is removed with a small margin of normal tissue surrounding it. If an excisional biopsy cannot be performed, *incisional biopsy* is the procedure of second choice. A wedge of tissue is removed, and an effort is made to include the majority of the cross-sectional diameter of the tumor in the biopsy to minimize sampling error. Biopsy techniques that involve cutting into tumor

TABLE 69-2 Diagnostic Biopsy: Standard of Care Molecular and Special Studies

Breast cancer: primary and suspected metastatic

Hormone receptors: estrogen, progesterone

HER2/neu oncogene

Lung cancer: primary and suspected metastatic

If nonsquamous non-small cell: epidermal growth factor receptor mutation; alk oncogene gene fusion; programmed cell death ligand-1

Colon cancer: suspected metastatic

Ki-ras mutation

Gastrointestinal stromal tumor

c-kit oncogene mutation

Melanoma

B-raf oncogene mutation

c-kit expression and mutation

Brain tumor gliomas

1p/19q co-deletion

Alkyguanine alkyltransferase promoter methylation

Leukemia (peripheral blood mononuclear cells and/or bone marrow)

Cytogenetics

Flow cytometry

Treatment-defining chromosomal translocations

Bcr-Abl fusion protein

t(15,17)

inversion 16

t(8,21)

Lymphoma

Immunohistochemistry for CD20, CD30, T cell markers

Treatment defining chromosomal translocations:

t(14,18)

t(8,14)

carry with them a risk of facilitating the spread of the tumor, and consideration with a surgeon of whether the biopsy might be the prelude to a curative surgery if certain diagnoses are established should inform the actual approach taken. *Core-needle biopsy* usually obtains considerably less tissue, but this procedure often provides enough information to plan a definitive surgical procedure. *Fine-needle aspiration* generally obtains only a suspension of cells from within a mass. This procedure is minimally invasive, and if positive for cancer, it may allow inception of systemic treatment when metastatic disease is evident, or it can provide a basis for planning a more meticulous and extensive surgical procedure. However, a negative fine-needle aspiration for a neoplastic diagnosis cannot be taken as definitive evidence that a tumor is absent or make a definitive diagnosis in someone not known to have a cancer.

CANCER STAGING

An essential component of correct patient management in many cancer types is defining the extent of disease, because this information critically informs whether localized treatments, “combined-modality” approaches, or systemic treatments should initially be considered. Radiographic and other imaging tests can be helpful in defining the clinical stage; however, pathologic staging requires defining the extent of involvement by documenting the histologic presence of tumor in tissue biopsies obtained through a surgical procedure. Axillary lymph node sampling in breast cancer and lymph node sampling at laparotomy for testicular, colon, and other intraabdominal cancers may provide crucial information for treatment planning and may determine the extent and nature of primary cancer treatment.

For tumors associated with a potential “primary site,” staging systems have evolved to define a “T” component related to the size of the tumor or its invasion into local structures, an “N” component related to the number and nature of lymph node groups adjacent to the tumor with evidence of tumor spread, and an “M” component, based on the presence of local or distant metastatic sites. The various “TNM” components are then aggregated to stages, usually stage I to III or IV, depending on the anatomic site. The numerical stages reflect similar long-term survival outcomes of the aggregated TNM groupings in a numeric stage after treatment tailored to the stage. In general, stage I tumors are T1 (reflecting small size), N0 or N1 (reflecting no or minimal node spread), and M0 (no metastases). Such early-stage tumors are amenable to curative approaches with local treatments. On the other hand, stage IV tumors usually have metastasized to distant sites or locally invaded viscera in a nonresectable way and are dealt with using techniques that have palliative intent, except for those diseases with exceptional sensitivity to systemic treatments such as chemotherapy or immunotherapy. Also, the TNM staging system is not useful in diseases such as leukemia, where bone marrow infiltration is never really localized, or central nervous system (CNS) tumors, where tumor histology and the extent of anatomically feasible resection are more important in driving prognosis.

CANCER TREATMENT

The goal of cancer treatment is first to eradicate the cancer. If this primary goal cannot be accomplished, the goal of cancer treatment shifts to palliation, the amelioration of symptoms, and preservation of quality of life while striving to extend life. The dictum *primum non nocere* may not always be the guiding principle of cancer therapy. When cure of cancer is possible, cancer treatments may be considered despite the certainty of severe and perhaps life-threatening toxicities. Every cancer treatment has the potential to cause harm, and treatment may be given that produces toxicity with no benefit. The therapeutic index of many interventions may be quite narrow, with treatments given to the point of toxicity. Conversely, when the clinical goal is palliation, careful attention to minimizing the toxicity of treatments becomes a significant goal.

Cancer treatments are divided into two main types: local and systemic. Local treatments include surgery, radiation therapy (including photodynamic therapy), and ablative approaches, including radiofrequency and cryosurgical approaches. Systemic treatments include chemotherapy (including hormonal therapy and molecularly targeted therapy) and biologic therapy (including immunotherapy). The modalities are often

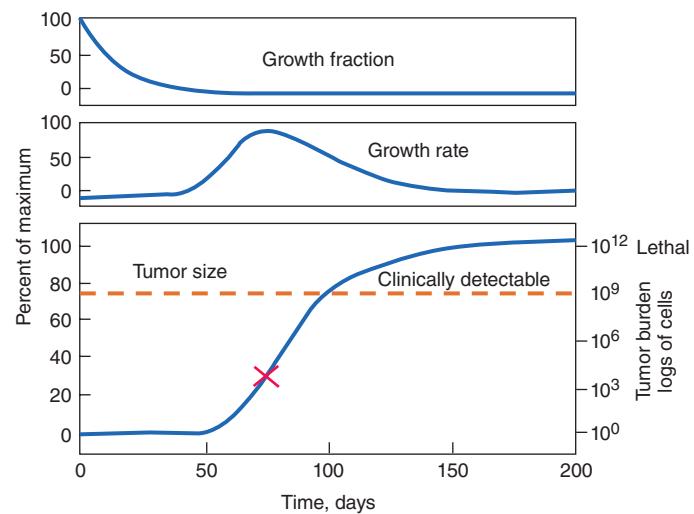


FIGURE 69-1 Gompertzian tumor growth. The growth fraction of a tumor declines exponentially over time (top). The growth rate of a tumor peaks before it is clinically detectable (middle). Tumor size increases slowly, goes through an exponential phase, and slows again as the tumor reaches the size at which limitation of nutrients or autoregulatory or host regulatory influences can occur. The maximum growth rate occurs at $1/e$, the point at which the tumor is about 37% of its maximum size (marked with an X). Tumor becomes detectable at a burden of about 10^9 (1 cm^3) cells and kills the patient at a tumor cell burden of about 10^{12} (1 kg). Efforts to treat the tumor and reduce its size can result in an increase in the growth fraction and an increase in growth rate.

used in combination, and agents in one category can act by several mechanisms. For example, cancer chemotherapy agents can induce differentiation, and antibodies (a form of immunotherapy) can be used to deliver radiation therapy. *Oncology*, the study of tumors including treatment approaches, is a multidisciplinary effort with surgical, radiation, and internal medicine-related areas of oncologic expertise. Treatments for patients with hematologic malignancies are often shared by hematologists and medical oncologists.

Normal organs and cancers share the property of having a population of cells actively progressing through the cell cycle with their division providing a basis for tumor growth, and a population of cells not in cycle; these include *cancer stem cells*, whose properties are being elucidated, as they may serve as a basis for giving rise to tumor initiating or repopulating cells. The stem cell fraction may define new targets for therapies that will retard their ability to reenter the cell cycle.

Tumors follow a Gompertzian growth curve (Fig. 69-1), with the apparent growth fraction of a neoplasm being high with small tumor burdens and declining until, at the time of diagnosis, with a tumor burden of $1-5 \times 10^9$ tumor cells, the growth fraction is usually 1–4% for many solid tumors. By this view, the most rapid growth rate occurs before the tumor is detectable. An alternative explanation for such growth properties may also emerge from the ability of tumors at metastatic sites to recruit circulating tumor cells from the primary tumor or other metastases. An additional key feature of a successful tumor is the ability to stimulate the development of a new supporting stroma through angiogenesis and production of proteases to allow invasion through basement membranes and normal tissue barriers (Chap. 68).

LOCALIZED CANCER TREATMENTS

SURGERY

Surgery is unquestionably the most effective means of treating cancer. Today at least 40% of cancer patients are cured by surgery. Unfortunately, a large fraction of patients with solid tumors (perhaps 60%) have metastatic disease that is not accessible for removal. Even when cancer is not curable by surgery alone, the removal of tumor can obtain important benefits, including local control of tumor, preservation of organ function, debulking that permits subsequent therapy to be more effective, and staging information on extent of involvement. Cancer surgery aiming for cure is usually planned to excise the tumor

completely with an adequate margin of normal tissue (the margin varies with the tumor and the anatomy), touching the tumor as little as possible to prevent vascular and lymphatic spread, and minimizing operative risk. Such a resection is defined as an R0 resection. R1 and R2 resections, in contrast, are imprecisely defined pathologically as having microscopic or macroscopic, respectively, tumor at resection margins. Such outcomes may be necessitated by proximity of the tumor to vital structures or recognition only in the resected specimen of the extent of tumor involvement, and may be the basis for reoperation to obtain optimal margins if feasible. Extending the procedure to resect draining lymph nodes obtains prognostic information and may, in some anatomic locations, improve survival.

Increasingly, laparoscopic approaches are being used to address primary abdominal and pelvic tumors. Lymph node spread may be assessed using the sentinel node approach, in which the first draining lymph node a spreading tumor would encounter is defined by injecting a dye or radioisotope into the tumor site at operation and then resecting the first node to turn blue or collect isotope. The sentinel node assessment is continuing to undergo clinical evaluation but appears to provide reliable information without the risks (lymphedema, lymphangiosarcoma) associated with resection of all the regional nodes. Advances in adjuvant chemotherapy (chemotherapy given systemically after removal of all local disease by operation and without evidence of active metastatic disease) and radiation therapy following surgery have permitted a substantial decrease in the extent of primary surgery necessary to obtain the best outcomes. Thus, lumpectomy with radiation therapy is as effective as modified radical mastectomy for breast cancer, and limb-sparing surgery followed or preceded by adjuvant radiation therapy and chemotherapy has replaced radical primary surgical procedures involving amputation and disarticulation for childhood rhabdomyosarcomas and osteosarcomas. More limited surgery is also being used to spare organ function, as in larynx and bladder cancer. In some settings (e.g., bulky testicular cancer or stage III breast cancer), surgery is not the first treatment modality used. After an initial diagnostic biopsy, chemotherapy and/or radiation therapy is delivered to reduce the size of the tumor and clinically control undetected metastatic disease. Such therapy is followed by a surgical procedure to remove residual masses; this is called *neoadjuvant therapy*. Because the sequence of treatment is critical to success and is different from the standard surgery-first approach, coordination among the surgical oncologist, radiation oncologist, and medical oncologist is crucial.

Surgery may be curative in a subset of patients with metastatic disease. Patients with lung metastases from osteosarcoma may be cured by resection of the lung lesions. In patients with colon cancer who have fewer than five liver metastases restricted to one lobe and no extrahepatic metastases, hepatic lobectomy may produce long-term disease-free survival in 25% of selected patients. Surgery can also be associated with systemic antitumor effects. In the setting of hormonally responsive tumors, oophorectomy and/or adrenalectomy may eliminate estrogen production, and orchietomy may reduce androgen production, hormones that drive certain breast and all prostate cancers, respectively; both procedures can have useful effects on metastatic tumor growth. In selecting a surgeon or center for primary cancer treatment, consideration must be given to the volume of cancer surgeries undertaken by the site. Studies in a variety of cancers have shown that increased annual procedure volume appears to correlate with outcome. In addition, facilities with extensive support systems—e.g., for joint thoracic and abdominal surgical teams with cardiopulmonary bypass, if needed—may allow resection of certain tumors that would otherwise not be possible.

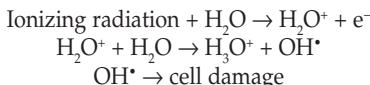
Surgery is used in a number of ways for palliative or supportive care of the cancer patient, not related to the goal of curing the cancer. These include insertion and care of central venous catheters, control of pleural and pericardial effusions and ascites, caval interruption for recurrent pulmonary emboli, stabilization of cancer-weakened weight-bearing bones, and control of hemorrhage, among others. Surgical bypass of gastrointestinal, urinary tract, or biliary tree obstruction can alleviate symptoms and prolong survival. Surgical procedures may provide

relief of otherwise intractable pain or reverse neurologic dysfunction (cord decompression). Splenectomy may relieve symptoms and reverse hypersplenism. Intrathecal or intrahepatic therapy relies on surgical placement of appropriate infusion portals. Surgery may correct other treatment-related toxicities such as adhesions or strictures. Surgical procedures are also valuable in rehabilitative efforts to restore health or function. Orthopedic procedures may be necessary to ensure proper ambulation. Breast reconstruction can make an enormous impact on the patient's perception of successful therapy. Plastic and reconstructive surgery can correct the effects of disfiguring primary treatment.

Surgery is also a tool valuable in the prevention of cancers in high-risk populations. Prophylactic mastectomy, colectomy, oophorectomy, and thyroidectomy are mainstays of prevention of genetic cancer syndromes. Resection of premalignant skin and uterine cervix lesions and colonic polyps prevents progression to frank malignancy.

RADIATION

Radiation Biology and Medicine Therapeutic radiation is ionizing, causing breaks in DNA and generation of free radicals from cell water that may damage cell membranes, proteins, and organelles. Radiation damage is augmented by oxygen; hypoxic cells are more resistant. Augmentation of oxygen presence is one basis for radiation sensitization. X-rays and gamma rays are the forms of ionizing radiation most commonly used to treat cancer. They are both electromagnetic, nonparticulate waves that cause the ejection of an orbital electron when absorbed. This orbital electron ejection results in ionization. These waves behave biologically as packets of energy, called *photons*. Particulate ionizing radiation using protons has also become available. Most radiation-induced cell damage is due to the formation of hydroxyl radicals from tissue water:



Radiation is quantitated based on the amount of radiation absorbed by the tumor in the patient; it is not based on the amount of radiation generated by the machine. The International System (SI) unit for radiation absorbed is the Gray (Gy): 1 Gy refers to 1 J/kg of tissue; 1 Gy equals 100 centigrays (cGy) of absorbed dose. A historically used unit appearing in the oncology literature, the *rad* (radiation absorbed dose), is defined as 100 ergs of energy absorbed per gram of tissue and is equivalent to 1 cGy. Radiation dosage is defined by the energy absorbed per mass of tissue. Radiation dose is measured by placing detectors at the body surface or based on radiating phantoms that resemble human form and substance, containing internal detectors. The features that make a particular cell more sensitive or more resistant to the biologic effects of radiation are not completely defined and critically involve DNA repair proteins that, in their physiologic role, protect against environmentally related DNA damage.

Localized Radiation Therapy Radiation effect is influenced by three determinants: total absorbed dose, number of fractions, and time of treatment. A frequent error is to omit the number of fractions and the duration of treatment. Thus, a typical course of radiation therapy should be described as 4500 cGy delivered to a particular target (e.g., mediastinum) over 5 weeks in 180-cGy fractions. Most curative radiation treatment programs are delivered once a day, 5 days a week, in 150- to 200-Gy fractions. Nondividing cells are more resistant than dividing cells, and this is one rationale for delivering radiation in repeated fractions, to ultimately expose a larger number of tumor cells that have entered the division cycle. In addition to these biologic parameters, physical parameters of the radiation are also crucial. The energy of the radiation determines its ability to penetrate tissue. Low-energy x-rays (150–400 kV) scatter when they strike the body, much like light diffuses when it strikes particles in the air. Such beams result in more damage to adjacent normal tissues and less radiation delivered to the tumor. Megavoltage radiation (>1 MeV) has very low lateral scatter; this produces a skin-sparing effect, more homogeneous

distribution of the radiation energy, and greater deposit of the energy in the tumor, or *target volume*. The tissues that the beam passes through to get to the tumor are called the *transit volume*. The maximum dose in the target volume is often the cause of complications to tissues in the transit volume, and the minimum dose in the target volume influences the likelihood of tumor recurrence. Dose homogeneity in the target volume is the goal. Computational approaches and delivery of many beams to converge on a target lesion are the basis for “gamma knife” and related approaches to deliver high doses to small volumes of tumor, sparing normal tissue.

Therapeutic radiation is delivered in three ways: (1) *teletherapy*, with focused beams of radiation generated at a distance and aimed at the tumor within the patient; (2) *brachytherapy*, with encapsulated sources of radiation implanted directly into or adjacent to tumor tissues; and (3) *systemic therapy*, with radionuclides administered, for example, intravenously but targeted by some means to a tumor site. Teletherapy with x-ray or gamma-ray photons is the most commonly used form of radiation therapy. Particulate forms of radiation are also used in certain circumstances, such as the use of proton beams. The difference between photons and protons relates to the volume in which the greatest delivery of energy occurs. Typically, protons have a much narrower range of energy deposition, theoretically resulting in more precise delivery of radiation with improvement in the degree to which adjacent structures may be affected, in comparison to photons. Electron beams are a particulate form of radiation that, in contrast to photons and protons, have a very low tissue penetrance and are used to treat cutaneous tumors. Certain drugs used in cancer treatment may also act as radiation sensitizers. For example, compounds that incorporate into DNA and alter its stereochemistry (e.g., halogenated pyrimidines, cisplatin) augment radiation effects at local sites, as does hydroxyurea, another DNA synthesis inhibitor. These are important adjuncts to the local treatment of certain tumors, such as squamous head and neck, uterine cervix, and rectal cancers.

Toxicity of Radiation Therapy Although radiation therapy is most often administered to a local region, systemic effects, including fatigue, anorexia, nausea, and vomiting, may develop that are related in part to the volume of tissue irradiated, dose fractionation, radiation fields, and individual susceptibility. Injured tissues release cytokines that act systemically to produce these effects. Bone is among the most radioresistant organs, with radiation effects being manifested mainly in children through premature fusion of the epiphyseal growth plate. By contrast, the male testis, female ovary, and bone marrow are the most sensitive organs. Any bone marrow in a radiation field will be eradicated by therapeutic irradiation. Organs with less need for cell renewal, such as heart, skeletal muscle, and nerves, are more resistant to radiation effects. In radiation-resistant organs, the vascular endothelium is the most sensitive component. Organs with more self-renewal as a part of normal homeostasis, such as the hematopoietic system and mucosal lining of the intestinal tract, are more sensitive. Acute toxicities include mucositis, skin erythema (ulceration in severe cases), and bone marrow toxicity. Often these can be alleviated by interruption of treatment.

Chronic toxicities are more serious. Radiation of the head and neck region often produces thyroid failure. Cataracts and retinal damage can lead to blindness. Salivary glands stop making saliva, which leads to dental caries and poor dentition. Taste and smell can be affected. Mediastinal irradiation leads to a threefold increased risk of fatal myocardial infarction. Other late vascular effects include chronic constrictive pericarditis, lung fibrosis, viscous stricture, spinal cord transection, and radiation enteritis. A serious late toxicity is the development of second solid tumors in or adjacent to the radiation fields. Such tumors can develop in any organ or tissue and occur at a rate of ~1% per year beginning in the second decade after treatment. Some organs vary in susceptibility to radiation carcinogenesis. A woman who receives mantle field radiation therapy for Hodgkin’s disease at age 25 years has a 30% risk of developing breast cancer by age 55 years. This is comparable in magnitude to genetic breast cancer syndromes. Women treated after age 30 years have little or no increased risk of breast cancer. No data suggest that a threshold dose of therapeutic radiation exists below which

the incidence of second cancers is decreased. High rates of second tumors occur in people who receive as little as 1000 cGy.

■ OTHER LOCALIZED CANCER TREATMENTS

Endoscopy techniques may allow the placement of stents to unblock viscera by mechanical means, palliating, for example, gastrointestinal or biliary obstructions. Radiofrequency ablation (RFA) refers to the use of focused microwave radiation to induce thermal injury within a volume of tissue. RFA can be useful in the control of metastatic lesions, particularly in liver, that may threaten biliary drainage (as one example) and threaten quality and duration of useful life in patients with otherwise unresectable disease. Cryosurgery uses extreme cold to sterilize lesions in certain sites, such as prostate and kidney, when at a very early stage, eliminating the need for modalities with more side effects such as surgery or radiation.

Some chemicals (porphyrins, phthalocyanines) are preferentially taken up by cancer cells by mechanisms not fully defined. When light, usually delivered by a laser, is shone on cells containing these compounds, free radicals are generated and the cells die. Hematoporphyrins and light (phototherapy) are being used with increasing frequency to treat skin cancer; ovarian cancer; and cancers of the lung, colon, rectum, and esophagus. Palliation of recurrent locally advanced disease can sometimes be dramatic and last many months.

Infusion of chemotherapeutic or biologic agents or radiation-bearing delivery devices such as isotope-coated glass spheres into local sites through catheters inserted into specific vascular sites such as liver or an extremity have been used in an effort to control disease limited to that site; in selected cases, prolonged control of truly localized disease has been possible.

SYSTEMIC CANCER TREATMENTS

The concept that systemically administered agents may have a useful effect on cancers was historically derived from three sets of observations. Paul Ehrlich in the nineteenth century observed that different dyes reacted with different cell and tissue components. He hypothesized the existence of compounds that would be “magic bullets” that might bind to tumors, owing to the affinity of the agent for the tumor. A second observation was the toxic effects of certain mustard gas derivatives on the bone marrow during World War I, leading to the idea that smaller doses of these agents might be used to treat tumors of marrow-derived cells. Finally, the observation that certain tumors from hormone-responsive tissues, e.g., breast tumors, could shrink after oophorectomy led to the idea that endogenous substances promoting the growth of a tumor might be antagonized. Chemicals achieving each of the goals are actually or intellectually the forbearers of the currently used cancer chemotherapy agents.

Systemic cancer treatments are of four broad types. *Conventional “cytotoxic” chemotherapy agents* were historically derived by the empirical observation that these “small molecules” (generally with molecular mass <1500 Da) could cause major regression of experimental tumors growing in animals. These agents mainly target DNA structure or segregation of DNA as chromosomes in mitosis. *Targeted agents* refer to small molecules or “biologics” (generally macromolecules such as antibodies or cytokines) designed and developed to interact with a defined molecular target important in maintaining the malignant state or expressed by the tumor cells. As described in Chap. 68, successful tumors have activated biochemical pathways that lead to uncontrolled proliferation through the action of, e.g., oncogene products, loss of cell cycle inhibitors, or loss of cell death regulation, and have acquired the capacity to replicate chromosomes indefinitely, invade, metastasize, and evade the immune system. Targeted therapies seek to capitalize on the biology behind the aberrant cellular behavior as a basis for therapeutic effects. *Hormonal therapies* (the first form of targeted therapy) capitalize on the biochemical pathways underlying estrogen and androgen function and action as a therapeutic basis for approaching patients with tumors of breast, prostate, and uterus. *Biologic therapies* are often macromolecules that have a particular target (e.g., anti-growth factor receptor or cytokine antibodies) or may have the capacity to induce a host immune response to kill tumor cells.

Principles The usefulness of any drug is governed by the extent to which a given dose causes a therapeutic effect (in the case of anticancer agents, toxicity to tumor cells) as opposed to a toxic effect to the host. The *therapeutic index* is the degree of separation between toxic and therapeutic doses. Really useful drugs have large therapeutic indices, and this usually occurs when the drug target is expressed in the disease-causing compartment as opposed to the normal compartment. Currently used chemotherapeutic agents have the unfortunate property that their targets are present in both normal and tumor tissues. Therefore, they have relatively narrow therapeutic indices.

Figure 69-2 illustrates steps in cancer drug development. Following demonstration of antitumor activity in animal models, potentially useful anticancer agents are further evaluated to define an optimal schedule of administration and arrive at a drug formulation designed

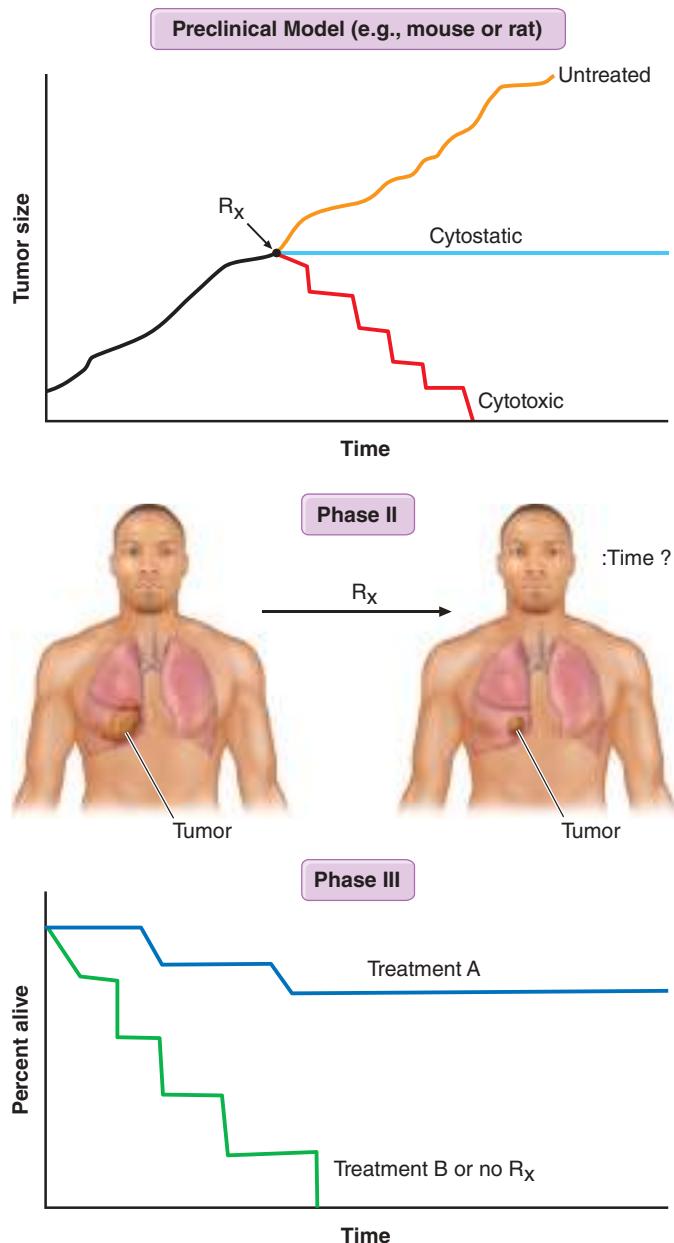


FIGURE 69-2 Steps in cancer drug discovery and development. Preclinical activity (top) in animal models of cancers may be used as evidence to support the entry of the drug candidate into phase 1 trials in humans to define a correct dose and observe any clinical antitumor effect that may occur. The drug may then be advanced to phase 2 trials directed against specific cancer types, with rigorous quantitation of antitumor effects (middle). Phase 3 trials then may reveal activity superior to standard or no treatment (bottom).

for a given route of administration and schedule. Safety testing in two species on an analogous schedule of administration defines the starting dose for a phase 1 trial in humans, usually but not always in patients with cancer who have exhausted "standard" (already approved) treatments. The initial dose is usually one-sixth to one-tenth of the dose just causing easily reversible toxicity in the more sensitive animal species. Escalating doses of the drug are then given during the human phase 1 trial until reversible toxicity is observed. Dose-limiting toxicity (DLT) defines a dose that conveys greater toxicity than would be acceptable in routine practice, allowing definition of a lower maximum-tolerated dose (MTD). The occurrence of toxicity is, if possible, correlated with plasma drug concentrations. The MTD or a dose just lower than the MTD is usually the dose suitable for phase 2 trials, where a fixed dose is administered to a relatively homogeneous set of patients with a particular tumor type in an effort to define whether the drug causes regression of tumors. In a phase 3 trial, evidence of improved overall survival or improvement in the time to progression of disease on the part of the new drug is sought in comparison to an appropriate control population, which is usually receiving an acceptable "standard of care" approach. A favorable outcome of a phase 3 trial is the basis for application to a regulatory agency for approval of the new agent for commercial marketing as safe and possessing a measure of clinical effectiveness.

Response, defined as tumor shrinkage, is the most immediate indicator of drug effect. To be clinically valuable, responses must translate into clinical benefit. This is conventionally established by a beneficial effect on overall survival, or at least an increased time to further progression of disease. Karnofsky was among the first to champion the evaluation of a chemotherapeutic agent's benefit by carefully quantitating its effect on tumor size and using these measurements to objectively decide the basis for further treatment of a particular patient or further clinical evaluation of a drug's potential. A partial response (PR) is defined conventionally as a decrease by at least 50% in a tumor's bidimensional area; a complete response (CR) connotes disappearance of all tumor; progression of disease signifies an increase in size of existing lesions by >25% from baseline or best response or development of new lesions; and stable disease fits into none of the above categories. Newer evaluation systems, such as Response Evaluation Criteria in Solid Tumors (RECIST), use unidimensional measurement, but the intent is similar in rigorously defining evidence for the activity of the agent in assessing its value to the patient. An active chemotherapy agent conventionally has PR rates of at least 20–25% with reversible non-life-threatening side effects, and it may then be suitable for study in phase 3 trials to assess efficacy in comparison to standard or no therapy. Active efforts are being made to quantitate effects of anticancer agents on quality of life. Cancer drug clinical trials conventionally use a toxicity grading scale where grade 1 toxicities do not require treatment, grade 2 toxicities may require symptomatic treatment but are not life-threatening, grade 3 toxicities are potentially life-threatening if untreated, grade 4 toxicities are actually life-threatening, and grade 5 toxicities are those that result in the patient's death.

Development of targeted agents may proceed quite differently. While phase 1–3 trials are still conducted, molecular analysis of human tumors may allow the precise definition of target expression in a patient's tumor that is necessary for or relevant to the drug's action. This information might then allow selection of patients expressing the drug target for participation in all trial phases. These patients may then have a greater chance of developing a useful response to the drug by virtue of expressing the target in the tumor. Clinical trials may be designed to incorporate an assessment of the behavior of the target in relation to the drug (pharmacodynamic studies). Ideally, the plasma concentration that affects the drug target is known, so escalation to MTD may not be necessary. Rather, the correlation of host toxicity while achieving an "optimal biologic dose" becomes a more relevant endpoint for phase 1 and early phase 2 trials with targeted agents.

Useful cancer drug treatment strategies using conventional chemotherapy agents, targeted agents, hormonal treatments, or biologics have one of two valuable outcomes. They can induce cancer cell death, resulting in tumor shrinkage with corresponding improvement in

patient survival, or increase the time until the disease progresses. Another potential outcome is to induce cancer cell *differentiation* or *dormancy* with loss of tumor cell replicative potential and reacquisition of phenotypic properties resembling normal cells. A general view of how cancer treatments work is that the interaction of a chemotherapeutic drug with its target induces a “cascade” of further signaling steps. These signals ultimately lead to cell death by triggering an “execution phase” where proteases, nucleases, and endogenous regulators of the cell death pathway are activated (Fig. 69-3).

Targeted agents differ from chemotherapy agents in that they do not indiscriminately cause macromolecular lesions but regulate the action of particular pathways. For example, the p210^{bcr-abl} fusion protein tyrosine kinase drives chronic myeloid leukemia (CML), and HER2/neu stimulates the proliferation of certain breast cancers. The tumor has been described as “addicted” to the function of these molecules in the sense that without the pathway’s continued action, the tumor cell cannot survive. In this way, targeted agents directed at p210^{bcr-abl} or HER2/neu may alter the “threshold” tumors driven by these molecules may have for undergoing cell death without actually creating any molecular lesions such as direct DNA strand breakage or altered membrane function.

Chemotherapy agents may be used for the treatment of active, clinically apparent cancer. The goal of such treatment in some cases is cure of the cancer, that is, elimination of all clinical and pathologic evidence

of cancer and return of the patient to an expected survival no different than the general population. **Table 69-3, A** lists those tumors considered curable by conventionally available chemotherapeutic agents when used to address disseminated or metastatic cancers. If a tumor is localized to a single site, serious consideration of surgery or primary radiation therapy should be given, because these treatment modalities may be curative as local treatments. Chemotherapy may then be used after the failure of these modalities to eradicate a local tumor or as part of multimodality approaches to offer primary treatment to a clinically localized tumor. In this event, it can allow organ preservation when given with radiation, as in the larynx or other upper airway sites, or sensitize tumors to radiation when given, e.g., to patients concurrently receiving radiation for lung or cervix cancer (**Table 69-3, B**). Chemotherapy can be administered as an *adjuvant*, i.e., in addition to surgery or radiation (**Table 69-3, C**), even after all clinically apparent disease has been removed. This use of chemotherapy has curative potential in breast and colorectal neoplasms, as it attempts to eliminate clinically unapparent tumor that may have already disseminated. *Neoadjuvant* chemotherapy refers to administration of chemotherapy before any surgery or radiation to a local tumor in an effort to enhance the effect of the local treatment.

Chemotherapy is routinely used in “conventional” dose regimens. In general, these doses produce reversible acute side effects, primarily consisting of transient myelosuppression with or without gastrointest-

inal toxicity (usually nausea), which are readily managed. “High-dose” chemotherapy regimens are predicated on the observation that the dose-response curve for many anticancer agents is rather steep, and increased dose can produce markedly increased therapeutic effect, although at the cost of potentially life-threatening complications that require intensive support, usually in the form of hematopoietic stem cell support from the patient (*autologous*) or from donors matched for histocompatibility loci (*allogeneic*), or pharmacologic “rescue” strategies to repair the effect of the high-dose chemotherapy on normal tissues. High-dose regimens have definite curative potential in defined clinical settings (**Table 69-3, D**).

If cure is not possible, chemotherapy may be undertaken with the goal of palliating some aspect of the tumor’s effect on the host. In this usage, value is perceived by the demonstration of improved symptom relief, progression-free survival, or overall survival at a certain time from the inception of treatment in the treated population, compared to a relevant control population established as the result of a clinical research protocol as a basis for U.S. Food and Drug Administration (FDA) approval of a particular cancer treatment as safe and effective. Common tumors that may be meaningfully addressed by chemotherapy with palliative intent are listed in **Table 69-3, E**.

Usually, tumor-related symptoms manifest as pain, weight loss, or some local symptom related to the tumor’s effect on normal structures. Patients treated with palliative intent should be aware of their diagnosis and the limitations of the proposed treatments, have access to supportive care, and have suitable “performance status,” according to assessment algorithms such as the one developed by Karnofsky

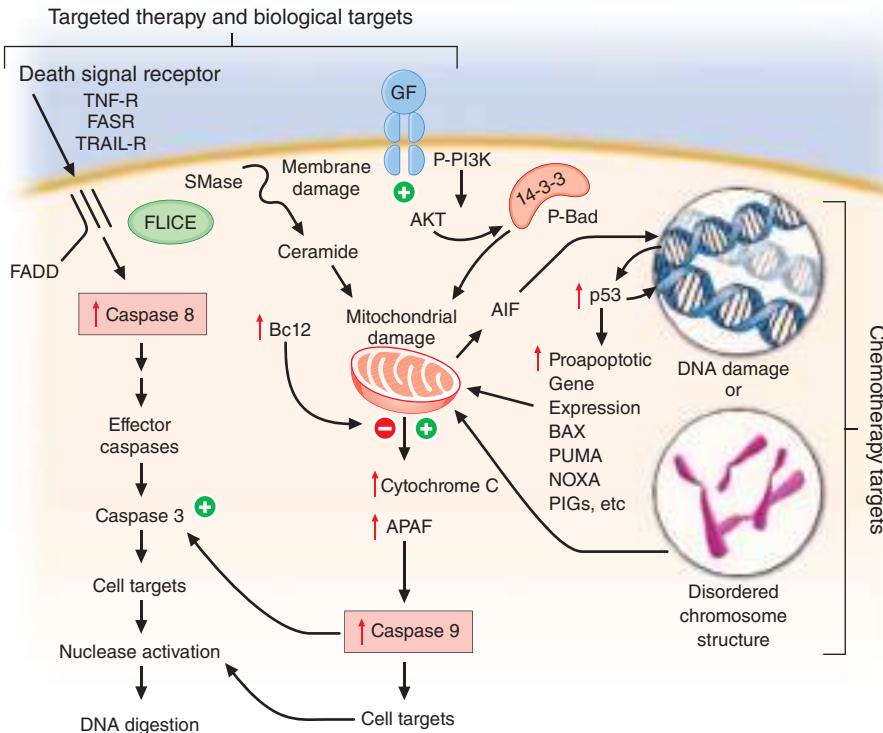


FIGURE 69-3 Integration of cell death responses. Cell death through apoptosis requires active participation of the cell. In response to interruption of growth factor (GF) or propagation of certain cytokine death signals (e.g., tumor necrosis factor receptor [TNF-R]), there is activation of “upstream” cysteine aspartyl proteases (caspases), which then directly digest cytoplasmic and nuclear proteins, resulting in activation of “downstream” caspases; these cause activation of nucleases, resulting in the characteristic DNA fragmentation that is a hallmark of apoptosis. Chemotherapy agents that create lesions in DNA or alter mitotic spindle function seem to activate aspects of this process by damage ultimately conveyed to the mitochondria, perhaps by activating the transcription of genes whose products can produce or modulate the toxicity of free radicals. In addition, membrane damage with activation of sphingomyelinases results in the production of ceramides that can have a direct action at mitochondria. The antiapoptotic protein bcl2 attenuates mitochondrial toxicity, while proapoptotic gene products such as bax antagonize the action of bcl2. Damaged mitochondria release cytochrome C and apoptosis-activating factor (APAF), which can directly activate caspase 9, resulting in propagation of a direct signal to other downstream caspases through protease activation. Apoptosis-inducing factor (AIF) is also released from the mitochondria and then can translocate to the nucleus, bind to DNA, and generate free radicals to further damage DNA. An additional proapoptotic stimulus is the bad protein, which can heterodimerize with bcl2 gene family members to antagonize apoptosis. Importantly, though, bad protein function can be retarded by its sequestration as phospho-bad through the 14-3-3 adapter proteins. The phosphorylation of bad is mediated by the action of the AKT kinase in a way that defines how growth factors that activate this kinase can retard apoptosis and promote cell survival.

TABLE 69-3 Curability of Cancers with Chemotherapy

A. Advanced Cancers with Possible Cure	D. Cancers Possibly Cured with "High-Dose" Chemotherapy with Stem Cell Support
Acute lymphoid and acute myeloid leukemia (pediatric/adult)	Relapsed leukemias, lymphoid and myeloid
Hodgkin's disease (pediatric/adult)	Relapsed lymphomas, Hodgkin's and non-Hodgkin's
Lymphomas—certain types (pediatric/adult)	Chronic myeloid leukemia
Germ cell neoplasms	Multiple myeloma
Embryonal carcinoma	
Teratocarcinoma	
Seminoma or dysgerminoma	
Choriocarcinoma	
Gestational trophoblastic neoplasia	
Pediatric neoplasms	
Wilms' tumor	
Embryonal rhabdomyosarcoma	
Ewing's sarcoma	
Peripheral neuroepithelioma	
Neuroblastoma	
Small-cell lung carcinoma	
Ovarian carcinoma	
B. Advanced Cancers Possibly Cured by Chemotherapy and Radiation	
Squamous carcinoma (head and neck)	
Squamous carcinoma (anus)	
Breast carcinoma	
Carcinoma of the uterine cervix	
Non-small-cell lung carcinoma (stage III)	
Small-cell lung carcinoma	
C. Cancers Possibly Cured with Chemotherapy as Adjuvant to Surgery	
Breast carcinoma	
Colorectal carcinoma ^a	
Osteogenic sarcoma	
Soft tissue sarcoma	
E. Cancers Responsive with Useful Palliation, But Not Cure, by Chemotherapy	
Bladder carcinoma	
Chronic myeloid leukemia	
Hairy cell leukemia	
Chronic lymphocytic leukemia	
Lymphoma—certain types	
Multiple myeloma	
Gastric carcinoma	
Cervix carcinoma	
Endometrial carcinoma	
Soft tissue sarcoma	
Head and neck cancer	
Adrenocortical carcinoma	
Islet cell neoplasms	
Breast carcinoma	
Colorectal carcinoma	
Renal carcinoma	
F. Tumors Poorly Responsive in Advanced Stages to Chemotherapy	
Pancreatic carcinoma	
Biliary tract neoplasms	
Thyroid carcinoma	
Carcinoma of the vulva	
Non-small-cell lung carcinoma	
Prostate carcinoma	
Melanoma (subsets)	
Hepatocellular carcinoma	
Salivary gland cancer	

^aRectum also receives radiation therapy.

(see Table 65-4) or by the Eastern Cooperative Oncology Group (ECOG) (see Table 65-5). ECOG performance status 0 (PS0) patients are without symptoms; PS1 patients are ambulatory but restricted in strenuous physical activity; PS2 patients are ambulatory but unable to work and are up and about 50% or more of the time; PS3 patients are capable of limited self-care and are up <50% of the time; and PS4 patients are totally confined to bed or chair and incapable of self-care. Only PS0, PS1, and PS2 patients are generally considered suitable for palliative (noncurative) treatment. If there is curative potential, even poor-performance status patients may be treated, but their prognosis is usually inferior to that of good-performance status patients treated with similar regimens.

An important perspective the primary care provider may bring to patients and their families facing incurable cancer is that, given the limited value of chemotherapeutic approaches at some point in the natural history of most metastatic cancers, *palliative care* or *hospice-based* approaches, with meticulous and ongoing attention to symptom relief and with family, psychological, and spiritual support, should receive prominent attention as a valuable therapeutic plan (Chaps. 9 and 65). Optimizing the quality of life rather than attempting to extend it becomes a valued intervention. Patients facing the impending progression of disease in a life-threatening way frequently choose to undertake toxic treatments of little to no potential value, and support provided by

the primary caregiver in accessing palliative and hospice-based options in contrast to receiving toxic and ineffective regimen can be critical in providing a basis for patients to make sensible choices.

Cytotoxic Chemotherapy Agents Table 69-4 lists commonly used cytotoxic cancer chemotherapy agents and pertinent clinical aspects of their use, with particular reference to adverse effects that might be encountered by the generalist in the care of patients. The drugs listed may be usefully grouped into two general categories: those affecting DNA and those affecting microtubules.

DNA-INTERACTIVE AGENTS DNA replication occurs during the synthesis or S-phase of the cell cycle, with chromosome segregation of the replicated DNA occurring in the M, or mitosis, phase. The G₁ and G₂ "gap phases" precede S and M, respectively. Historically, chemotherapeutic agents have been divided into "phase-nonspecific" agents, which can act in any phase of the cell cycle, and "phase-specific" agents, which require the cell to be at a particular cell cycle phase to cause greatest effect. "Checkpoints" in the cell cycle exist where the drug-related damage may be assessed and either repaired or cell death initiated.

Alkylating agents as a class are cell cycle phase-nonspecific agents. They break down, either spontaneously or after normal organ or tumor cell metabolism, to reactive intermediates that covalently modify bases in DNA. This leads to cross-linkage of DNA strands or the appearance of breaks in DNA as a result of repair efforts. "Broken" or cross-linked DNA is intrinsically unable to complete normal replication or cell division; in addition, it is a potent activator of cell cycle checkpoints and further activates cell-signaling pathways that can precipitate apoptosis. Alkylating agents share similar toxicities: myelosuppression, alopecia, gonadal dysfunction, mucositis, and pulmonary fibrosis. They differ greatly in a spectrum of normal organ toxicities. They also share the capacity to cause "second" neoplasms, particularly leukemia, many years after use, particularly when used in low doses for protracted periods.

Cyclophosphamide is inactive unless metabolized by the liver to 4-hydroxy-cyclophosphamide, which decomposes into an alkylating species, as well as to chloroacetaldehyde and acrolein. The latter causes chemical cystitis; therefore, excellent hydration must be maintained while using cyclophosphamide. If severe, the cystitis may be attenuated or prevented altogether (if expected from the dose of cyclophosphamide to be used) by mesna (*2-mercaptopethanesulfonate*). Liver disease impairs cyclophosphamide activation. Sporadic interstitial pneumonitis leading to pulmonary fibrosis can accompany the use of cyclophosphamide, and high doses used in conditioning regimens for bone marrow transplant can cause cardiac dysfunction. Ifosfamide is a cyclophosphamide analogue also activated in the liver, but more slowly, and it requires coadministration of mesna to prevent bladder injury. CNS effects, including somnolence, confusion, and psychosis, can follow ifosfamide use; the incidence appears related to low body surface area or decreased creatinine clearance.

Several alkylating agents are less commonly used. Bendamustine is a nitrogen mustard derivative with evidence of activity in chronic lymphocytic leukemia and certain lymphomas. Busulfan can cause profound myelosuppression, alopecia, and pulmonary toxicity but is relatively "lymphocyte sparing." Its routine use in treatment of CML has been curtailed in favor of imatinib (Gleevec) or dasatinib, but it is still used in transplant preparation regimens. Melphalan shows variable oral bioavailability and undergoes extensive binding to albumin and α_1 -acidic glycoprotein. Mucositis appears more prominently; however, it has prominent activity in multiple myeloma.

Nitrosoureas break down to carbamylating species that not only cause a distinct pattern of DNA base pair-directed toxicity but also can covalently modify proteins. They share the feature of causing relatively delayed bone marrow toxicity, which can be cumulative and long-lasting. Procarbazine is metabolized in the liver and possibly in tumor cells to yield a variety of free radical and alkylating species. In addition to myelosuppression, it causes hypnotic and other CNS effects, including vivid nightmares. It can cause a disulfiram-like syndrome on ingestion of ethanol. Dacarbazine (DTIC) is activated in the liver to yield the highly reactive methyl diazonium cation. It causes only modest

TABLE 69-4 Cytotoxic Chemotherapy Agents

DRUG	TOXICITY	INTERACTIONS, ISSUES
Direct DNA-Interacting Agents		
Alkylator		
Cyclophosphamide	Marrow (relative platelet sparing) Cystitis Common alkylator ^a Cardiac (high dose)	Liver metabolism required to activate to phosphoramide mustard + acrolein Mesna protects against “high-dose” bladder damage
Melphalan	Marrow (delayed nadir) GI (high dose)	Decreased renal function delays clearance
Carmustine (BCNU)	Marrow (delayed nadir) GI, liver (high dose)	
Lomustine (CCNU)	Renal	
Ifosfamide	Marrow (delayed nadir) Myelosuppressive Bladder Neurologic Metabolic acidosis	Analogue of cyclophosphamide Must use mesna Greater activity vs testicular neoplasms and sarcomas
Procarbazine	Marrow Nausea Neurologic Common alkylator ^a	Liver and tissue metabolism required Disulfiram-like effect with ethanol Acts as MAOI
Dacarbazine (DTIC)	Marrow Nausea Flulike	HBP after tyrosinase-rich foods Metabolic activation
Temozolomide	Nausea/vomiting Headache/fatigue Constipation	Infrequent myelosuppression
Cisplatin	Nausea Neuropathy Auditory	Maintain high urine flow; osmotic diuresis, monitor intake/output K ⁺ , Mg ²⁺ Emetogenic—prophylaxis needed Full dose if CrCl >60 mL/min and tolerate fluid push
Carboplatin	Marrow platelets > WBCs Renal Mg ²⁺ , Ca ²⁺	Reduce dose according to CrCl: to AUC of 5–7 mg/mL per min [AUC = dose/(CrCl + 25)]
Oxaliplatin	Marrow platelets > WBCs Nausea Renal (high dose)	Acute reversible neurotoxicity; chronic sensory neurotoxicity cumulative with dose; reversible laryngopharyngeal spasm
Antitumor Antibiotics and Topoisomerase Poisons		
Bleomycin	Pulmonary Skin effects Raynaud's Hypersensitivity	Inactivate by bleomycin hydrolase (decreased in lung/skin) O ₂ enhances pulmonary toxicity Cisplatin-induced decrease in CrCl may increase skin/lung toxicity
Dactinomycin	Marrow Nausea Mucositis Vesicant Alopecia	Reduce dose if CrCl <60 mL/min Radiation recall
Etoposide (VP16-213)	Marrow (WBCs > platelet) Alopecia Hypotension Hypersensitivity (rapid IV) Nausea Mucositis (high dose)	Hepatic metabolism—renal 30% Reduce doses with renal failure Schedule-dependent (5-day schedule better than 1-day) Late leukemogenic Accentuate antimetabolite action
Topotecan	Marrow Mucositis Nausea Mild alopecia	Reduce dose with renal failure No liver toxicity

(Continued)

TABLE 69-4 Cytotoxic Chemotherapy Agents (Continued)

DRUG	TOXICITY	INTERACTIONS, ISSUES
Irinotecan	Diarrhea: “early onset” with cramping, flushing, vomiting; “late onset” after several doses Marrow Alopecia Nausea Vomiting Pulmonary	Prodrug requires enzymatic clearance to active drug “SN 38” Early diarrhea due to acetylcholine release Late diarrhea, use “high-dose” loperamide (2 mg q2–4 h)
Doxorubicin and daunorubicin	Marrow Mucositis Alopecia Cardiovascular acute/chronic Vesicant	Heparin aggregate; coadministration increases clearance Acetaminophen, BCNU increase liver toxicity Radiation recall
Idarubicin	Marrow Cardiac (less than doxorubicin)	None established
Epirubicin	Marrow Cardiac	None established
Mitoxantrone	Marrow Cardiac (less than doxorubicin) Vesicant (mild) Blue urine, sclerae, nails	Interacts with heparin Less alopecia, nausea than doxorubicin Radiation recall Less alopecia, nausea than doxorubicin

Indirectly DNA-Interacting Agents**Antimetabolites**

6-Mercaptopurine (6-MP)	Marrow Liver Nausea	Variable bioavailability Metabolize by xanthine oxidase Decrease dose with allopurinol Increased toxicity with thiopurine methyltransferase deficiency
6-Thioguanine	Marrow Liver Nausea	Variable bioavailability Increased toxicity with thiopurine methyltransferase deficiency
2-Chlorodeoxyadenosine	Marrow Renal Fever	Notable use in hairy cell leukemia
Hydroxyurea	Marrow Nausea Mucositis Skin changes Rare renal, liver, lung, CNS	Decrease dose with renal failure Augments antimetabolite effect
Methotrexate	Marrow Liver/lung Renal tubular Mucositis	Toxicity lessened by “rescue” with leucovorin Excreted in urine Decrease dose in renal failure; NSAIDs increase renal toxicity
Pemetrexed	Anemia Neutropenia	Supplement folate/B ₁₂ Caution in renal failure
Pralatrexate	Thrombocytopenia Myelosuppression Mucositis	Active in peripheral T cell lymphoma
5-Fluorouracil (5FU)	Marrow Mucositis Neurologic Skin changes	Toxicity enhanced by leucovorin by increasing “ternary complex” with thymidylate synthase; dihydropyrimidine dehydrogenase deficiency increases toxicity; metabolism in tissue
Capecitabine	Diarrhea Hand-foot syndrome	Prodrug of 5FU due to intratumoral metabolism
Cytosine arabinoside	Marrow Mucositis Neurologic (high dose) Conjunctivitis (high dose) Noncardiogenic pulmonary edema	Enhances activity of alkylating agents Metabolizes in tissues by deamination but renal excretion prominent at doses >500 mg; therefore, dose reduce in “high-dose” regimens in patients with decreased CrCl

(Continued)

TABLE 69-4 Cytotoxic Chemotherapy Agents (Continued)

DRUG	TOXICITY	INTERACTIONS, ISSUES
Azacitidine	Marrow	Use limited to leukemia/myelodysplastic syndrome
Decitabine	Nausea	Altered methylation of DNA alters gene expression
Gemcitabine	Liver	
	Neurologic	
	Myalgia	
	Marrow	
	Nausea	
	Hepatic	
Fludarabine phosphate	Fever/"flu syndrome"	
	Marrow	Dose reduction with renal failure
	Neurologic	Metabolized to Fara converted to Fara ATP in cells by deoxycytidine kinase
	Lung	
Asparaginase	Decrease protein synthesis; indirect inhibition of DNA synthesis by decreased histone synthesis	Blocks methotrexate action
	Clotting factors	
	Glucose	
	Albumin	
	Hypersensitivity	
	CNS	
	Pancreatitis	
	Hepatic	
Antimitotic Agents		
Vincristine	Vesicant	Hepatic clearance
	Marrow	Dose reduction for bilirubin >1.5 mg/dL
	Neurologic	Prophylactic bowel regimen
	GI: ileus/constipation; bladder hypotoxicity; SIADH	
Vinblastine	Cardiovascular	
	Vesicant	Hepatic clearance
	Marrow	Dose reduction as with vincristine
	Neurologic (less common but similar spectrum to other vincas)	
	Hypertension	
	Raynaud's	
Vinorelbine	Vesicant	Hepatic clearance
	Marrow	
	Allergic/bronchospasm (immediate)	
	Dyspnea/cough (subacute)	
	Neurologic (less prominent but similar spectrum to other vincas)	
Paclitaxel	Hypersensitivity	Premedicate with steroids, H ₁ and H ₂ blockers
	Marrow	
	Mucositis	Hepatic clearance
	Alopecia	Dose reduction as with vincas
	Sensory neuropathy	
	CV conduction disturbance	
	Nausea—infrequent	
Docetaxel	Hypersensitivity	Premedicate with steroids, H ₁ and H ₂ blockers
	Fluid retention syndrome	
	Marrow	
	Dermatologic	
	Sensory neuropathy	
	Nausea infrequent	
	Some stomatitis	
Nab-paclitaxel (protein bound)	Neuropathy	Caution in hepatic insufficiency
	Anemia	
	Neutropenia	
	Thrombocytopenia	
Ixabepilone	Myelosuppression	
	Neuropathy	

^aCommon alkylator: alopecia, pulmonary, infertility, plus teratogenesis.

Abbreviations: ALL, acute lymphocytic leukemia; AUC, area under the curve; CHF, congestive heart failure; CNS, central nervous system; CrCl, creatinine clearance; CV, cardiovascular; GI, gastrointestinal; HBP, high blood pressure; MAOI, monoamine oxidase inhibitor; NSAID, nonsteroidal anti-inflammatory drug; SIADH, syndrome of inappropriate antidiuretic hormone secretion.

myelosuppression 21–25 days after a dose but causes prominent nausea on day 1. Temozolomide is structurally related to dacarbazine but was designed to be activated by nonenzymatic hydrolysis in tumors and is bioavailable orally. Brain tumors with alkylguanine alkyl transferase deficiency are selectively susceptible to temozolomide, which alkylates the O⁶ position of guanine.

Cisplatin was discovered fortuitously by observing that bacteria present in electrolysis solutions with platinum electrodes could not divide. Only the *cis* diamine configuration is active as an antitumor agent. In the intracellular environment, a chloride is lost from each position, being replaced by a water molecule. The resulting positively charged species is an efficient bifunctional interactor with DNA, forming Pt-based cross-links. Cisplatin requires administration with adequate hydration, including forced diuresis with mannitol to prevent kidney damage; even with the use of hydration, gradual decrease in kidney function is common, along with noteworthy anemia. Hypomagnesemia frequently attends cisplatin use and can lead to hypocalcemia and tetany. Other common toxicities include neurotoxicity with stocking-and-glove sensorimotor neuropathy. Hearing loss occurs in 50% of patients treated with conventional doses. Cisplatin is intensely emetogenic, requiring prophylactic antiemetics. Myelosuppression is less evident than with other alkylating agents. Chronic vascular toxicity (Raynaud's phenomenon, coronary artery disease) is a more unusual toxicity. Carboplatin displays less nephro-, oto-, and neurotoxicity. However, myelosuppression is more frequent, and because the drug is exclusively cleared through the kidney, adjustment of dose for creatinine clearance must be accomplished through use of various dosing nomograms. Oxaliplatin is a platinum analogue with noteworthy activity in colon cancers refractory to other treatments. It is prominently neurotoxic.

ANTITUMOR ANTIBIOTICS AND TOPOISOMERASE POISONS Antitumor antibiotics are substances produced by bacteria that in nature appear to provide a chemical defense against other hostile microorganisms. As a class, they bind to DNA directly and can frequently undergo electron transfer reactions to generate free radicals in close proximity to DNA, leading to DNA damage in the form of single-strand breaks or cross-links. Topoisomerase poisons include natural products or semisynthetic species derived ultimately from plants, and they modify enzymes that regulate the capacity of DNA to unwind to allow normal replication or transcription. These include topoisomerase I, which creates single-strand breaks that then rejoin following the passage of the other DNA strand through the break. Topoisomerase II creates double-strand breaks through which another segment of DNA duplex passes before rejoining. Owing to the role of topoisomerase I in the procession of the replication fork, topoisomerase I poisons cause lethality if the topoisomerase I-induced lesions are made in S-phase.

Doxorubicin can intercalate into DNA, thereby altering DNA structure, replication, and topoisomerase II function. It can also undergo reduction reactions by accepting electrons into its quinone ring system, with the capacity to undergo reoxidation to form reactive oxygen radicals after reoxidation. It causes predictable myelosuppression, alopecia, nausea, and mucositis. In addition, it causes acute cardiotoxicity in the form of atrial and ventricular dysrhythmias, but these are rarely of clinical significance. In contrast, cumulative doses >550 mg/m² are associated with a 10% incidence of chronic cardiomyopathy. The incidence of cardiomyopathy appears to be related to schedule (peak serum concentration), with low-dose, frequent treatment or continuous infusions better tolerated than intermittent higher-dose exposures. Cardiotoxicity has been related to iron-catalyzed oxidation and reduction of doxorubicin. Cardiotoxicity is related to peak plasma dose; thus, lower doses and continuous infusions are less likely to cause heart damage. Doxorubicin's cardiotoxicity is increased when given together with trastuzumab (Herceptin), the anti-HER2/neu antibody. Radiation recall or interaction with concomitantly administered radiation to cause local site complications is frequent. The drug is a powerful vesicant, with necrosis of tissue apparent 4–7 days after an extravasation; therefore, it should be administered into a rapidly flowing intravenous line. Dexrazoxane is an antidote to doxorubicin-induced extravasation. Doxorubicin is metabolized by the liver, so doses must be reduced by

50–75% in the presence of liver dysfunction. Daunorubicin is closely related to doxorubicin and was actually introduced first into leukemia treatment, where it remains part of curative regimens and has been shown preferable to doxorubicin owing to less mucositis and colonic damage. Idarubicin is also used in acute myeloid leukemia treatment and may be preferable to daunorubicin in activity. Encapsulation of daunorubicin into a liposomal formulation has attenuated cardiac toxicity and antitumor activity in Kaposi's sarcoma, other sarcomas, multiple myeloma, and ovarian cancer.

Bleomycin refers to a mixture of glycopeptides that have the unique feature of forming complexes with Fe²⁺ while also bound to DNA. It remains an important component of curative regimens for Hodgkin's disease and germ cell neoplasms. Oxidation of Fe²⁺ gives rise to superoxide and hydroxyl radicals. The drug causes little, if any, myelosuppression. The drug is cleared rapidly, but augmented skin and pulmonary toxicity in the presence of renal failure has led to the recommendation that doses be reduced by 50–75% in the face of a creatinine clearance <25 mL/min. Bleomycin is not a vesicant and can be administered intravenously, intramuscularly, or subcutaneously. Common side effects include fever and chills, facial flush, and Raynaud's phenomenon. The most feared complication of bleomycin treatment is pulmonary fibrosis, which increases in incidence at >300 cumulative units administered and is minimally responsive to treatment (e.g., glucocorticoids). The earliest indicator of an adverse effect is usually a decline in the carbon monoxide diffusing capacity (DLco) or coughing, although cessation of drug immediately upon documentation of a decrease in DLco may not prevent further decline in pulmonary function. Bleomycin is inactivated by a bleomycin hydrolase, whose concentration is diminished in skin and lung. Because bleomycin-dependent electron transport is dependent on O₂, bleomycin toxicity may become apparent after exposure to transient very high fraction of inspired oxygen (FIO₂). Thus, during surgical procedures, patients with prior exposure to bleomycin should be maintained on the lowest FIO₂ consistent with maintaining adequate tissue oxygenation.

Mitoxantrone is a synthetic compound that was designed to recapitulate features of doxorubicin but with less cardiotoxicity. It is quantitatively less cardiotoxic (comparing the ratio of cardiotoxic to therapeutically effective doses), but is still associated with a 10% incidence of cardiotoxicity at cumulative doses of >150 mg/m². It also causes alopecia. Etoposide binds directly to topoisomerase II and DNA in a reversible ternary complex. It stabilizes the covalent intermediate in the enzyme's action where the enzyme is covalently linked to DNA. Prominent clinical effects include myelosuppression, nausea, and transient hypotension related to the speed of administration of the agent. Etoposide is a mild vesicant but is relatively free from other large-organ toxicities. Camptothecins target topoisomerase I. Topotecan is a camptothecin-derivative approved for use in gynecologic tumors and small-cell lung cancer. Toxicity is limited to myelosuppression and mucositis. CPT-11, or irinotecan, is a camptothecin with evidence of activity in colon carcinoma. In addition to myelosuppression, it causes a secretory diarrhea related to the toxicity of a metabolite called SN-38. Levels of SN-38 are particularly high in the setting of Gilbert's disease, characterized by defective glucuronyl transferase and indirect hyperbilirubinemia, a condition that affects about 10% of the white population in the United States. The diarrhea can be treated effectively with loperamide or octreotide.

ANTIMETABOLITES A broad definition of antimetabolites would include compounds with structural similarity to precursors of purines or pyrimidines, or compounds that interfere with purine or pyrimidine synthesis. Some antimetabolites can cause DNA damage indirectly, through misincorporation into DNA, abnormal timing or progression through DNA synthesis, or altered function of pyrimidine and purine biosynthetic enzymes. They tend to convey greatest toxicity to cells in S-phase, and the degree of toxicity increases with duration of exposure. Common toxic manifestations include stomatitis, diarrhea, and myelosuppression. Second malignancies are not associated with their use.

Methotrexate inhibits dihydrofolate reductase, which regenerates reduced folates from the oxidized folates produced when thymidine monophosphate is formed from deoxyuridine monophosphate.

Without reduced folates, cells die a “thymine-less” death. N5-Tetrahydrofolate or N5-formyltetrahydrofolate (leucovorin) can bypass this block and rescue cells from methotrexate, which is maintained in cells by polyglutamylation. The drug and other reduced folates are transported into cells by a membrane carrier, and high concentrations of drug can bypass this carrier and allow diffusion of drug directly into cells. These properties have suggested the design of “high-dose” methotrexate regimens with leucovorin rescue of normal marrow and mucosa as part of curative approaches to osteosarcoma in the adjuvant setting and hematopoietic neoplasms of children and adults. Methotrexate is cleared by the kidney via both glomerular filtration and tubular secretion, and toxicity is augmented by renal dysfunction and drugs such as salicylates, probenecid, and nonsteroidal anti-inflammatory agents that undergo tubular secretion. With normal renal function, 15 mg/m^2 leucovorin will rescue $10^{-8}\text{--}10^{-6}\text{ M}$ methotrexate in 3–4 doses. However, with decreased creatinine clearance, doses of $50\text{--}100\text{ mg/m}^2$ are continued until methotrexate levels are $<5 \times 10^{-8}\text{ M}$. In addition to bone marrow suppression and mucosal irritation, methotrexate can cause renal failure itself at high doses owing to crystallization in renal tubules; therefore, high-dose regimens require alkalinization of urine with increased flow by hydration. Methotrexate can be sequestered in third-space collections and diffuse back into the general circulation, causing prolonged myelosuppression. Less frequent adverse effects include reversible increases in transaminases and hypersensitivity-like pulmonary syndrome. Chronic low-dose methotrexate can cause hepatic fibrosis. When administered to the intrathecal space, methotrexate can cause chemical arachnoiditis and CNS dysfunction.

Pemetrexed is a folate-directed antimetabolite. It inhibits the activity of several enzymes, including thymidylate synthetase (TS), dihydrofolate reductase, and glycynamide ribonucleotide formyltransferase, thereby affecting the synthesis of both purine and pyrimidine nucleic acid precursors. To avoid significant toxicity to the normal tissues, patients receiving pemetrexed should also receive low-dose folate and vitamin B₁₂ supplementation. Pemetrexed has notable activity against certain lung cancers and, in combination with cisplatin, also against mesotheliomas. Pralatrexate is an antifolate approved for use in T-cell lymphoma that is very efficiently transported into cancer cells.

5-Fluorouracil (5FU) represents an early example of “rational” drug design in that it originated from the observation that tumor cells incorporate radiolabeled uracil more efficiently into DNA than normal cells, especially gut. 5FU is metabolized in cells to 5'FdUMP, which inhibits TS. In addition, misincorporation can lead to single-strand breaks, and RNA can aberrantly incorporate FUMP. 5FU is metabolized by dihydropyrimidine dehydrogenase, and deficiency of this enzyme can lead to excessive toxicity from 5FU. Oral bioavailability varies unreliable, but prodrugs such as capecitabine have been developed that allow at least equivalent activity to many parenteral 5FU-based approaches. Intravenous administration of 5FU leads to bone marrow suppression after short infusions but to stomatitis after prolonged infusions. Leucovorin augments the activity of 5FU by promoting formation of the ternary covalent complex of 5FU, the reduced folate, and TS. Less frequent toxicities include CNS dysfunction, with prominent cerebellar signs, and endothelial toxicity manifested by thrombosis, including pulmonary embolus and myocardial infarction.

Cytosine arabinoside (ara-C) is incorporated into DNA after formation of ara-CTP, resulting in S-phase-related toxicity. Continuous infusion schedules allow maximal efficiency, with uptake maximal at $5\text{--}7\text{ }\mu\text{M}$. Ara-C can be administered intrathecally. Adverse effects include nausea, diarrhea, stomatitis, chemical conjunctivitis, and cerebellar ataxia. Gemcitabine is a cytosine derivative that is similar to ara-C in that it is incorporated into DNA after anabolism to the triphosphate, rendering DNA susceptible to breakage and repair synthesis, which differs from that in ara-C in that gemcitabine-induced lesions are very inefficiently removed. In contrast to ara-C, gemcitabine appears to have useful activity in a variety of solid tumors, with limited nonmyelosuppressive toxicities.

6-Thioguanine and 6-mercaptopurine (6MP) are used in the treatment of acute lymphoid leukemia. Although administered orally, they display variable bioavailability. 6MP is metabolized by xanthine

oxidase and therefore requires dose reduction when used with allopurinol. 6MP is also metabolized by thiopurine methyltransferase; genetic deficiency of thiopurine methyltransferase results in excessive toxicity.

Fludarabine phosphate is a prodrug of F-adenine arabinoside (F-ara-A), which in turn was designed to diminish the susceptibility of ara-A to adenosine deaminase. F-ara-A is incorporated into DNA and can cause delayed cytotoxicity even in cells with low growth fraction, including chronic lymphocytic leukemia and follicular B-cell lymphoma. CNS and peripheral nerve dysfunction and T-cell depletion leading to opportunistic infections can occur in addition to myelosuppression. 2-Chlorodeoxyadenosine is a similar compound with activity in hairy cell leukemia. Hydroxyurea inhibits ribonucleotide reductase, resulting in S-phase block. It is orally bioavailable and useful for the acute management of myeloproliferative states.

Asparaginase is a bacterial enzyme that causes breakdown of extracellular asparagine required for protein synthesis in certain leukemic cells. This effectively stops tumor cell DNA synthesis, as DNA synthesis requires concurrent protein synthesis. The outcome of asparaginase action is therefore very similar to the result of the small-molecule antimetabolites. Because asparaginase is a foreign protein, hypersensitivity reactions are common, as are effects on organs such as pancreas and liver that normally require continuing protein synthesis. This may result in decreased insulin secretion with hyperglycemia, with or without hyperamylasemia and clotting function abnormalities. Close monitoring of clotting functions should accompany use of asparaginase. Paradoxically, owing to depletion of rapidly turning over anticoagulant factors, thromboses particularly affecting the CNS may also be seen with asparaginase.

MITOTIC SPINDLE INHIBITORS Microtubules are cellular structures that form the mitotic spindle, and in interphase cells, they are responsible for the cellular “scaffolding” along which various motile and secretory processes occur. Microtubules are composed of repeating noncovalent multimers of a heterodimer of α and β isoform of the protein tubulin. Vincristine binds to the tubulin dimer with the result that microtubules are disaggregated. This results in the block of growing cells in M-phase; however, toxic effects in G₁ and S-phase are also evident, reflecting effects on normal cellular activities of microtubules. Vincristine is metabolized by the liver, and dose adjustment in the presence of hepatic dysfunction is required. It is a powerful vesicant, and infiltration can be treated by local heat and infiltration of hyaluronidase. At clinically used intravenous doses, neurotoxicity in the form of glove-and-stock neuropathy is frequent. Acute neuropathic effects include jaw pain, paralytic ileus, urinary retention, and the syndrome of inappropriate antidiuretic hormone secretion. Myelosuppression is not seen. Vinblastine is similar to vincristine, except that it tends to be more myelotoxic, with more frequent thrombocytopenia and also mucositis and stomatitis. Vinorelbine is a vinca alkaloid that appears to have differences in resistance patterns in comparison to vincristine and vinblastine; it may be administered orally.

The taxanes include paclitaxel and docetaxel. These agents differ from the vinca alkaloids in that the taxanes stabilize microtubules against depolymerization. The “stabilized” microtubules function abnormally and are not able to undergo the normal dynamic changes of microtubule structure and function necessary for cell cycle completion. Taxanes are among the most broadly active antineoplastic agents for use in solid tumors, with evidence of activity in ovarian cancer, breast cancer, Kaposi’s sarcoma, and lung tumors. They are administered intravenously, and paclitaxel requires use of a Cremophor-containing vehicle that can cause hypersensitivity reactions. Premedication with dexamethasone (8–16 mg orally or intravenously 12 and 6 h before treatment) and diphenhydramine (50 mg) and cimetidine (300 mg), both 30 min before treatment, decreases but does not eliminate the risk of hypersensitivity reactions to the paclitaxel vehicle. A protein-bound formulation of paclitaxel (called *nab-paclitaxel*) has at least equivalent antineoplastic activity and decreased risk of hypersensitivity reactions. Paclitaxel may also cause hypersensitivity reactions, myelosuppression, neurotoxicity in the form of glove-and-stock numbness, and paresthesia. Docetaxel causes comparable degrees of

myelosuppression and neuropathy. Docetaxel uses a polysorbate 80 formulation that can cause fluid retention in addition to hypersensitivity reactions; dexamethasone premedication with or without antihistamines is frequently used. Cabazitaxel is a taxane with somewhat better activity in prostate cancers than earlier generations of taxanes, perhaps due to superior delivery to sites of disease.

Epothilones represent a class microtubule-stabilizing agents that have been conscientiously optimized for activity in taxane-resistant tumors. Ixabepilone has clear evidence of activity in breast cancers resistant to taxanes and anthracyclines such as doxorubicin. It retains acceptable expected side effects, including myelosuppression, and can also cause peripheral sensory neuropathy. Eribulin is a microtubule-directed agent with activity in patients who have had progression of disease on taxanes. It alters dynamics of microtubule re-modeling in cells.

Targeted Chemotherapy • HORMONE RECEPTOR-DIRECTED THERAPY

Steroid hormone receptor-related molecules have emerged as prominent targets for small molecules useful in cancer treatment. When bound to their cognate ligands, these receptors can alter gene transcription and, in certain tissues, induce apoptosis. The pharmacologic effect is a mirror or parody of the normal effects of the agents acting on nontransformed normal tissues. While in some cases, such as breast cancer, demonstration of the target hormone receptor is necessary, in other cases such prostate cancer (androgen receptor) and lymphoid neoplasms (glucocorticoid receptor), the relevant receptor is always present in the tumor.

Glucocorticoids are generally given in “pulsed” high doses in leukemias and lymphomas, where they induce cell death in tumor cells. Cushing’s syndrome and inadvertent adrenal suppression on withdrawal from high-dose glucocorticoids can be significant complications, along with infections common in immunosuppressed patients, in particular *Pneumocystis pneumonia*, which classically appears a few days after completing a course of high-dose glucocorticoids.

Tamoxifen is a partial estrogen receptor antagonist; it has a tenfold greater antitumor activity in breast cancer patients whose tumors express estrogen receptors than in those who have low or no levels of expression. It might be considered the prototypic “molecularly targeted” agent. Owing to its agonistic activities in vascular and uterine tissue, side effects include a somewhat increased risk of cardiovascular complications, such as thromboembolic phenomena, and a small increased incidence of endometrial carcinoma, which appears after chronic use (usually >5 years). Progestational agents—including medroxyprogesterone acetate, androgens including fluoxymesterone (Halotestin), and, paradoxically, estrogens—have approximately the same degree of activity in primary hormonal treatment of breast cancers that have elevated expression of estrogen receptor protein. Estrogen itself is not used often owing to prominent cardiovascular and uterotrophic activity.

Aromatase refers to a family of enzymes that catalyze the formation of estrogen in various tissues, including the ovary and peripheral adipose tissue and some tumor cells. Aromatase inhibitors are of two types, the irreversible steroid analogues such as exemestane and the reversible inhibitors such as anastrozole or letrozole. Anastrozole is superior to tamoxifen in the adjuvant treatment of breast cancer in postmenopausal patients with estrogen receptor-positive tumors. Letrozole treatment affords benefit following tamoxifen treatment. Adverse effects of aromatase inhibitors may include an increased risk of osteoporosis.

Metastatic prostate cancer is treated by androgen deprivation. Orchiectomy causes responses in 80% of patients. In the event that orchiectomy is not accepted by the patient, testicular androgen suppression can also be effected by luteinizing hormone-releasing hormone (LHRH) agonists such as leuprolide and goserelin. These agents cause tonic stimulation of the LHRH receptor, with the loss of its normal pulsatile activation resulting in decreased output of LH by the anterior pituitary. Therefore, as primary hormonal manipulation in prostate cancer, one can choose orchiectomy or leuprolide, but not both. The addition of androgen receptor blockers, including flutamide

or bicalutamide, is of uncertain additional benefit in extending overall response duration, although pretreatment with these agents before LHRH agonists is important to avoid a surge in testosterone after initial LH release. Enzalutamide also binds to the androgen receptor and antagonizes androgen action in a mechanistically distinct way. Somewhat analogous to inhibitors of aromatase, agents have been derived that inhibit testosterone and other androgen synthesis in the testis, adrenal gland, and prostate tissue. Abiraterone inhibits 17 α -hydroxylase/C17,20 lyase (CYP 17A1) and has been shown to be active in prostate cancer patients experiencing progression despite androgen blockade.

Tumors that respond to a primary hormonal manipulation may frequently respond to second and third hormonal manipulations. Thus, breast tumors that had previously responded to tamoxifen have, on relapse, notable response rates to withdrawal of tamoxifen itself or to subsequent addition of an aromatase inhibitor or progestin. Likewise, initial treatment of prostate cancers with leuprolide plus flutamide may be followed after disease progression by response to withdrawal of flutamide. These responses may result from the removal of antagonists from mutant steroid hormone receptors that have come to depend on the presence of the antagonist as a growth-promoting influence.

DIAGNOSTICALLY GUIDED TARGETED THERAPY The basis for discovery of drugs of this type was the prior knowledge of oncogene directed pathways driving tumor growth. **Figure 69-4** summarizes how FDA-approved targeted agents act. In the case of diagnostically guided targeted chemotherapy, prior demonstration of a specific target is necessary to guide the rational use of the agent, while in the case of targeted agents directed at oncogenic pathways, specific diagnosis of pathway activation is not yet necessary or in some cases feasible, although this is an area of ongoing clinical research. **Table 69-5** lists currently approved targeted chemotherapy agents, with features of their use.

In hematologic tumors, the prototypic agent of this type is imatinib, which targets the ATP binding site of the p210^{bcr-abl} protein tyrosine kinase that is formed as the result of the chromosome 9;22 translocation producing the Philadelphia chromosome in CML. Imatinib is superior to interferon (IFN) plus chemotherapy in the initial treatment of the chronic phase of this disorder. It has lesser activity in the blast phase of CML, where the cells may have acquired additional mutations in p210^{bcr-abl} itself or other genetic lesions. Its side effects are relatively tolerable in most patients and include hepatic dysfunction, diarrhea, and fluid retention. Rarely, patients receiving imatinib have decreased cardiac function, which may persist after discontinuation of the drug. The quality of response to imatinib enters into the decision about when to refer patients with CML for consideration of transplant approaches. Nilotinib is a tyrosine protein kinase inhibitor with a similar spectrum of activity to imatinib, but with increased potency and perhaps better tolerance by certain patients. Dasatinib, another inhibitor of the p210^{bcr-abl} oncogenes, is active in certain mutant variants of p210^{bcr-abl} that are refractory to imatinib and arise during therapy with imatinib or are present de novo. Dasatinib also has inhibitory action against kinases belonging to the src tyrosine protein kinase family; this activity may contribute to its effects in hematopoietic tumors and suggest a role in solid tumors where src kinases are active. The T315I mutant of p210^{bcr-abl} is resistant to imatinib, nilotinib, bosutinib, and dasatinib; ponatinib has activity in patients with this p210^{bcr-abl} variant, but ponatinib has noteworthy associated thromboembolic toxicity. Use of this class of targeted agents is thus critically guided not only by the presence of the p210^{bcr-abl} tyrosine kinase, but also by the presence of different mutations in the ATP binding site.

All-*trans*-retinoic acid (ATRA) targets the PML-retinoic acid receptor (RAR) α fusion protein, which is the result of the chromosome 15;17 translocation pathogenic for most forms of APL. Administered orally, it causes differentiation of the neoplastic promyelocytes to mature granulocytes and attenuates the rate of hemorrhagic complications. Adverse effects include headache with or without pseudotumor cerebri and gastrointestinal and cutaneous toxicities.

In epithelial solid tumors, the small-molecule epidermal growth factor (EGF) antagonists act at the ATP binding site of the EGF receptor

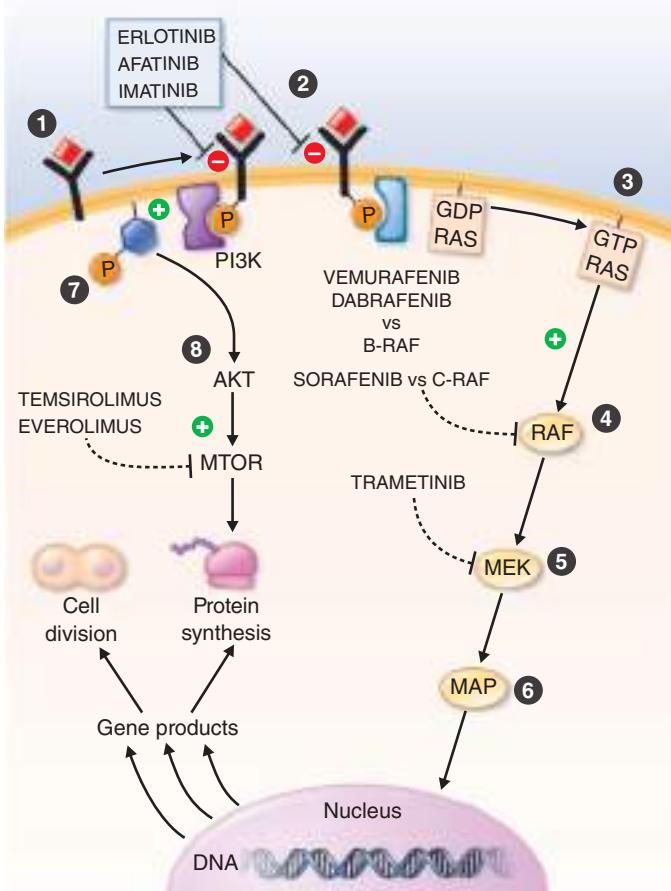


FIGURE 69-4 Targeted chemotherapeutic agents act in most instances by interrupting cell growth factor-mediated signaling pathways. After a growth factor binds to cognate receptor (1), in many cases there is activation of tyrosine kinase activity particularly after dimerization of the receptors (2). This leads to autophosphorylation of the receptor and docking of “adaptor” proteins. One important pathway activated occurs after exchange of GDP for GTP in the RAS family of protooncogene products (3). GTP-RAS activates the RAF proto-oncogene kinase (4), leading to a phosphorylation cascade of kinases (5, 6) that ultimately impart signals to regulators of gene function to produce transcripts which activate cell cycle progression and increase protein synthesis. In parallel, tyrosine phosphorylated receptors can activate the phosphatidylinositol-3-kinase to produce the phosphorylated lipid phosphatidyl-inositol-3-phosphate (7). This leads to the activation of the AKT kinase (8) which in turn stimulates the mammalian “Target of Rapamycin” kinase (mTOR), which directly increases the translation of key mRNAs for gene products regulating cell growth. Erlotinib and afatinib, are examples of Epidermal Growth Factor receptor tyrosine kinase inhibitors; imatinib can act on the nonreceptor tyrosine kinase bcr-abl or c-KIT membrane bound tyrosine kinase. Vemurafenib and Dabrafenib act on the B isoform of RAF uniquely in melanoma, and c-RAF is inhibited by sorafenib. Trametinib acts on MEK. Temsirolimus and everolimus inhibit mTOR kinase to downregulate translation of oncogenic mRNAs.

tyrosine kinase. In early clinical trials, gefitinib showed evidence of responses in a small fraction of patients with non-small-cell lung cancer (NSCLC). Side effects were generally acceptable, consisting mostly of acneiform rash (treated with glucocorticoid creams and clindamycin gel) and diarrhea. Subsequent analysis of responding patients revealed a high frequency of activating mutations in the EGF receptor. Patients with such activating mutations who initially responded to gefitinib but who then had progression of the disease then acquired additional mutations in the enzyme, analogous functionally to mutational variants responsible for imatinib resistance in CML. Erlotinib is another EGF receptor tyrosine kinase antagonist where the presence of EGF receptor tyrosine kinase mutations has recently been shown to be a basis for recommending erlotinib and afatinib for first-line treatment of advanced NSCLC. Osimertinib is uniquely active in lung cancers with the T790M mutation. Likewise, crizotinib targeting the

alk protooncogene fusion protein has value in the initial treatment of *alk*-positive NSCLC. Lapatinib is a tyrosine kinase inhibitor with both EGF receptor and HER2/neu antagonist activity, which is important in the treatment of breast cancers expressing the HER2/neu oncogene.

In addition to the p210^{bcr-abl} kinase, imatinib also has activity against the c-kit tyrosine kinase (the receptor for the *steel* growth factor, also called stem cell factor) and the platelet-derived growth factor receptor (PDGFR), both of which can be expressed in gastrointestinal stromal sarcoma (GIST). Imatinib has found clinical utility in GIST, a tumor previously notable for its refractoriness to chemotherapeutic approaches. Imatinib's degree of activity varies with the specific mutational variant of kit or PDGFR present in a particular patient's tumor.

The *BRAF* V600E mutation has been detected in a notable fraction of melanomas, thyroid tumors, and hairy cell leukemia, and preclinical models supported the concept that *BRAF* V600E drives oncogenic signaling in these tumors. Vemurafenib and dabrafenib, with selective capacity to inhibit the *BRAF* V600E serine kinase activity, were each shown to cause noteworthy responses in patients with *BRAF* V600E-mutated melanomas, although early relapse occurred in many patients treated with the drugs as single agents. Trametinib, acting downstream of *BRAF* V600E by directly inhibiting the MEK serine kinase by a non-ATP binding site mechanism, also displayed noteworthy responses in *BRAF* V600E-mutated melanomas, and the combination of trametinib and dabrafenib is even more active, by targeting the *BRAF* V600E-driven pathway at two points in the pathway leading to gene activation.

ONCOGENICALLY ACTIVATED PATHWAYS Agents in this class also target specific regulatory molecules in promoting the viability of tumor cells, but they do not require the diagnostically verified presence of a particular target or target variant at this time.

“Multitargeted” kinase antagonists are small-molecule ATP site-directed antagonists that inhibit more than one protein kinase and have value in the treatment of several solid tumors. Drugs of this type with prominent activity against the vascular endothelial growth factor receptor (VEGFR) tyrosine kinase have activity in renal cell carcinoma. Sorafenib is a VEGFR antagonist with activity against the *raf* serine-threonine protein kinase, and regorafenib is a closely related drug with value in relapsed advanced colon cancer. Pazopanib also prominently targets VEGFR and has activity in renal carcinoma and soft tissue sarcomas. Sunitinib has anti-VEGFR, anti-PDGFR, and anti-c-kit activity. It causes prominent responses and stabilization of disease in renal cell cancers and GISTS. Side effects for agents with anti-VEGFR activity prominently include hypertension, proteinuria, and, more rarely, bleeding and clotting disorders and perforation of scarred gastrointestinal lesions. Also encountered are fatigue, diarrhea, and the hand-foot syndrome, with erythema and desquamation of the distal extremities, in some cases requiring dose modification, particularly with sorafenib.

Tensirolimus and everolimus are mammalian target of rapamycin (mTOR) inhibitors with activity in renal cancers. They produce stomatitis, fatigue, and some hyperlipidemia (10%), myelosuppression (10%), and rare lung toxicity. Everolimus is also useful in patients with hormone receptor-positive breast cancers displaying resistance to hormonal inhibition and in certain neuroendocrine and brain tumors, the latter arising in patients with sporadic or inherited mutations in the pathway activating mTOR. Cyclin dependent kinases (CDKs) are activated as the result of oncogene pathway activity. Palbociclib, a selective inhibitor of CDKs 4 and 6, has noteworthy activity in conjunction with the mTOR inhibitors in advanced breast cancers also expressing the estrogen receptor.

In hematologic neoplasms, bortezomib is an inhibitor of the proteasome, the multisubunit assembly of protease activities responsible for the selective degradation of proteins important in regulating activation of transcription factors, including nuclear factor- κ B (NF- κ B) and proteins regulating cell cycle progression. It has activity in multiple myeloma and certain lymphomas. Adverse effects include neuropathy, orthostatic hypotension with or without hyponatremia, and reversible thrombocytopenia. Carfilzomib is a proteasome inhibitor chemically unrelated to bortezomib without prominent neuropathy, but with

TABLE 69-5 Molecularly Targeted Agents

DRUG	TARGET	ADVERSE EVENTS	NOTES
Diagnostically Guided Protein Kinase Antagonists			
Imatinib	Bcr-Abl fusion protein (CML/ALL); c-kit mutants, PDGFR variants (GI stromal tumor; eosinophilic syndromes)	Nausea Periorbital edema Rare CHF QTc prolongation	Myelosuppression not frequent in solid tumor indications
Nilotinib	Bcr-Abl fusion protein (CML) and some imatinib-resistant variants	Interaction with CYP3A4-metabolized drugs CHF Hepatotoxicity Hypothyroidism	Chronic phase and in patients resistant to imatinib
Dasatinib	Bcr-Abl fusion protein (CML/ALL); wild-type and imatinib-resistant mutants	Myelosuppression (bleeding, infection) Pulmonary hypertension CHF Fluid retention QTc prolongation	Chronic phase and imatinib or nilotinib resistant
Bosutinib	Bcr-Abl fusion protein (CML); wild-type and imatinib-resistant mutants	Myelosuppression Hepatic QTc prolongation	Chronic phase and imatinib or nilotinib resistant
Ponatinib	T315I mutation of Bcr-Abl fusion protein (CML)	Clotting Hepatic CHF Pancreatitis Neuropathy Rash	
Gefitinib	First-line treatment of NSCLC with ATP site mutation of EGFR	Diarrhea Interstitial pneumonitis	In the United States, only with prior documented benefit in second-line treatment of NSCLC
Erlotinib	First-line treatment of NSCLC with ATP site mutation of EGFR; second-line treatment of wild-type EGFR NSCLC	Rash Diarrhea Rare interstitial pneumonitis	1 h before, 2 h after meals
Afatinib	First-line treatment of NSCLC with ATP site mutation of EGFR	Diarrhea Cutaneous	Interacts with Pgp inhibitors
Crizotinib	EML4-Alk fusion protein	Interstitial pneumonitis Hepatic QTc prolongation Bradycardia	
Vemurafenib	BRAF V600E in melanoma	Nausea Rash Cutaneous	
Dabrafenib	BRAF V600E in melanoma	Second cutaneous neoplasms Cutaneous	
Trametinib	BRAF V600E in melanoma (both as single agent and in combination with dabrafenib)	Second cutaneous neoplasms Rash Diarrhea Lymphedema	In combination with dabrafenib, second neoplasms, hemorrhage, venous thrombosis, CHF, ocular, hyperglycemia
DRUG	INDICATION	ADVERSE EVENTS	NOTES
Diagnostically Guided Retinoid			
Tretinoin	APL t(15,17)	Teratogenic Cutaneous	APL differentiation syndrome: pulmonary dysfunction/infiltrate, pleural/pericardial effusion, fever
Multikinase Inhibitors			
Sorafenib	Renal cell, hepatocellular, differentiated thyroid carcinoma	Diarrhea Hand-foot syndrome Other rash Hypertension CHF	Targets c-raf, VEGFR
Pazopanib	Renal cell carcinoma, soft tissue sarcoma	Fatigue Diarrhea/GI Hypertension Thromboses QTc	Target VEGFR, c-kit, PDGFR

(Continued)

TABLE 69-5 Molecularly Targeted Agents (Continued)

DRUG	INDICATION	ADVERSE EVENTS	NOTES
Regorafenib	Second-line colorectal cancer; GI stromal tumor	Hypertension Hand-foot syndrome Thromboses Perforations	VEGFR/TIE2
Sunitinib	Renal cell carcinoma, pancreatic neuroendocrine tumor, GI stromal tumor	Fatigue Diarrhea Neutropenia	Target VEGFR
Vandetanib	Medullary thyroid cancer	Diarrhea Rash Hypertension Prolonged QTc Thromboses	Target VEGFR, ret, EGFR
Cabozantinib	Medullary thyroid cancer	Hypertension Wound healing Fistulas Osteonecrosis Proteinuria	Target VEGFR, c-met
Axitinib	Renal cell carcinoma, second line	Diarrhea/other GI Fatigue	Target VEGFR, PDGFR, c-kit
Osimertinib	Non-small cell lung cancer, EGFR T790M mutation	Hand-foot syndrome Interstitial lung disease QTc prolongation Cardiomyopathy	
Proteasome Inhibitors			
Bortezomib	Multiple myeloma, mantle cell lymphoma	Neuropathy Thrombocytopenia GI	
Carfilzomib	Multiple myeloma, second line	Infusion reaction CHF Thrombocytopenia Pulmonary Tumor lysis	
Histone Deacetylase Inhibitors			
Vorinostat	Cutaneous T-cell lymphoma, second line	Fatigue Diarrhea Thrombocytopenia Embolism	
Romidepsin	Cutaneous T-cell lymphoma, second line	Nausea Vomiting Cytopenias Cardiac conduction	
mTOR Inhibitors			
Temsirolimus	Renal cell carcinoma, second line or poor prognosis	Stomatitis Thrombocytopenia Nausea Anorexia, fatigue Metabolic (glucose, lipid)	
Everolimus	Renal cell carcinoma, advanced; subependymal giant-cell astrocytoma; breast cancer, hormone receptor positive, resistant to antiestrogen; pancreatic neuroendocrine	Stomatitis Fatigue	
Miscellaneous			
Arsenic trioxide	APL	↑ QT _c	APL differentiation syndrome (see under tretinoin)
Vismodegib	Metastatic basal cell carcinoma	GI Hair loss Fatigue Muscle spasm Dysgeusia	Target smoothened receptor in hedgehog pathway

Abbreviations: APL, acute promyelocytic leukemia; ALL, acute lymphocytic leukemia; CHF, congestive heart failure; CML, chronic myeloid leukemia; EGFR, epidermal growth factor receptor; GI, gastrointestinal; mTOR, mammalian target of rapamycin kinase; NSCLC, non-small-cell lung cancer; PDGFR, platelet-derived growth factor receptor; Pgp, P-glycoprotein; VEGFR, vascular endothelial growth factor receptor.

evidence of a cytokine release syndrome, which can be a cardiopulmonary stress. Other agents active in multiple myeloma and certain other hematologic neoplasms include the immunomodulatory agents related to thalidomide, including lenalidomide and pomalidomide. All these agents collectively inhibit aberrant angiogenesis in the bone marrow microenvironment, as well as influence stromal cell immune functions to alter the cytokine milieu supporting the growth of myeloma cells. Thalidomide, although clinically active, has prominent cytopenic, neuropathic, procoagulant, and CNS toxicities that have been somewhat attenuated in the other drugs of the class, although use of these agents frequently entails concomitant anticoagulant prophylaxis.

Ibrutinib and idelalisib are representative of novel classes of inhibitors directed at Bruton's tyrosine kinase and phosphatidyl inositide-3 kinase- δ , respectively, expressed in normal and neoplastic B cells. Initially approved for use in mantle cell lymphoma and chronic lymphocytic leukemia, respectively, they are potentially applicable to a number of B-cell neoplasms that depend on signals through the B-cell antigen receptor. Janus kinases likewise function downstream of a variety of cytokine receptors to amplify cytokine signals, and Janus kinase inhibitors including ruxolitinib have approved activity in myelofibrosis to ameliorate splenomegaly and systemic symptoms.

Vorinostat is an inhibitor of histone deacetylases, which are responsible for maintaining the proper orientation of histones on DNA, with resulting capacity for transcriptional readiness. Acetylated histones allow access of transcription factors to target genes and therefore increase expression of genes that are selectively repressed in tumors. The result can be differentiation with the emergence of a more normal cellular phenotype, or cell cycle arrest with expression of endogenous regulators of cell cycle progression. Vorinostat is approved for clinical use in cutaneous T-cell lymphoma, with dramatic skin clearing and very few side effects. Romidepsin is a distinct molecular class of histone deacetylase inhibitor also active in cutaneous T-cell lymphoma. Panobinostat has activity in multiple myeloma. DNA methyltransferase inhibitors, including 5-aza-cytidine and 2'-deoxy-5-azacytidine (decitabine), can also increase transcription of genes "silenced" during the pathogenesis of a tumor by causing demethylation of the methylated cytosines that are acquired as an "epigenetic" (i.e., after the DNA is replicated) modification of DNA. These drugs were originally considered antimetabolites but have clinical value in myelodysplastic syndromes and certain leukemias when administered at low doses.

Additional toxicities with several therapies affecting oncogene-activated pathways include poorly predicted hepatic and cardiac toxicities (imatinib, dasatinib, sorafenib, pazopanib) or cardiac conduction deficits including prolonged QT interval (pazopanib), and atrial fibrillation (ibrutinib). The occurrence of new cardiac or liver abnormalities in a patient receiving treatment with a protein kinase antagonist should lead to a consideration of the risk versus benefit and the possible relation of the agent to the new adverse event. The existence of prior cardiac dysfunction is a relative contraindication to the use of certain targeted therapies (e.g., trastuzumab), although each patient's needs should be individualized.

CANCER BIOLOGIC THERAPY

Principles The goal of biologic therapy is to manipulate the host-tumor interaction in favor of the host, potentially at an optimum biologic dose that might be different than the MTD. As a class, biologic therapies may be distinguished from molecularly targeted agents in that many biologic therapies require an active response (e.g., reexpression of silenced genes or antigen expression) on the part of the tumor cell or on the part of the host (e.g., immunologic effects) to allow therapeutic effect. This may be contrasted with the more narrowly defined antiproliferative or apoptotic response that is the ultimate goal of molecularly targeted agents discussed above. However, there is much commonality in the strategies to evaluate and use molecularly targeted and biologic therapies.

Antibody-Mediated Therapeutic Approaches In general, antibodies are not very effective at killing cancer cells. Because the tumor seems to influence the host toward making antibodies rather

than generating cellular immunity, it is inferred that antibodies are easier for the tumor to fend off. Many patients can be shown to have serum antibodies directed at their tumors, but these do not appear to influence disease progression. However, the ability to grow very large quantities of high-affinity antibody directed at a tumor has led to the application of antibodies in the treatment of cancer. In this approach, antibodies are derived where the antigen-combining regions are grafted onto human immunoglobulin gene products (chimerized or humanized) or derived de novo from mice bearing human immunoglobulin gene loci. Three general strategies have emerged using antibodies. *Tumor-regulatory antibodies* target tumor cells directly or indirectly to modulate intracellular functions or attract immune or stromal cells. *Immunoregulatory antibodies* target antigens expressed on the tumor cells or host immune cells to modulate primarily the host's immune responsiveness to the tumor. Finally, *antibody conjugates* can be made with the antibody linked to drugs, toxins, or radioisotopes to target these "warheads" for delivery to the tumor. Table 69-6 lists features of currently used or promising antibodies for cancer treatment.

TUMOR-REGULATORY ANTIBODIES Humanized antibodies against the CD20 molecule expressed on B-cell lymphomas (rituximab and ofatumumab) are exemplary of antibodies that affect both signaling events driving lymphomagenesis as well as activating immune responses against B-cell neoplasms. They are used as single agents and in combination with chemotherapy and radiation in the treatment of B-cell neoplasms. Obinutuzumab is an antibody with an altered glycosylation that enhances its ability to activate killer cells; it is also directed against CD20 and is of value in chronic lymphocytic leukemia. It seems to be more effective in this setting than rituximab.

The HER2/neu receptor overexpressed on epithelial cancers, especially breast cancer, was initially targeted by trastuzumab, with noteworthy activity in potentiating the action of chemotherapy in breast cancer as well as some evidence of single-agent activity. Trastuzumab also appears to interrupt intracellular signals derived from HER2/neu and to stimulate immune mechanisms. The anti-HER2 antibody pertuzumab, specifically targeting the domain of HER2/neu responsible for dimerization with other HER2 family members, is more specifically directed against HER2 signaling function and augments the action of trastuzumab.

EGF receptor (EGFR)-directed antibodies (such as cetuximab and panitumumab) have activity in colorectal cancer refractory to chemotherapy, particularly when used to augment the activity of an additional chemotherapy program, and in the primary treatment of head and neck cancers treated with radiation therapy. The mechanism of action is unclear. Direct effects on the tumor may mediate an antiproliferative effect as well as stimulate the participation of host mechanisms involving immune cell or complement-mediated response to tumor cell-bound antibody. Alternatively, the antibody may alter the release of paracrine factors promoting tumor cell survival.

The anti-VEGF antibody bevacizumab shows little evidence of antitumor effect when used alone, but when combined with chemotherapeutic agents, it improves tumor shrinkage and time to disease progression in colorectal and nonsquamous lung cancers. The mechanism for the effect is unclear and may relate to the capacity of the antibody to alter delivery and tumor uptake of the active chemotherapeutic agent. Ziv-aflibercept is not an antibody, but a solubilized VEGF receptor VEGF binding domain, and therefore may have a distinct mechanism of action with comparable side effects.

Unintended side effects of any antibody use include infusion-related hypersensitivity reactions, usually limited to the first infusion, which can be managed with glucocorticoid and/or antihistamine prophylaxis. In addition, distinct syndromes have emerged with different antibodies. Anti-EGFR antibodies produce an acneiform rash that poorly responds to glucocorticoid cream treatment. Trastuzumab (anti-HER2) can inhibit cardiac function, particularly in patients with prior exposure to anthracyclines. Bevacizumab has a number of side effects of medical significance, including hypertension, thrombosis, proteinuria, hemorrhage, and gastrointestinal perforations with or without prior surgeries; these adverse events also occur with small-molecule drugs modulating VEGFR function.

TABLE 69-6 Antibodies Used in Cancer Treatment

DRUG	TARGET	INDICATIONS AND FEATURES OF USE
Tumor Regulatory Antibodies		
Rituximab	CD20	B cell neoplasms (also emerging role in autoimmune disease); chimeric antibody with frequent mouse-derived sequences; frequent infusion reactions, particularly on initial doses; reactivation of infections, particularly hepatitis; progressive multifocal leukoencephalopathy; tumor lysis syndrome
Ofatumumab	CD20	active in CLL; fully human antibody with distinct binding site compared to rituximab; decreased intensity infusion reactions
Trastuzumab	HER2/neu	Active in breast cancer and GI cancers expressing HER2/neu; cardiotoxicity, particularly in setting of prior anthracyclines, requires monitoring; infusion reactions
Pertuzumab	HER2/neu	Breast cancer; targets distinct binding site from trastuzumab, inhibiting dimerization of HER2 family members; infusion reactions; cardiac toxicity
Cetuximab	EGFR	Colorectal cancers with wild-type Ki-ras oncogene; head and neck cancers with radiation; rash, diarrhea, infusion reactions
Panitumumab	EGFR	Colorectal cancers with wild-type Ki-ras oncogene; fully humanized; decreased infusion reactions; different IgG subtype than cetuximab
Bevacizumab	VEGF	Metastatic colorectal cancer and non-small-cell lung cancer (nonsquamous) with chemotherapy; renal cancer and glioblastoma as single agents; prominent HBP, proteinuria, GI perforations, hemorrhage, thrombosis (venous and arterial)
Daratumumab	CD38	Multiple myeloma
Elotuzumab	CD319	Multiple myeloma, with revlimid and dexamethasone
Olaratumab	PDGF-R	Soft tissue sarcoma, in conjunction with doxorubicin
Immunoregulatory Antibodies		
Alemtuzumab	CD52	CLL, T-cell lymphomas; activates complement after binding to cell surface; infusion reactions, hypersensitivity, tumor lysis, activation of infections, cytopenias
Ipilimumab	CTLA4	Melanoma; inhibits the negative proliferative signal to T cells acting through CTLA4, resulting in prominent T cell activation; side effects include immune-mediated toxicity to liver, skin, pituitary, gut, which if severe calls for steroids, which inhibit antineoplastic effect
Pembrolizumab	PD-1	Non-small cell lung cancer as a first- or second-line treatment if PDL1(+) and no actionable mutations; and as a second-line treatment for head and neck squamous cell carcinoma, after platinum-based chemotherapy; can cause immune-related colitis, hepatitis, hypophysitis, nephritis, and altered thyroid function; also consider steroids for treatment of severe adverse events
Nivolumab	PD-1	Metastatic melanoma in combination with ipilimumab if BRAF mutation negative; melanoma following treatment with ipilimumab and after a BRAF inhibitor if relevant; second-line treatment for squamous non-small cell lung cancer, renal cancer and for relapsed and refractory Hodgkin's Disease; side effects similar to pembrolizumab
Atezolizumab	PD-L1	Locally advanced or metastatic urothelial carcinoma treatment after failure of chemo- or radiotherapy; metastatic non-small cell lung cancer (NSCLC) whose disease progressed during or following platinum-containing chemotherapy, without actionable mutations

Abbreviations: CLL, chronic lymphocytic leukemia; EGFR, epidermal growth factor receptor; GI, gastrointestinal; HBP, high blood pressure; VEGF, vascular endothelial growth factor.

IMMUNOREGULATORY ANTIBODIES Purely immunoregulatory antibodies stimulate immune responses to mediate tumor-directed cytotoxicity. First-generation approaches sought to activate complement and are exemplified by alemtuzumab against CD52; these are active in chronic lymphoid leukemia and T-cell malignancies. A more refined understanding of the tumor-host interface has defined that cytotoxic tumor-directed T cells are frequently inhibited by ligands upregulated in the tumor cells. The programmed death ligand 1 (PD-L1; also known as B7-homolog 1) was initially recognized as an entity that induced T cell death through a receptor present on T cells, termed the PD receptor (**Fig. 69-5**), which physiologically exists to regulate the intensity of the immune response. The PD family of ligands and receptors also regulates macrophage function, present in tumor stroma. These actions raised the hypothesis that antibodies directed against the PD signaling axis (both anti-PD-L1 and anti-PD) might be useful in cancer treatment by allowing reactivation of the immune response against tumors. Nivolumab, directed against the PD-1 receptor, is approved for use in renal cancer, metastatic melanoma, and non-small cell lung cancer, as well as in relapsed Hodgkin's disease. Pembrolizumab is approved for first-line treatment of metastatic non-small cell lung cancer whose tumors express the PD-L1 ligand. This development was a milestone in cancer therapeutics, replacing chemotherapy in this patient subset.

Ipilimumab, an antibody directed against the anti-CTLA4 (cytotoxic T lymphocyte antigen 4), which is expressed on T cells (not tumor cells), responds to signals from antigen-presenting cells (**Fig. 69-5**), and also

downregulates the intensity of the T-cell proliferative response to antigens derived from tumor cells. Indeed, manipulation of the CTLA4 axis was the first demonstration that purely immunoregulatory antibody strategies directed at T-cell physiology could be safe and effective in the treatment of cancer, although it acts at a very early stage in T-cell activation and can be considered somewhat nonspecific in its basis for T-cell stimulation. Ipilimumab alone or in combination with PD1-directed antibodies, is approved for initial treatment of metastatic melanoma.

Prominent activation of autoimmune hepatic, endocrine, cutaneous, neurologic, and gastrointestinal responses is a basis for adverse events with the use of ipilimumab and the PD-1-directed antibodies; the emergent use of glucocorticoids may be required to attenuate severe toxicities, which unfortunately can theoretically attenuate antitumor effect. Importantly for the general internist, these events may occur late after exposure to ipilimumab while the patient may otherwise be enjoying sustained control of tumor growth owing to the beneficial actions of ipilimumab.

Another class of immunoregulatory antibody is the "bispecific" antibody blinatumomab, which was constructed to have an anti-CD19 antigen combining site as one valency of an antibody with anti-CD3 binding site as the other valency. This antibody thus can bring T cells (with its anti-CD3 activity) close to B cells bearing the CD19 determinant. Blinatumomab is active in B-cell neoplasms such as acute lymphocytic leukemia, which may not have prominent expression of the CD20 targeted by rituximab.

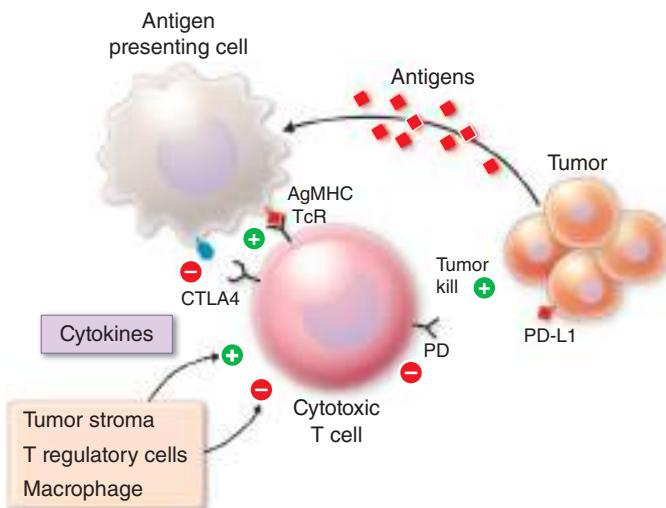


FIGURE 69-5 Tumors possess a microenvironment (tumor stroma) with immune cells including both helper T cells, suppressor T cells (both “regulatory” of other immune cell function), macrophages, and cytotoxic T cells. Cytokines found in the stroma and deriving from macrophages and regulatory T cells modulate the activities of cytotoxic T cells, which have the potential to kill tumor cells. Antigens released by tumor cells are taken up by Antigen Presenting Cells (APCs), also in the stroma. Antigens are processed by the APCs to peptides presented by the Major Histocompatibility Complex to T-cell antigen receptors, thus providing an (+) activation signal for the cytotoxic tumor cells to kill tumor cells bearing that antigen. Negative (−) signals inhibiting cytotoxic T cell action include the CTLA4 receptor (on T cells), interacting with the B7 family of negative regulatory signals from APCs, and the PD receptor (on T cells), interacting with the PD-L1 (−) signal coming from tumor cells expressing the PD-1 ligand (PD-L1). As both CTLA4 and PD1 signals attenuate the anti-tumor T cell response, strategies which inhibit CTLA4 and PD1 function are a means of stimulating cytotoxic T cell activity to kill tumor cells. Cytokines from other immune cells and macrophages can provide both (+) and (−) signals for T cell action, and are under investigation as novel immunoregulatory therapeutics.

ANTIBODY CONJUGATES Conjugates of antibodies with drugs and isotopes have also been shown to be effective in the treatment of cancer and have the intent of increasing the therapeutic index of the drug or isotope by delivering the toxic “warhead” directly to the tumor cell or tumor microenvironment. Ado-trastuzumab is a conjugate of the HER2/neu-directed trastuzumab and a highly toxic microtubule targeted drug (emtansine), which by itself is too toxic for human use; the antibody-drug conjugate shows valuable activity in patients with breast cancer who have developed resistance to the “naked” antibody. Brentuximab vedotin is an anti-CD30 antibody drug conjugate with a distinct microtubule poison with activity in neoplasms such as Hodgkin’s lymphoma where the tumor cells frequently express CD30. Radioconjugates targeting CD20 on lymphomas have been approved for use (ibritumomab tiuxetan [Zevalin], using yttrium-90 or ^{131}I -sotyamomab). Toxicity concerns have limited their use.

Cytokines Only IFN- α and interleukin 2 (IL-2)-related molecules are in routine clinical use. The two recombinant interferons commercially available are IFN- α 2a and α 2b. IFN is not curative for any tumor but can induce partial responses in follicular lymphoma, hairy cell leukemia, CML, melanoma, and Kaposi’s sarcoma. It has been used in the adjuvant setting in stage II melanoma, multiple myeloma, and follicular lymphoma. It produces fever, fatigue, a flulike syndrome, malaise, myelosuppression, and depression and can induce clinically significant autoimmune disease.

IL-2 exerts its antitumor effects indirectly through augmentation of immune function. Its biologic activity is to promote the growth and activity of T cells and natural killer (NK) cells. High doses of IL-2 can produce tumor regression in certain patients with metastatic melanoma and renal cell cancer. About 2–5% of patients may experience complete remissions that are durable, unlike any other treatment for these tumors. IL-2 is associated with intravascular volume depletion, capillary leak syndrome, adult respiratory distress syndrome, hypotension, fever, chills, skin rash, and impaired renal and liver function. Patients

may require blood pressure support and intensive care to manage the toxicity. However, once the agent is stopped, most of the toxicities reverse completely within 3–6 days. IL2 has been fused to translate in frame with a fragment of diphtheria toxin. A commercially available construct has activity against certain T-cell lymphomas. The drug’s utility derives from the internalization of the targeted receptor and cleavage of the active drug or toxin moiety.

Immune Cell-Mediated Therapies Tumors have a variety of means of avoiding the immune system: (1) they are often only subtly different from their normal counterparts; (2) they are capable of downregulating their major histocompatibility complex antigens, effectively masking them from recognition by T cells; (3) they are inefficient at presenting antigens to the immune system; (4) they can cloak themselves in a protective shell of fibrin to minimize contact with surveillance mechanisms; and (5) they can produce a range of soluble molecules, including potential immune targets, that can distract the immune system from recognizing the tumor cell or can kill or inactivate the immune effector cells. Prominent mediators of this effect are the PD receptors and their ligands described above. Some of the cell products initially polarize the immune response away from cellular immunity (shifting from $T_{H}1$ to $T_{H}2$ responses; [Chap. 342](#)) and ultimately lead to defects in T cells that prevent their activation and cytotoxic activity. A variety of strategies are being tested to overcome these barriers.

Cell-Mediated Immunity The strongest evidence that the immune system can exert clinically meaningful antitumor effects comes from allogeneic bone marrow transplantation. Adoptively transferred T cells from the donor expand in the tumor-bearing host, recognize the tumor as being foreign, and can mediate impressive anti-tumor effects (graft-versus-tumor effects). Three types of experimental interventions are being developed to take advantage of the ability of T cells to kill tumor cells.

1. *Transfer of allogeneic T cells.* This occurs in three major settings: in allogeneic bone marrow transplantation; as purified lymphocyte transfusions following bone marrow recovery after allogeneic bone marrow transplantation; and as pure lymphocyte transfusions following immunosuppressive (nonmyeloablative) therapy (also called reduced intensity or minitransplants). In each of these settings, the effector cells are donor T cells that recognize the tumor as being foreign, probably through minor histocompatibility differences. The main risk of such therapy is the development of graft-versus-host disease because of the minimal difference between the cancer and the normal host cells. This approach has been highly effective in certain hematologic cancers.
2. *Transfer of autologous T cells.* In this approach, the patient’s own T cells are removed from the tumor-bearing host, manipulated in several ways in vitro, and given back to the patient. There are three major classes of autologous T-cell manipulation. First, tumor antigen-specific T cells can be developed and expanded to large numbers over many weeks ex vivo before administration. Second, the patient’s T cells can be activated by exposure to polyclonal stimulators such as anti-CD3 and anti-CD28 after a short period ex vivo, and then amplified in the host after transfer by stimulation with IL-2, for example. Short periods removed from the patient permit the cells to overcome the tumor-induced T-cell defects, and such cells traffic and home to sites of disease better than cells that have been in culture for many weeks. In a third approach, genes that encode for a T-cell receptor specific for an antigen expressed by the tumor along with genes that facilitate T-cell activation can be introduced into subsets of a patient’s T cells, which, after transfer back into the patient, allow homing of cytotoxic T cells to tumor cells expressing the antigen.
3. *Tumor vaccines aimed at boosting T-cell immunity.* The finding that mutant oncogenes that are expressed only intracellularly can be recognized as targets of T cell killing greatly expanded the possibilities for tumor vaccine development. However, major difficulties remain in getting the tumor-specific peptides presented in a fashion to prime the T cells. Tumors themselves are very poor at presenting their own antigens to T cells at the first antigen exposure (*priming*).

Priming is best accomplished by professional antigen-presenting cells (dendritic cells). Thus, a number of experimental strategies are aimed at priming host T cells against tumor-associated peptides. Vaccine adjuvants such as granulocyte-macrophage colony-stimulating factor (GM-CSF) appear capable of attracting antigen-presenting cells to a skin site containing a tumor antigen. Purified antigen-presenting cells can be pulsed with tumor, its membranes, or particular tumor antigens and delivered as a vaccine. One such vaccine, Sipuleucel-T, is approved for use in patients with hormone-independent prostate cancer. In this approach, the patient undergoes leukapheresis, wherein mononuclear cells (that include antigen-presenting cells) are removed from the patient's blood. The cells are pulsed in a laboratory with an antigenic fusion protein comprising a protein frequently expressed by prostate cancer cells, prostate acid phosphatase, fused to GM-CSF, and matured to increase their capacity to present the antigen to immune effector cells. The cells are then returned to the patient in a well-tolerated treatment. Although no objective tumor response was documented in clinical trials, median survival was increased by about 4 months. Tumor cells can also be transfected with genes that attract antigen-presenting cells.

Another important vaccine strategy is directed at infectious agents whose action ultimately is tied to the development of human cancer. Hepatitis B vaccine in an epidemiologic sense prevents hepatocellular carcinoma, and a tetravalent human papillomavirus vaccine prevents infection by virus types currently accounting for 70% of cervical cancer. Unfortunately, these vaccines are ineffective at treating patients who have developed a virus-induced cancer.

■ SYSTEMIC RADIATION THERAPY

Although total-body irradiation has a role in preparing a patient to receive allogeneic stem cells, and antibodies as described above can specifically target radioisotopes, systemically administered isotopes of iodide salts have an important role in the treatment of thyroid neoplasms, owing to the selective upregulation of the iodide transporter in the tumor cell compartment. Likewise, isotopes of samarium and radium have been found useful in the palliation of symptoms from advanced bony metastases of prostate cancer owing to their selective deposition at the tumor–bone matrix interface, thereby potentially affecting the function of both tumor and stromal cells in the progressive growth of the metastatic deposit.

RESISTANCE TO CANCER TREATMENTS

Resistance mechanisms to the conventional cytotoxic agents were initially characterized in the late twentieth century as defects in drug uptake, metabolism, or export by tumor cells. The *multidrug resistance* (*mdr*) gene defined in vitro in cell lines exposed to increasing concentrations of drugs led to the definition of a family of transport proteins that efficiently excrete the drug from the tumor cells; no clinically useful modulator of this process has yet emerged. Drug-metabolizing enzymes such as cytidine deaminase are upregulated in resistant tumor cells, and this is the basis for so-called “high-dose cytarabine” regimens in the treatment of leukemia. Another resistance mechanism defined during this era involved increased expression of a drug’s target, exemplified by amplification of the dihydrofolate reductase gene, in patients who had lost responsiveness to methotrexate, or mutation of topoisomerase II in tumors that relapsed after topoisomerase II modulator treatment.

A second class of resistance mechanisms involves loss of the cellular apoptotic mechanism activated after the engagement of a drug’s target by the drug. This occurs in a way that is heavily influenced by the biology of the particular tumor type. For example, decreased alkylguanine alkyltransferase expression defines a subset of glioblastoma patients with the prospect of enhanced benefit from treatment with temozolamide, but has no predictive value for benefit from temozolamide in epithelial neoplasms. Likewise, ovarian cancers resistant to platinating agents have decreased expression of the proapoptotic gene *bax*. These types of findings have prompted the idea that responsive tumors to chemotherapeutic agents are populated by cells that express

drug-related cell death controlling genes, creating in effect a state of “synthetic lethality” with the drug (Chap. 68). When drug is not present, absence or mutation in these genes is tolerated, but become lethal in the presence of the drug.

A third class of resistance mechanisms emerged from sequencing of the targets of agents directed at oncogenic kinases. Thus, patients with CML resistant to imatinib have acquired mutations in the ATP binding domain of p210^{bcr-abl} in some cases, leading to the screening and design of agents with activity against the mutant proteins. Entirely analogous resistance mechanisms have emerged in patients with lung cancer treated with the EGFR antagonists gefitinib and erlotinib.

A final category of tumor resistance mechanisms to targeted agents includes the upregulation of alternate means of activating the pathway targeted by the agent. Thus melanomas initially responsive to *BRAF* V600E antagonists such as vemurafenib may reactivate raf signaling by upregulating isoforms that can bypass the variant blocked by the drug. Likewise, inhibition of HER2/neu signaling in breast cancer cells can lead to the emergence of variants with distinct oncogenic signaling pathways such as PI3 kinase. The susceptibility of a tumor to different treatments as a function of its expression of potential drug targets or their mutational profile has led to efforts to define the dominant pathways driving a patient’s tumor by genomic techniques including whole exome sequencing. The difficulty with applying such data to patient treatment is recognizing that these pathways may change during the natural history of a tumor and that different sites in a single patient may have tumors with different patterns of gene mutation.

SUPPORTIVE CARE DURING CANCER TREATMENT

■ MYELOSUPPRESSION

The common cytotoxic chemotherapeutic agents almost invariably affect bone marrow function. Titration of this effect determines the tolerated dose of the agent on a given schedule. The normal kinetics of blood cell turnover influences the sequence and sensitivity of each of the formed elements. Polymorphonuclear leukocytes (PMNs; $t_{1/2} = 6\text{--}8\text{ h}$), platelets ($t_{1/2} = 5\text{--}7\text{ days}$), and red blood cells (RBCs; $t_{1/2} = 120\text{ days}$) have most, less, and least susceptibility, respectively, to usually administered cytotoxic agents. The nadir count of each cell type in response to classes of agents is characteristic. Maximal neutropenia occurs 6–14 days after conventional doses of anthracyclines, antifolates, and antimetabolites. Alkylating agents differ from each other in the timing of cytopenias. Nitrosoureas, DTIC, and procarbazine can display delayed marrow toxicity, first appearing 6 weeks after dosing.

Complications of myelosuppression result from the predictable sequelae of the missing cells’ function. *Febrile neutropenia* refers to the clinical presentation of fever (one temperature $\geq 38.5^\circ\text{C}$ or three readings $\geq 38^\circ\text{C}$ but $\leq 38.5^\circ\text{C}$ per 24 h) in a neutropenic patient with an uncontrolled neoplasm involving the bone marrow or, more usually, in a patient undergoing treatment with cytotoxic agents. Mortality from uncontrolled infection varies inversely with the neutrophil count. If the nadir neutrophil count is $>1000/\mu\text{L}$, there is little risk; if $<500/\mu\text{L}$, risk of death is markedly increased. Management of febrile neutropenia has conventionally included empirical coverage with antibiotics for the duration of neutropenia (Chap. 70). Selection of antibiotics is governed by the expected association of infections with certain underlying neoplasms; careful physical examination (with scrutiny of catheter sites, dentition, mucosal surfaces, and perirectal and genital orifices by gentle palpation); chest x-ray; and Gram stain and culture of blood, urine, and sputum (if any) to define a putative site of infection. In the absence of any originating site, a broadly acting β -lactam with anti-*Pseudomonas* activity, such as ceftazidime, is begun empirically. The addition of vancomycin to cover potential cutaneous sites of origin (until these are ruled out or shown to originate from methicillin-sensitive organisms) or metronidazole or imipenem for abdominal or other sites favoring anaerobes reflects modifications tailored to individual patient presentations. Febrile neutropenic patients can be stratified broadly into two prognostic groups. The first, with expected short duration of neutropenia and no evidence of hypotension or abdominal or other localizing

symptoms, may be expected to do well even with oral regimens, e.g., ciprofloxacin or moxifloxacin, or amoxicillin plus clavulanic acid. A less favorable prognostic group is patients with expected prolonged neutropenia, evidence of sepsis, and end organ compromise, particularly pneumonia. Empirical addition of antifungal agents if fever and neutropenia persist for 7 days without identification of an adequately treated organism or site is frequent.

Transfusion of granulocytes has no role in the management of febrile neutropenia, owing to their exceedingly short half-life, mechanical fragility, and clinical syndromes of pulmonary compromise with leukostasis after their use. Instead, colony-stimulating factors (CSFs) are used to augment bone marrow production of PMNs. The American Society of Clinical Oncology has developed practice guidelines for the use of G-CSF and GM-CSF (Table 69-7).

Primary prophylaxis (i.e., shortly after completing chemotherapy to reduce the nadir) administers G-CSF to patients receiving cytotoxic regimens associated with a 20% incidence of febrile neutropenia. "Dose-dense" regimens, where cycling of chemotherapy is intended to be completed without delay of administered doses, may also benefit, but such patients should be on a clinical trial. Administration

of G-CSF in these circumstances has reduced the incidence of febrile neutropenia in several studies by about 50%. Most patients, however, receive regimens that do not have such a high risk of expected febrile neutropenia, and therefore most patients initially should not receive G-CSF or GM-CSF. Special circumstances—such as a documented history of febrile neutropenia with the regimen in a particular patient or categories of patients at increased risk, such as patients aged >65 years with aggressive lymphoma treated with curative chemotherapy regimens; extensive compromise of marrow by prior radiation or chemotherapy; or active, open wounds or deep-seated infection—may support primary treatment with G-CSF or GM-CSF. Administration of G-CSF or GM-CSF to afebrile neutropenic patients or to patients with low-risk febrile neutropenia is not recommended, and patients receiving concomitant chemoradiation treatment, particularly those with thoracic neoplasms, likewise are not generally recommended for treatment. In contrast, administration of G-CSF to high-risk patients with febrile neutropenia and evidence of organ compromise including sepsis syndrome, invasive fungal infection, concurrent hospitalization at the time fever develops, pneumonia, profound neutropenia ($<0.1 \times 10^9/L$), or age >65 years is reasonable.

Secondary prophylaxis refers to the administration of CSFs in patients who have experienced a neutropenic complication from a prior cycle of chemotherapy; dose reduction or delay may be a reasonably considered alternative. G-CSF or GM-CSF is conventionally started 24–72 h after completion of chemotherapy and continued until a PMN count of $10,000/\mu L$ is achieved, unless a "depot" preparation of G-CSF such as pegfilgrastim is used, where one dose is administered at least 14 days before the next scheduled administration of chemotherapy. Also, patients with myeloid leukemias undergoing induction therapy may have a slight reduction in the duration of neutropenia if G-CSF is commenced after completion of therapy, but the influence on long-term outcome has not been defined. GM-CSF probably has a more restricted utility than G-CSF, with its use currently limited to patients after autologous bone marrow transplants, although proper head-to-head comparisons with G-CSF have not been conducted in most instances. GM-CSF may be associated with more systemic side effects.

Dangerous degrees of thrombocytopenia do not frequently complicate the management of patients with solid tumors receiving cytotoxic chemotherapy (with the possible exception of certain carboplatin-containing regimens), but they are frequent in patients with certain hematologic neoplasms where marrow is infiltrated with tumor. Severe bleeding related to thrombocytopenia occurs with increased frequency at platelet counts $<20,000/\mu L$ and is very prevalent at counts $<5000/\mu L$.

The precise "trigger" point at which to transfuse patients has been defined as a platelet count of $10,000/\mu L$ or less in patients without medical comorbidities that may increase the risk of bleeding. This issue is important not only because of the costs of frequent transfusion, but unnecessary platelet transfusions expose the patient to the risks of alloimmunization and loss of value from subsequent transfusion owing to rapid platelet clearance, as well as the infectious and hypersensitivity risks inherent in any transfusion. Prophylactic transfusions to keep platelets $>20,000/\mu L$ are reasonable in patients with leukemia who are stressed by fever or concomitant medical conditions (the threshold for transfusion is $10,000/\mu L$ in patients with solid tumors and no other bleeding diathesis or physiologic stressors such as fever or hypotension, a level that might also be reasonably considered for leukemia patients who are thrombocytopenic but not stressed or bleeding). Careful review of medication lists to prevent exposure to nonsteroidal anti-inflammatory agents and maintenance of clotting factor levels adequate to support near-normal prothrombin and partial thromboplastin time tests are important in minimizing the risk of bleeding in the thrombocytopenic patient.

Anemia associated with chemotherapy can be managed by transfusion of packed RBCs. Transfusion is not undertaken until the hemoglobin falls to $<80\text{ g/L}$ (8 g/dL), compromise of end organ function occurs, or an underlying condition (e.g., coronary artery disease) calls for maintenance of hemoglobin $>90\text{ g/L}$ (9 g/dL). Randomized trials in certain tumors have raised the possibility that erythropoietin (EPO) use may promote tumor-related adverse events.

TABLE 69-7 Indications for the Clinical Use of G-CSF or GM-CSF

Preventive Uses

- With the first cycle of chemotherapy (so-called *primary CSF administration*)
 - Not needed on a routine basis
 - Use if the probability of febrile neutropenia is $\geq 20\%$
 - Use if patient has preexisting neutropenia or active infection
 - Age >65 years treated for lymphoma with curative intent or other tumors treated by similar regimens
 - Poor performance status
 - Extensive prior chemotherapy
 - Dose-dense regimens in a clinical trial or with strong evidence of benefit
- With subsequent cycles if febrile neutropenia has previously occurred (so-called *secondary CSF administration*)
 - Not needed after short-duration neutropenia without fever
 - Use if patient had febrile neutropenia in previous cycle
 - Use if prolonged neutropenia (even without fever) delays therapy

Therapeutic Uses

- Afebrile neutropenic patients
 - No evidence of benefit
- Febrile neutropenic patients
 - No evidence of benefit
 - May feel compelled to use in the face of clinical deterioration from sepsis, pneumonia, or fungal infection, but benefit unclear
- In bone marrow or peripheral blood stem cell transplantation
 - Use to mobilize stem cells from marrow
 - Use to hasten myeloid recovery
- In acute myeloid leukemia
 - G-CSF of minor or no benefit
 - GM-CSF of no benefit and may be harmful
- In myelodysplastic syndromes
 - Not routinely beneficial
 - Use intermittently in subset with neutropenia and recurrent infection

What Dose and Schedule Should Be Used?

- G-CSF: 5 mg/kg per day subcutaneously
- GM-CSF: 250 mg/m² per day subcutaneously
- Pegfilgrastim: one dose of 6 mg 24 h after chemotherapy

When Should Therapy Begin and End?

- When indicated, start 24–72 h after chemotherapy
- Continue until absolute neutrophil count is $10,000/\mu L$
- Do not use concurrently with chemotherapy or radiation therapy

Abbreviations: CSF, cerebrospinal fluid; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor.

Source: From the American Society of Clinical Oncology: J Clin Oncol 24:3187, 2006.

■ NAUSEA AND VOMITING

The most common side effect of chemotherapy administration is nausea, with or without vomiting. Nausea may be acute (within 24 h of chemotherapy), delayed (>24 h), or anticipatory of the receipt of chemotherapy. Patients may be likewise stratified for their risk of susceptibility to nausea and vomiting, with increased risk in young, female, heavily pretreated patients without a history of alcohol or drug use but with a history of motion or morning sickness. Antineoplastic agents vary in their capacity to cause nausea and vomiting. Highly emetogenic drugs (>90%) include mechlorethamine, streptozotocin, DTIC, cyclophosphamide at >1500 mg/m², and cisplatin; moderately emetogenic drugs (30–90% risk) include carboplatin, cytosine arabinoside (>1 mg/m²), ifosfamide, conventional-dose cyclophosphamide, and anthracyclines; low-risk (10–30%) agents include 5FU, taxanes, etoposide, and bortezomib, with minimal risk (<10%) afforded by treatment with antibody-drugs, bleomycin, busulfan, fludarabine, and vinca alkaloids.

Serotonin antagonists (5-HT₃) and neurokinin 1 (NK1) receptor antagonists are useful in “high-risk” chemotherapy regimens. The combination acts at both peripheral gastrointestinal and CNS sites that control nausea and vomiting. For example, the 5-HT₃ blocker dolasetron, 100 mg intravenously or orally; dexamethasone, 12 mg; and the NK1 antagonist aprepitant, 125 mg orally are combined on the day of administration of severely emetogenic regimens, with repetition of dexamethasone (8 mg) and aprepitant (80 mg) on days 2 and 3 for delayed nausea. Alternate 5-HT₃ antagonists include ondansetron, given as 0.15 mg/kg intravenously for three doses just before and at 4 and 8 h after chemotherapy; palonosetron at 0.25 mg over 30 s, 30 min before chemotherapy; and granisetron, given as a single dose of 0.01 mg/kg just before chemotherapy. Emesis from moderately emetic chemotherapy regimens may be prevented with a 5-HT₃ antagonist and dexamethasone alone for patients not receiving doxorubicin and cyclophosphamide combinations; the latter combination requires the 5-HT₃/dexamethasone/aprepitant on day 1, but aprepitant alone on days 2 and 3. Emesis from low-emetic-risk regimens may be prevented with 8 mg of dexamethasone alone or with non-5-HT₃ non-NK1 antagonist approaches including the following.

Antidopaminergic phenothiazines act directly at the chemoreceptor trigger zone (CTZ) in the brainstem medulla and include prochlorperazine (Compazine), 10 mg intramuscularly or intravenously, 10–25 mg orally, or 25 mg per rectum every 4–6 h for up to four doses; and thiethylperazine, 10 mg by potentially all of the above routes every 6 h. Haloperidol is a butyrophенone dopamine antagonist given at 1 mg intramuscularly or orally every 8 h. Metoclopramide acts on peripheral dopamine receptors to augment gastric emptying and is used in high doses for highly emetogenic regimens (1–2 mg/kg intravenously 30 min before chemotherapy and every 2 h for up to three additional doses as needed); intravenous doses of 10–20 mg every 4–6 h as needed or 50 mg orally 4 h before and 8 and 12 h after chemotherapy are used for moderately emetogenic regimens. 5-9-Tetrahydrocannabinol (Marinol) is a rather weak antiemetic compared to other available agents, but it may be useful for persisting nausea and is used orally at 10 mg every 3–4 h as needed.

■ DIARRHEA

Regimens that include 5FU infusions and/or irinotecan may produce severe diarrhea. Similar to the vomiting syndromes, chemotherapy-induced diarrhea may be immediate or can occur in a delayed fashion up to 48–72 h after the drugs. Careful attention to maintained hydration and electrolyte repletion, intravenously if necessary, along with antimotility treatments such as “high-dose” loperamide, commenced with 4 mg at the first occurrence of diarrhea, with 2 mg repeated every 2 h until 12 h without loose stools, not to exceed a total daily dose of 16 mg. Octreotide (100–150 µg), a somatostatin analogue, or opiate-based preparations may be considered for patients not responding to loperamide.

■ MUCOSITIS

Irritation and inflammation of the mucous membranes particularly afflicting the oral and anal mucosa, but potentially involving the gastrointestinal tract, may accompany cytotoxic chemotherapy. Mucositis

is due to damage to the proliferating cells at the base of the mucosal squamous epithelia or in the intestinal crypts. Topical therapies, including anesthetics and barrier-creating preparations, may provide symptomatic relief in mild cases. Palifermin or keratinocyte growth factor, a member of the fibroblast growth factor family, is effective in preventing severe mucositis in the setting of high-dose chemotherapy with stem cell transplantation for hematologic malignancies. It may also prevent or ameliorate mucositis from radiation.

■ ALOPECIA

Chemotherapeutic agents vary widely in causing alopecia, with anthracyclines, alkylating agents, and topoisomerase inhibitors reliably causing near-total alopecia when given at therapeutic doses. Antimetabolites are more variably associated with alopecia. Psychological support and the use of cosmetic resources are to be encouraged, and “chemo caps” that reduce scalp temperature to decrease the degree of alopecia are controversial during treatment with curative intent of neoplasms, such as leukemia or lymphoma, or in adjuvant breast cancer therapy. The richly vascularized scalp can certainly harbor micrometastatic or disseminated disease.

■ GONADAL DYSFUNCTION AND PREGNANCY

Cessation of ovulation and azoospermia reliably result from alkylating agent—and topoisomerase poison-containing regimens. The duration of these effects varies with age and sex. Sperm banking before treatment may be considered. Females experience amenorrhea with anovulation after alkylating agent therapy; egg preservation may be considered, but may delay inception of urgent treatment. Recovery of normal menses is frequent if treatment is completed before age 30 but unlikely to recover menses after age 35. Even those who regain menses usually experience premature menopause. Because the magnitude and extent of decreased fertility can be difficult to predict, patients should be counseled to maintain effective contraception, preferably by barrier means, during and after therapy. Resumption of efforts to conceive should be considered in the context of the patient’s likely prognosis. Hormone replacement therapy should be undertaken in women who do not have a hormonally responsive tumor. For patients who have had a hormone-sensitive tumor primarily treated by a local modality, conventional practice would counsel against hormone replacement, but this issue is under investigation.

Chemotherapy agents have variable effects on the success of pregnancy. All agents tend to have increased risk of adverse outcomes when administered during the first trimester, and strategies to delay chemotherapy, if possible, until after this milestone should be considered if the pregnancy is to continue to term. Patients in their second or third trimester can be treated with most regimens for the common neoplasms afflicting women in their childbearing years, with the exception of antimetabolites, particularly antifolates, which have notable teratogenic or fetotoxic effects throughout pregnancy. The need for anticancer chemotherapy per se is infrequently a clear basis to recommend termination of a concurrent pregnancy, although each treatment strategy in this circumstance must be tailored to the individual needs of the patient.

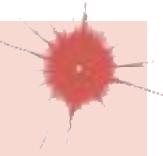
Late effects of cancer and its treatment are reviewed in Chap. 91.

■ FURTHER READING

- JAYSON GC et al: Antiangiogenic therapy in oncology: Current status and future directions. *Lancet* 388:518, 2016.
- MAUS MV et al: Antibody-modified T cells: CARs take the front seat for hematologic malignancies. *Blood* 123:2625, 2014.
- ROSENBERG SA, RESTIFO NP: Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 348:62, 2015.
- SOCINSKI MA, VILLARUZ LC, Ross J: Understanding mechanisms of resistance in the epithelial growth factor receptor in non-small cell lung cancer and the role of biopsy at progression. *Oncologist* 22:3, 2017.
- SWANTON C, GOVINDAN R: Clinical implications of genomic discoveries in lung cancer. *N Engl J Med* 374:1864, 2016.
- TOPALIAN SL et al: Immune checkpoint blockade: A common denominator approach to cancer therapy. *Cancer Cell* 27:450, 2015.

70

Infections in Patients with Cancer



Robert W. Finberg

Infections are a common cause of death and an even more common cause of morbidity in patients with a wide variety of neoplasms. Autopsy studies show that most deaths from acute leukemia and half of deaths from lymphoma are caused directly by infection. With more intensive chemotherapy, patients with solid tumors have also become more likely to die of infection. Fortunately, an evolving approach to prevention and treatment of infectious complications of cancer has decreased infection-associated mortality rates and will probably continue to do so. This accomplishment has resulted from three major steps:

- 1. Early treatment:** The practice of using “early empirical” antibiotics reduced mortality rates among patients with leukemia and bacteremia from 84% in 1965 to 44% in 1972. The mortality rate due to infection in febrile neutropenic patients dropped to <10% by 2013. This dramatic improvement is attributed to early intervention with appropriate antimicrobial therapy.
- 2. Empirical treatment:** “Empirical” antifungal therapy has also lowered the incidence of disseminated fungal infection, with dramatic decreases in mortality rates. An antifungal agent is administered—on the basis of likely fungal infection—to neutropenic patients who, after 4–7 days of antibiotic therapy, remain febrile but have no positive cultures.
- 3. Prophylaxis:** Use of antibiotics for afebrile neutropenic patients as broad-spectrum prophylaxis against infections has decreased both mortality and morbidity even further. The current approach to treatment of severely neutropenic patients (e.g., those receiving high-dose chemotherapy for leukemia or high-grade lymphoma) is based on initial prophylactic therapy at the onset of neutropenia, subsequent “empirical” antibacterial therapy targeting the organisms whose involvement is likely in light of physical findings (most often fever alone), and finally “empirical” antifungal therapy based on the known likelihood that fungal infection will become a serious issue after 4–7 days of broad-spectrum antibacterial therapy.

A physical predisposition to infection in patients with cancer (**Table 70-1**) can be a result of the neoplasm’s production of a break

in the skin. For example, a squamous cell carcinoma may cause local invasion of the epidermis, which allows bacteria to gain access to subcutaneous tissue and permits the development of cellulitis. The artificial closing of a normally patent orifice can also predispose to infection; for example, obstruction of a ureter by a tumor can cause urinary tract infection, and obstruction of the bile duct can cause cholangitis. Part of the host’s normal defense against infection depends on the continuous emptying of a viscous; without emptying, a few bacteria that are present as a result of bacteremia or local transit can multiply and cause disease.

A similar problem can affect patients whose lymph node integrity has been disrupted by radical surgery, particularly patients who have had radical node dissections. A common clinical problem following radical mastectomy is the development of cellulitis (usually caused by streptococci or staphylococci) because of lymphedema and/or inadequate lymph drainage. In most cases, this problem can be addressed by local measures designed to prevent fluid accumulation and breaks in the skin, but antibiotic prophylaxis has been used in refractory cases.

A life-threatening problem common to many cancer patients is the loss of the reticuloendothelial capacity to clear microorganisms after splenectomy, which may be performed as part of the management of hairy cell leukemia, chronic lymphocytic leukemia (CLL), and chronic myelogenous leukemia (CML) and in Hodgkin’s disease. Even after curative therapy for the underlying disease, the lack of a spleen predisposes such patients to rapidly fatal infections. The loss of the spleen through trauma similarly predisposes the normal host to overwhelming infection throughout life. The splenectomized patient should be counseled about the risks of infection with certain organisms, such as the protozoan *Babesia* (Chap. 220) and *Capnocytophaga canimorsus*, a bacterium carried in the mouths of animals (Chaps. 136 and 153). Because encapsulated bacteria (*Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*) are the organisms most commonly associated with postsplenectomy sepsis, splenectomized persons should be vaccinated (and revaccinated; Table 70-2 and Chap. 118) against the capsular polysaccharides of these organisms. Many clinicians recommend giving splenectomized patients a small supply of antibiotics effective against *S. pneumoniae*, *N. meningitidis*, and *H. influenzae* to avert rapid, overwhelming sepsis in the event that they cannot present for medical attention immediately after the onset of fever or other signs or symptoms of bacterial infection. A few tablets of amoxicillin/clavulanic acid (or levofloxacin if resistant strains of *S. pneumoniae* are prevalent locally) are a reasonable choice for this purpose.

The level of suspicion of infections with certain organisms should depend on the type of cancer diagnosed (Table 70-3). Diagnosis of multiple myeloma or CLL should alert the clinician to the possibility

TABLE 70-1 Disruption of Normal Barriers in Patients with Cancer That May Predispose Them to Infections

TYPE OF DEFENSE	SPECIFIC LESION	CELLS INVOLVED	ORGANISM	CANCER ASSOCIATION	DISEASE
Physical barrier	Breaks in skin	Skin epithelial cells	Staphylococci, streptococci	Head and neck, squamous cell carcinoma	Cellulitis, extensive skin infection
Emptying of fluid collections	Occlusion of orifices: ureters, bile duct, colon	Luminal epithelial cells	Gram-negative bacilli	Renal, ovarian, biliary tree, metastatic diseases of many cancers	Rapid, overwhelming bacteremia; urinary tract infection
Lymphatic function	Node dissection	Lymph nodes	Staphylococci, streptococci	Breast cancer surgery	Cellulitis
Splenic clearance of microorganisms	Splenectomy	Splenic reticuloendothelial cells	<i>Streptococcus pneumoniae</i> , <i>Haemophilus influenzae</i> , <i>Neisseria meningitidis</i> , <i>Babesia</i> , <i>Capnocytophaga canimorsus</i>	Hodgkin’s disease, leukemia	Rapid, overwhelming sepsis
Phagocytosis	Lack of granulocytes	Granulocytes (neutrophils)	Staphylococci, streptococci, enteric organisms, fungi	Acute myeloid and acute lymphocytic leukemias, hairy cell leukemia	Bacteremia
Humoral immunity	Lack of antibody	B cells	<i>S. pneumoniae</i> , <i>H. influenzae</i> , <i>N. meningitidis</i>	Chronic lymphocytic leukemia, multiple myeloma	Infections with encapsulated organisms, sinusitis, pneumonia
Cellular immunity	Lack of T cells	T cells and macrophages	<i>Mycobacterium tuberculosis</i> , <i>Listeria</i> , herpesviruses, fungi, intracellular parasites	Hodgkin’s disease, leukemia, T cell lymphoma	Infections with intracellular bacteria, fungi, parasites; virus reactivation

TABLE 70-2 Vaccination of Cancer Patients Receiving Chemotherapy^a

VACCINE	USE IN INDICATED PATIENTS		
	INTENSIVE CHEMOTHERAPY	HODGKIN'S DISEASE	HEMATOPOIETIC STEM CELL TRANSPLANTATION
Diphtheria-tetanus-pertussis ^b	Primary series and boosters as necessary	No special recommendation	3 doses given 6–12 months after transplantation
Poliomyelitis ^c	Complete primary series and boosters	No special recommendation	3 doses given 6–12 months after transplantation
<i>Haemophilus influenzae</i> type b conjugate	Primary series and booster for children	Single dose for adults	3 doses given 6–12 months after transplantation (separated by 1 month)
Human papillomavirus (HPV)	HPV vaccine is approved for males and females 9–26 years of age. Check Centers for Disease Control and Prevention (CDC) website (www.cdc.gov/vaccines) for updated recommendations.	HPV vaccine is approved for males and females 9–26 years of age. Check CDC website (www.cdc.gov/vaccines) for updated recommendations.	HPV vaccine is approved for males and females 9–26 years of age. Check CDC website (www.cdc.gov/vaccines) for updated recommendations.
Hepatitis A	As indicated for normal hosts on the basis of occupation and lifestyle	As indicated for normal hosts on the basis of occupation and lifestyle	As indicated for normal hosts on the basis of occupation and lifestyle
Hepatitis B	Same as for normal hosts	As indicated for normal hosts on the basis of occupation and lifestyle	3 doses given 6–12 months after transplantation
Pneumococcal conjugate vaccine (PCV13) Pneumococcal polysaccharide vaccine (PPSV23) ^d	Finish series prior to chemotherapy if possible.	Patients with splenectomy should receive both PCV13 and PPSV23.	Three doses of PCV13, beginning 3–6 months after transplantation, are followed by a dose of PPSV23 at least 8 weeks later. A second PPSV23 dose can be given 5 years later.
Quadrivalent meningococcal vaccine ^e	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories. An additional dose can be given after 5 years.	Should be administered to splenectomized patients and to patients living in endemic areas, including college students in dormitories. An additional dose can be given after 5 years.
Meningococcal B vaccine	See above.	See above.	See above (see www.cdc.gov/vaccines for updated recommendations).
Influenza	Seasonal immunization	Seasonal immunization	Seasonal immunization (A seasonal dose is recommended and can be given as early as 4 months after transplantation; if given <6 months after transplantation, an additional dose is recommended.)
Measles/mumps/rubella	Contraindicated	Contraindicated during chemotherapy	After 24 months in patients without graft-versus-host disease
Varicella-zoster virus ^f	Contraindicated ^g	Contraindicated	Contraindicated (CDC recommends use on a case-by-case basis following reevaluation.)

^aThe latest recommendations by the Advisory Committee on Immunization Practices and the CDC guidelines can be found at www.cdc.gov/vaccines. ^bA single dose of TdAp (tetanus–diphtheria–acellular pertussis), followed by a booster dose of Td (tetanus–diphtheria) every 10 years, is recommended for adults. ^cLive-virus vaccine is contraindicated; inactivated vaccine should be used. ^dTwo types of vaccines are used to prevent pneumococcal disease. A conjugate vaccine active against 13 serotypes (13-valent pneumococcal conjugate vaccine, or PCV13) is currently administered in three separate doses to all children. A polysaccharide vaccine active against 23 serotypes (23-valent pneumococcal polysaccharide vaccine, or PPSV23) elicits titers of antibody lower than those achieved with the conjugate vaccine, and immunity may wane more rapidly. Because the ablative chemotherapy given to recipients of hematopoietic stem cell transplants (HSCTs) eradicates immunologic memory, revaccination is recommended for all such patients. Vaccination is much more effective once immunologic reconstitution has occurred; however, because of the need to prevent serious disease, pneumococcal vaccine should be administered 6–12 months after transplantation in most cases. Because PPSV23 includes serotypes not present in PCV13, HSCT recipients should receive a dose of PPSV23 at least 8 weeks after the last dose of PCV13. Although antibody titers from PPSV23 clearly decay, experience with multiple doses of PPSV23 is limited, as are data on the safety, toxicity, or efficacy of such a regimen. For this reason, the CDC currently recommends the administration of one additional dose of PPSV23 at least 5 years after the last dose to immunocompromised patients, including transplant recipients, as well as patients with Hodgkin's disease, multiple myeloma, lymphoma, or generalized malignancies. Beyond this single additional dose, further doses are not recommended at this time. ^eMeningococcal conjugate vaccine (MenACWY) is recommended for adults ≤55 years old, and meningococcal polysaccharide vaccine (MPSV4) is recommended for those ≥56 years old. ^fIncludes both varicella vaccine for children and zoster vaccine for adults. ^gContact the manufacturer for more information on use in children with acute lymphocytic leukemia.

of hypogammaglobulinemia. While immunoglobulin replacement therapy can be effective, in most cases prophylactic antibiotics are a cheaper, more convenient method of eliminating bacterial infections in CLL patients with hypogammaglobulinemia. Patients with acute lymphocytic leukemia (ALL), patients with non-Hodgkin's lymphoma, and all cancer patients treated with high-dose glucocorticoids (or glucocorticoid-containing chemotherapy regimens) should receive antibiotic prophylaxis for *Pneumocystis* infection (Table 70-3) for the duration of their chemotherapy. In addition to exhibiting susceptibility to certain infectious organisms, patients with cancer are likely to manifest their infections in characteristic ways. For example, fever—generally a sign of infection in normal hosts—continues to be a reliable indicator in neutropenic patients. In contrast, patients receiving glucocorticoids

and agents that impair T cell function and cytokine secretion may have serious infections in the absence of fever. Similarly, neutropenic patients commonly present with cellulitis without purulence and with pneumonia without sputum or even x-ray findings (see below).

The use of monoclonal antibodies that target B and T cells as well as drugs that interfere with lymphocyte signal transduction events is associated with reactivation of latent infections. The use of rituximab, the antibody to CD20 (a B cell surface protein), is associated with the development of reactivation tuberculosis as well as other latent viral infections, including hepatitis B and cytomegalovirus (CMV) infection. Like organ transplant recipients (Chap. 138), patients with latent bacterial disease (like tuberculosis) and latent viral disease (like herpes simplex or zoster) should be carefully monitored for reactivation disease.

TABLE 70-3 Infections Associated with Specific Types of Cancer

CANCER	UNDERLYING IMMUNE ABNORMALITY	ORGANISM(S) CAUSING INFECTION
Multiple myeloma	Hypogammaglobulinemia	<i>Streptococcus pneumoniae</i> , <i>Haemophilus influenzae</i> , <i>Neisseria meningitidis</i>
Chronic lymphocytic leukemia	Hypogammaglobulinemia	<i>S. pneumoniae</i> , <i>H. influenzae</i> , <i>N. meningitidis</i>
Acute myeloid or lymphocytic leukemia	Granulocytopenia, skin and mucous membrane lesions	Extracellular gram-positive and gram-negative bacteria, fungi
Hodgkin's disease	Abnormal T cell function	Intracellular pathogens (<i>Mycobacterium tuberculosis</i> , <i>Listeria</i> , <i>Salmonella</i> , <i>Cryptococcus</i> , <i>Mycobacterium avium</i>); herpesviruses
Non-Hodgkin's lymphoma and acute lymphocytic leukemia	Glucocorticoid chemotherapy, T and B cell dysfunction	<i>Pneumocystis</i>
Colon and rectal tumors	Local abnormalities ^a	<i>Streptococcus bovis</i> biotype 1 (bacteremia)
Hairy cell leukemia	Abnormal T cell function	Intracellular pathogens (<i>M. tuberculosis</i> , <i>Listeria</i> , <i>Cryptococcus</i> , <i>M. avium</i>)

^aThe reason for this association is not well defined.

SYSTEM-SPECIFIC SYNDROMES

SKIN-SPECIFIC SYNDROMES

Skin lesions are common in cancer patients, and the appearance of these lesions may permit the diagnosis of systemic bacterial or fungal infection. While cellulitis caused by skin organisms such as *Streptococcus* or *Staphylococcus* is common, neutropenic patients—that is, those with <500 functional polymorphonuclear leukocytes (PMNs)/μL—and patients with impaired blood or lymphatic drainage may develop infections with unusual organisms. Innocent-looking macules or papules may be the first sign of bacterial or fungal sepsis in immunocompromised patients (Fig. 70-1). In the neutropenic host, a macule progresses rapidly to ecthyma gangrenosum (see Fig. A1-34), a usually painless, round, necrotic lesion consisting of a central black or gray-black eschar with surrounding erythema. Ecthyma gangrenosum, which is located in nonpressure areas (as distinguished from necrotic lesions associated with lack of circulation), is often associated with *Pseudomonas aeruginosa* bacteremia (Chap. 159) but may be caused by other bacteria.

Candidemia (Chap. 211) is also associated with a variety of skin conditions (see Fig. A1-37) and commonly presents as a maculopapular rash. Punch biopsy of the skin may be the best method for diagnosis.

Cellulitis, an acute spreading inflammation of the skin, is most often caused by infection with group A *Streptococcus* or *Staphylococcus aureus*, virulent organisms normally found on the skin (Chap. 124). Although cellulitis tends to be circumscribed in normal hosts, it may spread rapidly in neutropenic patients. A tiny break in the skin may lead to spreading cellulitis, which is characterized by pain and erythema; in the affected patients, signs of infection (e.g., purulence) are often lacking. What might be a furuncle in a normal host may require amputation because of uncontrolled infection in a patient presenting with leukemia. A dramatic response to an infection that might be trivial in a normal host can mark the first sign of leukemia. Fortunately, granulocytopenic patients are likely to be infected with certain types of organisms (Table 70-4); thus the selection of an antibiotic regimen is somewhat easier than it might otherwise be (see “Antibacterial Therapy,” below). It is essential to recognize cellulitis early and to treat it aggressively. Patients who are neutropenic or who have previously received antibiotics for other reasons may develop cellulitis with unusual organisms (e.g., *Escherichia coli*, *Pseudomonas*, or fungi). Early treatment, even of innocent-looking lesions, is essential to prevent necrosis and loss of



A



B

FIGURE 70-1 **A.** Papules related to *Escherichia coli* bacteremia in a patient with acute lymphocytic leukemia. **B.** The same lesions on the following day.

TABLE 70-4 Organisms Likely to Cause Infections in Granulocytopenic Patients

Gram-Positive Coccii	
<i>Staphylococcus epidermidis</i>	<i>Staphylococcus aureus</i>
<i>Viridans Streptococcus</i>	<i>Enterococcus faecalis</i>
<i>Streptococcus pneumoniae</i>	
Gram-Negative Bacilli	
<i>Escherichia coli</i>	<i>Serratia</i> spp.
<i>Klebsiella</i> spp.	<i>Acinetobacter</i> spp. ^a
<i>Pseudomonas aeruginosa</i>	<i>Stenotrophomonas</i> spp.
<i>Enterobacter</i> spp.	<i>Citrobacter</i> spp.
Non-aeruginosa <i>Pseudomonas</i> spp. ^a	
Gram-Positive Bacilli	
Diphtheroids	JK bacillus ^a
Fungi	
<i>Candida</i> spp.	<i>Mucor/Rhizopus</i>
<i>Aspergillus</i> spp.	

^aOften associated with intravenous catheters.

tissue. Debridement to prevent spread may sometimes be necessary early in the course of disease, but it can often be performed after chemotherapy, when the PMN count increases.

Sweet syndrome, or *febrile neutrophilic dermatosis*, was originally described in women with elevated white blood cell (WBC) counts. The disease is characterized by the presence of leukocytes in the lower dermis, with edema of the papillary body. Ironically, this disease now is usually seen in neutropenic patients with cancer, most often in association with acute myeloid leukemia (AML) but also in association with a variety of other malignancies. Sweet syndrome usually presents as red or bluish-red papules or nodules that may coalesce and form sharply bordered plaques (see Fig. A1-40). The edema may suggest vesicles, but on palpation the lesions are solid, and vesicles probably never arise in this disease. The lesions are most common on the face, neck, and arms. On the legs, they may be confused with erythema nodosum (see Fig. A1-39). The development of lesions is often accompanied by high fevers and an elevated erythrocyte sedimentation rate. Both the lesions and the temperature elevation respond dramatically to glucocorticoid administration. Treatment begins with high doses of glucocorticoids (prednisone, 60 mg/d) followed by tapered doses over the next 2–3 weeks.

Data indicate that *erythema multiforme* (see Fig. A1-24) with mucous membrane involvement is often associated with herpes simplex virus (HSV) infection and is distinct from Stevens-Johnson syndrome, which is associated with drugs and tends to have a more widespread distribution. Because cancer patients are both immunosuppressed (and therefore susceptible to herpes infections) and heavily treated with drugs (and therefore subject to Stevens-Johnson syndrome [see Fig. A2-4]), both of these conditions are common in this population.

Cytokines, which are used as adjuvants or primary treatments for cancer, can themselves cause characteristic rashes, further complicating the differential diagnosis. This phenomenon is a particular problem in bone marrow transplant recipients (Chap. 138), who, in addition to having the usual chemotherapy-, antibiotic-, and cytokine-induced rashes, are plagued by graft-versus-host disease.

CATHETER-RELATED INFECTIONS

Because IV catheters are commonly used in cancer chemotherapy and are prone to cause infection (Chap. 137), they pose a major problem in the care of patients with cancer. Some catheter-associated infections can be treated with antibiotics, whereas in others the catheter must be removed (Table 70-5). If the patient has a “tunneled” catheter (which consists of an entrance site, a subcutaneous tunnel, and an exit site), a red streak over the subcutaneous part of the line (the tunnel) is grounds

for immediate device removal. Failure to remove catheters under these circumstances may result in extensive cellulitis and tissue necrosis.

More common than tunnel infections are exit-site infections, often with erythema around the area where the line penetrates the skin. Most authorities (Chap. 142) recommend treatment (usually with vancomycin) for an exit-site infection caused by coagulase-negative *Staphylococcus*. Treatment of coagulase-positive staphylococcal infection is associated with a poorer outcome, and it is advisable to remove the catheter if possible. Similarly, most clinicians remove catheters associated with infections due to *P. aeruginosa* and *Candida* species, because such infections are difficult to treat and bloodstream infections with these organisms are likely to be deadly. Catheter infections caused by *Burkholderia cepacia*, *Stenotrophomonas* species, *Agrobacterium* species, *Acinetobacter baumannii*, *Pseudomonas* species other than *aeruginosa*, and carbapenem-resistant Enterobacteriaceae are likely to be very difficult to eradicate with antibiotics alone. Similarly, isolation of *Bacillus*, *Corynebacterium*, and *Mycobacterium* species should prompt removal of the catheter.

GASTROINTESTINAL TRACT-SPECIFIC SYNDROMES

Upper Gastrointestinal Tract Disease • INFECTIONS OF THE MOUTH The oral cavity is rich in aerobic and anaerobic bacteria (Chap. 172) that normally live in a commensal relationship with the host. The antimetabolic effects of chemotherapy cause a breakdown of mucosal host defenses, leading to ulceration of the mouth and the potential for invasion by resident bacteria. Mouth ulcerations afflict most patients receiving cytotoxic chemotherapy and have been associated with viridans streptococcal bacteremia. *Candida* infections of the mouth are very common. Fluconazole is clearly effective in the treatment of both local infections (thrush) and systemic infections (esophagitis) due to *Candida albicans*. Other azoles (e.g., voriconazole) as well as echinocandins offer similar efficacy as well as activity against the fluconazole-resistant organisms that are associated with chronic fluconazole treatment (Chap. 211).

Noma (*cancrum oris*), commonly seen in malnourished children, is a penetrating disease of the soft and hard tissues of the mouth and adjacent sites, with resulting necrosis and gangrene. It has a counterpart in immunocompromised patients and is thought to be due to invasion of the tissues by *Bacteroides*, *Fusobacterium*, and other normal inhabitants of the mouth. Noma is associated with debility, poor oral hygiene, and immunosuppression.

Viruses, particularly HSV, are a prominent cause of morbidity in immunocompromised patients, in whom they are associated with

TABLE 70-5 Approach to Catheter Infections in Immunocompromised Patients

CLINICAL PRESENTATION OR ISOLATED PATHOGEN	CATHETER REMOVAL	ANTIBIOTICS	COMMENTS
Evidence of Infection, Negative Blood Cultures			
Exit-site erythema	Not necessary if infection responds to treatment	Usually, begin treatment for gram-positive cocci.	Coagulase-negative staphylococci are most common.
Tunnel-site erythema	Required	Treat for gram-positive cocci pending culture results.	Failure to remove the catheter may lead to necrosis of the involved area requiring skin grafts in the future.
Blood Culture-Positive Infections			
Coagulase-negative staphylococci	Line removal optimal but may be unnecessary if patient is clinically stable and responds to antibiotics	Usually, start with vancomycin. Linezolid, quinupristin/dalfopristin, and daptomycin are alternative agents.	If there are no contraindications to line removal, this course of action is optimal. If the line is removed, antibiotics may not be necessary.
Other gram-positive cocci (e.g., <i>Staphylococcus aureus</i> , <i>Enterococcus</i>); gram-positive rods (<i>Bacillus</i> , <i>Corynebacterium</i> spp.)	Recommended	Treat with antibiotics to which the organism is sensitive, with duration based on the clinical setting.	The incidence of metastatic infections following <i>S. aureus</i> infection and the difficulty of treating enterococcal infection make line removal the recommended course of action. In addition, gram-positive rods do not respond readily to antibiotics alone.
Gram-negative bacteria	Recommended	Use an agent to which the organism is shown to be sensitive.	Organisms like <i>Stenotrophomonas</i> , <i>Pseudomonas</i> , and <i>Burkholderia</i> are notoriously hard to treat, as are carbapenem-resistant organisms.
Fungi	Recommended	—	Fungal infections of catheters are extremely difficult to treat.

severe mucositis. The use of acyclovir, either prophylactically or therapeutically, is of value.

ESOPHAGEAL INFECTIONS The differential diagnosis of esophagitis (usually presenting as substernal chest pain upon swallowing) includes herpes simplex and candidiasis, both of which are readily treatable.

Lower Gastrointestinal Tract Disease Hepatic candidiasis (*Chap. 211*) results from seeding of the liver (usually from a gastrointestinal source) in neutropenic patients. It is most common among patients being treated for AML and usually presents symptomatically around the time neutropenia resolves. The characteristic picture is that of persistent fever unresponsive to antibiotics, abdominal pain and tenderness or nausea, and elevated serum levels of alkaline phosphatase in a patient with hematologic malignancy who has recently recovered from neutropenia. The diagnosis of this disease (which may present in an indolent manner and persist for several months) is based on the finding of yeasts or pseudohyphae in granulomatous lesions. Hepatic ultrasound or CT may reveal bull's-eye lesions. MRI scans reveal small lesions not visible by other imaging modalities. The pathology (a granulomatous response) and the timing (with resolution of neutropenia and an elevation in granulocyte count) suggest that the host response to *Candida* is an important component of the manifestations of disease. In many cases, although organisms are visible, cultures of biopsied material may be negative. The designation *hepatosplenic candidiasis* or *hepatic candidiasis* is a misnomer because the disease often involves the kidneys and other tissues; the term *chronic disseminated candidiasis* may be more appropriate. Because of the risk of bleeding with liver biopsy, diagnosis is often based on imaging studies (MRI, CT). Treatment should be directed to the causative agent (usually *C. albicans* but sometimes *Candida tropicalis* or other less common *Candida* species).

Typhlitis *Typhlitis* (also referred to as necrotizing colitis, neutropenic colitis, necrotizing enteropathy, ileocecal syndrome, and cecitis) is a clinical syndrome of fever and right-lower-quadrant (or generalized abdominal) tenderness in an immunosuppressed host. This syndrome is classically seen in neutropenic patients after chemotherapy with cytotoxic drugs. It may be more common among children than among adults and appears to be much more common among patients with AML or ALL than among those with other types of cancer. Physical examination reveals right-lower-quadrant tenderness, with or without rebound tenderness. Associated diarrhea (often bloody) is common, and the diagnosis can be confirmed by the finding of a thickened cecal wall on CT, MRI, or ultrasonography. Plain films may reveal a right-lower-quadrant mass, but CT with contrast or MRI is a much more sensitive means of diagnosis. Although surgery is sometimes attempted to avoid perforation from ischemia, most cases resolve with medical therapy alone. The disease is sometimes associated with positive blood cultures (which usually yield aerobic gram-negative bacilli), and therapy is recommended for a broad spectrum of bacteria (particularly gram-negative bacilli, which are likely to be found in the bowel flora).

Clostridium difficile-Induced Diarrhea Patients with cancer are predisposed to the development of *C. difficile* diarrhea (*Chap. 129*) as a consequence of chemotherapy alone. Thus, they may test positive for *C. difficile* even without receiving antibiotics. Obviously, such patients are also subject to *C. difficile*-induced diarrhea as a result of antibiotic pressure. *C. difficile* should always be considered as a possible cause of diarrhea in cancer patients who have received either chemotherapy or antibiotics. New approaches to treatment of *C. difficile*-induced diarrhea and to prevention of *C. difficile* expansion as part of the gut microbiota may make this disease less troublesome in the future.

CENTRAL NERVOUS SYSTEM-SPECIFIC SYNDROMES

Meningitis The presentation of meningitis in patients with lymphoma or CLL and in patients receiving chemotherapy (particularly with glucocorticoids) for solid tumors suggests a diagnosis of cryptococcal or listerial infection. As noted previously, splenectomized

TABLE 70-6 Differential Diagnosis of Central Nervous System Infections in Patients with Cancer

FINDINGS ON CT OR MRI	UNDERLYING PREDISPOSITION	
	PROLONGED NEUTROPENIA	DEFECTS IN CELLULAR IMMUNITY ^a
Mass lesions	<i>Aspergillus</i> , <i>Nocardia</i> , or <i>Cryptococcus</i> brain abscess	Toxoplasmosis, Epstein-Barr virus lymphoma (rare)
Diffuse encephalitis	Progressive multifocal leukoencephalopathy (JC virus)	Infection with varicella-zoster virus, cytomegalovirus, herpes simplex virus, human herpesvirus type 6, JC virus, <i>Listeria</i>

^aHigh-dose glucocorticoid therapy, cytotoxic chemotherapy.

patients are susceptible to rapid, overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Similarly, patients who are antibody-deficient (e.g., those with CLL, those who have received intensive chemotherapy, or those who have undergone bone marrow transplantation) are likely to have infections caused by these bacteria. Other cancer patients, however, because of their defective cellular immunity, are likely to be infected with other pathogens (Table 70-3). Central nervous system (CNS) tuberculosis should be considered, especially in patients from countries where tuberculosis is highly prevalent in the population.

Encephalitis The spectrum of disease resulting from viral encephalitis is expanded in immunocompromised patients. A predisposition to infections with intracellular organisms similar to those encountered in patients with AIDS (*Chap. 197*) is seen in cancer patients receiving (1) high-dose cytotoxic chemotherapy, (2) chemotherapy affecting T cell function (e.g., fludarabine), or (3) antibodies that eliminate T cells (e.g., anti-CD3, alemtuzumab, anti-CD52) or cytokine activity (anti-tumor necrosis factor agents or interleukin 1 receptor antagonists). Infection with varicella-zoster virus (VZV) has been associated with encephalitis that may be caused by VZV-related vasculitis. Chronic viral infections may also be associated with dementia and encephalitic presentations. A diagnosis of progressive multifocal leukoencephalopathy (*Chap. 133*) should be considered when a patient who has received chemotherapy (rituximab in particular) presents with dementia (*Table 70-6*). Other abnormalities of the CNS that may be confused with infection include normal-pressure hydrocephalus and vasculitis resulting from CNS irradiation. It may be possible to differentiate these conditions by MRI.

Brain Masses Mass lesions of the brain most often present as headache with or without fever or neurologic abnormalities. Infections associated with mass lesions may be caused by bacteria (particularly *Nocardia*), fungi (particularly *Cryptococcus* or *Aspergillus*), or parasites (*Toxoplasma*). Epstein-Barr virus (EBV)-associated lymphoma may also present as single—or sometimes multiple—mass lesions of the brain. A biopsy may be required for a definitive diagnosis.

PULMONARY INFECTIONS

Pneumonia (*Chap. 121*) in immunocompromised patients may be difficult to diagnose because conventional methods of diagnosis depend on the presence of neutrophils. Bacterial pneumonia in neutropenic patients may present without purulent sputum—or, in fact, without any sputum at all—and may not produce physical findings suggestive of chest consolidation (rales or egophony).

In granulocytopenic patients with persistent or recurrent fever, the chest x-ray pattern may help to localize an infection and thus to determine which investigative tests and procedures should be undertaken and which therapeutic options should be considered (*Table 70-7*). In this setting, a simple chest x-ray is a screening tool; because the impaired host response results in less evidence of consolidation or infiltration, high-resolution CT is recommended for the diagnosis of pulmonary infections. The difficulties encountered in the management of pulmonary infiltrates relate in part to the difficulties of performing diagnostic procedures on the patients involved. When platelet counts

TABLE 70-7 Differential Diagnosis of Chest Infiltrates in Immunocompromised Patients

INFILTRATE	CAUSE OF PNEUMONIA	
	INFECTIOUS	NONINFECTIOUS
Localized	Bacteria (including <i>Legionella</i> , mycobacteria)	Local hemorrhage or embolism, tumor
Nodular	Fungi (e.g., <i>Aspergillus</i> or <i>Mucor</i>), <i>Nocardia</i>	Recurrent tumor
Diffuse	Viruses (especially cytomegalovirus), <i>Chlamydia</i> , <i>Pneumocystis</i> , <i>Toxoplasma gondii</i> , mycobacteria	Congestive heart failure, radiation pneumonitis, drug-induced lung injury, lymphangitic spread of cancer

can be increased to adequate levels by transfusion, microscopic and microbiologic evaluation of the fluid obtained by endoscopic bronchial lavage is often diagnostic. Lavage fluid should be cultured for *Mycoplasma*, *Chlamydia*, *Legionella*, *Nocardia*, more common bacterial pathogens, fungi, and viruses. In addition, the possibility of *Pneumocystis* pneumonia should be considered, especially in patients with ALL or lymphoma who have not received prophylactic trimethoprim-sulfamethoxazole (TMP-SMX). The characteristics of the infiltrate may be helpful in decisions about further diagnostic and therapeutic maneuvers. Nodular infiltrates suggest fungal pneumonia (e.g., that caused by *Aspergillus* or *Mucor*). Such lesions may best be approached by visualized biopsy procedures. It is worth noting that while bacterial pneumonias classically present as lobar infiltrates in normal hosts, bacterial pneumonias in granulocytopenic hosts present with a paucity of signs, symptoms, or radiographic abnormalities; thus, the diagnosis is difficult.

Aspergillus species (Chap. 212) can colonize the skin and respiratory tract or cause fatal systemic illness. Although this fungus may cause aspergillomas in a previously existing cavity or may produce allergic bronchopulmonary disease in some patients, the major problem posed by this genus in neutropenic patients is invasive disease, primarily due to *Aspergillus fumigatus* or *Aspergillus flavus*. The organisms enter the host following colonization of the respiratory tract, with subsequent invasion of blood vessels. The disease is likely to present as a thrombotic or embolic event because of this ability of the fungi to invade blood vessels. The risk of infection with *Aspergillus* correlates directly with the duration of neutropenia. In prolonged neutropenia, positive surveillance cultures for nasopharyngeal colonization with *Aspergillus* may predict the development of disease.

Patients with *Aspergillus* infection often present with pleuritic chest pain and fever, which are sometimes accompanied by cough. Hemoptysis may be an ominous sign. Chest x-rays may reveal new focal infiltrates or nodules. Chest CT may reveal a characteristic halo consisting of a mass-like infiltrate surrounded by an area of low attenuation. The presence of a "crescent sign" on chest x-ray or chest CT, in which the mass progresses to central cavitation, is characteristic of invasive *Aspergillus* infection but may develop as the lesions are resolving.

In addition to causing pulmonary disease, *Aspergillus* may invade through the nose or palate, with deep sinus penetration. The appearance of a discolored area in the nasal passages or on the hard palate should prompt a search for invasive *Aspergillus*. This situation is likely to require surgical debridement. Catheter infections with *Aspergillus* usually require both removal of the catheter and antifungal therapy.

Diffuse interstitial infiltrates suggest viral, parasitic, or *Pneumocystis* pneumonia. If the patient has a diffuse interstitial pattern on chest x-ray, it may be reasonable, while considering invasive diagnostic procedures, to institute empirical treatment for *Pneumocystis* with TMP-SMX and for *Chlamydia*, *Mycoplasma*, and *Legionella* with a quinolone or azithromycin. Noninvasive procedures, such as staining of induced sputum smears for *Pneumocystis*, serum cryptococcal antigen tests, and urine testing for *Legionella* antigen, may be helpful. Serum galactomannan and β -D-glucan tests may be of value in diagnosing *Aspergillus* infection, but their utility is limited by their lack of sensitivity and specificity. The presence of an elevated level of β -D-glucan in the serum

of a patient being treated for cancer who is not receiving prophylaxis against *Pneumocystis* suggests the diagnosis of *Pneumocystis* pneumonia. Infections with viruses that cause only upper respiratory symptoms in immunocompetent hosts, such as respiratory syncytial virus (RSV), influenza viruses, and parainfluenza viruses, may be associated with fatal pneumonitis in immunocompromised hosts. CMV reactivation occurs in cancer patients receiving chemotherapy, but CMV pneumonia is most common among hematopoietic stem cell transplant (HSCT) recipients (Chap. 138). Polymerase chain reaction testing now allows rapid diagnosis of viral pneumonia, which can lead to treatment in some cases (e.g., influenza). Multiplex studies that can detect a wide array of viruses in the lung and upper respiratory tract are now available and will lead to specific diagnoses of viral pneumonias.

Bleomycin is the most common cause of chemotherapy-induced lung disease. Other causes include alkylating agents (such as cyclophosphamide, chlorambucil, and melphalan), nitrosoureas (carmustine [BCNU], lomustine [CCNU], and methyl-CCNU), busulfan, procarbazine, methotrexate, and hydroxyurea. Both infectious and noninfectious (drug- and/or radiation-induced) pneumonitis can cause fever and abnormalities on chest x-ray; thus, the differential diagnosis of an infiltrate in a patient receiving chemotherapy encompasses a broad range of conditions (Table 70-7). The treatment of radiation pneumonitis (which may respond dramatically to glucocorticoids) or drug-induced pneumonitis is different from that of infectious pneumonia, and a biopsy may be important in the diagnosis. Unfortunately, no definitive diagnosis can be made in ~30% of cases, even after bronchoscopy.

Open-lung biopsy is the gold standard of diagnostic techniques. Biopsy via a visualized thoracostomy can replace an open procedure in many cases. When a biopsy cannot be performed, empirical treatment can be undertaken; a quinolone or an erythromycin derivative (azithromycin) and TMP-SMX are used in the case of diffuse infiltrates, and an antifungal agent is administered in the case of nodular infiltrates. The risks should be weighed carefully in these cases. If inappropriate drugs are administered, empirical treatment may prove toxic or ineffective; either of these outcomes may be riskier than biopsy.

■ CARDIOVASCULAR INFECTIONS

Patients with Hodgkin's disease are prone to persistent infections by *Salmonella*, sometimes (and particularly often in elderly patients) affecting a vascular site. The use of IV catheters deliberately lodged in the right atrium is associated with a high incidence of bacterial endocarditis, presumably related to valve damage followed by bacteremia. Nonbacterial thrombotic endocarditis (marantic endocarditis) has been described in association with a variety of malignancies (most often solid tumors) and may follow bone marrow transplantation as well. The presentation of an embolic event with a new cardiac murmur suggests this diagnosis. Blood cultures are negative in this disease of unknown pathogenesis.

■ ENDOCRINE SYNDROMES

Infections of the endocrine system have been described in immunocompromised patients. *Candida* infection of the thyroid may be difficult to diagnose during the neutropenic period. It can be defined by indium-labeled WBC scans or gallium scans after neutrophil counts increase. CMV infection can cause adrenalitis with or without resulting adrenal insufficiency. The presentation of a sudden endocrine anomaly in an immunocompromised patient can be a sign of infection in the involved end organ.

■ MUSCULOSKELETAL INFECTIONS

Infection that is a consequence of vascular compromise, resulting in gangrene, can occur when a tumor restricts the blood supply to muscles, bones, or joints. The process of diagnosis and treatment of such infection is similar to that in normal hosts, with the following caveats:

1. *In terms of diagnosis*, a lack of physical findings resulting from a lack of granulocytes in the granulocytopenic patient should make the clinician more aggressive in obtaining tissue rather than more willing to rely on physical signs.

2. In terms of therapy, aggressive debridement of infected tissues may be required. However, it is usually difficult to operate on patients who have recently received chemotherapy, both because of a lack of platelets (which results in bleeding complications) and because of a lack of WBCs (which may lead to secondary infection). A blood culture positive for *Clostridium perfringens*—an organism commonly associated with gas gangrene—can have a number of meanings (Chap. 149). *Clostridium septicum* bacteremia is associated with the presence of an underlying malignancy. Bloodstream infections with intestinal organisms such as *Streptococcus bovis* biotype 1 and *C. perfringens* may arise spontaneously from lower gastrointestinal lesions (tumor or polyps); alternatively, these lesions may be harbingers of invasive disease. The clinical setting must be considered in order to define the appropriate treatment for each case.

■ RENAL AND URETERAL INFECTIONS

Infections of the urinary tract are common among patients whose ureteral excretion is compromised (Table 70-1). *Candida*, which has a predilection for the kidney, can invade either from the bloodstream or in a retrograde manner (via the ureters or bladder) in immunocompromised patients. The presence of “fungus balls” or persistent candiduria suggests invasive disease. Persistent funguria (with *Aspergillus* as well as *Candida*) should prompt a search for a nidus of infection in the kidney.

Certain viruses are typically seen only in immunosuppressed patients. BK virus (polyomavirus hominis 1) has been documented in the urine of bone marrow transplant recipients and, like adenovirus, may be associated with hemorrhagic cystitis.

■ ABNORMALITIES THAT PREDISPOSE TO INFECTION

(Table 70-1)

■ THE LYMPHOID SYSTEM

It is beyond the scope of this chapter to detail how all the immunologic abnormalities that result from cancer or from chemotherapy for cancer lead to infections. Disorders of the immune system are discussed in other sections of this book. As has been noted, patients with antibody deficiency are predisposed to overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Infections that result from the lack of a functional cellular immune system are described in Chap. 197. It is worth mentioning, however, that patients undergoing intensive chemotherapy for any form of cancer will have not only defects due to granulocytopenia but also lymphocyte dysfunction, which may be profound. Thus, these patients—especially those receiving glucocorticoid-containing regimens or drugs that inhibit either T cell activation (calcineurin inhibitors or drugs like fludarabine, which affect lymphocyte function) or cytokine induction—should be given prophylaxis for *Pneumocystis* pneumonia.

Patients receiving treatment that eliminates B cells (e.g., with anti-CD20 antibodies or rituximab) are especially vulnerable to intercurrent viral infections. The incidence of progressive multifocal leukoencephalopathy (caused by JC virus) is elevated among these patients.

■ THE HEMATOPOIETIC SYSTEM

 Initial studies in the 1960s revealed a dramatic increase in the incidence of infections (fatal and nonfatal) among cancer patients with a granulocyte count of $<500/\mu\text{L}$. The use of prophylactic antibacterial agents has reduced the number of bacterial infections, but 35–78% of febrile neutropenic patients being treated for hematologic malignancies develop infections at some time during chemotherapy. Aerobic pathogens (both gram-positive and gram-negative) predominate in all series, but the exact organisms isolated vary from center to center. Infections with anaerobic organisms are uncommon. Geographic patterns affect the types of fungi isolated. Tuberculosis and malaria are common causes of fever in the developing world and may present in this setting as well.

Neutropenic patients are unusually susceptible to infection with a wide variety of bacteria; thus, antibiotic therapy should be initiated

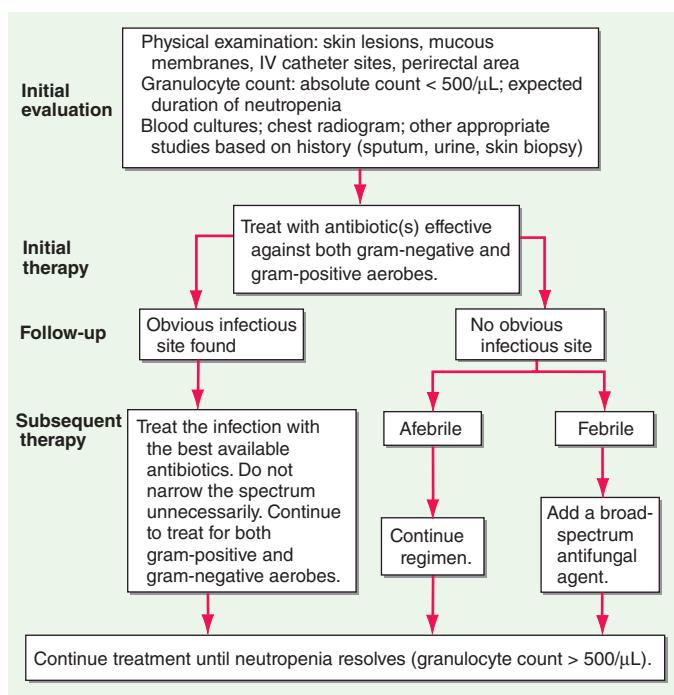


FIGURE 70-2 Algorithm for the diagnosis and treatment of fever and neutropenia.

promptly to cover likely pathogens if infection is suspected. Indeed, early initiation of antibacterial agents is mandatory to prevent deaths. Like most immunocompromised patients, neutropenic patients are threatened by their own microbial flora, including gram-positive and gram-negative organisms found commonly on the skin and mucous membranes and in the bowel (Table 70-4). Because treatment with narrow-spectrum agents leads to infection with organisms not covered by the antibiotics used, the initial regimen should target all pathogens likely to be the initial causes of bacterial infection in neutropenic hosts. As noted in the algorithm shown in Fig. 70-2, administration of antimicrobial agents is routinely continued until neutropenia resolves—that is, the granulocyte count is sustained above $500/\mu\text{L}$ for at least 2 days. Fever may not resolve prior to granulocyte recovery. In some cases, patients remain febrile after resolution of neutropenia. In these instances, the risk of sudden death from overwhelming bacteraemia is greatly reduced, and the following diagnoses should be seriously considered: (1) fungal infection, (2) bacterial abscesses or undrained foci of infection, and (3) drug fever (including reactions to antimicrobial agents as well as to chemotherapy or cytokines). In the proper setting, viral infection or graft-versus-host disease should be considered. In clinical practice, antibacterial therapy is usually discontinued when the patient is no longer neutropenic and all evidence of bacterial disease has been eliminated. Antifungal agents are then discontinued if there is no evidence of fungal disease. If the patient remains febrile, a search for viral diseases or unusual pathogens is conducted while unnecessary cytokines and other drugs are systematically eliminated from the regimen.

TREATMENT

Infections in Cancer Patients

ANTIBACTERIAL THERAPY

Hundreds of antibacterial regimens have been tested for use in patients with cancer. The major risk of infection is related to the degree of neutropenia seen as a consequence of either the disease or the therapy. Many of the relevant studies have involved small populations in which the outcomes have generally been good, and most have lacked the statistical power to detect differences among the regimens studied. Each febrile neutropenic patient should be approached as a unique problem, with particular attention given to

previous infections and recent antibiotic exposures. Several general guidelines are useful in the initial treatment of neutropenic patients with fever (Fig. 70-2):

1. In the initial regimen, it is necessary to use antibiotics active against both gram-negative and gram-positive bacteria (Table 70-4).
2. Monotherapy with an aminoglycoside or an antibiotic lacking good activity against gram-positive organisms (e.g., ciprofloxacin or aztreonam) is not adequate in this setting.
3. The agents used should reflect both the epidemiology and the antibiotic resistance pattern of the hospital.
4. If the pattern of resistance justifies its use, a single third-generation cephalosporin constitutes an appropriate initial regimen in many hospitals.
5. Most standard regimens are designed for patients who have not previously received prophylactic antibiotics. The development of fever in a patient who has received antibiotics affects the choice of subsequent therapy, which should target resistant organisms and organisms known to cause infections in patients being treated with the antibiotics already administered.
6. Randomized trials have indicated the safety of oral antibiotic regimens in the treatment of “low-risk” patients with fever and neutropenia. Outpatients who are expected to remain neutropenic for <10 days and who have no concurrent medical problems (such as hypotension, pulmonary compromise, or abdominal pain) can be classified as low risk and treated with a broad-spectrum oral regimen.
7. Several large-scale studies indicate that prophylaxis with a fluoroquinolone (ciprofloxacin or levofloxacin) decreases morbidity and mortality rates among afebrile patients who are anticipated to have neutropenia of long duration.

Commonly used antibiotic regimens for the treatment of febrile patients in whom prolonged neutropenia (>7 days) is anticipated include (1) ceftazidime or cefepime, (2) piperacillin/tazobactam, or (3) imipenem/cilastatin or meropenem. All three regimens have shown equal efficacy in large trials. All three are active against *P. aeruginosa* and a broad spectrum of aerobic gram-positive and gram-negative organisms. Imipenem/cilastatin has been associated with an elevated rate of *C. difficile* diarrhea, and many centers reserve carbapenem antibiotics for treatment of gram-negative bacteria that produce extended-spectrum β-lactamases; these limitations make carbapenems less attractive as an initial regimen. Despite the frequent involvement of coagulase-negative staphylococci, the initial use of vancomycin or its automatic addition to the initial regimen has not resulted in improved outcomes, and the antibiotic does exert toxic effects. For these reasons, only judicious use of vancomycin is recommended—for example, when there is good reason to suspect the involvement of coagulase-negative staphylococci (e.g., the appearance of erythema at the exit site of a catheter or a positive culture for methicillin-resistant *S. aureus* or coagulase-negative staphylococci). Because the sensitivities of bacteria vary from hospital to hospital, clinicians are advised to check their local sensitivities and to be aware that resistance patterns can change quickly, necessitating a change in approach to patients with fever and neutropenia. Similarly, infection control services should monitor for basic antibiotic resistance and for fungal infections. The appearance of a large number of *Aspergillus* infections, in particular, suggests the possibility of an environmental source that requires further investigation and remediation.

The initial antibacterial regimen should be refined on the basis of culture results (Fig. 70-2). Blood cultures are the most relevant basis for selection of therapy; surface cultures of skin and mucous membranes may be misleading. In the case of gram-positive bacteremia or another gram-positive infection, it is important that the antibiotic be optimal for the organism isolated. Once treatment with broad-spectrum antibiotics has begun, it is not desirable to discontinue all antibiotics because of the risk of failing to treat a potentially fatal bacterial infection; the addition of more and more antibacterial agents to the regimen is not appropriate unless there is a clinical

or microbiologic reason to do so. Planned progressive therapy (the serial, empirical addition of one drug after another without culture data) is not efficacious in most settings and may have unfortunate consequences. Simply adding another antibiotic for fear that a gram-negative infection is present is a dubious practice. The synergy exhibited by β-lactams and aminoglycosides against certain gram-negative organisms (especially *P. aeruginosa*) provides the rationale for using two antibiotics in this setting, but recent analyses suggest that efficacy is not enhanced by the addition of aminoglycosides, while toxicity may be increased. Mere “double coverage,” with the addition of a quinolone or another antibiotic that is not likely to exhibit synergy, has not been shown to be of benefit and may cause additional toxicities and side effects. Cephalosporins can cause bone marrow suppression, and vancomycin is associated with neutropenia in some healthy individuals. Furthermore, the addition of multiple cephalosporins may induce β-lactamase production by some organisms; cephalosporins and double β-lactam combinations should probably be avoided altogether in *Enterobacter* infections.

ANTIFUNGAL THERAPY

Fungal infections in cancer patients are most often associated with neutropenia. Neutropenic patients are predisposed to the development of invasive fungal infections, most commonly those due to *Candida* and *Aspergillus* species and occasionally those caused by *Mucor*, *Rhizopus*, *Fusarium*, *Trichosporon*, *Bipolaris*, and others. Cryptococcal infection, which is common among patients taking immunosuppressive agents, is uncommon among neutropenic patients receiving chemotherapy for AML. Invasive candidal disease is usually caused by *C. albicans* or *C. tropicalis* but can be caused by *C. krusei*, *C. parapsilosis*, and *C. glabrata*.

For decades, it has been common clinical practice to add amphotericin B to antibacterial regimens if a neutropenic patient remains febrile despite 4–7 days of treatment with antibacterial agents. The rationale for this empirical addition is that it is difficult to culture fungi before they cause disseminated disease and that mortality rates from disseminated fungal infections in granulocytopenic patients are high. Before the introduction of newer azoles into clinical practice, amphotericin B was the mainstay of antifungal therapy. The insolubility of amphotericin B has resulted in the marketing of several lipid formulations that are less toxic than the amphotericin B deoxycholate complex. Echinocandins (e.g., caspofungin) are useful in the treatment of infections caused by azole-resistant *Candida* strains as well as in therapy for aspergillosis and have been shown to be equivalent to liposomal amphotericin B for the empirical treatment of patients with prolonged fever and neutropenia. Newer azoles have also been demonstrated to be effective in this setting. Although fluconazole is efficacious in the treatment of infections due to many *Candida* species, its use against serious fungal infections in immunocompromised patients is limited by its narrow spectrum: it has no activity against *Aspergillus* or against several non-albicans *Candida* species. The broad-spectrum azoles (e.g., voriconazole and posaconazole) provide another option for the treatment of *Aspergillus* infections (Chap. 212), including CNS infection. Clinicians should be aware that the spectrum of each azole is somewhat different and that no drug can be assumed to be efficacious against all fungi. *Aspergillus terreus* is resistant to amphotericin B. Although voriconazole is active against *Pseudallescheria boydii*, amphotericin B is not; however, voriconazole has no activity against *Mucor*. Posaconazole, which is administered orally, is useful as a prophylactic agent in patients with prolonged neutropenia. Studies in progress are assessing the use of these agents in combinations. **For a full discussion of antifungal therapy, see Chap. 206.**

ANTIVIRAL THERAPY

The availability of a variety of agents active against herpes-group viruses, including some new agents with a broader spectrum of activity, has heightened focus on the treatment of viral infections, which pose a major problem in cancer patients. Viral diseases caused by the herpes group are prominent. Serious (and sometimes fatal) infections

due to HSV and VZV are well documented in patients receiving chemotherapy. CMV may also cause serious disease, but fatalities from CMV infection are more common in hematopoietic stem cell transplant recipients. The roles of human herpesvirus (HHV)-6, HHV-7, and HHV-8 (Kaposi's sarcoma-associated herpesvirus) in cancer patients are still being defined (*Chap. 190*). EBV lymphoproliferative disease (LPD) can occur in patients receiving chemotherapy but is much more common among transplant recipients (*Chap. 138*). While clinical experience is most extensive with acyclovir, which can be used therapeutically or prophylactically, a number of derivative drugs offer advantages over this agent (*Chap. 186*).

In addition to the herpes group, several respiratory viruses (especially RSV) may cause serious disease in cancer patients. Although influenza vaccination is recommended (see below), it may be ineffective in this patient population. The availability of antiviral drugs with activity against influenza viruses gives the clinician additional options for the prophylaxis and treatment of these patients (*Chaps. 186 and 195*).

OTHER THERAPEUTIC MODALITIES

Another way to address the problems posed by the febrile neutropenic patient is to replenish the neutrophil population. Although granulocyte transfusions may be effective in the treatment of refractory gram-negative bacteremia, they do not have a documented role in prophylaxis. Because of the expense, the risk of leukoagglutinin reactions (which has probably been decreased by improved cell-separation procedures), and the risk of transmission of CMV from unscreened donors (which has been reduced by the use of filters), granulocyte transfusion is reserved for patients whose condition is unresponsive to antibiotics. This modality is efficacious for documented gram-negative bacteremia refractory to antibiotics, particularly in situations where granulocyte numbers will be depressed for only a short period. The demonstrated usefulness of granulocyte colony-stimulating factor in mobilizing neutrophils and advances in preservation techniques may make this option more useful than in the past.

A variety of cytokines, including granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor, enhance granulocyte recovery after chemotherapy and consequently shorten the period of maximal vulnerability to fatal infections. The role of these cytokines in routine practice is still a matter of some debate. Most authorities recommend their use only when neutropenia is both severe and prolonged, and they should be used only in the appropriate setting (i.e., when stem cells are likely to be responsive) and not as an adjunct to antimicrobial agents. The cytokines themselves may have adverse effects, including fever, hypoxemia, and pleural effusions or serositis in other areas (*Chap. 342*).

Once neutropenia has resolved, the risk of infection decreases dramatically. However, depending on what drugs they receive, patients who continue on chemotherapeutic protocols remain at high risk for certain diseases. Any patient receiving more than a maintenance dose of glucocorticoids (e.g., in many treatment regimens for diffuse lymphoma) should also receive prophylactic TMP-SMX because of the risk of *Pneumocystis* infection; those with ALL should receive such prophylaxis for the duration of chemotherapy.

PREVENTION OF INFECTION IN CANCER PATIENTS

EFFECT OF THE ENVIRONMENT

Outbreaks of fatal *Aspergillus* infection have been associated with construction projects and materials in several hospitals. The association between spore counts and risk of infection suggests the need for a high-efficiency air-handling system in hospitals that care for large numbers of neutropenic patients. The use of laminar-flow rooms and prophylactic antibiotics has decreased the number of infectious episodes in severely neutropenic patients. However, because of the expense of such a program and the failure to show that it dramatically affects mortality rates, most centers do not routinely use laminar flow

to care for neutropenic patients. Some centers use "reverse isolation," in which health care providers and visitors to a patient who is neutropenic wear gowns and gloves. Since most of the infections these patients develop are due to organisms that colonize the patients' own skin and bowel, the validity of such schemes is dubious, and limited clinical data do not support their use. Hand washing by all staff caring for neutropenic patients should be required to prevent the spread of resistant organisms.

The presence of large numbers of bacteria (particularly *P. aeruginosa*) in certain foods, especially fresh vegetables, has led some authorities to recommend a special "low-bacteria" diet. A diet consisting of cooked and canned food is satisfactory to most neutropenic patients and does not involve elaborate disinfection or sterilization protocols. However, there are no studies to support even this type of dietary restriction. Counseling of patients to avoid leftovers, deli foods, undercooked meat, and unpasteurized dairy products is recommended since these foods have been associated with outbreaks of listeriosis.

PHYSICAL MEASURES

Although few studies address this issue, patients with cancer are predisposed to infections resulting from anatomic compromise (e.g., lymphedema resulting from node dissections after radical mastectomy). Surgeons who specialize in cancer surgery can provide specific guidelines for the care of such patients, and patients benefit from common-sense advice about how to prevent infections in vulnerable areas.

IMMUNOGLOBULIN REPLACEMENT

Many patients with multiple myeloma or CLL have immunoglobulin deficiencies as a result of their disease, and all allogeneic bone marrow transplant recipients are hypogammaglobulinemic for a period after transplantation. However, current recommendations reserve intravenous immunoglobulin replacement therapy for those patients with severe, prolonged hypogammaglobulinemia (<400 mg of total IgG/dL) and a history of repeated infections. Antibiotic prophylaxis has been shown to be cheaper and is efficacious in preventing infections in most CLL patients with hypogammaglobulinemia. Routine use of immunoglobulin replacement is not recommended.

SEXUAL PRACTICES

The use of condoms is recommended for severely immunocompromised patients. Any sexual practice that results in oral exposure to feces is not recommended. Neutropenic patients should be advised to avoid any practice that results in trauma, as even microscopic cuts may result in bacterial invasion and fatal sepsis.

ANTIBIOTIC PROPHYLAXIS

Several studies indicate that the use of oral fluoroquinolones prevents infection and decreases mortality rates among severely neutropenic patients. Prophylaxis for *Pneumocystis* is mandatory for patients with ALL and for all cancer patients receiving glucocorticoid-containing chemotherapy regimens.

VACCINATION OF CANCER PATIENTS

In general, patients undergoing chemotherapy respond less well to vaccines than do normal hosts. Their greater need for vaccines thus leads to a dilemma in their management. Purified proteins and inactivated vaccines are almost never contraindicated and should be given to patients even during chemotherapy. For example, all adults should receive diphtheria-tetanus toxoid boosters at the indicated times as well as seasonal influenza vaccine. However, if possible, vaccination should not be undertaken concurrent with cytotoxic chemotherapy. If patients are expected to be receiving chemotherapy for several months and vaccination is indicated (e.g., influenza vaccination in the fall), the vaccine should be given midcycle—as far apart in time as possible from the antimetabolic agents that will prevent an immune response. The meningococcal and pneumococcal polysaccharide vaccines should be given to patients before splenectomy, if possible. The *H. influenzae* type b conjugate vaccine should be administered to all splenectomized patients.

In general, live virus (or live bacterial) vaccines should not be given to patients during intensive chemotherapy because of the risk of disseminated infection. Recommendations on vaccination are summarized in Table 70-2 (see <https://www.cdc.gov/vaccines/hcp/index.html> for updated recommendations).

FURTHER READING

- KLASTERSKY J et al: The MASCC Neutropenia, Infection and Myelosuppression Study Group evaluates recent new concepts for the use of granulocyte colony-stimulating factors for the prevention of febrile neutropenia. *Support Care Cancer* 21:1793, 2013.
- PAPPAS PG et al: Clinical practice guideline for the management of candidiasis: 2016 update by the Infectious Diseases Society of America. *Clin Infect Dis* 62:e1, 2016.
- PATTERSON TF et al: Practice guidelines for the diagnosis and management of aspergillosis: 2016 update by the Infectious Diseases Society of America. *Clin Infect Dis* 63:e1, 2016.
- TAUR Y, PAMER EG: Microbiome mediation of infections in the cancer setting. *Genome Med* 8:40, 2016.

WEBSITE

Prevention and Treatment of Cancer-Related Infections; National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology Version 2.2016 (<https://www.nccn.org>)

71

Oncologic Emergencies

Rasim Gucalp, Janice P. Dutcher



Emergencies in patients with cancer may be classified into three groups: pressure or obstruction caused by a space-occupying lesion, metabolic or hormonal problems (paraneoplastic syndromes, [Chap. 89](#)), and treatment-related complications.

STRUCTURAL-OBSTRUCTIVE ONCOLOGIC EMERGENCIES

SUPERIOR VENA CAVA SYNDROME

Superior vena cava syndrome (SVCS) is the clinical manifestation of superior vena cava (SVC) obstruction, with severe reduction in venous return from the head, neck, and upper extremities. Malignant tumors, such as lung cancer, lymphoma, and metastatic tumors, are responsible for the majority of SVCS cases. With the expanding use of intravascular devices (e.g., permanent central venous access catheters, pacemaker/defibrillator leads), the prevalence of benign causes of SVCS is increasing now, accounting for at least 40% of cases. Lung cancer, particularly of small-cell and squamous cell histologies, accounts for ~85% of all cases of malignant origin. In young adults, malignant lymphoma is a leading cause of SVCS. Hodgkin's lymphoma involves the mediastinum more commonly than other lymphomas but rarely causes SVCS. When SVCS is noted in a young man with a mediastinal mass, the differential diagnosis is lymphoma versus primary mediastinal germ cell tumor. Metastatic cancers to the mediastinal lymph nodes, such as testicular and breast carcinomas, account for a small proportion of cases. Other causes include benign tumors, aortic aneurysm, thyromegaly, thrombosis, and fibrosing mediastinitis from prior irradiation, histoplasmosis, or Behcet's syndrome. SVCS as the initial manifestation of Behcet's syndrome may be due to inflammation of the SVC associated with thrombosis.

Patients with SVCS usually present with neck and facial swelling (especially around the eyes), dyspnea, and cough. Other symptoms include hoarseness, tongue swelling, headaches, nasal congestion, epistaxis, hemoptysis, dysphagia, pain, dizziness, syncope, and lethargy. Bending forward or lying down may aggravate the symptoms. The characteristic physical findings are dilated neck veins; an increased number of collateral veins covering the anterior chest wall; cyanosis;

and edema of the face, arms, and chest. Facial swelling and plethora are typically exacerbated when the patient is supine. More severe cases include proptosis, glossal and laryngeal edema, and obtundation. The clinical picture is milder if the obstruction is located above the azygos vein. Symptoms are usually progressive, but in some cases, they may improve as collateral circulation develops.

Signs and symptoms of cerebral and/or laryngeal edema, though rare, are associated with a poorer prognosis and require urgent evaluation. Seizures are more likely related to brain metastases than to cerebral edema from venous occlusion. Patients with small-cell lung cancer and SVCS have a higher incidence of brain metastases than those without SVCS.

Cardiorespiratory symptoms at rest, particularly with positional changes, suggest significant airway and vascular obstruction and limited physiologic reserve. Cardiac arrest or respiratory failure can occur, particularly in patients receiving sedatives or undergoing general anesthesia.

Rarely, esophageal varices may develop, particularly in the setting of SVC syndrome due to hemodialysis catheter. These are "downhill" varices based on the direction of blood flow from cephalad to caudad (in contrast to "uphill" varices associated with caudad to cephalad flow from portal hypertension). If the obstruction to the SVC is proximal to the azygous vein, varices develop in the upper one-third of the esophagus. If the obstruction involves or is distal to the azygous vein, varices occur in the entire length of the esophagus. Variceal bleeding may be a late complication of chronic SVCS.

SVC obstruction may lead to bilateral breast edema with bilateral enlarged breast. Unilateral breast dilation may be seen as a consequence of axillary or subclavian vein blockage.

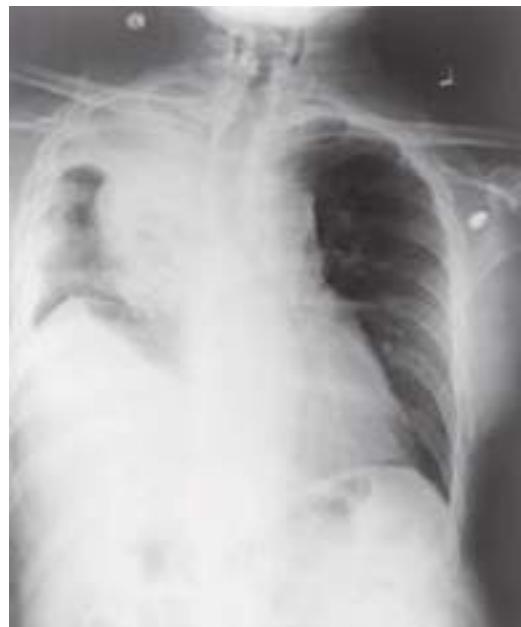
The diagnosis of SVCS is a clinical one. The most significant chest radiographic finding is widening of the superior mediastinum, most commonly on the right side. Pleural effusion occurs in only 25% of patients, often on the right side. The majority of these effusions are exudative and occasionally chylous. However, a normal chest radiograph is still compatible with the diagnosis if other characteristic findings are present. Computed tomography (CT) provides the most reliable view of the mediastinal anatomy. The diagnosis of SVCS requires diminished or absent opacification of central venous structures with prominent collateral venous circulation. Magnetic resonance imaging (MRI) is increasingly being used to diagnose SVC obstruction with a 100% sensitivity and specificity, but dyspneic SVCS patients may have difficulty remaining supine for the entire imaging process. Invasive procedures, including bronchoscopy, percutaneous needle biopsy, mediastinoscopy, and even thoracotomy, can be performed by a skilled clinician without any major risk of bleeding. Endobronchial or esophageal ultrasound-guided needle aspiration may establish the diagnosis safely. For patients with a known cancer, a detailed workup usually is not necessary, and appropriate treatment may be started after obtaining a CT scan of the thorax. For those with no history of malignancy, a detailed evaluation is essential to rule out benign causes and determine a specific diagnosis to direct the appropriate therapy.

TREATMENT

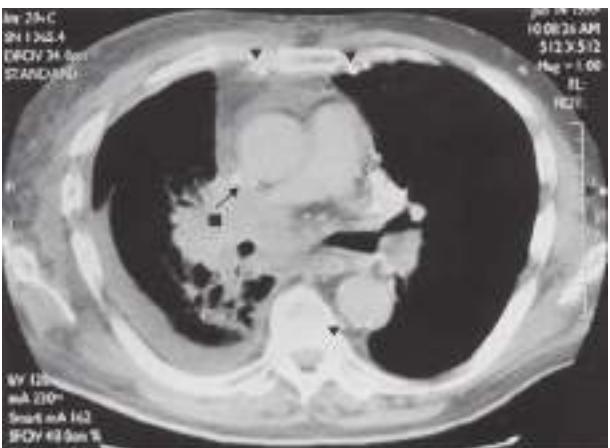
Superior Vena Cava Syndrome

The one potentially life-threatening complication of a superior mediastinal mass is tracheal obstruction. Upper airway obstruction demands emergent therapy. Diuretics with a low-salt diet, head elevation, and oxygen may produce temporary symptomatic relief. Glucocorticoids have a limited role except in the setting of mediastinal lymphoma masses.

Radiation therapy is the primary treatment for SVCS caused by non-small-cell lung cancer and other metastatic solid tumors. Chemotherapy is effective when the underlying cancer is small-cell carcinoma of the lung, lymphoma, or germ cell tumor. SVCS recurs in 10–30% of patients; it may be palliated with the use of intravascular self-expanding stents ([Fig. 71-1](#)). Early stenting may be necessary in patients with severe symptoms; however, the prompt increase in venous return after stenting may precipitate heart failure and



A



B



C

FIGURE 71-1 Superior vena cava syndrome (SVCS). A. Chest radiographs of a 59-year-old man with recurrent SVCS caused by non-small-cell lung cancer showing right paratracheal mass with right pleural effusion. B. Computed tomography of same patient demonstrating obstruction of the superior vena cava with thrombosis (arrow) by the lung cancer (square) and collaterals (arrowheads). C. Balloon angioplasty (arrowhead) with Wallstent (arrow) in same patient.

pulmonary edema. Other complications of stent placement include hematoma at the insertion site, SVC perforation, stent migration in the right ventricle, stent fracture, and pulmonary embolism.

Clinical improvement occurs in most patients, although this improvement may be due to the development of adequate collateral circulation. The mortality associated with SVCS does not relate to caval obstruction but rather to the underlying cause.

SVCS AND CENTRAL VENOUS CATHETERS IN ADULTS

The use of long-term central venous catheters has become common practice in patients with cancer. Major vessel thrombosis may occur. In these cases, catheter removal should be combined with anticoagulation to prevent embolization. SVCS in this setting, if detected early, can be treated by fibrinolytic therapy without sacrificing the catheter. When managing patients with transvenous lead-related SVC syndrome, anticoagulation, local and systemic thrombolytic therapy, and surgical intervention can be effective therapy in select patients. Endovascular stenting has also been shown to be safe and promising, with minimal procedural or clinical complications. The role of anticoagulation after SVC stent placement is controversial.

PERICARDIAL EFFUSION/TAMPOONADE

Malignant pericardial disease is found at autopsy in 5–10% of patients with cancer, most frequently with lung cancer, breast cancer, leukemias, and lymphomas. Cardiac tamponade as the initial presentation of extrathoracic malignancy is rare. The origin is not malignancy in ~50% of cancer patients with symptomatic pericardial disease, but it can be related to irradiation, drug-induced pericarditis, including chemotherapeutic agents such as all-trans retinoic acid, arsenic trioxide, imatinib and other abl kinase inhibitors, hypothyroidism, idiopathic pericarditis, infection, or autoimmune diseases. Two types of radiation pericarditis occur: an acute inflammatory, effusive pericarditis occurring within months of irradiation, which usually resolves spontaneously, and a chronic effusive pericarditis that may appear up to 20 years after radiation therapy and is accompanied by a thickened pericardium.

Most patients with pericardial metastasis are asymptomatic. However, the common symptoms are dyspnea, cough, chest pain, orthopnea, and weakness. Pleural effusion, sinus tachycardia, jugular venous distention, hepatomegaly, peripheral edema, and cyanosis are the most frequent physical findings. Relatively specific diagnostic findings, such as paradoxical pulse, diminished heart sounds, pulsus alternans (pulse waves alternating between those of greater and lesser amplitude with successive beats), and friction rub are less common than with nonmalignant pericardial disease. Chest radiographs and electrocardiogram (ECG) reveal abnormalities in 90% of patients, but half of these abnormalities are nonspecific. Echocardiography is the most helpful diagnostic test. Pericardial fluid may be serous, serosanguineous, or hemorrhagic, and cytologic examination of pericardial fluid is diagnostic in most patients. Measurements of tumor markers in the pericardial fluid are not helpful in the diagnosis of malignant pericardial fluid. Pericardiectomy with targeted pericardial and epicardial biopsy may differentiate neoplastic and benign pericardial disease. A combination of cytology, pericardial and epicardial biopsy, and guided pericardiectomy gives the best diagnostic yield. CT scan of chest may also reveal the presence of a concomitant thoracic neoplasm. Cancer patients with pericardial effusion containing malignant cells on cytology have a very poor survival, ~7 weeks.

TREATMENT

Pericardial Effusion/Tamponade

Pericardiocentesis with or without the introduction of sclerosing agents, the creation of a pericardial window, complete pericardial stripping, cardiac irradiation, or systemic chemotherapy are effective treatments. Acute pericardial tamponade with life-threatening hemodynamic instability requires immediate drainage of fluid. This can be quickly achieved by pericardiocentesis. The recurrence rate after percutaneous catheter drainage is ~20%. Sclerotherapy (pericardial instillation of bleomycin, mitomycin C, or tetracycline) may

decrease recurrences. Alternatively, subxiphoid pericardiotomy can be performed in 45 min under local anesthesia. Thoracoscopic pericardial fenestration can be employed for benign causes; however, 60% of malignant pericardial effusions recur after this procedure. In a subset of patients, drainage of the pericardial effusion is paradoxically followed by worsening hemodynamic instability. This so-called “postoperative low cardiac output syndrome” occurs in up to 10% of patients undergoing surgical drainage and carries poor short-term survival.

■ INTESTINAL OBSTRUCTION

Intestinal obstruction and reobstruction are common problems in patients with advanced cancer, particularly colorectal or ovarian carcinoma. However, other cancers, such as lung or breast cancer and melanoma, can metastasize within the abdomen, leading to intestinal obstruction. Metastatic disease from colorectal, ovarian, pancreatic, gastric, and occasionally breast cancer can lead to peritoneal carcinomatosis, with infiltration of the omentum and peritoneal surface, thus limiting bowel motility. Typically, obstruction occurs at multiple sites in peritoneal carcinomatosis. Melanoma has a predilection to involve the small bowel; this involvement may be isolated, and resection may result in prolonged survival. Intestinal pseudoobstruction is caused by infiltration of the mesentery or bowel muscle by tumor, involvement of the celiac plexus, or paraneoplastic neuropathy in patients with small-cell lung cancer. Paraneoplastic neuropathy is associated with IgG antibodies reactive to neurons of the myenteric and submucosal plexuses of the jejunum and stomach. Ovarian cancer can lead to authentic luminal obstruction or to pseudoobstruction that results when circumferential invasion of a bowel segment arrests the forward progression of peristaltic contractions.

The onset of obstruction is usually insidious. Pain is the most common symptom and is usually colicky in nature. Pain can also be due to abdominal distention, tumor masses, or hepatomegaly. Vomiting can be intermittent or continuous. Patients with complete obstruction usually have constipation. Physical examination may reveal abdominal distention with tympany, ascites, visible peristalsis, high-pitched bowel sounds, and tumor masses. Erect plain abdominal films may reveal multiple air-fluid levels and dilation of the small or large bowel. Acute cecal dilation to >12–14 cm is considered a surgical emergency because of the high likelihood of rupture. CT scan is useful in defining the extent of disease and the exact nature of the obstruction and differentiating benign from malignant causes of obstruction in patients who have undergone surgery for malignancy. Malignant obstruction is suggested by a mass at the site of obstruction or prior surgery, adenopathy, or an abrupt transition zone and irregular bowel thickening at the obstruction site. Benign obstruction is more likely when CT shows mesenteric vascular changes, a large volume of ascites, or a smooth transition zone and smooth bowel thickening at the obstruction site. In challenging patients with obstructive symptoms, particularly low-grade small-bowel obstruction (SBO), CT enteroclysis often can help establish the diagnosis by providing distention of small-bowel loops. In this technique, water-soluble contrast is infused through a nasoenteric tube into the duodenum or proximal small bowel followed by CT images. The prognosis for the patient with cancer who develops intestinal obstruction is poor; median survival is 3–4 months. About 25–30% of patients are found to have intestinal obstruction due to causes other than cancer. Adhesions from previous operations are a common benign cause. Ileus induced by vinca alkaloids, narcotics, or other drugs is another reversible cause.

TREATMENT

Intestinal Obstruction

The management of intestinal obstruction in patients with advanced malignancy depends on the extent of the underlying malignancy, options for further antineoplastic therapy, estimated life expectancy, the functional status of the major organs, and the extent of the obstruction. The initial management should include surgical

evaluation. Operation is not always successful and may lead to further complications with a substantial mortality rate (10–20%). Laparoscopy can diagnose and treat malignant bowel obstruction in some cases. Self-expanding metal stents placed in the gastric outlet, duodenum, proximal jejunum, colon, or rectum may palliate obstructive symptoms at those sites without major surgery. Patients known to have advanced intraabdominal malignancy should receive a prolonged course of conservative management, including nasogastric decompression. Percutaneous endoscopic or surgical gastrostomy tube placement is an option for palliation of nausea and vomiting, the so-called “venting gastrostomy.” Treatment with antiemetics, antispasmodics, and analgesics may allow patients to remain outside the hospital. Octreotide may relieve obstructive symptoms through its inhibitory effect on gastrointestinal secretion. Glucocorticoids have anti-inflammatory effects and may help the resolution of bowel obstruction. They also have antiemetic effects.

■ URINARY OBSTRUCTION

Urinary obstruction may occur in patients with prostatic or gynecologic malignancies, particularly cervical carcinoma; metastatic disease from other primary sites such as carcinomas of the breast, stomach, lung, colon, and pancreas; or lymphomas. Radiation therapy to pelvic tumors may cause fibrosis and subsequent ureteral obstruction. Bladder outlet obstruction is usually due to prostate and cervical cancers and may lead to bilateral hydronephrosis and renal failure.

Flank pain is the most common symptom. Persistent urinary tract infection, persistent proteinuria, or hematuria in patients with cancer should raise suspicion of ureteral obstruction. Total anuria and/or anuria alternating with polyuria may occur. A slow, continuous rise in the serum creatinine level necessitates immediate evaluation. Renal ultrasound is the safest and cheapest way to identify hydronephrosis. The function of an obstructed kidney can be evaluated by a nuclear scan. CT scan can reveal the point of obstruction and identify a retroperitoneal mass or adenopathy.

TREATMENT

Urinary Obstruction

Obstruction associated with flank pain, sepsis, or fistula formation is an indication for immediate palliative urinary diversion. Internal ureteral stents can be placed under local anesthesia. Percutaneous nephrostomy offers an alternative approach for drainage. The placement of a nephrostomy is associated with a significant rate of pyelonephritis. In the case of bladder outlet obstruction due to malignancy, a suprapubic cystostomy can be used for urinary drainage. An aggressive intervention with invasive approaches to improve the obstruction should be weighed against the likelihood of antitumor response, and the ability to reverse renal insufficiency should be evaluated.

■ MALIGNANT BILIARY OBSTRUCTION

This common clinical problem can be caused by a primary carcinoma arising in the pancreas, ampulla of Vater, bile duct, or liver or by metastatic disease to the periductal lymph nodes or liver parenchyma. The most common metastatic tumors causing biliary obstruction are gastric, colon, breast, and lung cancers. Jaundice, light-colored stools, dark urine, pruritus, and weight loss due to malabsorption are usual symptoms. Pain and secondary infection are uncommon in malignant biliary obstruction. Ultrasound, CT scan, or percutaneous transhepatic or endoscopic retrograde cholangiography will identify the site and nature of the biliary obstruction.

TREATMENT

Malignant Biliary Obstruction

Palliative intervention is indicated only in patients with disabling pruritus resistant to medical treatment, severe malabsorption, or

infection. Stenting under radiographic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. The choice of therapy should be based on the site of obstruction (proximal vs distal), the type of tumor (sensitive to radiotherapy, chemotherapy, or neither), and the general condition of the patient. Stenting under radiographic or endoscopic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. Photodynamic therapy and radiofrequency ablation are promising endoscopic therapies for malignant biliary obstruction.

■ SPINAL CORD COMPRESSION

Malignant spinal cord compression (MSCC) is defined as compression of the spinal cord and/or cauda equina by an extradural tumor mass. The minimum radiologic evidence for cord compression is indentation of the theca at the level of clinical features. Spinal cord compression occurs in 5–10% of patients with cancer. Epidural tumor is the first manifestation of malignancy in ~10% of patients. The underlying cancer is usually identified during the initial evaluation; lung cancer is the most common cause of MSCC.

Metastatic tumor involves the vertebral column more often than any other part of the bony skeleton. Lung, breast, and prostate cancers are the most frequent offenders. Multiple myeloma also has a high incidence of spine involvement. Lymphomas, melanoma, renal cell cancer, and genitourinary cancers also cause cord compression. The thoracic spine is the most common site (70%), followed by the lumbosacral spine (20%) and the cervical spine (10%). Involvement of multiple sites is most frequent in patients with breast and prostate carcinoma. Cord injury develops when metastases to the vertebral body or pedicle enlarge and compress the underlying dura. Another cause of cord compression is direct extension of a paravertebral lesion through the intervertebral foramen. These cases usually involve a lymphoma, myeloma, or pediatric neoplasm. Parenchymal spinal cord metastasis due to hematogenous spread is rare. Intramedullary metastases can be seen in lung cancer, breast cancer, renal cancer, melanoma, and lymphoma, and are frequently associated with brain metastases and leptomeningeal disease.

Expanding extradural tumors induce injury through several mechanisms. Expanding extradural tumors induce mechanical injury to axons and myelin. Compression compromises blood flow, leading to ischemia and/or infarction.

The most common initial symptom in patients with spinal cord compression is localized back pain and tenderness due to involvement of vertebrae by tumor. Pain is usually present for days or months before other neurologic findings appear. It is exacerbated by movement and by coughing or sneezing. It can be differentiated from the pain of disk disease by the fact that it worsens when the patient is supine. Radicular pain is less common than localized back pain and usually develops later. Radicular pain in the cervical or lumbosacral areas may be unilateral or bilateral. Radicular pain from the thoracic roots is often bilateral and is described by patients as a feeling of tight, band-like constriction around the thorax and abdomen. Typical cervical radicular pain radiates down the arm; in the lumbar region, the radiation is down the legs. *Lhermitte's sign*, a tingling or electric sensation down the back and upper and lower limbs upon flexing or extending the neck, may be an early sign of cord compression. Loss of bowel or bladder control may be the presenting symptom but usually occurs late in the course. Occasionally patients present with ataxia of gait without motor and sensory involvement due to involvement of the spinocerebellar tract.

On physical examination, pain induced by straight leg raising, neck flexion, or vertebral percussion may help to determine the level of cord compression. Patients develop numbness and paresthesias in the extremities or trunk. Loss of sensibility to pinprick is as common as loss of sensibility to vibration or position. The upper limit of the zone of sensory loss is often one or two vertebrae below the site of compression. Motor findings include weakness, spasticity, and abnormal muscle stretching. An extensor plantar reflex reflects significant compression. Deep tendon reflexes may be brisk. Motor and sensory

loss usually precedes sphincter disturbance. Patients with autonomic dysfunction may present with decreased anal tonus, decreased perineal sensibility, and a distended bladder. The absence of the anal wink reflex or the bulbocavernosus reflex confirms cord involvement. In doubtful cases, evaluation of postvoiding urinary residual volume can be helpful. A residual volume of >150 mL suggests bladder dysfunction. Autonomic dysfunction is an unfavorable prognostic factor. Patients with progressive neurologic symptoms should have frequent neurologic examinations and rapid therapeutic intervention. Other illnesses that may mimic cord compression include osteoporotic vertebral collapse, disk disease, pyogenic abscess or vertebral tuberculosis, radiation myelopathy, neoplastic leptomeningitis, benign tumors, epidural hematoma, and spinal lipomatosis.

Cauda equina syndrome is characterized by low back pain; diminished sensation over the buttocks, posterior-superior thighs, and perineal area in a saddle distribution; rectal and bladder dysfunction; sexual impotence; absent bulbocavernous, patellar, and Achilles' reflexes; and variable amount of lower-extremity weakness. This reflects compression of nerve roots as they form the cauda equina after leaving the spinal cord. The majority of cauda equine tumors are primary tumors of glial or nerve sheath origin; metastases are very rare.

Patients with cancer who develop back pain should be evaluated for spinal cord compression as quickly as possible (Fig. 71-2). Treatment is more often successful in patients who are ambulatory and still have sphincter control at the time treatment is initiated. Patients should have a neurologic examination and plain films of the spine. Those whose physical examination suggests cord compression should receive dexamethasone starting immediately.

Erosion of the pedicles (the "winking owl" sign) is the earliest radiologic finding of vertebral tumor. Other radiographic changes include increased intrapedicular distance, vertebral destruction, lytic or sclerotic lesions, scalloped vertebral bodies, and vertebral body collapse. Vertebral collapse is not a reliable indicator of the presence of tumor; ~20% of cases of vertebral collapse, particularly those in older patients and postmenopausal women, are due not to cancer but to osteoporosis. Also, a normal appearance on plain films of the spine does not exclude the diagnosis of cancer. The role of bone scans in the detection of cord compression is not clear; this method is sensitive but less specific than spinal radiography.

The full-length image of the cord provided by MRI is the imaging procedure of choice. Multiple epidural metastases are noted in 25% of patients with cord compression, and their presence influences treatment plans. On T1-weighted images, good contrast is noted between the cord, cerebrospinal fluid (CSF), and extradural lesions. Owing to its sensitivity in demonstrating the replacement of bone marrow by tumor, MRI can show which parts of a vertebra are involved by tumor. MRI also visualizes intraspinal extradural masses compressing the cord. T2-weighted images are most useful for the demonstration of intramedullary pathology. Gadolinium-enhanced MRI can help to delineate intramedullary disease. MRI is as good as or better than myelography plus postmyelogram CT scan in detecting metastatic epidural disease with cord compression. Myelography should be reserved for patients who have poor MRIs or who cannot undergo MRI promptly. CT scan in conjunction with myelography enhances the detection of small areas of spinal destruction.

In patients with cord compression and an unknown primary tumor, a simple workup including chest radiography, mammography, measurement of prostate-specific antigen, and abdominal CT usually reveals the underlying malignancy.

TREATMENT

Spinal Cord Compression

The treatment of patients with spinal cord compression is aimed at relief of pain and restoration/preservation of neurologic function (Fig. 71-2). Management of MSCC requires a multidisciplinary approach.

Radiation therapy plus glucocorticoids is generally the initial treatment of choice for most patients with spinal cord compression.

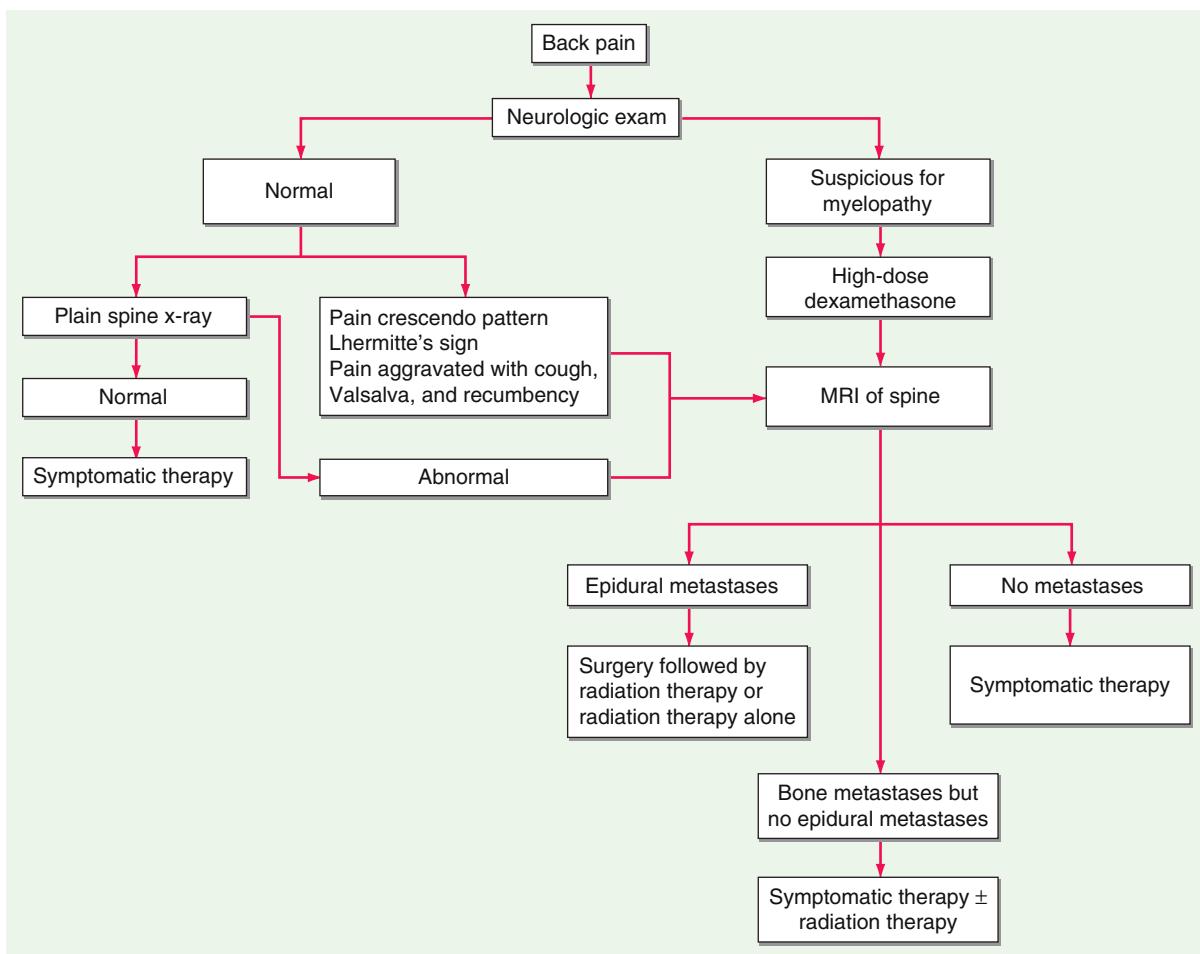


FIGURE 71-2 Management of cancer patients with back pain.

The management decision of SCC involves assessment of neurologic (N), oncologic (O), mechanical (M), and systemic factors (S). NOMS was developed by Memorial Sloan Kettering Cancer Center (MSKCC) researchers to provide an algorithm for management of SCC. The neurologic assessment is based on the degree of epidural SCC, myelopathy, and/or functional radiculopathy. Oncologic assessment involves the radio-sensitivity of the tumor type. In patients with radio-resistant tumors, stereotactic body radiotherapy (SRS) is the preferred approach if radiation is appropriate. Safe delivery of SRS requires a 2- to 3-mm margin away from the spinal cord. Separation surgery followed by SRS is necessary in patients with high-grade SCC due to radio-resistant tumors. In patients with mechanical instability or retropulsion of bone fragments into the spinal canal or cord, a surgical approach is the treatment of choice. Systemic factors that need to be considered are the extent of disease and medical comorbidities that determine the patient's ability to tolerate planned therapy. Chemotherapy may have a role in patients with chemosensitive tumors who have had prior radiotherapy to the same region and who are not candidates for surgery. Patients who previously received radiotherapy for MSKCC with an in-field tumor progression can be treated with reirradiation if they are not surgical candidates.

Patients with painful pathologic compression fractures without spinal instability may benefit from percutaneous vertebroplasty or kyphoplasty, the injection of acrylic cement into a collapsed vertebra to stabilize the fracture. Pain palliation is common, and local antitumor effects have been noted. Cement leakage may cause symptoms in ~10% of patients. Bisphosphonates and/or denosumab may be helpful in prevention of SCC in patients with bony involvement.

The histology of the tumor is an important determinant of both recovery and survival. Rapid onset and progression of signs and symptoms are poor prognostic features.

■ INCREASED INTRACRANIAL PRESSURE

About 25% of patients with cancer die with intracranial metastases. The cancers that most often metastasize to the brain are lung and breast cancers and melanoma. Brain metastases often occur in the presence of systemic disease, and they frequently cause major symptoms, disability, and early death. The initial presentation of brain metastases from a previously unknown primary cancer is common. Lung cancer is most commonly the primary malignancy. CT scans of the chest/abdomen and MRI of the brain as the initial diagnostic studies can identify a biopsy site in most patients.

The signs and symptoms of a metastatic brain tumor are similar to those of other intracranial expanding lesions: headache, nausea, vomiting, behavioral changes, seizures, and focal, progressive neurologic changes. Occasionally the onset is abrupt, resembling a stroke, with the sudden appearance of headache, nausea, vomiting, and neurologic deficits. This picture is usually due to hemorrhage into the metastasis. Melanoma, germ cell tumors, and renal cell cancers have a particularly high incidence of intracranial bleeding. The tumor mass and surrounding edema may cause obstruction of the circulation of CSF, with resulting hydrocephalus. Patients with increased intracranial pressure may have papilledema with visual disturbances and neck stiffness. As the mass enlarges, brain tissue may be displaced through the fixed cranial openings, producing various herniation syndromes.

MRI is superior to CT scan. Gadolinium-enhanced MRI is more sensitive than CT at revealing meningeal involvement and small lesions, particularly in the brainstem or cerebellum. The MRI of the brain shows brain metastases as multiple enhancing lesions of various sizes with surrounding areas of low-density edema.

Intracranial hypertension ("pseudotumor cerebri") secondary to tretinoin therapy for acute promyelocytic leukemia has been reported, as another cause of intracranial pressure in the setting of a malignancy.

TREATMENT**Increased Intracranial Pressure**

Dexamethasone is the best initial treatment for all symptomatic patients with brain metastases. Patients with multiple lesions should usually receive whole-brain radiation. Patients with a single-brain metastasis and with controlled extracranial disease may be treated with surgical excision followed by whole-brain radiation therapy, especially if they are aged <60 years. Radioresistant tumors should be resected if possible. Stereotactic radiosurgery (SRS) is recommended in patients with a limited number of brain metastases (one to four) who have stable, systemic disease or reasonable systemic treatment options and in patients who have a small number of metastatic lesions in whom whole-brain radiation therapy has failed. With a gamma knife or linear accelerator, multiple small, well-collimated beams of ionizing radiation destroy lesions seen on MRI. Some patients with increased intracranial pressure associated with hydrocephalus may benefit from shunt placement. If neurologic deterioration is not reversed with medical therapy, ventriculotomy to remove CSF or craniotomy to remove tumors or hematomas may be necessary.

Targeted agents and checkpoint inhibitors have significant activity in brain metastases from non-small-cell lung cancer, breast cancer, renal cancer, and melanoma.

NEOPLASTIC MENINGITIS

Tumor involving the leptomeninges is a complication of both primary central nervous system (CNS) tumors and tumors that metastasize to the CNS. The incidence is estimated at 3–8% of patients with cancer. Melanoma, breast and lung cancer, lymphoma (including AIDS-associated), and acute leukemia are the most common causes. Synchronous intraparenchymal brain metastases are evident in 11–31% of patients with neoplastic meningitis. Leptomeningeal seeding is frequent in patients undergoing resection of brain metastases or receiving stereotactic radiotherapy for brain metastases.

Patients typically present with multifocal neurologic signs and symptoms, including headache, gait abnormality, mental changes, nausea, vomiting, seizures, back or radicular pain, and limb weakness. Signs include cranial nerve palsies, extremity weakness, paresthesia, and decreased deep tendon reflexes.

Diagnosis is made by demonstrating malignant cells in the CSF; however, up to 40% of patients may have false-negative CSF cytology. An elevated CSF protein level is nearly always present (except in HTLV-1-associated adult T-cell leukemia). Patients with neurologic signs and symptoms consistent with neoplastic meningitis who have a negative CSF cytology should have the spinal tap repeated at least one more time for cytologic examination. MRI findings suggestive of neoplastic meningitis include leptomeningeal, subependymal, dural, or cranial nerve enhancement; superficial cerebral lesions; intradural nodules; and communicating hydrocephalus. Spinal cord imaging by MRI is a necessary component of the evaluation of nonleukemia neoplastic meningitis because ~20% of patients have cord abnormalities, including intradural enhancing nodules that are diagnostic for leptomeningeal involvement. Cauda equina lesions are common, but lesions may be seen anywhere in the spinal canal. The value of MRI for the diagnosis of leptomeningeal disease is limited in patients with hematopoietic malignancy. Radiolabeled CSF flow studies are abnormal in up to 70% of patients with neoplastic meningitis; ventricular outlet obstruction, abnormal flow in the spinal canal, or impaired flow over the cerebral convexities may affect distribution of intrathecal chemotherapy, resulting in decreased efficacy or increased toxicity. Radiation therapy may correct CSF flow abnormalities before use of intrathecal chemotherapy. Neoplastic meningitis can also lead to intracranial hypertension and hydrocephalus. Placement of a ventriculoperitoneal shunt may effectively palliate symptoms in these patients.

The development of neoplastic meningitis usually occurs in the setting of uncontrolled cancer outside the CNS; thus, prognosis is poor

(median survival 10–12 weeks). However, treatment of the neoplastic meningitis may successfully alleviate symptoms and control the CNS spread.

TREATMENT**Neoplastic Meningitis**

Intrathecal chemotherapy, usually methotrexate, cytarabine, or thiotepa, is delivered by lumbar puncture or by an intraventricular reservoir (Ommaya). Among solid tumors, breast cancer responds best to therapy. Focal radiotherapy may have role in bulky disease, and in symptomatic or obstructive lesions. Targeted therapy such as systemically administered epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in non-small-cell lung cancer may improve in subgroups of cancer patients with leptomeningeal spread. Patients with neoplastic meningitis from either acute leukemia or lymphoma may be cured of their CNS disease if the systemic disease can be eliminated.

SEIZURES

Seizures occurring in a patient with cancer can be caused by the tumor itself, by metabolic disturbances, by radiation injury, by cerebral infarctions, by chemotherapy-related encephalopathies, or by CNS infections. Metastatic disease to the CNS is the most common cause of seizures in patients with cancer. However, seizures occur more frequently in primary brain tumors than in metastatic brain lesions. Seizures are a presenting symptom of CNS metastasis in 6–29% of cases. Approximately 10% of patients with CNS metastasis eventually develop seizures. Tumors that affect the frontal, temporal, and parietal lobes are more commonly associated with seizures than are occipital lesions. Both early and late seizures are uncommon in patients with posterior fossa and sellar lesions. Seizures are common in patients with CNS metastases from melanoma and low-grade primary brain tumors. Very rarely, cytotoxic drugs such as etoposide, busulfan, ifosfamide, and chlorambucil cause seizures. Another cause of seizures related to drug therapy is reversible posterior leukoencephalopathy syndrome (RPLS). Chemotherapy, targeted therapy, and immunotherapies have been associated with the development of RPLS. RPLS occurs in patients undergoing allogeneic bone marrow or solid-organ transplantation. RPLS is characterized by headache, altered consciousness, generalized seizures, visual disturbances, hypertension, and symmetric posterior cerebral white matter vasogenic edema on CT/MRI. Seizures may begin focally but are typically generalized.

TREATMENT**Seizures**

Patients in whom seizures due to CNS metastases have been demonstrated should receive anticonvulsive treatment with phenytoin or levetiracetam. If this is not effective, valproic acid can be added. Prophylactic anticonvulsant therapy is not recommended. In postcraniotomy patients, prophylactic antiepileptic drugs should be withdrawn during the first week after surgery. Most antiseizure medications including phenytoin induce cytochrome P450 (CYP450), which alters the metabolism of many antitumor agents, including irinotecan, taxanes, and etoposide as well as molecular targeted agents, including imatinib, gefitinib, erlotinib, tipifarnib, sorafenib, sunitinib, temsirolimus, everolimus, and vemurafenib. Levetiracetam and topiramate are anticonvulsant agents not metabolized by the hepatic CYP450 system and do not alter the metabolism of antitumor agents. They have become the preferred drugs. Surgical resection and other antitumor treatments such as radiotherapy and chemotherapy may improve seizure control.

PULMONARY AND INTRACEREBRAL LEUKOSTASIS

Hyperleukocytosis and the leukostasis syndrome associated with it is a potentially fatal complication of acute leukemia (particularly myeloid leukemia) that can occur when the peripheral blast cell count is $>100,000/\text{mL}$. The frequency of hyperleukocytosis is 5–13% in acute myeloid leukemia (AML) and 10–30% in acute lymphoid leukemia; however, leukostasis is rare in lymphoid leukemia. At such high blast cell counts, blood viscosity is increased, blood flow is slowed by aggregates of tumor cells, and the primitive myeloid leukemic cells are capable of invading through the endothelium and causing hemorrhage. Brain and lung are most commonly affected. Patients with brain leukostasis may experience stupor, headache, dizziness, tinnitus, visual disturbances, ataxia, confusion, coma, or sudden death. On examination, papilledema, retinal vein distension, retinal hemorrhages, and focal deficit may be present. Pulmonary leukostasis may present as respiratory distress and hypoxemia and progress to respiratory failure. Chest radiographs may be normal but usually show interstitial or alveolar infiltrates. Hyperleukocytosis rarely may cause acute leg ischemia, renal vein thrombosis, myocardial ischemia, bowel infarction, and priapism. Arterial blood gas results should be interpreted cautiously. Rapid consumption of plasma oxygen by the markedly increased number of white blood cells can cause spuriously low arterial oxygen tension. Pulse oximetry is the most accurate way of assessing oxygenation in patients with hyperleukocytosis. Hydroxyurea can rapidly reduce a high blast cell count while the diagnostic workup is in progress. After the diagnosis is established, the patient should start quickly with effective induction chemotherapy. Leukapheresis should be used in patients with symptoms of hyperleukocytosis. Patients with hyperleukocytosis are also at the risk for disseminated intravascular coagulation and tumor lysis syndrome. The clinician should monitor the patient for these complications and take preventive and therapeutic actions during induction therapy. Intravascular volume depletion and unnecessary blood transfusions may increase blood viscosity and worsen the leukostasis syndrome. Leukostasis is very rarely a feature of the high white cell counts associated with chronic lymphoid or chronic myeloid leukemia.

When acute promyelocytic leukemia is treated with differentiating agents like tretinoin and arsenic trioxide, cerebral or pulmonary leukostasis may occur as tumor cells differentiate into mature neutrophils. This complication can be largely avoided by using cytotoxic chemotherapy together with the differentiating agents.

HEMOPTYSIS

Hemoptysis may be caused by nonmalignant conditions, but lung cancer accounts for a large proportion of cases. Up to 20% of patients with lung cancer have hemoptysis some time in their course. Endobronchial metastases from carcinoid tumors, breast cancer, colon cancer, kidney cancer, and melanoma may also cause hemoptysis. The volume of bleeding is often difficult to gauge. Massive hemoptysis is defined as $>200\text{--}600 \text{ mL}$ of blood produced in 24 h. However, any hemoptysis should be considered massive if it threatens life. When respiratory difficulty occurs, hemoptysis should be treated emergently. The first priorities are to maintain the airway, optimize oxygenation, and stabilize the hemodynamic status. If the bleeding side is known, the patient should be placed in a lateral decubitus position, with the bleeding side down to prevent aspiration into the unaffected lung, and given supplemental oxygen. If large-volume bleeding continues or the airway is compromised, the patient should be intubated and undergo emergency bronchoscopy. If the site of bleeding is detected, either the patient undergoes a definitive surgical procedure or the lesion is treated with a neodymium:yttrium-aluminum-garnet (Nd:YAG) laser, argon plasma coagulation, or electrocautery. In stable patients, multidetector CT angiography delineates bronchial and nonbronchial systemic arteries and identifies the source of bleeding and underlying pathology with high sensitivity. Massive hemoptysis usually originates from the high-pressure bronchial circulation. Bronchial artery embolization is considered a first-line definite procedure for managing hemoptysis. Bronchial artery embolization may control brisk bleeding in 75–90% of patients,

permitting the definitive surgical procedure to be done more safely if it is appropriate.

Embolization without definitive surgery is associated with rebleeding in 20–50% of patients. Recurrent hemoptysis usually responds to a second embolization procedure. A postembolization syndrome characterized by pleuritic pain, fever, dysphagia, and leukocytosis may occur; it lasts 5–7 days and resolves with symptomatic treatment. Bronchial or esophageal wall necrosis, myocardial infarction, and spinal cord infarction are rare complications. Surgery, as a salvage strategy, is indicated after failure of embolization and is associated with better survival when performed in a nonurgent setting.

Pulmonary hemorrhage with or without hemoptysis in hematologic malignancies is often associated with fungal infections, particularly *Aspergillus* sp. After granulocytopenia resolves, the lung infiltrates in aspergillosis may cavitate and cause massive hemoptysis. Thrombocytopenia and coagulation defects should be corrected, if possible. Surgical evaluation is recommended in patients with aspergillosis-related cavitary lesions.

Bevacizumab, an antibody to vascular endothelial growth factor (VEGF) that inhibits angiogenesis, has been associated with life-threatening hemoptysis in patients with non-small-cell lung cancer, particularly of squamous cell histology. Non-small-cell lung cancer patients with cavitary lesions or previous hemoptysis ($\geq 2.5 \text{ mL}$) within the past 3 months have higher risk for pulmonary hemorrhage.

AIRWAY OBSTRUCTION

Airway obstruction refers to a blockage at the level of the mainstem bronchi or above. It may result either from intraluminal tumor growth or from extrinsic compression of the airway. The most common cause of malignant upper airway obstruction is invasion from an adjacent primary tumor, most commonly lung cancer, followed by esophageal, thyroid, and mediastinal malignancies including lymphomas. Extrathoracic primary tumors such as renal, colon, or breast cancer can cause airway obstruction through endobronchial and/or mediastinal lymph node metastases. Patients may present with dyspnea, hemoptysis, stridor, wheezing, intractable cough, postobstructive pneumonia, or hoarseness. Chest radiographs usually demonstrate obstructing lesions. CT scans reveal the extent of tumor. Cool, humidified oxygen, glucocorticoids, and ventilation with a mixture of helium and oxygen (Heliox) may provide temporary relief. If the obstruction is proximal to the larynx, a tracheostomy may be lifesaving. For more distal obstructions, particularly intrinsic lesions incompletely obstructing the airway, bronchoscopy with mechanical debulking and dilation or ablational treatments including laser treatment, photodynamic therapy, argon plasma coagulation, electrocautery, or stenting can produce immediate relief in most patients (Fig. 71-3). However, radiation therapy (either external-beam irradiation or brachytherapy) given together with glucocorticoids may also open the airway. Symptomatic extrinsic compression may be palliated by stenting. Patients with primary airway tumors such as squamous cell carcinoma, carcinoid tumor, adenocystic carcinoma, or non-small-cell lung cancer, if resectable, should have surgery.

METABOLIC EMERGENCIES

HYPERCALCEMIA

Hypercalcemia is the most common paraneoplastic syndrome. Its pathogenesis and management are discussed fully in Chaps. 89 and 403.

SYNDROME OF INAPPROPRIATE SECRETION OF ANTIDIURETIC HORMONE

Hyponatremia is a common electrolyte abnormality in cancer patients, and syndrome of inappropriate secretion of antidiuretic hormone (SIADH) is the most common cause among patients with cancer. SIADH is discussed fully in Chaps. 89 and 374.

LACTIC ACIDOSIS

Lactic acidosis is a rare and potentially fatal metabolic complication of cancer. Lactic acidosis associated with sepsis and circulatory failure is



A



B

FIGURE 71-3 Airway obstruction. **A.** Computed tomography scan of a 62-year-old man with tracheal obstruction caused by renal carcinoma showing paratracheal mass with tracheal invasion/obstruction (arrow). **B.** Chest x-ray of same patient after stent (arrows) placement.

a common preterminal event in many malignancies. Lactic acidosis in the absence of hypoxemia may occur in patients with leukemia, lymphoma, or solid tumors. In some cases, hypoglycemia also is present. Extensive involvement of the liver by tumor is often present. In most cases, decreased metabolism and increased production by the tumor both contribute to lactate accumulation. Tumor cell overexpression of certain glycolytic enzymes and mitochondrial dysfunction can contribute to its increased lactate production. HIV-infected patients have an increased risk of aggressive lymphoma; lactic acidosis that occurs in such patients may be related either to the rapid growth of the tumor or from toxicity of nucleoside reverse transcriptase inhibitors. Symptoms of lactic acidosis include tachypnea, tachycardia, change of mental status, and hepatomegaly. The serum level of lactic acid may reach 10–20 mmol/L (90–180 mg/dL). Treatment is aimed at the underlying disease. *The danger from lactic acidosis is from the acidosis, not the lactate.* Sodium bicarbonate should be added if acidosis is very severe or if hydrogen ion production is very rapid and uncontrolled. Other treatment options include renal replacement therapy, such as hemodialysis, and thiamine replacement. The prognosis is poor regardless of the treatment offered.

HYPOGLYCEMIA

Persistent hypoglycemia is occasionally associated with tumors other than pancreatic islet cell tumors. Usually these tumors are large; tumors of mesenchymal origin, hepatomas, or adrenocortical tumors may cause hypoglycemia. Mesenchymal tumors are usually located in the retroperitoneum or thorax. Obtundation, confusion,

and behavioral aberrations occur in the postabsorptive period and may precede the diagnosis of the tumor. These tumors often secrete incompletely processed insulin-like growth factor II (IGF-II), a hormone capable of activating insulin receptors and causing hypoglycemia. Tumors secreting incompletely processed big IGF-II are characterized by an increased IGF-II to IGF-I ratio, suppressed insulin and C-peptide level, and inappropriately low growth hormone and β -hydroxybutyrate concentrations. Rarely, hypoglycemia is due to insulin secretion by a non-islet cell carcinoma. The development of hepatic dysfunction from liver metastases and increased glucose consumption by the tumor can contribute to hypoglycemia. If the tumor cannot be resected, hypoglycemia symptoms may be relieved by the administration of glucose, glucocorticoids, recombinant growth hormone, or glucagon.

Hypoglycemia can be artifactual; hyperleukocytosis from leukemia, myeloproliferative diseases, leukemoid reactions, or colony-stimulating factor treatment can increase glucose consumption in the test tube after blood is drawn, leading to pseudohypoglycemia.

ADRENAL INSUFFICIENCY

In patients with cancer, adrenal insufficiency may go unrecognized because the symptoms, such as nausea, vomiting, anorexia, and orthostatic hypotension, are nonspecific and may be mistakenly attributed to progressive cancer or to therapy. Primary adrenal insufficiency may develop owing to replacement of both glands by metastases (lung, breast, colon, or kidney cancer; lymphoma), to removal of both glands, or to hemorrhagic necrosis in association with sepsis or anticoagulation. Impaired adrenal steroid synthesis occurs in patients being treated for cancer with mitotane, ketoconazole, or aminoglutethimide or undergoing rapid reduction in glucocorticoid therapy. Megestrol acetate, used to manage cancer and HIV-related cachexia, may suppress plasma levels of cortisol and adrenocorticotrophic hormone (ACTH). Patients taking megestrol may develop adrenal insufficiency, and even those whose adrenal dysfunction is not symptomatic may have inadequate adrenal reserve if they become seriously ill. Paradoxically, some patients may develop Cushing's syndrome and/or hyperglycemia because of the glucocorticoid-like activity of megestrol acetate. Ipilimumab, an anti-CTLA-4 antibody used for treatment of malignant melanoma, may cause autoimmunity including autoimmune-like enterocolitis, hypophysitis, (leading to secondary adrenal insufficiency), hepatitis, and, rarely, primary adrenal insufficiency. Autoimmune hypophysitis may present with headache, visual field defects, and pituitary hormone deficiencies manifesting as hypopituitarism, adrenal insufficiency (including adrenal crisis), or hypothyroidism. Ipilimumab-associated hypophysitis symptoms occur at an average of 6–12 weeks after initiation of therapy. An MRI usually shows homogeneous enhancement of pituitary gland. Early glucocorticoid treatment and hormone replacement are the initial treatment. The role of high-dose glucocorticoids in the treatment of hypophysitis is not clear. High-dose glucocorticoids may not improve the frequency of pituitary function recovery. Autoimmune adrenalitis can also be observed with anti-CTLA-4 antibody. Pituitary dysfunction is usually permanent, requiring long term hormone replacement therapy. Other checkpoint inhibitors, monoclonal antibodies targeting program death-1 (PD-1), an inhibitory receptor expressed by T cells or one of its ligands (PD-L1) may cause hypophysitis infrequently (~1%). Autoimmune adrenalitis is more frequent with use of PD/PD-L1 than with CTLA-4 inhibitors, but incidence is low. Cranial irradiation for childhood brain tumors may affect the hypothalamus-pituitary-adrenal axis, resulting in secondary adrenal insufficiency. Rarely, metastatic replacement causes primary adrenal insufficiency as the first manifestation of an occult malignancy. Metastasis to the pituitary or hypothalamus is found at autopsy in up to 5% of patients with cancer, but associated secondary adrenal insufficiency is rare.

Acute adrenal insufficiency is potentially lethal. Treatment of suspected adrenal crisis is initiated after the sampling of serum cortisol and ACTH levels ([Chap. 379](#)).

TREATMENT-RELATED EMERGENCIES

TUMOR LYYSIS SYNDROME

Tumor lysis syndrome (TLS) is characterized by hyperuricemia, hyperkalemia, hyperphosphatemia, and hypocalcemia, and is caused by the destruction of a large number of rapidly proliferating neoplastic cells. Acidosis may also develop. Acute renal failure occurs frequently.

TLS is most often associated with the treatment of Burkitt's lymphoma, acute lymphoblastic leukemia, and other rapidly proliferating lymphomas, but it also may be seen with chronic leukemias and, rarely, with solid tumors. This syndrome has been seen in patients with chronic lymphocytic leukemia after treatment with nucleosides like fludarabine and is increased in frequency in lymphoid neoplasms treated with venetoclax, a bcl-2 antagonist. TLS has been observed with administration of glucocorticoids, hormonal agents such as letrozole and tamoxifen, and monoclonal antibodies such as rituximab and gemtuzumab. TLS usually occurs during or shortly (1–5 days) after chemotherapy. Rarely, spontaneous necrosis of malignancies causes TLS.

Hyperuricemia may be present at the time of chemotherapy. Effective treatment kills malignant cells and leads to increased serum uric acid levels from the turnover of nucleic acids. Owing to the acidic local environment, uric acid can precipitate in the tubules, medulla, and collecting ducts of the kidney, leading to renal failure. Lactic acidosis and dehydration may contribute to the precipitation of uric acid in the renal tubules. The finding of uric acid crystals in the urine is strong evidence for uric acid nephropathy. The ratio of urinary uric acid to urinary creatinine is >1 in patients with acute hyperuricemic nephropathy and <1 in patients with renal failure due to other causes.

Hyperphosphatemia, which can be caused by the release of intracellular phosphate pools by tumor lysis, produces a reciprocal depression in serum calcium, which causes severe neuromuscular irritability and tetany. Deposition of calcium phosphate in the kidney and hyperphosphatemia may cause renal failure. Potassium is the principal intracellular cation, and massive destruction of malignant cells may lead to hyperkalemia. Hyperkalemia in patients with renal failure may rapidly become life threatening by causing ventricular arrhythmias and sudden death.

The likelihood that TLS will occur in patients with Burkitt's lymphoma is related to the tumor burden and renal function. Hyperuricemia and high serum levels of lactate dehydrogenase (LDH >1500 U/L), both of which correlate with total tumor burden, also correlate with the risk of TLS. In patients at risk for TLS, pretreatment evaluations should include a complete blood count, serum chemistry evaluation, and urine analysis. High leukocyte and platelet counts may artificially elevate potassium levels ("pseudohyperkalemia") due to lysis of these cells after the blood is drawn. In these cases, plasma potassium instead of serum potassium should be followed. In pseudohyperkalemia, no electrocardiographic abnormalities are present. In patients with abnormal baseline renal function, the kidneys and retroperitoneal area should be evaluated by sonography and/or CT to rule out obstructive uropathy. Urine output should be watched closely.

TREATMENT

Tumor Lysis Syndrome

Recognition of risk and prevention are the most important steps in the management of this syndrome (Fig. 71-4). The standard preventive approach consists of allopurinol and aggressive hydration. Urinary alkalization with sodium bicarbonate is no longer recommended. It increases uric acid solubility, but a high pH decreases the solubility of xanthine, hypoxanthine, and calcium phosphate, potentially increasing the likelihood of intratubular crystallization. Intravenous allopurinol may be given in patients who cannot tolerate oral therapy. Febuxostat, a potent nonpurine selective xanthine oxidase inhibitor, is indicated for treatment of hyperuricemia. It has less hypersensitivity reactions than allopurinol. Febuxostat does not require dosage adjustment in patients with mild to moderate renal impairment. Febuxostat achieved significantly superior serum uric

acid control in comparison to allopurinol in patients with hematologic malignancies at intermediate to high TLS risk. In some cases, uric acid levels cannot be lowered sufficiently with the standard preventive approach. Rasburicase (recombinant urate oxidase) can be effective in these instances, particularly when renal failure is present. Urate oxidase is missing from primates and catalyzes the conversion of poorly soluble uric acid to readily soluble allantoin. Rasburicase acts rapidly, decreasing uric acid levels within hours; however, it may cause hypersensitivity reactions such as bronchospasm, hypoxemia, and hypotension. Rasburicase should also be administered to high-risk patients for TLS prophylaxis. Rasburicase is contraindicated in patients with glucose-6-phosphate dehydrogenase deficiency who are unable to break down hydrogen peroxide, an end product of the urate oxidase reaction. Rasburicase is known to cause ex vivo enzymatic degradation of uric acid in test tube at room temperature. This leads to spuriously low uric acid levels during laboratory monitoring of the patient with TLS. Samples must be cooled immediately to deactivate the urate oxidase. Despite aggressive prophylaxis, TLS and/or oliguric or anuric renal failure may occur. Dialysis is often necessary and should be considered early in the course. Hemodialysis is preferred. Hemofiltration offers a gradual, continuous method of removing cellular by-products and fluid.

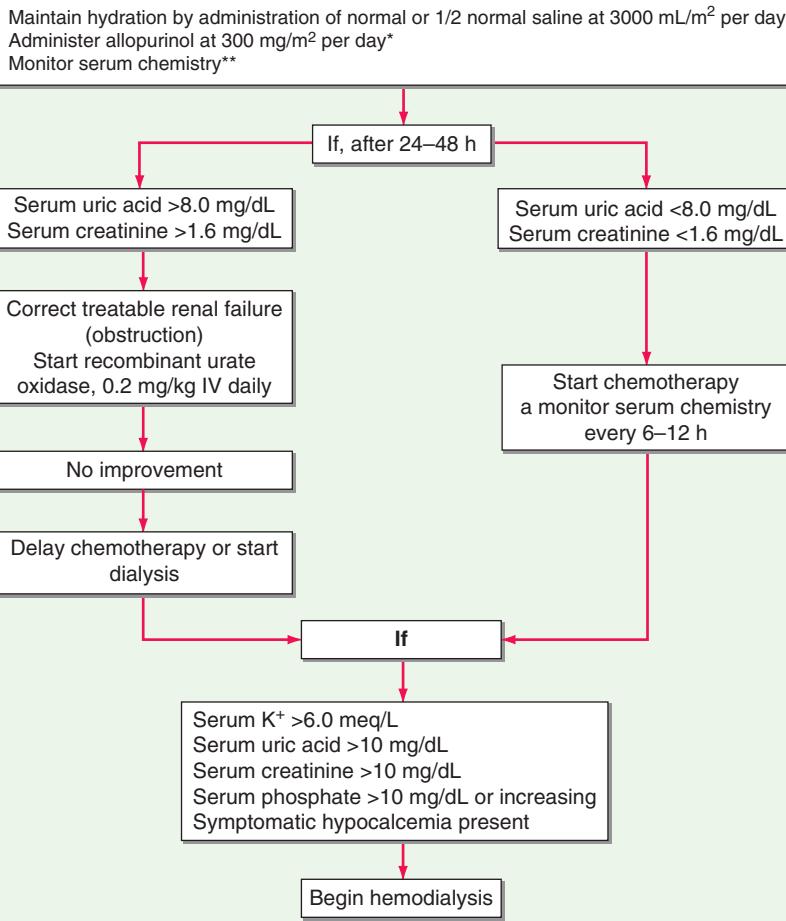
HUMAN ANTIBODY INFUSION REACTIONS

The initial infusion of human or humanized antibodies (e.g., rituximab, gemtuzumab, trastuzumab, alemtuzumab, panitumumab, brentuximab vedotin, blinatumomab) is associated with fever, chills, nausea, asthenia, and headache in up to half of treated patients. Bronchospasm and hypotension occur in 1% of patients. Severe manifestations including pulmonary infiltrates, acute respiratory distress syndrome (ARDS), and cardiogenic shock occur rarely. Laboratory manifestations include elevated hepatic aminotransferase levels, thrombocytopenia, and prolongation of prothrombin time. The pathogenesis is thought to be activation of immune effector processes (cells and complement) and release of inflammatory cytokines, such as tumor necrosis factor α , interferon gamma, interleukin 6, and interleukin 10 (cytokine release syndrome [CRS]). Although its origins are not completely understood, CRS is believed to be due to activation of a variety of cell types including monocytes/macrophages and T and B lymphocytes. Severe reactions from rituximab have occurred with high numbers ($>50 \times 10^9$ lymphocytes) of circulating cells bearing the target antigen (CD20) and have been associated with a rapid fall in circulating tumor cells, mild electrolyte evidence of TLS, and, very rarely, death. In addition, increased liver enzymes, D-dimer, and LDH and prolongation of the prothrombin time may occur. Diphenhydramine, hydrocortisone, and acetaminophen can often prevent or suppress the infusion-related symptoms. If they occur, the infusion is stopped and restarted at half the initial infusion rate after the symptoms have abated. Severe CRS may require intensive support for ARDS and resistant hypotension. Emerging clinical experience at several institutions has concluded that tocilizumab is an effective treatment for severe or life-threatening CRS. Tocilizumab prevents IL-6 binding to both cell-associated and soluble IL-6Rs and therefore inhibits both classical and trans-IL-6 signaling.

Adoptive transfer of chimeric antigen receptor (CAR)-engineered T cells is a promising therapy for cancers. The most common acute toxicity of CAR T cells is CRS. CAR T-cell-associated CRS may be associated cardiac dysfunction and neurotoxicity. The management includes supportive care and tocilizumab.

HEMOLYTIC-UREMIC SYNDROME

Hemolytic-uremic syndrome (HUS) and, less commonly, thrombotic thrombocytopenic purpura (TTP) (Chap. 311) may rarely occur after treatment with antineoplastic drugs, including mitomycin, gemcitabine, cisplatin, and bleomycin, and with VEGF inhibitors. Mitomycin and gemcitabine are the most common offenders. Unlike mitomycin, there is no clear-cut relationship between the cumulative dose of gemcitabine and risk of HUS. It occurs most often in patients with gastric, colorectal, pancreatic, and breast carcinoma. In one series, 35%

PATIENT MANAGEMENT FOR TUMOR LYSIS**FIGURE 71-4 Management of patients at high risk for the tumor lysis syndrome.**

of patients were without evident cancer at the time this syndrome appeared. Secondary HUS/TTP has also been reported as a rare but sometimes fatal complication of BMT.

HUS usually has its onset 4–8 weeks after the last dose of chemotherapy, but it is not rare to detect it several months later. HUS is characterized by microangiopathic hemolytic anemia, thrombocytopenia, and renal failure. Dyspnea, weakness, fatigue, oliguria, and purpura are also common initial symptoms and findings. Systemic hypertension and pulmonary edema frequently occur. Severe hypertension, pulmonary edema, and rapid worsening of hemolysis and renal function may occur after a blood or blood product transfusion. Cardiac findings include atrial arrhythmias, pericardial friction rub, and pericardial effusion. Raynaud's phenomenon is part of the syndrome in patients treated with bleomycin.

Laboratory findings include severe to moderate anemia associated with red blood cell fragmentation and numerous schistocytes on peripheral smear. Reticulocytosis, decreased plasma haptoglobin, and an LDH level document hemolysis. The serum bilirubin level is usually normal or slightly elevated. The Coombs' test is negative. The white cell count is usually normal, and thrombocytopenia (<100,000/ μ L) is almost always present. Most patients have a normal coagulation profile, although some have mild elevations in thrombin time and in levels of fibrin degradation products. The serum creatinine level is elevated at presentation and shows a pattern of subacute worsening within weeks of the initial azotemia. The urinalysis reveals hematuria, proteinuria, and granular or hyaline casts, and circulating immune complexes may be present.

The basic pathologic lesion appears to be deposition of fibrin in the walls of capillaries and arterioles, and these deposits are similar to those seen in HUS due to other causes. These microvascular

abnormalities involve mainly the kidneys and rarely occur in other organs. The pathogenesis of cancer treatment-related HUS is not completely understood, but probably the most important factor is endothelial damage. Primary forms of HUS/TTP are related to a decrease in processing of von Willebrand factor by a protease called ADAMTS13.

The case fatality rate is high; most patients die within a few months. There is no consensus on the optimal treatment for chemotherapy-induced HUS. Treatment modalities for HUS/TTP including immunocomplex removal (plasmapheresis, immunoabsorption, or exchange transfusion), antiplatelet/anticoagulant therapies, immunosuppressive therapies, and plasma exchange have varying degrees of success. The outcome with plasma exchange is generally poor, as in many other cases of secondary TTP. Rituximab is successfully used in patients with chemotherapy-induced HUS as well as in ADAMTS13-deficient TTP.

■ NEUTROPENIA AND INFECTION

These remain the most common serious complications of cancer therapy. They are covered in detail in Chap. 70.

■ PULMONARY INFILTRATES

Patients with cancer may present with dyspnea associated with diffuse interstitial infiltrates on chest radiographs. Such infiltrates may be due to progression of the underlying malignancy, treatment-related toxicities, infection, and/or unrelated diseases. The cause may be multifactorial; however, most commonly they occur as a consequence of treatment. Infiltration of the lung by malignancy has been described in patients with leukemia, lymphoma, and breast and other solid cancers. Pulmonary lymphatics may be involved diffusely by neoplasm (pulmonary lymphangitic carcinomatosis), resulting in a diffuse increase in interstitial markings on chest radiographs.

The patient is often mildly dyspneic at the onset, but pulmonary failure develops over a period of weeks. In some patients, dyspnea precedes changes on the chest radiographs and is accompanied by a nonproductive cough. This syndrome is characteristic of solid tumors. In patients with leukemia, diffuse microscopic neoplastic peribronchial and peribronchiolar infiltration is frequent but may be asymptomatic. However, some patients present with diffuse interstitial infiltrates, an alveolar capillary block syndrome, and respiratory distress. Thickening of bronchovascular bundles and prominence of peripheral arteries are CT findings suggestive of leukemic infiltration. In these situations, glucocorticoids can provide symptomatic relief, but specific chemotherapy should always be started promptly.

Several cytotoxic agents, such as bleomycin, methotrexate, busulfan, nitrosoureas, gemcitabine, mitomycin, vinorelbine, docetaxel, paclitaxel, fludarabine, pentostatin, and ifosfamide may cause pulmonary damage. The most frequent presentations are interstitial pneumonitis, alveolitis, and pulmonary fibrosis. Some cytotoxic agents, including methotrexate and procarbazine, may cause an acute hypersensitivity reaction. Cytosine arabinoside has been associated with noncardiogenic pulmonary edema. Administration of multiple cytotoxic drugs, as well as radiotherapy and preexisting lung disease, may potentiate the pulmonary toxicity. Supplemental oxygen may potentiate the effects of drugs and radiation injury. Patients should always be managed with the lowest F_{IO_2} that is sufficient to maintain hemoglobin saturation.

The onset of symptoms may be insidious, with symptoms including dyspnea, nonproductive cough, and tachycardia. Patients may have bibasilar crepitant rales, end-inspiratory crackles, fever, and cyanosis. The chest radiograph generally shows an interstitial and sometimes an

intraalveolar pattern that is strongest at the lung bases and may be symmetric. A small effusion may occur. Hypoxemia with decreased carbon monoxide diffusing capacity is always present. Glucocorticoids may be helpful in patients in whom pulmonary toxicity is related to radiation therapy or to chemotherapy. Treatment is otherwise supportive.

Molecular targeted agents, imatinib, erlotinib, and gefitinib are potent inhibitors of tyrosine kinases. These drugs may cause interstitial lung disease (ILD). In the case of gefitinib, preexisting fibrosis, poor performance status, and prior thoracic irradiation are independent risk factors; this complication has a high fatality rate. In Japan, incidence of ILD associated with gefitinib was ~4.5% compared to 0.5% in the United States. Temsirolimus and everolimus, both esters of a derivative of rapamycin, are agents that block the effects of mammalian target of rapamycin (mTOR), an enzyme that has an important role in regulating the synthesis of proteins that control cell division. It may cause ground-glass opacities in the lung with or without diffuse interstitial disease and lung parenchymal consolidation. Patients may be asymptomatic with only radiologic findings or may be symptomatic. Symptoms include cough, dyspnea, and/or hypoxemia, and sometimes patients present with systemic symptoms such as fever and fatigue. The incidence of everolimus-induced ILD also appears to be higher in Japanese patients. Treatment includes dose reduction or withdrawal and, in some cases, the addition of glucocorticoids.

The Food and Drug Administration (FDA)-approved immune checkpoint inhibitors of the PD-1 and PD-L1 pathway, including nivolumab, pembrolizumab, durvalumab, avelumab, and atezolizumab, enhance antitumor activity by blocking negative regulators of T cell function. Immune-mediated pneumonitis is rare (10%) but a life-threatening complication of these drugs. Pneumonitis symptoms include cough, shortness of breath, dyspnea, and fever, and often involve only asymptomatic radiographic changes. Pneumonitis shows ground-glass patchy lesions and/or disseminated nodular infiltrates, predominantly in the lower lobes. Treatment includes temporary or permanent withdrawal of drug and the addition of high-dose glucocorticoids.

Radiation pneumonitis and/or fibrosis are relatively frequent side effects of thoracic radiation therapy. It may be acute or chronic. Radiation-induced lung toxicity is a function of the irradiated lung volume, dose per fraction, and radiation dose. The larger the irradiated lung field, the higher is the risk for radiation pneumonitis. The use of concurrent chemoradiation, particularly regimens including paclitaxel, increases pulmonary toxicity. Radiation pneumonitis usually develops 2–6 months after completion of radiotherapy. The clinical syndrome, which varies in severity, consists of dyspnea, cough with scanty sputum, low-grade fever, and an initial hazy infiltrate on chest radiographs. The infiltrate and tissue damage usually are confined to the radiation field. The CT scan may show ground-glass opacities, consolidation, fibrosis, atelectatic cicatrization, pleural volume loss, or pleural thickening. The patients subsequently may develop a patchy alveolar infiltrate and air bronchograms, which may progress to acute respiratory failure that is sometimes fatal. A lung biopsy may be necessary to make the diagnosis. Asymptomatic infiltrates found incidentally after radiation therapy need not be treated. However, prednisone should be administered to patients with fever or other symptoms. The dosage should be tapered slowly after the resolution of radiation pneumonitis, because abrupt withdrawal of glucocorticoids may cause an exacerbation of pneumonia. Delayed radiation fibrosis may occur years after radiation therapy and is signaled by dyspnea on exertion. Often it is mild, but it can progress to chronic respiratory failure. Therapy is supportive.

Classic radiation pneumonitis that leads to pulmonary fibrosis is due to radiation-induced production of local cytokines such as platelet-derived growth factor β , tumor necrosis factor, interleukins, and transforming growth factor β in the radiation field.

Stereotactic body radiation therapy (SBRT) is a radiotherapy treatment method that has been applied to the treatment of stage I lung cancers in medically inoperable patients. SBRT accurately delivers a high dose of irradiation in one or few treatment fractions to an image-defined lung mass. Most of the acute changes after SBRT occur

later than 3 months after treatment, and the shape of the SBRT-induced injury conforms more tightly to the tumor.

Pneumonia is a common problem in patients undergoing treatment for cancer (see Chap 70). In patients with pulmonary infiltrates who are febrile, heart failure and multiple pulmonary emboli are in the differential diagnosis.

■ NEUTROPENIC ENTEROCOLITIS

Neutropenic enterocolitis (typhlitis) is the inflammation and necrosis of the cecum and surrounding tissues that may complicate the treatment of acute leukemia. Nevertheless, it may involve any segment of the gastrointestinal tract including small intestine, appendix, and colon. This complication has also been seen in patients with other forms of cancer treated with taxanes, 5-fluorouracil, irinotecan, vinorelbine, cisplatin, carboplatin, and high-dose chemotherapy (Fig. 71-5). It also has been reported in patients with AIDS, aplastic anemia, cyclic neutropenia, idiosyncratic drug reactions involving antibiotics, and immunosuppressive therapies. The patient develops right lower quadrant abdominal pain, often with rebound tenderness and a tense, distended abdomen, in a setting of fever and neutropenia. Watery diarrhea (often containing sloughed mucosa) and bacteremia are common, and bleeding may occur. Plain abdominal films are generally of little value in the diagnosis; CT scan may show marked bowel wall thickening, particularly in the cecum, with bowel wall edema, mesenteric stranding, and



A

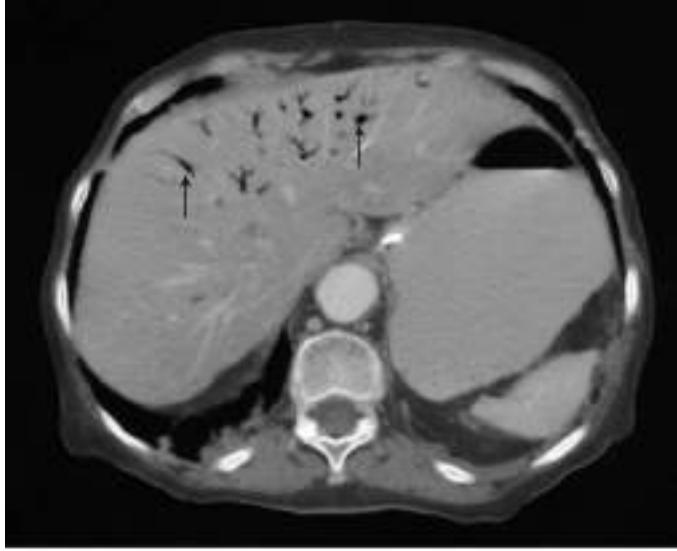


FIGURE 71-5 Abdominal computed tomography (CT) scans of a 72-year-old woman with neutropenic enterocolitis secondary to chemotherapy. **A.** Air in inferior mesenteric vein (arrow) and bowel wall with pneumatosis intestinalis. **B.** CT scan of upper abdomen demonstrating air in portal vein (arrows).

ascites, and may help to differentiate neutropenic colitis from other abdominal disorders such as appendicitis, diverticulitis, and *Clostridium difficile*-associated colitis in this high-risk population. Patients with bowel wall thickness >10 mm on ultrasonogram have higher mortality rates. However, bowel wall thickening is significantly more prominent in patients with *C. difficile* colitis. Pneumatosis intestinalis is a more specific finding, seen only in those with neutropenic enterocolitis and ischemia. The combined involvement of the small and large bowel suggests a diagnosis of neutropenic enterocolitis. Rapid institution of broad-spectrum antibiotics, bowel rest, and nasogastric suction may reverse the process. Use of myeloid growth factors improved outcome significantly. Surgical intervention is reserved for severe cases of neutropenic enterocolitis with evidence of perforation, peritonitis, gangrenous bowel, or gastrointestinal hemorrhage despite correction of any coagulopathy.

C. difficile colitis is increasing in incidence. Newer strains of *C. difficile* produce ~20 times more of toxins A and B compared to previously studied strains. *C. difficile* risk is also increased with chemotherapy. Antibiotic coverage for *C. difficile* should be added if pseudomembranous colitis cannot be excluded.

■ HEMORRHAGIC CYSTITIS

Hemorrhagic cystitis is characterized by diffuse bladder mucosal bleeding that develops secondary to chemotherapy (mostly cyclophosphamide or ifosfamide), radiation therapy, bone marrow transplantation (BMT), and/or opportunistic infections. Both cyclophosphamide and ifosfamide are metabolized to acrolein, which is a strong chemical irritant that is excreted in the urine. Prolonged contact or high concentrations may lead to bladder irritation and hemorrhage. Symptoms include gross hematuria, frequency, dysuria, burning, urgency, incontinence, and nocturia. The best management is prevention. Maintaining a high rate of urine flow minimizes exposure. In addition, 2-mercaptopethanesulfonate (mesna) detoxifies the metabolites and can be coadministered with the instigating drugs. Mesna usually is given three times on the day of ifosfamide administration in doses that are each 20% of the total ifosfamide dose. If hemorrhagic cystitis develops, the maintenance of a high urine flow may be sufficient supportive care. If conservative management is not effective, irrigation of the bladder with a 0.37–0.74% formalin solution for 10 min stops the bleeding in most cases. *N*-Acetylcysteine may also be an effective irrigant. Prostaglandin (carboprost) can inhibit the process. In extreme cases, ligation of the hypogastric arteries, urinary diversion, or cystectomy may be necessary.

In the BMT setting, early-onset hemorrhagic cystitis is related to drugs in the treatment regimen (e.g., cyclophosphamide), and late-onset hemorrhagic cystitis is usually due to the polyoma virus BKV or adenovirus type 11. BKV load in urine alone or in combination with acute graft-versus-host disease correlates with development of hemorrhagic cystitis. Viral causes are usually detected by polymerase chain reaction (PCR)-based diagnostic tests. Treatment of viral hemorrhagic cystitis is largely supportive, with reduction in doses of immunosuppressive agents, if possible. No antiviral therapy is approved, although cidofovir is reported to be effective in a small series. Hyperbaric oxygen therapy has been used successfully in patients with BKV-associated and cyclophosphamide-induced hemorrhagic cystitis during hematopoietic stem cell transplantation, as well as in hemorrhagic radiation cystitis.

■ HYPERSENSITIVITY REACTIONS TO ANTINEOPLASTIC DRUGS

Many antineoplastic drugs may cause hypersensitivity reaction. These reactions are unpredictable and potentially life threatening. Most reactions occur during or within hours of parenteral drug administration. Taxanes, platinum compounds, asparaginase, etoposide, procarbazine, and biologic agents, including rituximab, bevacizumab, trastuzumab, gemtuzumab, cetuximab, and alemtuzumab, are more commonly associated with acute hypersensitivity reactions than are other agents. Hypersensitivity reactions to some drugs, such as taxanes, occur during the first or second dose administered. Hypersensitivity to platinum compounds occurs after prolonged exposure. Skin testing may identify

patients with high risk for hypersensitivity after carboplatin exposure. Premedication with histamine H₁ and H₂ receptor antagonists and glucocorticoids reduces the incidence of hypersensitivity reaction to taxanes, particularly paclitaxel. Despite premedication, hypersensitivity reactions may still occur. In these cases, rapid desensitization in the intensive care unit setting or re-treatment may be attempted with care, but the use of alternative agents may be required. Skin testing is used to assess the involvement of IgE in the reaction. Tryptase levels measured at the time of the reaction help to explain the mechanism of the reaction and its severity. Increased tryptase levels indicate underlying mast cell activation. Candidate patients for desensitization include those who have mild to severe hypersensitivity type I, with mast cell-mediated and IgE-dependent reactions occurring during a chemotherapy infusion or shortly thereafter.

■ FURTHER READING

- BAUER R et al: Treatment of epileptic seizures in brain tumors: A critical review. *Neurosurg Rev* 37:381, 2014.
- BODNAR TW: Management of non-islet-cell tumor hypoglycemia: A clinical review. *J Clin Endocrinol Metab* 99:713, 2014.
- JONES GL et al: Guidelines for the management of tumour lysis syndrome in adults and children with haematological malignancies on behalf of the British Committee for Standards in Haematology. *Br J Haematol* 169:661, 2015.
- LAUFER I et al: The NOMS framework: Approach to the treatment of spinal metastatic tumors. *Oncologist* 18:744, 2013.
- LEE DW et al: Current concepts in the diagnosis and management of cytokine release syndrome. *Blood* 124:188, 2014.
- LIN X, DEANGELIS LM: Treatment of brain metastases. *J Clin Oncol* 33:3475, 2015.
- MACK F et al: Therapy of leptomeningeal metastasis in solid tumors. *Cancer Treat Rev* 43:83, 2016.
- NISHINO M et al: Anti-PD-1 inhibitor-related pneumonitis in non-small cell lung cancer. *Cancer Immunol Res* 4:289, 2016.
- RICE TW et al: The superior vena cava syndrome: Clinical characteristics and evolving etiology. *Medicine* 85:37, 2006.
- RUGGIERO A et al: Management of hyperleukocytosis. *Curr Treat Options Oncol* 17:7, 2016.

72

Cancer of the Skin

Brendan D. Curti, Sancy Leachman,
Walter J. Urba



MELANOMA

Pigmented lesions are among the most common findings on skin examination. The challenge for the physician is to distinguish cutaneous melanomas, which account for the overwhelming majority of deaths resulting from skin cancer, from the remainder, which are usually benign. Cutaneous melanoma can occur in adults of all ages, even young individuals, and people of all colors; its location on the skin and its distinct clinical features often permit detection at a time when complete surgical excision leads to cure. Examples of malignant and benign pigmented lesions are shown in Fig. 72-1.

■ EPIDEMIOLOGY

Melanoma is an aggressive malignancy of melanocytes, pigment-producing cells that originate from the neural crest and migrate to the skin, meninges, mucous membranes, upper esophagus, and eyes. Melanocytes in each of these locations have the potential for malignant transformation, but the vast majority arise in the skin. Melanomas can also arise in the mucosa of the head and neck (nasal cavity, paranasal sinuses, and oral cavity), the gastrointestinal tract, the CNS, the female genital tract (vulva, vagina), and the uveal tract of the eye. Cutaneous melanoma is predominantly a malignancy of white-skinned people

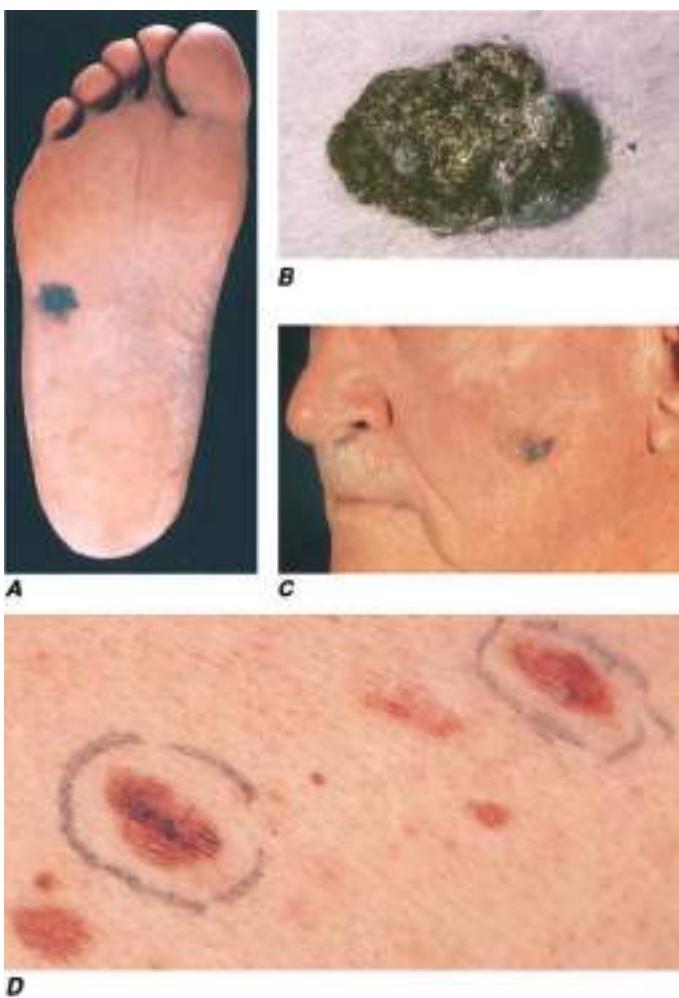


FIGURE 72-1 Atypical and malignant pigmented lesions. The most common melanoma is superficial spreading melanoma (not pictured). **A.** Acral lentiginous melanoma is the most common melanoma in blacks, Asians, and Hispanics and occurs as an enlarging hyperpigmented macule or plaque on the palms and soles. **B.** Nodular melanoma most commonly manifests as a rapidly growing, often ulcerated or crusted black nodule. **C.** Lentigo maligna melanoma occurs on sun-exposed skin as a large, hyperpigmented macule or plaque with irregular borders and variable pigmentation. **D.** Dysplastic nevi are benign, irregularly pigmented and shaped melanocytic hamartomas with some atypical cellular features and frequently associated with familial melanoma.

(98% of cases), and the incidence correlates with latitude of residence, providing strong evidence for the role of sun exposure. Men are affected slightly more than women (1.3:1), and the median age at diagnosis is the late fifties. In 2016, >76,000 individuals in the United States were expected to develop melanoma, and ~10,130 were expected to die. Mortality rates begin to rise at age 55, with the greatest increase in men age >65 years. Of particular concern is the increase in incidence among women <40 years of age, an increase believed to be associated with a greater emphasis on tanned skin as a marker of beauty, the increased availability and use of indoor tanning beds, and exposure to intense ultraviolet (UV) light in childhood. The latest Surveillance, Epidemiology and End Results (SEER) Registry data reveal that from 2004 to 2013, the rate of new melanoma cases has risen 1.4% each year, while death rates have remained stable. This is in the context of a 5-year relative survival improvement from 93.1% to 93.3% overall, despite a 17.9% survival rate for those diagnosed with distant metastases. These statistics highlight the need to promote prevention and early detection.

■ GLOBAL CONSIDERATIONS

The incidence of both non-melanoma and melanoma skin cancers around the world has been increasing. Every year between 2 and 3 million people will get non-melanoma skin

cancer and in 2012 there were 232,000 cases of melanoma. The highest incidence of melanoma is found in New Zealand and Australia consistent with Caucasians living in latitudes with increased UV exposure. The likelihood of developing melanoma is 25 per 100,000 in non-Hispanic whites, 4 per 100,000 in Hispanics, and 1 per 100,000 in African Americans.

Dark-skinned populations (such as those of India and Puerto Rico), blacks, and east Asians also develop melanoma, albeit at rates 10–20 times lower than those in whites. Cutaneous melanomas in these populations are more often diagnosed at a higher stage, and patients tend to have worse outcomes. Furthermore, in nonwhite populations, the frequency of acral (subungual, plantar, palmar) and mucosal melanomas is much higher. In China, about 20,000 new cases are reported each year and, in contrast to the United States where rates are stable, mortality is increasing. This may be due in part to the gap that remains in the diagnosis and treatment of melanoma between China and Western countries or to the fact that in Asians and dark-skinned populations, the melanomas that arise from the skin (comprising 50–70% of patients versus 90% in the West) arise from acral areas and the others from mucosal areas, all of which carry a poorer prognosis than cutaneous melanomas diagnosed in the West.

■ RISK FACTORS

Presence of Nevi The risk of developing melanoma is related to genetic, environmental, and host factors. The strongest risk factors for melanoma are the presence of multiple benign or atypical nevi and a family or personal history of melanoma. The presence of >40 melanocytic nevi, common or dysplastic, is a marker for increased risk of melanoma. Nevi have been referred to as precursor lesions because they can transform into melanomas; however, the actual risk of transformation for any individual nevus is exceedingly low. About one-quarter of melanomas are histologically associated with nevi, but the majority arise de novo. The number of clinically atypical moles may vary from one to several hundred, and they usually differ from one another in appearance, although individuals can develop multiple similar atypical nevi (signature nevi). The borders are often hazy and indistinct, and the pigment pattern is more highly varied than that in benign acquired nevi. Individuals with clinically atypical moles and a strong family history of melanoma have been reported to have a >50% lifetime risk for developing melanoma and warrant close follow-up with a dermatologist. Of the 90% of patients whose disease is sporadic (i.e., who lack a family history of melanoma), ~40% have clinically atypical moles, compared with an estimated 5–10% of the population at large.

Congenital melanocytic nevi, which are classified as small (<1.5 cm), medium (1.5–20 cm), and giant (>20 cm), can be precursors for melanoma. The risk is highest for the giant melanocytic nevus, also called the bathing trunk nevus, a rare malformation that affects 1 in 30,000–100,000 individuals. Since the lifetime risk of melanoma development is estimated to be as high as 6%, prophylactic excision early in life is prudent. This usually requires staged removal with coverage by split-thickness skin grafts. Surgery cannot remove all at-risk nevus cells, as some may penetrate into the muscles or central nervous system (CNS) below the nevus. Small- to medium-size congenital melanocytic nevi affect ~1% of persons; the lifetime risk of melanoma development in a typical nevus is low, estimated to be about 0.03% (1 in 3164) for men and 0.009% (1 in 10,800) for women. The management of small- to medium-size congenital melanocytic nevi remains controversial and is primarily based on histologic findings from biopsies of clinically atypical nevi.

Personal and Family History Once diagnosed, patients with melanoma require a lifetime of surveillance because their risk of developing another melanoma is 10 times that of the general population. First-degree relatives have a twofold higher risk of developing melanoma than do individuals without a family history, but only 5–10% of all melanomas are truly familial. In familial melanoma, patients tend to be younger at first diagnosis, lesions are thinner, and multiple primary melanomas are common.



Genetic Susceptibility Approximately 20–40% of cases of hereditary melanoma (0.2–2% of all melanomas) are due to germline mutations in the cell cycle regulatory gene cyclin-dependent kinase inhibitor 2A (*CDKN2A*). In fact, 70% of all cutaneous melanomas have mutations or deletions affecting the *CDKN2A* locus on chromosome 9p21. This locus encodes two distinct tumor-suppressor proteins from alternate reading frames: p16 and ARF (p14^{ARF}). The p16 protein inhibits CDK4/6-mediated phosphorylation and inactivation of the retinoblastoma (RB) protein, whereas ARF inhibits MDM2 ubiquitin-mediated degradation of p53. The end result of the loss of *CDKN2A* is inactivation of two critical tumor-suppressor pathways, RB and p53, which control entry of cells into the cell cycle. Several studies have shown an increased risk of pancreatic cancer among melanoma-prone families with *CDKN2A* mutations. A second high-risk locus for melanoma susceptibility, *CDK4*, is located on chromosome 12q13 and encodes the kinase inhibited by p16. *CDK4* mutations, which also inactivate the RB pathway, are much rarer than *CDKN2A* mutations. Germline mutations in the melanoma lineage-specific oncogene microphthalmia-associated transcription factor (*MITF*) and telomerase reverse transcriptase (TERT) mutations predispose to both familial and sporadic melanomas.

The melanocortin-1 receptor (*MC1R*) gene is a moderate-risk inherited melanoma susceptibility factor. Solar radiation stimulates the production of melanocortin (α -melanocyte-stimulating hormone [α -MSH]), the ligand for *MC1R*, which is a G-protein-coupled receptor that signals via cyclic AMP and regulates the amount and type of pigment produced. *MC1R* is highly polymorphic, and among its 80 variants are those that result in partial loss of signaling and lead to the production of red/yellow pheomelanins, which are not sun-protective and produce red hair, rather than brown/black eumelanins that are photoprotective. This red hair color (RHC) phenotype is associated with fair skin, red hair, freckles, increased sun sensitivity, and increased risk of melanoma. In addition to its weak UV shielding capacity relative to eumelanin, increased pheomelanin production in patients with inactivating polymorphisms of *MC1R* also provides a UV-independent carcinogenic contribution to melanomagenesis via oxidative damage and reduced DNA damage repair.

A number of other more common, low-penetrance polymorphisms that have small effects on melanoma susceptibility include other genes related to pigmentation, nevus count, immune responses, DNA repair, metabolism, and the vitamin D receptor. Approximately 50% of the genetic risk for hereditary melanoma can be ascribed to previously identified melanoma predisposition genes, with ~40% of the risk being due to *CDKN2A*. The missing inherited risk is most likely due to the inheritance of additional modifier genes and/or shared environmental exposures.

■ PREVENTION AND EARLY DETECTION

Primary prevention of melanoma and nonmelanoma skin cancer (NMSC) is based on protection from the sun. Public health initiatives, such as the SunSmart program that started in Australia and now is operative in Europe and the United States, have demonstrated that behavioral change can decrease the incidence of NMSC and melanoma. Preventive measures should start early in life because damage from UV light begins early despite the fact that cancers develop years later. Some individuals tan compulsively. There is greater understanding of tanning addiction and the biology of cutaneous-neural connections that may give rise to this behavior. Compulsive tanners exhibit differences in dopamine binding and reactivity in reward pathways in the brain, such as the basal striatum, resulting in cutaneous secretion of β -endorphins after UV exposure. Identifying individuals with tanning addiction may be another method for preventive intervention. Regular use of broad-spectrum sunscreens that block UVA and UVB with a sun protection factor (SPF) of at least 30 and protective clothing should be encouraged. Avoidance of sunburns, tanning beds, and midday sun exposure is recommended.

Secondary prevention comprises education, screening, and early detection. Patients should be taught to recognize the clinical features of melanoma (ABCDEs; see below) and advised to report any change in

a pigmented lesion. Brochures are available from the American Cancer Society, the American Academy of Dermatology, the National Cancer Institute, and the Skin Cancer Foundation. Self-examination at monthly intervals may enhance the likelihood of detecting change. Although the U.S. Preventive Services Task Force states that evidence is insufficient to recommend for or against skin cancer screening, a full-body skin exam seems to be a simple, practical way to approach reducing the mortality rate for skin cancer. Depending on the presence or absence of risk factors, strategies for early detection can be individualized. This is particularly true for patients with clinically atypical moles (dysplastic nevi) and those with a personal history of melanoma. For these individuals, surveillance should be performed by the dermatologist and include total-body photography and dermoscopy where appropriate. Individuals with three or more primary melanomas and families with at least one invasive melanoma and two or more cases of melanoma and/or pancreatic cancer among first- or second-degree relatives on the same side of the family may benefit from genetic testing. Precancerous and *in situ* lesions should be treated early. Early detection of small tumors allows the use of simpler treatment modalities with higher cure rates and lower morbidity.

■ DIAGNOSIS

The goal is to identify a melanoma before it invades and life-threatening metastases have occurred. Early detection may be facilitated by applying the ABCDEs: asymmetry (benign lesions are usually symmetric); border irregularity (most nevi have clear-cut borders); color variegation (benign lesions usually have uniform light or dark pigment); diameter >6 mm (the size of a pencil eraser); and evolving (any change in size, shape, color, or elevation or new symptoms such as bleeding, itching, and crusting). In addition, any nevus that appears atypical and different from the rest of the nevi on that individual (an “ugly duckling”) should be considered suspicious.

The entire skin surface, including the scalp and mucous membranes, as well as the nails should be examined in each patient. Bright room illumination is important, and a hand lens is helpful for evaluating variation in pigment pattern. Any suspicious lesions should be biopsied, evaluated by a specialist, or recorded by chart and/or photography for follow-up. A focused method for examining individual lesions, dermoscopy, employs low-level magnification of the epidermis with polarized light and may allow a more precise visualization of patterns of pigmentation than is possible with the naked eye. Additional technologies, including *in vivo* confocal microscopy, multi- and hyper-spectral imaging, optical coherence tomography, gene expression panels, tape stripping, and electrical conductance methods have been developed and are being refined for improved early detection of melanoma.

Biopsy Any pigmented cutaneous lesion that has changed in size or shape or has other features suggestive of malignant melanoma is a candidate for biopsy. An excisional biopsy with 1- to 3-mm margins is suggested though excision can be accomplished tangentially or in a fusiform fashion. This facilitates pathologic assessment of the lesion, permits accurate measurement of thickness if the lesion is melanoma, and constitutes definitive treatment if the lesion is benign. For lesions that are large or on anatomic sites where excisional biopsy may not be feasible (such as the face, hands, and feet), an incisional biopsy through the most nodular or darkest area of the lesion is acceptable. Incisional biopsy does not appear to facilitate the spread of melanoma. For suspicious lesions, every attempt should be made to preserve the ability to assess the deep and peripheral margins and to perform immunohistochemistry. Shave, saucerization or tangential biopsies are an acceptable alternative, particularly if the suspicion of malignancy is low. They should be deep enough to include the deepest component of the entire lesion and any pigment at the base of the lesion should be removed and included with the biopsy specimen. The biopsy should be read by a pathologist experienced in pigmented lesions, and the report should include Breslow thickness, mitotic rate, presence or absence of ulceration and lymphatic invasion, microsatellitosis and peripheral and deep margin status. Breslow thickness is the greatest thickness of

a primary cutaneous melanoma measured on the slide from the top of the epidermal granular layer, or from the ulcer base, to the bottom of the tumor. To distinguish melanomas from benign nevi in challenging cases, fluorescence *in situ* hybridization (FISH) with multiple probes and comparative genome hybridization (CGH) can be helpful. Gene expression profiling assays have been developed to enhance diagnosis but are not yet widely applied.

CLASSIFICATION AND PATHOGENESIS

Clinical The features of five major types of cutaneous melanoma are described in **Table 72-1**. In *superficial spreading melanoma*, *lentigo maligna melanoma*, and *acral lentiginous melanoma*, the lesion has a period of superficial (so-called radial) growth during which it increases in size but does not penetrate deeply. It is during this period that the melanoma is most capable of being cured by surgical excision. A fourth type—*nodular melanoma*—does not have a recognizable radial growth phase and usually presents as a deeply invasive lesion that is capable of early metastasis. Tumors that begin to penetrate deeply into the skin are in the so-called vertical growth phase. Melanomas with a radial growth phase are characterized by irregular and sometimes notched borders, variation in pigment pattern, and variation in color. A fifth type of melanoma, *desmoplastic melanoma*, is associated with a fibrotic response, neural invasion, and a greater tendency for local recurrence. Occasionally, melanomas appear clinically to be amelanotic, in which case the diagnosis is established microscopically after biopsy.

Although these subtypes are clinically and histopathologically distinct, this classification has minimal prognostic value and histologic subtype is not part of American Joint Committee on Cancer (AJCC) staging. Characterizing the genomic and mutational profiles of melanoma has become increasingly common and can reflect the mechanisms of tumorigenesis. These molecular classifications inform treatment and surveillance strategies.

Genomic Considerable evidence from epidemiologic and molecular studies indicate that cutaneous melanomas arise via multiple causal pathways. There are both environmental and genetic components

(susceptibility genes discussed earlier), and the major environmental factor in cutaneous melanogenesis is sun exposure. The major effect of UV solar radiation is to cause genetic changes in the skin. However, it also impairs cutaneous immune function, increases the production of growth factors, and induces the formation of DNA-damaging reactive oxygen species that affect keratinocytes and melanocytes.

The advent of next-generation sequencing (NGS) has led to whole exome sequencing of hundreds of cutaneous melanomas derived from non-glabrous skin. This has revealed a very complex genetic landscape with genetic changes resulting from both germline (described earlier) and somatic mutations. Cutaneous melanomas have one of the highest somatic mutation rates (>10 mutations/Mb) compared to other cancers; the majority (76% primary tumors and 84% of metastatic melanomas) exhibit a mutation signature indicating UVR exposure. The mutation rate varies based on body site; melanomas arising in chronic sun-damaged skin harbor substantially more mutations than melanomas from non-sun-damaged skin.

Melanoma tumors can harbor thousands of mutations, but only a few are driver mutations; a mutation that is causally implicated in oncogenesis by virtue of a conferred growth advantage on the cancer cell. The driver mutations that have been identified for cutaneous melanoma are depicted in **Fig. 72-2**. As more melanomas are sequenced, more driver mutations have been identified. These mutations tend to be found in a smaller fraction of patients. Driver mutations often affect pathways that promote cell proliferation or inhibit normal pathways of apoptosis in response to DNA repair. They are often found in combination with mutations to the genetic susceptibility genes described earlier. The altered melanocytes accumulate DNA damage, and selection occurs for all the attributes that constitute the malignant phenotype: invasion, metastasis, and angiogenesis.

A recent report from the Cancer Genome Atlas (TCGA) has proposed a genomic classification of cutaneous melanoma based on the pattern of the most prevalent significantly mutated genes: BRAF, RAS, NF-1, and triple-WT (wild type). Distinct patterns of DNA mutations can vary with the site of origin and can be independent of the histologic subtype of the tumor. Thus, although the genetic landscape of melanoma is complex, and continues to evolve, the overall pattern of mutation, amplification, and loss of cancer genes indicates they have convergent effects on key biochemical pathways involved in proliferation, senescence, and apoptosis. An advantage of this classification is that these mutations can be used to select therapy.

TABLE 72-1 Major Histologic Subtypes of Malignant Melanoma

TYPE	SITE	AVERAGE AGE AT DIAGNOSIS, YEARS	APPEARANCE
Lentigo maligna melanoma	Sun-exposed surfaces, particularly malar region and temple	70	In flat portions, brown and tan predominate, but whitish gray occasionally present; in nodules, reddish brown, bluish gray, bluish black
Superficial spreading melanoma	Any site (more common on upper back and, in women, lower legs)	40–50	Brown mixed with bluish red, bluish black, reddish brown, and often whitish pink, and the border of lesion is at least in part visibly and/or palpably elevated
Nodular melanoma	Any	40–50	Reddish blue (purple) or bluish black; either uniform in color or mixed with brown or black
Acral lentiginous melanoma	Palm, sole, nail bed, mucous membrane	60	In flat portions, dark brown predominantly; in raised lesions (plaques), brown-black or blue-black predominantly
Desmoplastic melanoma	Any site (more common head and neck)	60	Highly variable, mimics other lesions; pigmentation is frequently absent

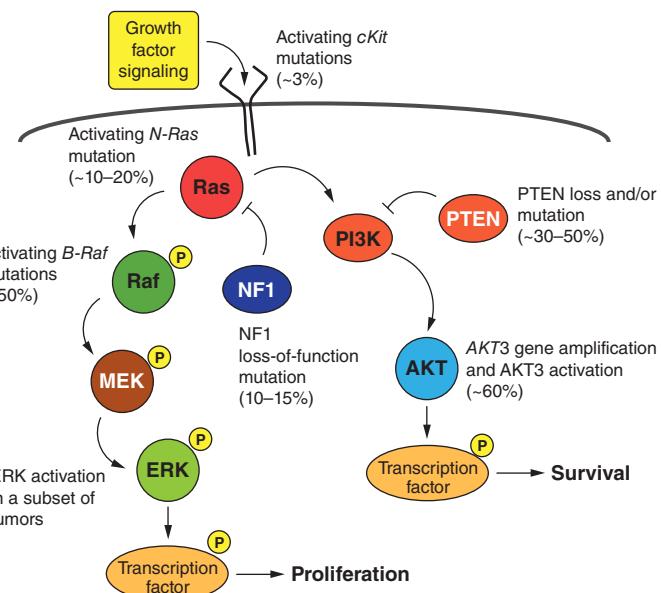


FIGURE 72-2 Major pathways involved in melanoma. The MAP kinase and PI3K/AKT pathways, which promote proliferation and inhibit apoptosis, respectively, are subject to mutations in melanoma. ERK, extracellular signal-regulated kinase; MEK, mitogen-activated protein kinase; NF-1, neurofibromatosis type 1 gene; PTEN, phosphatase and tensin homolog.

The *p16* mutation that affects cell cycle arrest and the *ARF* mutation that results in defective apoptotic responses to genotoxic damage were described earlier. The proliferative pathways affected were the mitogen-activated protein (MAP) kinase and phosphatidylinositol 3' kinase/AKT pathways (Fig. 72-2). *RAS* and *BRAF*, members of the MAP kinase pathway, which classically mediates the transcription of genes involved in cell proliferation and survival, undergo somatic mutation in melanoma and thereby generate potential therapeutic targets. *N-RAS* is mutated in ~20% of melanomas, and somatic activating *BRAF* mutations are found in most benign nevi and 40–50% of cutaneous melanomas. Neither mutation by itself appears to be sufficient to cause melanoma; thus, they often are accompanied by other mutations, such as *TERT*. The *BRAF* mutation is most commonly a point mutation (T→A nucleotide change) that results in a valine-to-glutamate amino acid substitution (V600E). V600E *BRAF* mutations are more common in younger patients and are present in most melanomas that arise on sites with intermittent sun exposure and are less common in melanomas from chronically sun-damaged skin. At present, *BRAF* mutations are the most important in therapeutic decision making in patients with advanced melanoma.

Melanomas also harbor mutations in *AKT* (primarily in *AKT3*) and *PTEN* (phosphatase and tensin homolog). *AKT* can be amplified, and

PTEN may be deleted or undergo epigenetic silencing that leads to constitutive activation of the PI3K/AKT pathway and enhanced cell survival by antagonizing the intrinsic pathway of apoptosis. Loss of *PTEN*, which dysregulates AKT activity, and mutation of *AKT3* both prolong cell survival through inactivation of *BAD*, *BCL-2*-antagonist of cell death, and activation of the forkhead transcription factor *FOXO1*, which leads to synthesis of prosurvival genes. A loss-of-function mutation in *NF1*, which can affect both MAP kinase and PI3K/AKT pathways, has been described in 10–15% of melanomas. In melanoma, these two signaling pathways (MAP kinase and PI3K/AKT) enhance tumorigenesis, chemoresistance, migration, and cell cycle dysregulation. Drugs that inhibit some of these pathways have been developed, and have proven to be effective therapeutic agents (see below).

PROGNOSTIC FACTORS

The most important prognostic factors for a newly diagnosed patient are incorporated in the staging classification (Table 72-2). The best predictor of metastatic risk is Breslow thickness. The anatomic site of the primary is also prognostic; favorable sites are the forearm and leg (excluding the feet), and unfavorable sites include the scalp, hands, feet, and mucous membranes. In general, women with stage I or II disease have better survival than men, perhaps in part because of earlier

TABLE 72-2 Staging Criteria for Melanoma

PATHOLOGIC AND TNM STAGE	THICKNESS, mm	ULCERATION	NO. OF INVOLVED LYMPH NODES	NODAL INVOLVEMENT	15-YEAR SURVIVAL ESTIMATE (%)
0 Tis	In situ	No	0	None	98
IA T1a	<1	No, mitosis <1/mm	0	None	92
IB T1b T2a	<1 1.01–2	Yes or mitosis >1/mm No	0 0	None None	80
IIA T2b T3a	1.01–2 2.01–4	Yes No	0 0	None None	62
IIB T3b T4a	2.01–4 >4	Yes No	0 0	None None	51
IIC T4b	>4	Yes	0	None	37
IIIA N1a N2a	T1-4a T1-4a	No No	1 2 or 3	Microscopic Microscopic	68
IIIB N1a N2a N1b N2b N2c	Any Any Any Any Any	Yes Yes Yes or no Yes or no Yes or no	1 2 or 3 1 2 or 3 In-transit metastases/satellites, no nodal involvement	Microscopic Microscopic Macroscopic Macroscopic	38
IIIC N1b N2b N2c N3	Any Any Any Any	Yes or no Yes or no Yes or no Yes or no	1 2 or 3 In-transit metastases/satellites, no nodal involvement 4+ metastatic nodes, matted nodes or in-transit metastases/ satellites, with metastatic nodes	Macroscopic Macroscopic	22
IV M1a M1b M1c		Distant metastasis Skin, subcutaneous Lung Other visceral site Elevated lactate dehydrogenase			<10

diagnosis; women frequently have melanomas on the lower leg, where self-recognition is more likely and the prognosis is better. The effect of age is not straightforward. Older individuals, especially men >60, have worse prognoses, a finding that has been explained in part by a tendency toward later diagnosis (and thus thicker tumors) and in part by a higher proportion of acral melanomas in men. However, there is a greater risk of lymph node metastasis in young patients. Other important adverse factors recognized via the staging classification include high mitotic rate, presence of ulceration, microsatellite lesions and/or in-transit metastases, evidence of nodal involvement, elevated serum lactate dehydrogenase (LDH), and presence and site of distant metastases.

■ STAGING

Once the diagnosis of melanoma has been made, the tumor is staged to determine the prognosis and aid in treatment selection. The current melanoma staging criteria and estimated 15-year survival by stage are depicted in Table 72-2. The clinical stage is determined after the microscopic evaluation of the melanoma skin lesion and clinical and radiologic assessment. Pathologic staging also includes the microscopic evaluation of the regional lymph nodes obtained at sentinel lymph node biopsy or completion lymphadenectomy as indicated. All patients should have a complete history, with attention to symptoms that may suggest metastatic disease, such as malaise, weight loss, headaches, visual changes, and pain, and physical examination directed to the site of the primary melanoma, looking for persistent disease or for dermal or subcutaneous nodules that could represent satellite or in-transit metastases, and to the regional draining lymph nodes, CNS, liver, and lungs. A complete blood count (CBC), complete metabolic panel, and LDH should be performed. Although these rarely help uncover occult metastatic disease, a microcytic anemia would raise the possibility of bowel metastases, and the LDH, if elevated, should prompt a more extensive evaluation, including computed tomography (CT) scan or possibly a positron emission tomography (PET) (or CT/PET combined) scan. If signs or symptoms of metastatic disease are present, appropriate diagnostic imaging should be performed. At initial presentation, >80% of patients will have disease confined to the skin and a negative history and physical examination, in which case imaging is not indicated.

TREATMENT

Melanoma

MANAGEMENT OF CLINICALLY LOCALIZED MELANOMA (STAGE I, II)

For a newly diagnosed cutaneous melanoma, wide surgical excision of the lesion with a margin of normal skin is necessary to remove all malignant cells and minimize possible local recurrence. The following margins are recommended for a primary melanoma: *in situ*, 0.5–1.0 cm; invasive up to 1 mm thick, 1 cm; >1.01–2 mm, 1–2 cm; and >2 mm, 2 cm. For lesions on the face, hands, and feet, strict adherence to these margins must give way to individual considerations about the constraints of surgery and minimization of morbidity. In all instances, however, inclusion of subcutaneous fat in the surgical specimen facilitates adequate thickness measurement and assessment of surgical margins by the pathologist. Topical imiquimod also has been used, particularly for lentigo maligna, in cosmetically sensitive locations.

Sentinel lymph node biopsy (SLNB) is a valuable staging tool that has replaced elective regional node dissection for the evaluation of regional nodal status. SLNB provides prognostic information and helps identify patients at high risk for relapse who may be candidates for adjuvant therapy. The initial (sentinel) draining node(s) from the primary site is (are) identified by injecting a blue dye and a radioisotope around the primary site. The sentinel node(s) then is (are) identified by inspection of the nodal basin for the blue-stained node and/or the node with high uptake of the radioisotope. The identified nodes are removed and subjected to careful histopathologic

analysis with serial section using hematoxylin and eosin stains as well as immunohistochemical stains (e.g., S100, HMB45, and MelanA) to identify melanocytes.

Not every patient requires a SLNB. Patients whose melanomas are ≤0.75 mm thick have <5% risk of sentinel lymph node (SLN) disease and do not require a SLNB. Patients with tumors >1 mm thick generally undergo SLNB. For melanomas 0.76–1.0 mm thick, SLNB may be considered for lesions with high-risk features such as ulceration, high mitotic index, or lymphovascular invasion, but wide excision alone is the usual definitive therapy. Most other patients with clinically negative lymph nodes should undergo a SLNB. Patients whose SLNB is negative are spared a complete node dissection and its attendant morbidities, and can simply be followed, or based on the features of the primary melanoma, be considered for adjuvant therapy or a clinical trial. The current standard of care for all patients with a positive SLN is to perform a complete lymphadenectomy; however, complete lymph node dissection is not necessary for patients with lymph node micrometastases <1 mm. Patients with positive lymph nodes should be considered for adjuvant therapy with ipilimumab, interferon alpha or enrollment in a clinical trial.

MANAGEMENT OF REGIONALLY METASTATIC MELANOMA (STAGE III)

Melanomas may recur at the edge of the scar or graft, as satellite metastases, which are separate from but within 2 cm of the scar; as in-transit metastases, which are recurrences >2 cm from the primary lesion but not beyond the regional nodal basin; or, most commonly, as metastasis to a draining lymph node basin. Each of these presentations is managed surgically, following which there is the possibility of long-term disease-free survival. Isolated limb perfusion or infusion with melphalan and hyperthermia are options for patients with extensive cutaneous regional recurrences in an extremity. High complete response rates have been reported and significant palliation of symptoms can be achieved, but there is no change in overall survival. Other options for in-transit disease and distant skin and soft tissue metastases include topical immunotherapy and direct injection of melanoma lesions. Topical therapy with imiquimod has been useful for patients with low-volume dermal lesions. Historically, intralesional bacille Calmette-Guerin (BCG) has been used with high rates of regression of injected lesions and occasional regression of a distant, uninjected lesion. Talimogene laherparepvec is an engineered, oncolytic herpes simplex virus type 1 that is U.S. Food and Drug Administration (FDA) approved for injection of melanoma lesions that cannot be completely removed by surgery.

Patients rendered free of disease after surgery may be at high risk for a local or distant recurrence and should be considered for adjuvant therapy. Radiotherapy can reduce the risk of local recurrence after lymphadenectomy, but does not affect overall survival. Patients with large nodes (>3–4 cm), four or more involved lymph nodes, or extranodal spread on microscopic examination should be considered for radiation. Systemic adjuvant therapy is indicated primarily for patients with stage III disease, but high-risk, node-negative patients (>4 mm thick or ulcerated lesions), and patients with completely resected stage IV disease also may benefit.

Current treatment options include ipilimumab, interferon α 2b (IFN- α 2b) or investigational therapy. Ipilimumab is a fully human monoclonal antibody that blocks the immune checkpoint cytotoxic T-lymphocyte antigen-4 (CTLA-4) and augments antitumor immune responses. Treatment with ipilimumab 10 mg/kg IV every three weeks for four doses, then every three months for up to three years, improved survival of patients with high-risk stage III disease compared to placebo. IFN- α 2b may be administered at high doses for one year or pegylated IFN can be administered at a lower dose for five years. The single study of ipilimumab documented a survival benefit whereas multiple trials of IFN have reported clear improvement in disease-free survival, but questionable improvement in overall survival. The two agents have not been compared directly. Ongoing clinical trials will address this issue as well as evaluate the potential value of other immunotherapies (e.g., PD-1/PD-L1).

blocking agents) and targeted therapies in patients with BRAF mutated tumors in the adjuvant setting.

Both IFN and ipilimumab are accompanied by significant toxicity. For IFN, this may include a flulike illness, decline in performance status, and the development of depression. Side effects can be managed in most patients by appropriate treatment of symptoms, dose reduction, and treatment interruption. IFN may need to be discontinued prematurely because of unacceptable toxicity. The major side effects of ipilimumab are discussed below.

TREATMENT

Metastatic Disease

At diagnosis, 84% patients with melanoma will have early-stage disease and 4% will present with metastases. Many others will develop metastases after initial therapy for loco-regional disease. The probability of recurrence is related to initial stage, ranging from <5% with stage IA to >90% for subsets of patients with stage IIIC disease at presentation. Patients with a history of melanoma who develop signs or symptoms suggesting recurrent disease should undergo restaging as described earlier. Distant metastases (stage IV) may involve any organ and commonly involve the skin and lymph nodes as well as viscera, bone, or the brain. The prognosis is better for patients with skin and subcutaneous metastases (M1a) than for lung (M1b) and worst for those with metastases to liver, bone, and brain (M1c). An elevated serum LDH is a poor prognostic factor and places the patient in stage M1c regardless of the site of the metastases (Table 72-2). Although historical data suggest that the 15-year survival of patients with melanoma is <10%; advances in targeted and immunotherapy have improved disease-free and overall survival, especially for patients with M1a and M1b disease.

The treatment for patients with stage IV melanoma has changed dramatically since 2011. FDA-approved agents include three immune T-cell checkpoint inhibitors, ipilimumab, nivolumab, and pembrolizumab, four oral agents that target the MAP kinase pathway: the BRAF inhibitors, vemurafenib and dabrafenib, the MEK inhibitors, trametinib and cobimetinib, and the oncolytic virus talimogene laherparepvec (Table 72-3).

Surgery should be considered for patients with oligometastatic disease because they may experience long-term disease-free survival after metastasectomy. Patients with solitary metastases are the best candidates, but surgery can also be used for patients with metastases at more than one site if a complete resection of all sites can be achieved. Patients rendered free of disease can be considered for adjuvant therapy or a clinical trial because their risk of developing additional metastases is very high. Surgery can also be used as an adjunct to systemic therapy when for example, a few of many metastatic lesions prove resistant to immunotherapy.

TABLE 72-3 Treatment Options for Metastatic Melanoma

Surgery: Metastasectomy for small number of lesions

Immunotherapy:

Interleukin 2

Immune checkpoint blockade

- Anti-CTLA-4: ipilimumab
- Anti-PD-1: nivolumab, pembrolizumab
- Combined ipilimumab and nivolumab

Experimental

- Anti-PD-L1

Molecular targeted therapy:

BRAF inhibitor: vemurafenib, dabrafenib

MEK inhibitor: trametinib, cobimetinib

Oncolytic virus: talimogene laherparepvec

Chemotherapy: dacarbazine, temozolomide, paclitaxel, albumin-bound paclitaxel, carboplatin

IMMUNOTHERAPY

Interleukin 2 (IL-2 or aldesleukin) is used effectively to treat stage IV patients who have a good performance status. High-dose IL-2, which requires hospitalization in an intensive care unit-like setting, is administered by intravenous bolus doses over a 1-week cycle mainly at centers with experience managing IL-2-related toxicity. Treatment is continued until maximal benefit is achieved, usually 4–6 cycles distributed over 4–6 months to allow for recovery from toxicities between cycles. Long-term disease-free survival (probable cure) is observed in 5% of treated patients.

Checkpoint Blockade Newer immunotherapies are based on an understanding of the control mechanisms of the normal immune response. Inhibitory receptors or checkpoints, including CTLA-4 and PD-1, are upregulated on T cells after engagement of the T-cell receptor by cognate tumor antigen in the context of the appropriate class I or II HLA molecules during the interaction between a T cell and antigen-presenting cell. An absolute requirement to ensure proper regulation of a normal immune response, the continued expression of inhibitory receptors during chronic infection (hepatitis, HIV) and in cancer patients leads to exhausted T cells with limited potential for proliferation, cytokine production, or cytotoxicity (Fig. 72-3). Checkpoint blockade with an antagonistic monoclonal antibody results in improved T-cell function and eradication of tumor cells in preclinical animal models. Ipilimumab, a fully human IgG1 antibody that binds CTLA-4 and blocks inhibitory signals, was the first drug shown in a randomized trial to improve survival in patients with metastatic melanoma. A full course of therapy is four outpatient infusions of ipilimumab 3 mg/kg every 3 weeks. Although response rates are low (~10%), overall survival is improved.

Chronic T-cell activation also leads to induction of PD-1 on the surface of T cells. Expression of one of its ligands, PD-L1, on tumor cells can protect them from immune destruction (Fig. 72-3). Blockade of the PD-1:PD-L1 axis by IV administration of anti-PD-1 or anti-PD-L1 has substantial clinical activity in patients with advanced melanoma (and lung, renal, bladder and oral head and neck cancers as well as Hodgkin lymphoma) with significantly less toxicity than ipilimumab. The PD-1 blockers, nivolumab and pembrolizumab, have been approved to treat patients with advanced melanoma. Combination T-cell checkpoint therapy, blocking both inhibitory pathways with ipilimumab and nivolumab, leads to superior antitumor activity compared to treatment with either agent alone. Combined therapy with intravenous ipilimumab and nivolumab is administered in the outpatient setting every 3 weeks for 4 doses (induction), followed by nivolumab given every 2 weeks (maintenance) for up to one year. This regimen produces an objective response rate of 56% and enhanced survival compared to

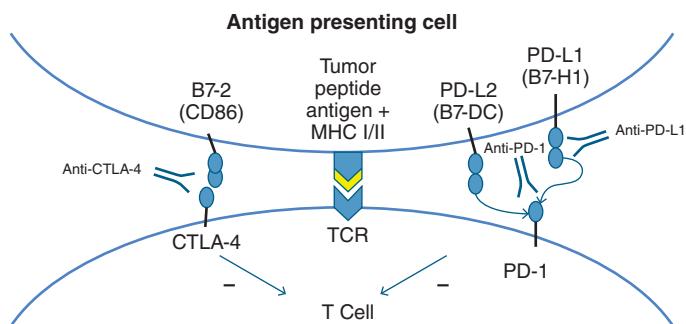


FIGURE 72-3 Inhibitory regulatory pathways that influence T-cell function, memory and lifespan after engagement of the T-cell receptor by tumor peptide antigen presented by antigen-presenting cells in the context of MHC I/II. CTLA-4 and PD-1 are members of the CD28 family and their inhibitory effects can be mitigated by antagonistic antibodies to the receptors or ligand resulting in enhanced T-cell function and anti-tumor effects. TCR: T-cell receptor, MHC: Major Histocompatibility Complex, CTLA-4: Cytotoxic T-Lymphocyte Antigen-4, PD-1: Programmed Death-1, PD-L1: Programmed Death Ligand-1, PD-L2: Programmed Death Ligand-2.

ipilimumab monotherapy. There may be subsets of patients, specifically those who have >5% expression of PD-1 on T cells in a melanoma biopsy sample, who derive a similar level of clinical benefit from nivolumab monotherapy.

The main benefit to patients from immune-based therapy is the durability of the responses achieved. The percentage of patients whose tumors regress following combination anti-CTLA-4 and anti-PD-1 immunotherapy is comparable to the response rate after targeted therapy (see below); however, the durability of immunotherapy-induced responses (>10 years in some cases with checkpoints and greater than 20 years in some patients after IL-2) appears to be superior to responses after targeted therapy and suggests that many of these patients have been cured.

T-cell checkpoint antibodies can also interfere with normal immune regulatory mechanisms, which may produce a novel spectrum of side effects. The most common immune-related adverse events were skin rash and diarrhea (sometimes severe, life-threatening colitis), but toxicity can involve most any organ (e.g., hypophysitis, hepatitis, nephritis, pneumonitis, myocarditis, neuritis). The severity and frequency of toxicity is greatest with combination T-cell checkpoint antibody therapy, followed by anti-CTLA-4 and then anti-PD-1 monotherapies. Vigilance, interruption of therapy and early intervention with steroids or other immunosuppressive agents, such as anti-tumor necrosis factor antibodies or mycophenolate mofetil, can mitigate toxicity and prevent permanent organ damage. The management of drug-induced toxicity with immunosuppressive agents does not appear to interfere with antitumor activity. The use of T-cell checkpoint antibodies for metastatic melanoma has become commonplace, but there is controversy about whether all patients need combined anti-CTLA-4 and anti-PD-1, whether biomarkers can be used to select patients who may benefit from anti-PD-1 alone and the best sequence of targeted and immunotherapy in patients who have a BRAF mutation. There is also a significant economic impact with the cost of combination anti-CTLA-4 and anti-PD-1, which must be placed in the context of the survival benefit.

TARGETED THERAPY

The high frequency of oncogenic mutations in the RAS-RAF-MEK-ERK pathway, which delivers proliferation and survival signals from the cell surface to the cytoplasm and nucleus, has led to the development of inhibitors to BRAF and MEK. RAF and MEK inhibitors of the MAP kinase pathway can induce regression of melanomas that harbor a BRAF mutation. Two BRAF inhibitors, vemurafenib and dabrafenib, have been approved for the treatment of patients whose stage IV melanomas harbor a mutation at position 600 in BRAF. Monotherapy with BRAF inhibitors has been supplanted with combined BRAF and MEK inhibition to address the rapid adaptation of the majority of melanomas that use MAP kinase pathway reactivation to facilitate growth when BRAF is inhibited. Combined therapy with BRAF and MEK inhibitors (dabrafenib and trametinib or vemurafenib with cobimetinib) improved progression-free survival compared to monotherapy with a BRAF inhibitor. The durability of responses following combined therapy is superior to monotherapy and survival is also enhanced. Long-term results of inhibition of the MAP kinase pathway are not yet available, but the major limitation of both monotherapy and combined therapy appears to be the acquisition of resistance; the vast majority of patients relapse and eventually die. The mechanisms of resistance are diverse and reflect the genomic heterogeneity of melanoma; however, most instances involve reactivation of the MAPK pathway, often through RAS mutations or mutant BRAF amplification. Patients who develop resistance to BRAF and MEK inhibition are candidates for immunotherapy or clinical trials.

Targeted therapy is accompanied by manageable side effects that differ from those experienced during immunotherapy or chemotherapy. A class-specific side effect of BRAF inhibition is the development of numerous skin lesions, some of which are well-differentiated squamous cell skin cancers (SCC) (seen in up to 25%

of patients). These hyperproliferative lesions are believed to be due to paradoxical activation of the MAPK pathway resulting from BRAF inhibitor-mediated changes in BRAF-wild type cells. The paradoxical activation is blocked by the MEK inhibitor, which explains why these lesions occur much less frequently during combined therapy. Patients should be co-managed with a dermatologist as these skin cancers will need excision. Metastases of the treatment-induced SCCs have not been reported, and BRAF and MEK inhibitors can be continued safely following simple excision. Cardiac and ocular toxicities, although infrequent, can occur with BRAF and MEK inhibitors and require medical evaluation and management.

Activating mutations in the c-kit receptor tyrosine kinase are found in a minority of cutaneous melanomas with chronic sun damage, but are more common in mucosal and acral lentiginous subtypes. Overall, the number of patients with *c-kit* mutations is exceedingly small, but when present, they are similar to those found in gastrointestinal stromal tumors; melanomas with activating *c-kit* mutations can have clinically meaningful responses to imatinib. The probability of objective response in patients whose melanomas harbor a *c-kit* mutation is 29%. *N-RAS* mutations occur in 15–20% of melanomas. At present, there are no effective targeted agents for these patients, but MEK inhibitors are being investigated in clinical trials.

CHEMOTHERAPY

No chemotherapy regimen has ever been shown to improve survival of patients with metastatic melanoma. The advances in immunotherapy and targeted therapy have relegated chemotherapy to the palliation of symptoms. Drugs with antitumor activity include dacarbazine (DTIC) or its orally administered analog temozolomide (TMZ), cisplatin and carboplatin, the taxanes (paclitaxel alone or albumin-bound), and carmustine (BCNU), which have reported response rates of 12–20%.

INITIAL APPROACH TO PATIENT WITH METASTATIC DISEASE

Upon diagnosis of stage IV disease, a sample of the patient's tumor should be submitted for molecular testing to determine whether a druggable mutation (e.g., BRAF and *c-kit*) is present. Analysis of a metastatic lesion biopsy (if possible) is preferred, but any sample will suffice because there is little discordance between primary and metastatic lesions. Treatment algorithms start with the tumor's BRAF status. For BRAF wild-type tumors, immunotherapy is recommended. For patients whose tumors harbor a BRAF mutation, initial therapy with either combination BRAF and MEK inhibitors or immunotherapy is acceptable. Combined therapy with BRAF and MEK inhibitors is favored for patients with rapidly growing and symptomatic disease when a BRAF mutation is present. The sequence of immunotherapy and targeted therapy that confers the greatest survival benefit in patients with minimally symptomatic melanoma is not yet known, but ongoing randomized phase III trials should answer this important question. Despite improvements in therapy, the majority of patients with metastatic melanoma are not cured so enrollment in a clinical trial is always an important consideration, even for previously untreated patients.

Since most patients with stage IV disease will eventually experience tumor progression despite therapy and many, because of extensive disease burden, poor performance status, or concomitant illness, will be poor candidates for therapy, the timely integration of palliative care and hospice should be a major focus of care. Future advances in the management of melanoma will likely include biomarkers to select the optimal combination and sequence of agents or to identify patients who are unlikely to respond to extant therapies and for whom clinical trials should be considered. New therapeutic agents could include T-cell co-stimulatory antibodies, engineered T cells, oncolytic viruses and possibly vaccines to prevent melanoma development or recurrence.

FOLLOW-UP

Skin examination and surveillance at least once a year are recommended for all patients with melanoma. Routine blood work and

imaging for patients with stage IA–IIA disease is not recommended unless symptoms are present. In general, because there is no survival benefit to patients, routine surveillance diagnostic imaging is not recommended for patients with higher stage disease and imaging should be reserved for patients with signs or symptoms of recurrent disease. For stage-specific recommendations, please consult the National Comprehensive Cancer Network (NCCN) guidelines (see Further Reading).

NONMELANOMA SKIN CANCER

Nonmelanoma skin cancer (NMSC) is the most common cancer in the United States. Although tumor registries do not routinely gather data on the incidence of basal cell and squamous cell skin cancers, it is estimated that the annual incidence is 1.5–2 million cases in the United States. Basal cell carcinomas (BCCs) account for 70–80% and squamous cell carcinomas (SCCs) ~20% of NMSCs, respectively. SCCs are more significant because they metastasize and account for 2400 deaths annually. There has also been an increase in the incidence of nonepithelial skin cancer, especially Merkel cell carcinoma, with nearly 5000 new diagnoses and 3000 deaths annually.

PATHOPHYSIOLOGY AND ETIOLOGY

The most significant cause of BCC and SCC is UV radiation, whether through direct exposure to sunlight or by artificial UV light sources (tanning beds). Both UVA and UVB light can induce DNA damage. The DNA damage can be repaired or lead to cell death. The mechanism for DNA repair involves excising damaged nucleotides. Inherited disorders of DNA repair, such as xeroderma pigmentosum, are associated with a greatly increased incidence of skin cancer and help to establish the link between UV-induced DNA damage, inadequate DNA repair, and skin cancer. The genes damaged most commonly by UV in BCC involve the hedgehog signaling pathway (Hh) and lead to basal cell proliferation. This is usually the result of loss of function of the tumor-suppressor patched homolog 1 (*PTCH1*), which normally inhibits the signaling of smoothened homolog (SMO). Aberrant *PTCH1* signaling is propagated by the nuclear transcription factors Gli1 and Gli2, which are salient in the development of BCC. Two oral SMO inhibitors, vismodegib and sonidegib, have been approved by the FDA to treat advanced inoperable or metastatic BCC and locally advanced BCC that has recurred following surgery or RT, respectively (Fig. 72-4). Vismodegib also reduces the incidence of BCC in patients with basal cell nevus syndrome who have *PTCH1* mutations, affirming the importance of Hh in the onset of BCC.

In SCC, *p53* and *N-RAS* are commonly affected. There is a dose-response relationship between tanning bed use and the incidence of skin cancer. As few as four tanning bed visits per year confers a 15% increase in BCC and an 11% increase in SCC and melanoma. Tanning bed use as a teenager or young adult confers greater risk than comparable exposure in older individuals. Other associations include blond or red hair, blue or green eyes, a tendency to sunburn easily, and an outdoor occupation. The incidence of NMSC increases with decreasing latitude. Most tumors develop on sun-exposed areas of the head and neck. The risk of lip or oral SCC is increased with cigarette smoking and, like SCC of the ear, has a worse prognosis than that seen on other body sites. Human papillomaviruses and UV radiation may act as co-carcinogens.

Chronically immunosuppressed solid organ transplant recipients have a 65-fold increase in SCC and a 10-fold increase in BCC. The frequency of skin cancer is proportional to the level and duration of immunosuppression and the extent of sun exposure before and after transplantation. SCCs in this population also demonstrate higher rates of local recurrence, metastasis, and mortality. Tumor necrosis factor (TNF) antagonist therapy of inflammatory bowel disease and autoimmune disorders, such as rheumatoid and psoriatic arthritis, may also confer an increased risk of NMSC.

Other risk factors include HIV infection, ionizing radiation, thermal burn scars, and chronic ulcerations. Albinism, xeroderma pigmentosum, Muir-Torre syndrome, Rombo's syndrome, Bazex-Dupré-Christol

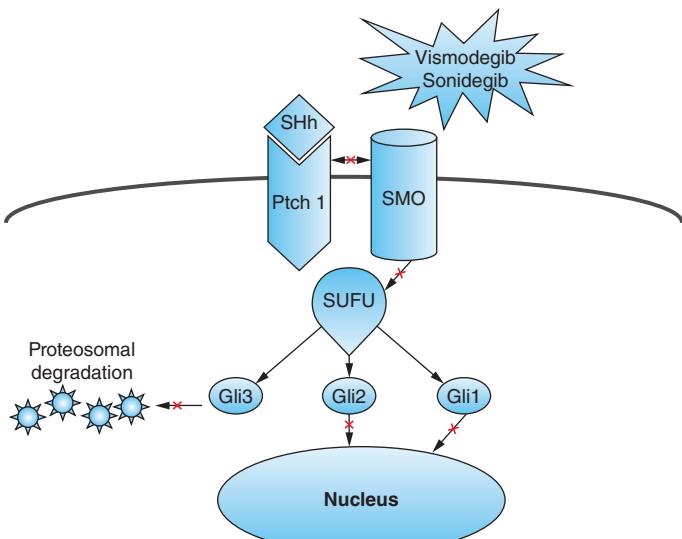


FIGURE 72-4 Inhibition of the hedgehog (Hh) pathway. The Hh pathway promotes gene transcription and is important in the pathogenesis of BCC. Normally, one of three Hh ligands (sonic [SHh], Indian, or desert) binds to patched homolog 1 (PTCH1), causing its degradation and release of smoothened homolog (SMO). SMO release represses another regulatory protein called suppressor of fused (SUFU). SUFU normally binds glioblastoma transcription factors Gli1, Gli2, and Gli3. SUFU repression allows Gli1 and Gli2 to translocate to the nucleus and promote gene transcription. Vismodegib and sonidegib are SMO antagonists. Antagonizing SMO decreases the interaction between SMO and PTCH1, resulting in decreased Hh pathway signaling, gene transcription, and cell division. The Hh pathway events inhibited by vismodegib and sonidegib are indicated in red.

syndrome, dyskeratosis congenita, and basal cell nevus syndrome (Gorlin syndrome) also increase the incidence of NMSC.

CLINICAL PRESENTATION

Basal Cell Carcinoma BCC arises from epidermal basal cells. The least invasive of BCC subtypes, superficial BCC, consists of often subtle, erythematous scaling plaques that slowly enlarge and are most commonly seen on the trunk and proximal extremities (Fig. 72-5). This BCC subtype may be confused with benign inflammatory dermatoses, especially nummular eczema and psoriasis or premalignant actinic keratoses. BCC also can present as a small, slowly growing pearly nodule, often with tortuous telangiectatic vessels on its surface, rolled borders, and a central crust (nodular BCC). The occasional presence of melanin in this variant of nodular BCC (pigmented BCC) may lead to confusion with melanoma. Morpheaform (fibrosing), infiltrative, and micronodular BCC, the most invasive and potentially aggressive subtypes, manifest as solitary, flat or slightly depressed, indurated whitish, yellowish, or pink scar-like plaques. Borders are typically indistinct, and lesions can be subtle; thus, delay in treatment is common, and tumors can be more extensive than expected clinically.

Squamous Cell Carcinoma Primary cutaneous SCC is a malignant neoplasm of keratinizing epidermal cells. SCC has a variable clinical course, ranging from indolent to rapid growth, with the potential to metastasize to regional and distant sites. Commonly, SCC appears as an ulcerated erythematous nodule or superficial erosion on sun-exposed skin of the head, neck, trunk, and extremities (Fig. 72-5). It may also appear as a banal, firm, dome-shaped papule or rough-textured plaque. It is commonly mistaken for a wart or callous when the inflammatory response to the lesion is minimal. Visible overlying telangiectasias are uncommon, although dotted or coiled vessels are a hallmark of SCC when viewed through a dermatoscope. The margins of this tumor may be ill defined, and fixation to underlying structures may occur ("tethering").

A very rapidly growing but low-grade form of SCC, called keratoacanthoma (KA), typically appears as a large dome-shaped papule with a central keratotic crater. Some KAs regress spontaneously without therapy, but because progression to metastatic SCC has been



FIGURE 72-5 Cutaneous neoplasms. **A.** Non-Hodgkin's lymphoma involves the skin with typical violaceous, "plum-colored" nodules. **B.** Squamous cell carcinoma is seen here as a hyperkeratotic crusted and somewhat eroded plaque on the lower lip. Sun-exposed skin in areas such as the head, neck, hands, and arms represent other typical sites of involvement. **C.** Actinic keratoses consist of hyperkeratotic erythematous papules and patches on sun-exposed skin. They arise in middle-aged to older adults and can undergo malignant transformation. **D.** Metastatic carcinoma to the skin is characterized by inflammatory, often ulcerated dermal nodules. **E.** Mycosis fungoides is a cutaneous T-cell lymphoma, and plaque-stage lesions are seen in this patient. **F.** Keratoacanthoma is a low-grade squamous cell carcinoma that presents as an exophytic nodule with central keratinous debris. **G.** This basal cell carcinoma shows central ulceration and a pearly, rolled telangiectatic tumor border.

documented, KAs should be treated in the same manner as other types of cutaneous SCC. KAs occur in 15–25% of patients receiving monotherapy with a BRAF inhibitor.

Actinic keratoses and *cheilitis* (actinic keratoses on the lip), both premalignant forms of SCC, present as hyperkeratotic papules on sun-exposed areas. Malignant transformation occurs in 0.25 to 20% of untreated lesions. SCC *in situ*, also called *Bowen's disease*, is the intraepidermal form of SCC and usually presents as a scaling, erythematous plaque. SCC *in situ* most commonly arises on sun-damaged skin, but can occur anywhere on the body. Bowen's disease occurring secondary to infection with human papillomavirus (HPV) can arise on skin with minimal or no prior sun exposure, such as the buttock or posterior thigh. Treatment of premalignant and *in situ* lesions reduces the subsequent risk of invasive disease.

NATURAL HISTORY

Basal Cell Carcinoma The natural history of BCC is that of a slowly enlarging, locally invasive neoplasm. The degree of local destruction and risk of recurrence vary with the size, duration, location, and histologic subtype of the tumor. Location on the central face, ears, or scalp may portend a higher risk. Small nodular, pigmented, cystic, or superficial BCCs respond well to most treatments. Large lesions and micronodular, infiltrative, and morpheaform subtypes may be more aggressive. The metastatic potential of BCC is low (0.0028–0.1%) in immunocompetent patients, but the risk of recurrence or a new primary NMSC is about 40% over 5 years.

Squamous Cell Carcinoma The natural history of SCC depends on tumor and host characteristics. Tumors arising on sun-damaged skin have a lower metastatic potential than do those on non-sun-exposed areas. Cutaneous SCC metastasizes in 0.3–5.2% of individuals, most frequently to regional lymph nodes. Tumors occurring on the lower lip and ear develop regional metastases in 13 and 11% of patients, respectively, whereas the metastatic potential of SCC arising in scars, chronic ulcerations, and genital or mucosal surfaces is higher. Recurrent SCC has a much higher potential for metastatic disease, approaching 30%. Large, poorly differentiated, deep tumors with perineural or lymphatic

invasion, multifocal tumors, and those arising in immunosuppressed patients often behave aggressively.

TREATMENT

Basal Cell and Squamous Cell Carcinoma

BASAL CELL CARCINOMA

Treatments used for BCC include electrodesiccation and curettage (ED&C), excision, cryosurgery, radiation therapy (RT), laser therapy, Mohs micrographic surgery (MMS), topical 5-fluorouracil, photodynamic therapy (PDT), and topical immunomodulators such as imiquimod. The choice of therapy depends on tumor characteristics including depth and location, patient age, medical status, and patient preference. ED&C remains the most commonly employed method for superficial, minimally invasive nodular BCCs and low-risk tumors (e.g., a small tumor of a less aggressive subtype in a favorable location). Wide local excision with standard margins is usually selected for invasive, ill-defined, and more aggressive subtypes of tumors, or for cosmetic reasons. MMS, a specialized type of surgical excision that provides the best method for tumor removal while preserving uninvolved tissue, is associated with cure rates >98%. It is the preferred modality for lesions that are recurrent, in high-risk or cosmetically sensitive locations (including recurrent tumors in these locations), and for which maximal tissue conservation is critical (e.g., the eyelids, lips, ears, nose, and digits). RT can cure patients not considered surgical candidates and can be used as a surgical adjunct in high-risk tumors. Imiquimod can be used to treat superficial and smaller nodular BCCs, although it is not FDA-approved for nodular BCC. Topical 5-fluorouracil therapy should be limited to superficial BCC. PDT, which uses selective activation of a photoactive drug by visible light, has been used in patients with numerous tumors. Intralesional therapy (5-fluorouracil or IFN) can also be employed. Like RT, it remains an option for selected patients who cannot or will not undergo surgery. Systemic therapy with an SMO inhibitor, vismodegib or sonidegib, is indicated for patients with metastatic or advanced BCC that has recurred after

local therapy and who are not candidates for surgery or radiation. Targeted therapy with SMO antagonists does not cure patients with BCC, but induces regression in approximately 50% of patients with a median duration of response greater than 9 months.

SQUAMOUS CELL CARCINOMA

Therapy for cutaneous SCC should be based on the size, location, histologic differentiation, patient age, and functional status. Surgical excision and MMS are standard treatments. Cryosurgery and ED&C have been used for premalignant lesions and small, superficial, *in situ* primary tumors. Lymph node metastases are treated with surgical resection, RT, or both. Combination chemotherapy that includes cisplatin, and intralesional and systemic 5-fluorouracil, and cetuximab are also options for palliation in patients with advanced disease. SCC and keratoacanthomas that develop in patients receiving BRAF-targeted therapy should be excised, after which BRAF therapy can be continued.

PREVENTION

The general principles for prevention are those described for melanoma earlier. Unique strategies for NMSC include active surveillance for patients on immunosuppressive medications or BRAF-targeted therapy. Chemoprophylaxis using synthetic retinoids and immunosuppression reduction when possible may be useful in controlling new lesions and managing patients with multiple tumors. Field therapy with topical 5-FU, ingenol mebutate, or imiquimod can reduce transformation to SCC in patients with severe sun damaged skin and numerous premalignant actinic keratoses.

OTHER NONMELANOMA CUTANEOUS MALIGNANCIES

Neoplasms of cutaneous adnexae and sarcomas of fibrous, mesenchymal, fatty, and vascular tissues make up the remaining 1–2% of NMSCs.

Merkel cell carcinoma (MCC) is a neural crest-derived highly aggressive malignancy with mortality rates approaching 33% at 3 years. An oncogenic Merkel cell polyomavirus (MCPyV) is present in 80% of tumors and UV exposure also increases the incidence of this malignancy. In patients with MCPyV+ tumors, there is inactivation of tumor suppressor genes, specifically the p53 transcription factor and retinoblastoma protein (Rb). In addition, the viral large T antigen is expressed on tumor cells and many patients have detectable cellular or humoral immune responses to polyoma viral proteins, although this immune response is insufficient to eradicate the malignancy. Survival depends on extent of disease: 90% survive with local disease, 52% with nodal involvement, but only 10% with distant disease. MCC incidence tripled over the last 20 years with an estimated 1600 cases per year in the United States. Immunosuppression increases the incidence and diminishes the prognosis compared to patients with no immunosuppression. MCC lesions typically present as an asymptomatic rapidly expanding bluish-red/violaceous tumor on sun-exposed skin of older white patients. Treatment is surgical excision with sentinel lymph node biopsy for accurate staging in patients with localized disease, often followed by adjuvant RT. Patients with extensive disease can be offered systemic chemotherapy; however, there is no survival benefit. Immunotherapy using anti-PD-1 (pembrolizumab) was associated with a 56% response rate with a progression-free survival at 6 months of 67%. Tumor regression occurred in MCPyV positive and negative tumors. A monoclonal antibody targeting anti-PD-L1 known as avelumab showed objective responses in 33% of patients with advanced MCC that was durable in 82% of the responders. The U.S. FDA approved avelumab for the treatment of patients with metastatic MCC in April 2017. Whenever possible a clinical trial should be considered for patients with this rare but aggressive NMSC.

Extramammary Paget's disease is an uncommon apocrine malignancy arising from stem cells of the epidermis that are characterized histologically by the presence of Paget cells. These tumors present as moist erythematous patches on anogenital or axillary skin of the elderly.

Outcomes are generally good with surgery, and 5-year disease-specific survival is ~95% with localized disease. Advanced age and extensive disease at presentation confer diminished prognosis. RT or topical imiquimod can be considered for more extensive disease. Local management may be challenging because these tumors often extend far beyond clinical margins; surgical excision with MMS has the highest cure rates. Similarly, MMS is the treatment of choice in other rare cutaneous tumors with extensive subclinical extension such as *dermatofibromas* *protuberans*.

Kaposi's sarcoma (KS) is a soft tissue sarcoma of vascular origin that is induced by the human herpesvirus 8. The incidence of KS increased dramatically during the AIDS epidemic, but has now decreased tenfold with the institution of highly active antiretroviral therapy.

ACKNOWLEDGMENT

Steven Kolker, MD, provided valued feedback and suggested improvements to this chapter.

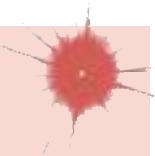
FURTHER READING

- THE CANCER GENOME ATLAS NETWORK: Genomic classification of cutaneous melanoma. *Cell* 161:1681, 2015.
- GUO J et al: Chinese guidelines on the diagnosis and treatment of melanoma (2015 edition). *Ann Transl Med* 3:322, 2015.
- INTERNATIONAL AGENCY FOR RESEARCH ON CANCER: GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012. Available from <http://globocan.iarc.fr/default.aspx>. Accessed December 19, 2016.
- LARKIN J et al: Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 373:23, 2015.
- LEITER U et al: Complete lymph node dissection versus no dissection in patients with sentinel lymph node biopsy positive melanoma (DeCOG-SLT): A multicentre, randomised, phase 3 trial. *Lancet Oncol* 17:757, 2016.
- NATIONAL COMPREHENSIVE CANCER NETWORK: NCCN clinical practice guidelines in oncology (NCCN guidelines): Melanoma. Available from https://www.nccn.org/professionals/physician_gls/pdf/melanoma.pdf. Accessed December 20, 2016.
- ROBERT C et al: Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* 372:30, 2015.
- SHAIN AH, BASTIAN BC: From melanocytes to melanomas. *Nat Rev Cancer* 16:345, 2016.
- WU YP et al: A systematic review of interventions to improve adherence to melanoma preventive behaviors for individuals at elevated risk. *Prev Med* 88:153, 2016.
- ZHANG T et al: The genomic landscape of cutaneous melanoma. *Pigment Cell Melanoma Res* 29:266, 2016.

73

Head and Neck Cancer

Everett E. Vokes



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. 378.*

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 48,330 in 2016, accounting for about 3% of adult malignancies; estimated deaths were 13,190. 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 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 in-door pollution.

In Western countries, the human papilloma virus (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 even 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 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.

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 non-keratinizing 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 history of premalignant lesions. Multiple synchronous or metachronous cancers can also be observed. In fact, over time, patients with treated early-stage 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. 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 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 non-healing 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 carcinoma of unknown primary (Fig. 73-1). 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

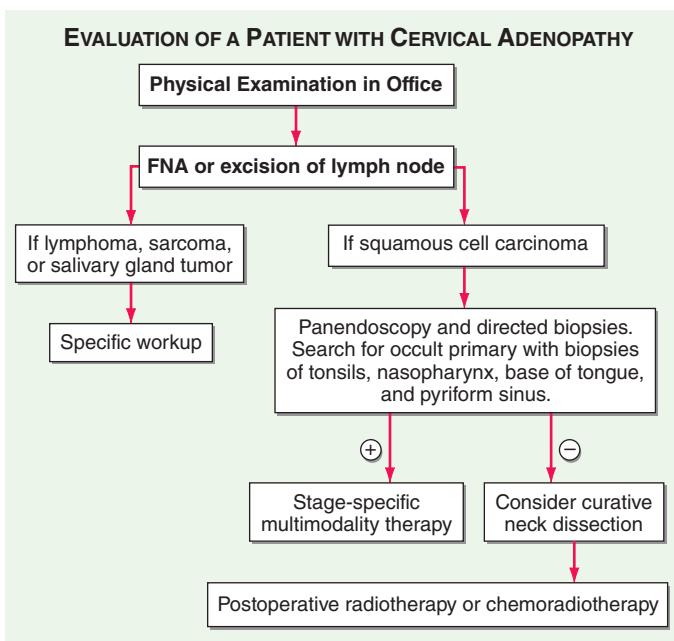


FIGURE 73-1 Evaluation of a patient with cervical adenopathy without a primary mucosal lesion; a diagnostic workup. FNA, fine-needle aspiration.

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.

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 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 may also be administered and 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 (**Fig. 73-2**). 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 upcoming 8th edition of the AJCC staging manual (active in 2018) 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 non-smoker or has limited lifelong tobacco exposure.

In patients with lymph node involvement and no visible primary, the diagnosis should be made by lymph node excision (Fig. 73-1). 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 dental decay. Randomized data suggest 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 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 is now performed for most patients at the time of diagnosis, and 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 either 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.

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

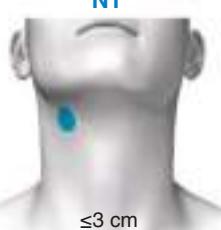
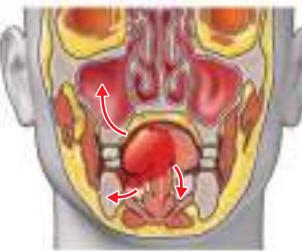
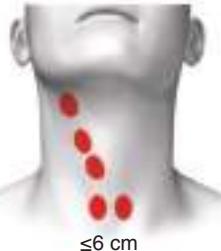
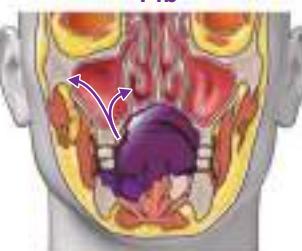
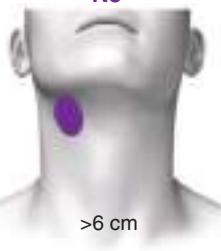
Definition of TNM					Stage groupings		
Stage I  T1	Tumor ≤ 2 cm in greatest dimension without extraparenchymal extension	 N0	N0- No regional lymph node metastasis		T1	N0	M0
Stage II  T2	Tumor ≥ 2 cm but not more than 4 cm in greatest dimension without extraparenchymal extension	 N0	N0- No regional lymph node metastasis		T2	N0	M0
Stage III  T3	Tumor ≥ 4 cm and/or tumor having extraparenchymal extension	 N1 ≤3 cm	N1- Metastasis in a single ipsilateral lymph node, ≤ 3 cm in greatest dimension		T3	N0	M0
					T1	N1	M0
					T2	N1	M0
					T3	N1	M0
Stage IVA  T4a	Tumor invades skin, mandible, ear canal, and/or fascial nerve	 N2 ≤6 cm	N2a- Metastasis in a single ipsilateral lymph node, >3 cm but ≤ 6 cm N2b- Metastasis in a multiple ipsilateral lymph node, none >6 cm N2c- Metastasis in a bilateral or contralateral lymph nodes, none >6 cm		T4a	N0	M0
					T4a	N1	M0
					T1	N2	M0
					T2	N2	M0
					T3	N2	M0
					T4a	N2	M0
Stage IVB  T4b	Tumor invades skull base and/or pterygoid plates and/or encases carotid artery	 N3 >6 cm	N3- Metastasis in a lymph node >6 cm in greatest dimension		T4b	Any N	M0
					Any T	N3	M0
Stage IVC		M1			Any T	Any N	M1

FIGURE 73-2 Tumor-node-metastasis (TNM) staging system.

preservation in patients with laryngeal and hypopharyngeal cancer, and it has been shown to result 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. Results seem more favorable in recent trials as more active drugs or more intensive radiotherapy schedules are used. 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 has also produced improved survival in patients with advanced nasopharyngeal cancer. The outcome of HPV-related cancers seems to be especially favorable following cisplatin-based chemoradiotherapy.

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, there has been a resurgence of interest in surgical approaches using robotic surgery which allows better visualization of the base of tongue and tonsil. While technically feasible, this approach remains investigational at this time since a large number of patients with disease involving multiple lymph nodes disease will still require post-operative 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 chemotherapy. Response rates to chemotherapy average only 30–50%; the durations of response are short, and the median survival time is 8–10 months. Therefore, chemotherapy provides transient symptomatic benefit. Drugs with single-agent activity in this setting include methotrexate, 5-FU, cisplatin, paclitaxel, and docetaxel. Combinations of cisplatin with 5-FU, carboplatin with 5-FU, and cisplatin or carboplatin with paclitaxel or docetaxel are frequently used.

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 was

shown to result in a significant increase in median survival. Drugs targeting specific mutations are under investigation, but no such strategy has yet been shown to be feasible in head and neck cancer.

IMMUNOTHERAPIES

Inhibitors of the immune suppressive lymphocyte-surface receptor PD-1 have shown activity in squamous cell cancers of the head and neck. A randomized trial evaluating the PD-1 inhibitor nivolumab vs traditional chemotherapy in second-line treatment of patients with current or metastatic disease showed a significant increase in survival time (7.5 vs 5.1 months) and 1-year survival rates with fewer severe treatment-related toxicities. Similarly, the PD-1 inhibitor pembrolizumab was shown to result in encouraging response rates and survival times in a single-arm phase II trial.

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

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

- D'CRUZ AK et al: Elective versus therapeutic neck dissection in node-negative oral cancer. *N Engl J Med* 373:521, 2015.
- FERRIS RL et al: Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med* 375:1856, 2016.
- FORASTIERE AA et al: Long-term results of RTOG 91-11: A comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol* 31:845, 2013.
- 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.
- HAYES DN, VAN WAES C, SEIWERT TY: Genetic landscape of human papillomavirus-associated head and neck cancer and comparison to tobacco-related tumors. *J Clin Oncol* 33:3227, 2015.
- KANG H et al: Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. *Clin Cancer Res* 23:283, 2017.
- VOKES EE, AGRAWAL N, SEIWERT TY: HPV-associated head and neck cancer. *J Natl Cancer Inst* 107:dvj344, 2015.

74

Neoplasms of the Lung

Leora Horn, Christine M. Lovly



Lung cancer, which was rare before 1900 with fewer than 400 cases described in the medical literature, is considered a disease of modern man. By the mid-twentieth century, lung cancer had become epidemic and firmly established as the leading cause of cancer-related death in North America and Europe, killing over three times as many men as prostate cancer and nearly twice as many women as breast cancer. Tobacco consumption is the primary cause of lung cancer, a reality firmly established in the mid-twentieth century and codified with the release of the U.S. Surgeon General's 1964 report on the health effects of tobacco smoking. Following the report, cigarette use started to decline in North America and parts of Europe, and with it, so did the incidence of lung cancer. Unfortunately, in many parts of the world cigarette use continues to increase, and along with it, the incidence of lung cancers is also rising. Although tobacco smoking remains the primary cause of lung cancer worldwide, approximately 60% of new lung cancers in the United States occur in former smokers (smoked ≥100 cigarettes per lifetime, quit ≥1 year), many of whom quit decades ago, or never smokers (smoked <100 cigarettes per lifetime). Moreover, one in five women and one in 12 men diagnosed with lung cancer have never smoked. Given the magnitude of the problem, it is incumbent that every internist has a general knowledge of lung cancer and its management.

EPIDEMIOLOGY

Lung cancer is the most common cause of cancer death among American men and women. Approximately 225,000 individuals will be diagnosed with lung cancer in the United States in 2017, and over 150,000 individuals will die from the disease. Lung cancer is uncommon below age 40, with rates increasing until age 80, after which the rate tapers off. The projected lifetime probability of developing lung cancer is estimated to be ~8% among males and ~6% among females. The incidence of lung cancer varies by racial and ethnic group, with the highest age-adjusted incidence rates among African Americans. The excess in age-adjusted rates among African Americans occurs only among men, but examinations of age-specific rates show that below age 50, mortality from lung cancer is more than 25% higher among African American than Caucasian women. Incidence and mortality rates among Hispanics and Native and Asian Americans are ~40–50% those of whites.

RISK FACTORS

Cigarette smokers have a 10-fold or greater increased risk of developing lung cancer compared to those who have never smoked. A large scale genomic study suggested that one genetic mutation is induced for every 15 cigarettes smoked. The risk of lung cancer is lower among persons who quit smoking than among those who continue smoking; former smokers have a ninefold increased risk of developing lung cancer compared to men who have never smoked versus the 20-fold excess in those who continue to smoke. The size of the risk reduction increases with the length of time the person has quit smoking, although generally even long-term former smokers have higher risks of lung cancer than those who never smoked. Cigarette smoking has been shown to increase the risk of all the major types of lung cancer. Environmental tobacco smoke (ETS) or second-hand smoke is also an established cause of lung cancer. The risk from ETS is less than from active smoking, with about a 20–30% increase in lung cancer observed among never smokers married for many years to smokers, in comparison to the 2000% increase among continuing active smokers.

Although cigarette smoking is the cause of the majority of lung cancers, several other risk factors have been identified, including occupational exposures to asbestos, arsenic, bischloromethyl ether, hexavalent chromium, mustard gas, nickel (as in certain nickel-refining processes), and polycyclic aromatic hydrocarbons. Occupational observations also have provided insight into possible mechanisms of lung

cancer induction. For example, the risk of lung cancer among asbestos-exposed workers is increased primarily among those with underlying asbestosis, raising the possibility that the scarring and inflammation produced by this fibrotic nonmalignant lung disease may in many cases (although likely not in all) be the trigger for asbestos-induced lung cancer. Several other occupational exposures have been associated with increased rates of lung cancer, but the causal nature of the association is not as clear.

The risk of lung cancer appears to be higher among individuals with low fruit and vegetable intake during adulthood. This observation led to hypotheses that specific nutrients, in particular retinoids and carotenoids, might have chemopreventative effects for lung cancer. However, randomized trials failed to validate this hypothesis. In fact, studies found that the incidence of lung cancer was increased among smokers with supplementation. Ionizing radiation is also an established lung carcinogen, most convincingly demonstrated from studies showing increased rates of lung cancer among survivors of the atom bombs dropped on Hiroshima and Nagasaki and large excesses among workers exposed to alpha irradiation from radon in underground uranium mining. Prolonged exposure to low-level radon in homes might impart a risk of lung cancer equal or greater than that of ETS. Prior lung diseases such as chronic bronchitis, emphysema, and tuberculosis have been linked to increased risks of lung cancer as well.

Smoking Cessation Given the undeniable link between cigarette smoking and lung cancer (not even addressing other tobacco-related illnesses), physicians must promote tobacco abstinence. Physicians also must help their patients who smoke to stop smoking. Smoking cessation, even well into middle age, can minimize an individual's subsequent risk of lung cancer. Stopping tobacco use before middle age avoids more than 90% of the lung cancer risk attributable to tobacco. However, there is little health benefit derived from just "cutting back." Importantly, smoking cessation can even be beneficial in individuals with an established diagnosis of lung cancer, as it is associated with improved survival, fewer side effects from therapy, and an overall improvement in quality of life. Moreover, smoking can alter the metabolism of many chemotherapy drugs, potentially adversely altering the toxicities and therapeutic benefits of the agents. Consequently, it is important to promote smoking cessation even *after* the diagnosis of lung cancer is established.

Physicians need to understand the essential elements of smoking cessation therapy. The individual must want to stop smoking and must be willing to work hard to achieve the goal of smoking abstinence. Self-help strategies alone only marginally affect quit rates, whereas individual and combined pharmacotherapies in combination with counseling can significantly increase rates of cessation. Therapy with an antidepressant (e.g., bupropion) and nicotine replacement therapy (varenicline, a $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist) are approved by the U.S. Food and Drug Administration (FDA) as first-line treatments for nicotine dependence. However, both drugs have been reported to increase suicidal ideation and must be used with caution. In a randomized trial, varenicline was shown to be more efficacious than bupropion or placebo. Prolonged use of varenicline beyond the initial induction phase proved useful in maintaining smoking abstinence. Clonidine and nortriptyline are recommended as second-line treatments. (Chap. 448).

Inherited Predisposition to Lung Cancer Exposure to environmental carcinogens, such as those found in tobacco smoke, induce or facilitate the transformation from bronchoepithelial cells to the malignant phenotype. The contribution of carcinogens on transformation is modulated by polymorphic variations in genes that affect aspects of carcinogen metabolism. Certain genetic polymorphisms of the P450 enzyme system, specifically CYP1A1, and chromosome fragility are associated with the development of lung cancer. These genetic variations occur at relatively high frequency in the population, but their contribution to an individual's lung cancer risk is generally low. However, because of their population frequency, the overall impact on lung cancer risk could be high. In addition, environmental factors,

as modified by inherited modulators, likely affect specific genes by deregulating important pathways to enable the cancer phenotype.

First-degree relatives of lung cancer probands have a two- to threefold excess risk of lung cancer and other cancers, many of which are not smoking-related. These data suggest that specific genes and/or genetic variants may contribute to susceptibility to lung cancer. However, very few such genes have yet been identified. Individuals with inherited mutations in *RB* (patients with retinoblastoma living to adulthood) and *p53* (Li-Fraumeni syndrome) genes may develop lung cancer. Common gene variants involved in lung cancer have been recently identified through large, collaborative, genome-wide association studies. These studies identified three separate loci that are associated with lung cancer (5p15, 6p21, and 15q25) and include genes that regulate acetylcholine nicotinic receptors and telomerase production. A rare germline mutation (T790M) involving the epidermal growth factor receptor (EGFR) maybe be linked to lung cancer susceptibility in never smokers. Likewise, a susceptibility locus on chromosome 6q greatly increases risk lung cancer risk among light and never smokers. Although progress has been made, there is a significant amount of work that remains to be done in identifying heritable risk factors for lung cancer. Currently no molecular criteria are suitable to select patients for more intense screening programs or for specific chemopreventative strategies.

PATHEOLOGY

The World Health Organization (WHO) defines lung cancer as tumors arising from the respiratory epithelium (bronchi, bronchioles, and alveoli). The WHO classification system divides epithelial lung cancers into four major cell types: small-cell lung cancer (SCLC), adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma; the latter three types are collectively known as non-small-cell carcinomas (NSCLCs) (Fig. 74-1). Small-cell carcinomas consist of small cells with scant cytoplasm, ill-defined cell borders, finely granular nuclear chromatin, absent or inconspicuous nucleoli, and a high mitotic count. SCLC may be distinguished from NSCLC by the presence of neuroendocrine markers including CD56, neural cell adhesion molecule (NCAM), synaptophysin, and chromogranin. Adenocarcinomas possess glandular differentiation or mucin production and may show acinar, papillary, lepidic, or solid features or a mixture of these patterns. Squamous cell carcinomas of the lung are morphologically identical to extrapulmonary squamous cell carcinomas and cannot be distinguished by immunohistochemistry alone. Squamous cell tumors show keratinization and/or intercellular bridges that arise from bronchial epithelium. The tumor tends to consists of sheets of cells rather than the three-dimensional groups of cells characteristic of adenocarcinomas. Large-cell carcinomas comprise less than 10% of lung carcinomas. These tumors lack the cytologic and architectural features of small-cell carcinoma and glandular or squamous differentiation. Together these four histologic types account for ~90% of all epithelial lung cancers.

All histologic types of lung cancer can develop in current and former smokers, although squamous and small-cell carcinomas are most commonly associated with heavy tobacco use. Through the first half of the twentieth century, squamous carcinoma was the most common subtype of NSCLC diagnosed in the United States. However, with the decline

in cigarette consumption over the past six decades, adenocarcinoma has become the most frequent histologic subtype of lung cancer in the United States as both squamous carcinoma and small-cell carcinoma are on the decline. In lifetime never smokers or former light smokers (<10 pack-year history), women, and younger adults (<60 years), adenocarcinoma tends to be the most common form of lung cancer.

In addition to distinguishing between SCLC and NSCLC, because these tumors have quite different natural histories and therapeutic approaches (see below), it is necessary to classify if NSCLC is squamous or nonsquamous because of the recognition that some active chemotherapy agents perform quite differently in squamous carcinomas versus adenocarcinomas and the different recommendations for molecular testing. The revised 2011 classification system, developed jointly by the International Association for the Study of Lung Cancer, the American Thoracic Society, and the European Respiratory Society, provides an integrated approach to the classification of lung adenocarcinoma that includes clinical, molecular, radiographic, and pathologic information.

It is recognized that most lung cancers present in an advanced stage and are often diagnosed based on small biopsies or cytologic specimens, rendering clear histologic distinctions difficult if not impossible. This was addressed by the WHO 2015 revised classification of lung tumors. The distinction between squamous and nonsquamous lung cancer is viewed as critical to optimal therapeutic decision making, a diagnosis of *non-small-cell carcinoma, not otherwise specified* is no longer considered acceptable. This distinction can be achieved using a single marker for adenocarcinoma (thyroid transcription factor-1 or napsin-A) plus a squamous marker (p40 or p63) and/or mucin stains. If tissue is limited and a clear morphological pattern is evident, a diagnosis can be made without immunohistochemistry staining. Both classification systems recommend preservation of sufficient specimen material for appropriate molecular testing necessary to help guide therapeutic decision making (see below).

The terms *adenocarcinoma in situ* and *minimally invasive adenocarcinoma* are now recommended for small solitary adenocarcinomas (≤ 3 cm) with either pure lepidic growth (term used to describe single-layered growth of atypical cuboidal cells coating the alveolar walls) or predominant lepidic growth with ≤ 5 mm invasion. Individuals with these entities experience 100% or near 100% 5-year disease-free survival with complete tumor resection. *Invasive adenocarcinomas*, representing more than 70–90% of surgically resected lung adenocarcinomas, are now classified by their predominant pattern: lepidic, acinar, papillary, and solid patterns. Lepidic-predominant subtype has a favorable prognosis, acinar and papillary have an intermediate prognosis, and solid-predominant has a poor prognosis. The terms *signet ring* and *clear cell adenocarcinoma* have been eliminated from the variants of invasive lung adenocarcinoma, whereas the term *micropapillary*, a subtype with a particularly poor prognosis, has been added. Because of prognostic implications, squamous cell carcinoma has also been modified to consist of keratinizing, nonkeratinizing and basaloid, analogous to head and neck cancers.

IMMUNOHISTOCHEMISTRY

The diagnosis of lung cancer most often rests on the morphologic or cytologic features correlated with clinical and radiographic findings. Immunohistochemistry may be used to verify neuroendocrine differentiation within a tumor, with markers such as neuron-specific enolase (NSE), CD56 or NCAM, synaptophysin, chromogranin, and Leu7. Immunohistochemistry is also helpful in differentiating primary from metastatic adenocarcinomas; thyroid transcription factor-1 (TTF-1), identified in tumors of thyroid and pulmonary origin, is positive in over 70% of pulmonary adenocarcinomas and is a reliable indicator of primary lung cancer, provided a thyroid primary has been excluded. A negative TTF-1, however, does not exclude the possibility of a lung primary. TTF-1 is also positive in neuroendocrine tumors of pulmonary and extrapulmonary origin. Napsin-A (Nap-A) is an aspartic protease that plays an important role in maturation of surfactant B7 and is expressed in cytoplasm of type II pneumocytes. In several studies, Nap-A has been reported in >90% of primary lung adenocarcinomas.

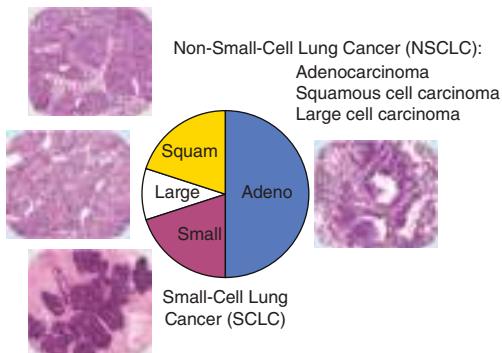


FIGURE 74-1 Traditional histologic view of lung cancer.

Notably, a combination of Nap-A and TTF-1 is useful in distinguishing primary lung adenocarcinoma (Nap-A positive, TTF-1 positive) from primary lung squamous cell carcinoma (Nap-A negative, TTF-1 negative) and primary SCLC (Nap-A negative, TTF-1 positive). Cytokeratins 7 and 20 used in combination can help narrow the differential diagnosis; nonsquamous NSCLC, SCLC, and mesothelioma may stain positive for CK7 and negative for CK20, whereas squamous cell lung cancer often will be both CK7 and CK20 negative. p63 is a useful marker for the detection of NSCLCs with squamous differentiation when used in cytologic pulmonary samples. Mesothelioma can be easily identified ultrastructurally, but it has historically been difficult to differentiate from adenocarcinoma through morphology and immunohistochemical staining. Several markers in the last few years have proven to be more helpful including CK5/6, calretinin, and Wilms tumor gene-1 (WT-1), all of which show positivity in mesothelioma.

MOLECULAR PATHOGENESIS

Cancer is a disease involving dynamic changes in the genome. As proposed by Hanahan and Weinberg, virtually all cancer cells acquire six hallmark capabilities: self-sufficiency in growth signals, insensitivity to antitumor signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The order in which these hallmark capabilities are acquired appears quite variable and can differ from tumor to tumor. Events leading to acquisition of these hallmarks can vary widely, although broadly, cancers arise as a result from accumulations of gain-of-function mutations in oncogenes and loss-of-function mutations in tumor-suppressor genes. Further complicating the study of lung cancer, the sequence of events that lead to disease is clearly different for the various histopathologic entities.

The exact cell of origin for lung cancers is not clearly defined. Whether one cell of origin leads to all histologic forms of lung cancer is unclear. However, for lung adenocarcinoma, evidence suggests that type II epithelial cells (or alveolar epithelial cells) have the capacity to give rise to tumors. For SCLC, cells of neuroendocrine origin have been implicated as precursors.

For cancers in general, one theory holds that a small subset of the cells within a tumor (i.e., "stem cells") are responsible for the full malignant behavior of the tumor. As part of this concept, the large bulk of the cells in a cancer are "offspring" of these cancer stem cells. While clonally related to the cancer stem cell subpopulation, most cells by themselves cannot regenerate the full malignant phenotype. The stem cell concept may explain the failure of standard medical therapies to eradicate lung cancers, even when there is a clinical complete response. Disease recurs because therapies do not eliminate the stem cell component, which may be more resistant to chemotherapy or targeted therapy. Precise human lung cancer stem cells have yet to be identified.

Lung cancer cells harbor multiple chromosomal abnormalities, including mutations, amplifications, insertions, deletions, and translocations. One of the earliest sets of oncogenes found to be aberrant was the MYC family of transcription factors (*MYC*, *MYCN*, and *MYCL*). *MYC* is most frequently activated via gene amplification or transcriptional dysregulation in both SCLC and NSCLC. Currently, there are no *MYC*-specific drugs.

Among lung cancer histologies, adenocarcinomas have been the most extensively catalogued for recurrent genomic gains and losses as well as for somatic mutations (Fig. 74-2). While multiple different kinds of aberrations have been found, a major class involves "driver mutations," which are mutations that occur in genes encoding signaling proteins that when aberrant, drive initiation and maintenance of tumor cells. Importantly, driver mutations can serve as potential Achilles' heels

for tumors, if their gene products can be targeted appropriately. For example, one set of mutations involves the epidermal growth factor receptor (EGFR), which belongs to the ERBB (HER) family of proto-oncogenes, including *EGFR* (ERBB1), *HER2/neu* (ERBB2), *HER3* (ERBB3), and *HER4* (ERBB4). These genes encode cell-surface receptors consisting of an extracellular ligand-binding domain, a transmembrane structure, and an intracellular tyrosine kinase (TK) domain. The binding of ligand to receptor activates receptor dimerization and TK autophosphorylation, initiating a cascade of intracellular events, and leading to increased cell proliferation, angiogenesis, metastasis, and a decrease in apoptosis. Lung adenocarcinomas can arise when tumors express mutant EGFR. These same tumors display high sensitivity to small-molecule EGFR TK inhibitors (TKIs). Additional examples of driver mutations in lung adenocarcinoma include the GTPase *KRAS*, the serine-threonine kinase *BRAF*, and the lipid kinase *PIK3CA*. More recently, additional subsets of lung adenocarcinoma have been identified as defined by the presence of specific chromosomal rearrangements resulting in the aberrant activation of the tyrosine kinases *ALK*, *ROS1*, *NTRK* and *RET*. Notably, most driver mutations in lung cancer appear to be mutually exclusive, suggesting that acquisition of one of these mutations is sufficient to drive tumorigenesis. Although driver mutations have mostly been found in adenocarcinomas, three potential molecular targets recently have been identified in squamous cell lung carcinomas: *FGFR1* amplification, *DDR2* mutations, and *PIK3CA* mutations/*PTEN* loss (Table 74-1).

A large number of tumor-suppressor genes have also been identified that are inactivated during the pathogenesis of lung cancer. These include *TP53*, *RB1*, *RASSF1A*, *CDKN2A/B*, *LKB1* (*STK11*), and *FHIT*. Nearly 90% of SCLCs harbor mutations in *TP53* and *RB1*. Several tumor-suppressor genes on chromosome 3p appear to be involved in nearly all lung cancers. Allelic loss for this region occurs very early in lung cancer pathogenesis, including in histologically normal smoking-damaged lung epithelium.

EARLY DETECTION AND SCREENING

In lung cancer, clinical outcome is related to the stage at diagnosis, and hence, it is generally assumed that early detection of occult tumors will

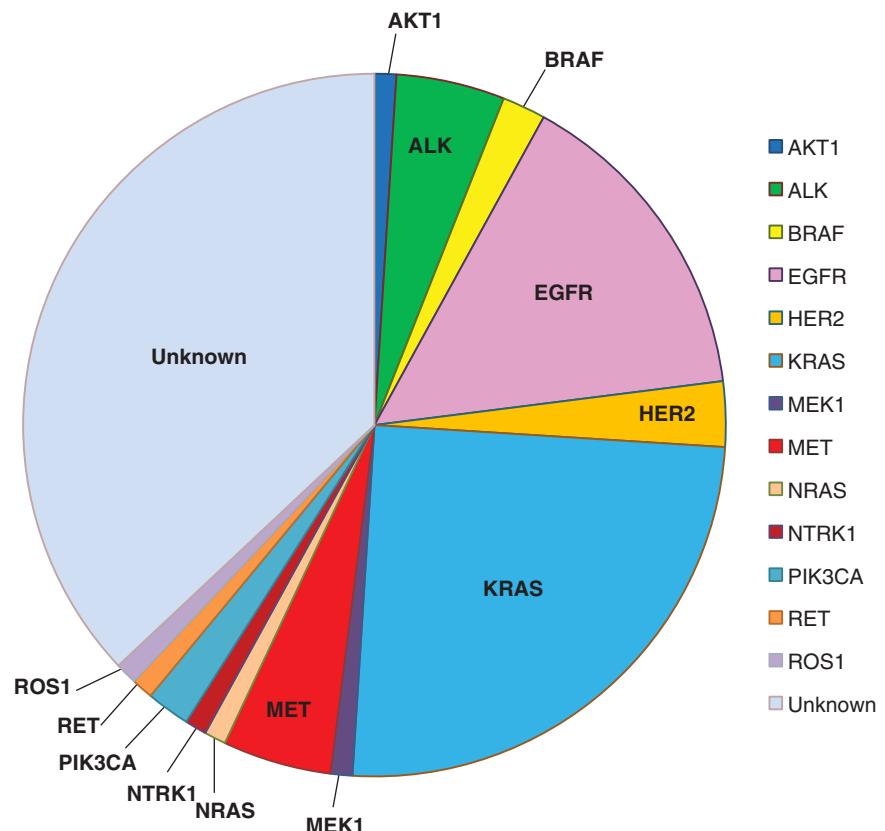


FIGURE 74-2 Driver mutations in lung adenocarcinomas.

TABLE 74-1 Driver Mutations in Non-Small-Cell Lung Cancer (NSCLC)

GENE	ALTERATION	FREQUENCY IN NSCLC	TYPICAL HISTOLOGY
AKT1	Mutation	1%	Adenocarcinoma, squamous
ALK	Rearrangement	3–7%	Adenocarcinoma
BRAF	Mutation	1–3%	Adenocarcinoma
DDR2	Mutation	~4%	Squamous
EGFR	Mutation	10–35%	Adenocarcinoma
FGFR1	Amplification	~20%	Squamous
HER2	Mutation	2–4%	Adenocarcinoma
KRAS	Mutation	15–25%	Adenocarcinoma
MEK1	Mutation	1%	Adenocarcinoma
MET	Amplification	2–4%	Adenocarcinoma
NRAS	Mutation	1%	Adenocarcinoma
NTRK	Rearrangement	1–2%	Adenocarcinoma
PIK3CA	Mutation	1–3%	Squamous
PTEN	Mutation	4–8%	Squamous
ROS1	Rearrangement	1–2%	Adenocarcinoma

lead to improved survival. Early detection is a process that involves screening tests, surveillance, diagnosis, and early treatment. Screening refers to the use of tests across a healthy population in order to identify individuals who harbor asymptomatic disease. For a screening program to be successful, there must be a high burden of disease within the target population; the test must be sensitive, specific, accessible, and cost effective; and there must be effective treatment that can reduce mortality. With any screening procedure, it is important to consider the possible influence of *lead-time bias* (detecting the cancer earlier without an effect on survival), *length-time bias* (indolent cancers are detected on screening and may not affect survival, whereas aggressive cancers are likely to cause symptoms earlier in patients and are less likely to be detected), and *overdiagnosis* (diagnosing cancers so slow growing that they are unlikely to cause the death of the patient).

Because a majority of lung cancer patients present with advanced disease beyond the scope of surgical resection, there is understandable skepticism about the value of screening in this condition. Indeed, randomized controlled trials conducted in the 1960s to 1980s using screening chest x-rays (CXR), with or without sputum cytology, reported no impact on lung cancer-specific mortality in patients characterized as high risk (males age ≥ 45 years with a smoking history). These studies have been criticized for their design, statistical analyses, and outdated imaging modalities. The results of the more recently conducted Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) are consistent with these earlier reports. Initiated in 1993, participants in the PLCO lung cancer screening trial received annual CXR screening for 4 years, whereas participants in the usual care group received no interventions other than their customary medical care. The diagnostic follow-up of positive screening results was determined by participants and their physicians. The PLCO trial differed from previous lung cancer screening studies in that women and never smokers were eligible. The study was designed to detect a 10% reduction in lung cancer mortality in the interventional group. A total of 154,901 individuals between 55 and 74 years of age were enrolled (77,445 assigned to annual CXR screenings; 77,456 assigned to usual care). Participant

demographics and tumor characteristics were well balanced between the two groups. Through 13 years of follow-up, cumulative lung cancer incidence rates (20.1 vs 19.2 per 10,000 person-years; rate ratio [RR], 1.05; 95% confidence interval [CI], 0.98–1.12) and lung cancer mortality ($n = 1213$ vs $n = 1230$) were identical between the two groups. The stage and histology of detected cancers in the two groups also were similar. These data corroborate previous recommendations against CXR screening for lung cancer.

In contrast to CXR, low-dose, noncontrast, thin-slice spiral chest computed tomography (LDCT) has emerged as an effective tool to screen for lung cancer. In nonrandomized studies conducted in the 1990s, LDCT scans were shown to detect more lung nodules and cancers than standard CXR in selected high-risk populations (e.g., age ≥ 60 years and a smoking history of ≥ 10 pack-years). Notably, up to 85% of the lung cancers discovered in these trials were classified as stage I disease and therefore considered potentially curable with surgical resection.

These data prompted the National Cancer Institute (NCI) to initiate the National Lung Screening Trial (NLST), a randomized study designed to determine if LDCT screening could reduce mortality from lung cancer in high-risk populations as compared with standard posterior anterior CXR. High-risk patients were defined as individuals between 55 and 74 years of age, with a ≥ 30 pack-year history of cigarette smoking; former smokers must have quit within the previous 15 years. Excluded from the trial were individuals with a previous lung cancer diagnosis, a history of hemoptysis, an unexplained weight loss of >15 lb in the preceding year, or a chest CT within 18 months of enrollment. A total of 53,454 persons were enrolled and randomized to annual screening yearly for three years (LDCT screening, $n = 26,722$; CXR screening, $n = 26,732$). Any noncalcified nodule measuring ≥ 4 mm in any diameter found on LDCT and CXR images with any noncalcified nodule or mass were classified as “positive.” Participating radiologists had the option of not calling a final screen positive if a noncalcified nodule had been stable on the three screening examinations. Overall, 39.1% of participants in the LDCT group and 16% in the CXR group had at least one positive screening result. Of those who screened positive, the false-positive rate was 96.4% in the LDCT group and 94.5% in the CXR group. This was consistent across all three rounds. In the LDCT group, 1060 cancers were identified compared with 941 cancers in the CXR group (645 vs 572 per 100,000 person-years; RR, 1.13; 95% CI, 1.03 to 1.23). Nearly twice as many early-stage 1A cancers were detected in the LDCT group compared with the CXR group (40% vs 21%). The overall rates of lung cancer death were 247 and 309 deaths per 100,000 participants in the LDCT and CXR groups, respectively, representing a 20% reduction in lung cancer mortality in the LDCT-screened population (95% CI, 6.8–26.7%; $p = 0.004$). Compared with the CXR group, the rate of death in the LDCT group from any cause was reduced by 6.7% (95% CI, 1.2–13.6; $p = 0.02$) (Table 74-2). The number needed to screen (NNTS) to prevent one lung cancer death was calculated to be 320.

LDCT screening for lung cancer comes with known risks including a high rate of false-positive results, false-negative results, potential for unnecessary follow-up testing, radiation exposure, overdiagnosis, changes in anxiety level and quality of life, and substantial financial costs. By far the biggest challenge confronting the use of CT screening is the high false-positive rate. False positives can have a substantial impact on patients through the expense and risk of unneeded further evaluation and emotional stress. The management of these patients

TABLE 74-2 Results of National Lung Screening Trial

	EVENT NUMBER		RATES OF EVENTS PER 100,000 PERSON-YEARS		RELATIVE RISK (95% CI)	P VALUE
	LDCT (N = 26,772)	CXR (N = 26,732)	LDCT	CXR		
Lung cancer mortality	356	443	247	309	0.80 (0.73–0.93)	.004
All-cause mortality	1877	2000	1303	1395	0.93 (0.86–0.99)	.02
Mortality not due to lung cancer	1521	1557	1056	1086	0.99 (0.95–1.02)	.51

Abbreviations: CI, confidence interval; CXR, chest x-ray; LDCT, low-dose computed tomography; RR, rate ratio.

Source: Modified from PB Bach et al: JAMA 307:2418, 2012.

TABLE 74-3 The Benefits and Harms of LDCT Screening for Lung Cancer Based on NLST Data

	LDCT	CXR
Benefits: How did CT scans help compared to CXR?		
4 in 1000 fewer died from lung cancer	13 in 1000	17 in 1000
5 in 1000 fewer died from all causes	70 in 1000	75 in 1000
Harms: What problems did CT scans cause compared to CXR?		
223 in 1000 had at least 1 false alarm	365 in 1000	142 in 1000
18 in 1000 had a false alarm leading to an invasive procedure	25 in 1000	7 in 1000
2 in 1000 had a major complication from an invasive procedure	3 in 1000	1 in 1000

Abbreviations: CXR, chest x-ray; LDCT, low-dose computed tomography; NLST, National Lung Screening Trial.

Source: Modified from S Woloshin et al: N Engl J Med 367:1677, 2012.

usually consists of serial CT scans over time to see if the nodules grow, attempted fine-needle aspirates, or surgical resection. At \$300 per scan (NCI estimated cost), the outlay for initial LDCT alone could run into the billions of dollars annually, an expense that only further escalates when factoring in various downstream expenditures an individual might incur in the assessment of positive findings. A formal cost-effectiveness analysis of the NLST demonstrated differences between sex, age, and current smoking status and the method of follow up. Despite some questions, low dose LDCT screening has been recommended for all patients meeting criteria for enrollment on NLST. When discussing the option of LDCT screening, use of absolute risks rather than relative risks is helpful because studies indicate the public can process absolute terminology more effectively than relative risk projections. A useful guide has been developed by the NCI to help patients and physicians assess the benefits and harms of LDCT screening for lung cancer (**Table 74-3**).

CLINICAL MANIFESTATIONS

Over half of all patients diagnosed with lung cancer present with locally advanced or metastatic disease at the time of diagnosis. The majority of patients present with signs, symptoms, or laboratory abnormalities that can be attributed to the primary lesion, local tumor growth, invasion or obstruction of adjacent structures, growth at distant metastatic sites, or a paraneoplastic syndrome (**Tables 74-4 and 74-5**). The prototypical lung cancer patient is a current or former smoker of either sex, usually in the seventh decade of life. A history of chronic cough with or without hemoptysis in a current or former smoker with chronic obstructive pulmonary disease (COPD) age 40 years or older should prompt a thorough investigation for lung cancer even in the face of a normal CXR. A persistent pneumonia without constitutional symptoms and

TABLE 74-5 Clinical Findings Suggestive of Metastatic Disease

Symptoms elicited in history	<ul style="list-style-type: none"> Constitutional: weight loss >10 lb Musculoskeletal: pain Neurologic: headaches, syncope, seizures, extremity weakness, recent change in mental status
Signs found on physical examination	<ul style="list-style-type: none"> Lymphadenopathy (>1 cm) Hoarseness, superior vena cava syndrome Bone tenderness Hepatomegaly (>13 cm span) Focal neurologic signs, papilledema Soft-tissue mass
Routine laboratory tests	<ul style="list-style-type: none"> Hematocrit, <40% in men; <35% in women Elevated alkaline phosphatase, GGT, SGOT, and calcium levels

Abbreviations: GGT, gamma-glutamyltransferase; SGOT, serum glutamic-oxaloacetic transaminase.

Source: Reproduced with permission from GA Silvestri et al: Chest 123(1 Suppl): 147S, 2003.

unresponsive to repeated courses of antibiotics also should prompt an evaluation for the underlying cause. Lung cancer arising in a lifetime never smoker is more common in women and East Asians. Such patients also tend to be younger than their smoking counterparts at the time of diagnosis. The clinical presentation of lung cancer in never smokers tends to mirror that of current and former smokers.

Patients with central or endobronchial growth of the primary tumor may present with cough, hemoptysis, wheeze, stridor, dyspnea, or postobstructive pneumonitis. Peripheral growth of the primary tumor may cause pain from pleural or chest wall involvement, dyspnea on a restrictive basis, and symptoms of a lung abscess resulting from tumor cavitation. Regional spread of tumor in the thorax (by contiguous growth or by metastasis to regional lymph nodes) may cause tracheal obstruction, esophageal compression with dysphagia, recurrent laryngeal paralysis with hoarseness, phrenic nerve palsy with elevation of the hemidiaphragm and dyspnea, and sympathetic nerve paralysis with Horner's syndrome (enophthalmos, ptosis, miosis, and anhydrosis). Malignant pleural effusions can cause pain, dyspnea, or cough. Pancoast (or superior sulcus tumor) syndromes result from local extension of a tumor growing in the apex of the lung with involvement of the eighth cervical and first and second thoracic nerves, and present with shoulder pain that characteristically radiates in the ulnar distribution of the arm, often with radiologic destruction of the first and second ribs. Often Horner's syndrome and Pancoast syndrome coexist. Other problems of regional spread include superior vena cava syndrome from vascular obstruction; pericardial and cardiac extension with resultant tamponade, arrhythmia, or cardiac failure; lymphatic obstruction with resultant pleural effusion; and lymphangitic spread through the lungs with hypoxemia and dyspnea. In addition, lung cancer can spread transbronchially, producing tumor growth along multiple alveolar surfaces with impairment of gas exchange, respiratory insufficiency, dyspnea, hypoxemia, and sputum production. Constitutional symptoms may include anorexia, weight loss, weakness, fever, and night sweats. Apart from the brevity of symptom duration, these parameters fail to clearly distinguish SCLC from NSCLC or even from neoplasms metastatic to lungs.

Extrathoracic metastatic disease is found at autopsy in more than 50% of patients with squamous carcinoma, 80% of patients with adenocarcinoma and large-cell carcinoma, and more than 95% of patients with SCLC. Approximately one-third of patients present with symptoms as a result of distant metastases. Lung cancer metastases may occur in virtually every organ system, and the site of metastatic involvement largely determines other symptoms. Patients with brain metastases may present with headache, nausea and vomiting, seizures, or neurologic deficits. Patients with bone metastases may present with pain, pathologic fractures, or cord compression. The latter may also occur with epidural metastases. Individuals with bone marrow invasion may present with cytopenias or leukoerythroblastosis. Those with liver metastases may present with hepatomegaly, right upper quadrant

TABLE 74-4 Presenting Signs and Symptoms of Lung Cancer

SYMPTOM AND SIGNS	RANGE OF FREQUENCY
Cough	8–75%
Weight loss	0–68%
Dyspnea	3–60%
Chest pain	20–49%
Hemoptysis	6–35%
Bone pain	6–25%
Clubbing	0–20%
Fever	0–20%
Weakness	0–10%
Superior vena cava obstruction	0–4%
Dysphagia	0–2%
Wheezing and stridor	0–2%

Source: Reproduced with permission from MA Beckles: Chest 123:97, 2003.

pain, fever, anorexia, and weight loss. Liver dysfunction and biliary obstructions are rare. Adrenal metastases are common but rarely cause pain or adrenal insufficiency unless they are large.

Paraneoplastic syndromes are common in patients with lung cancer, especially those with SCLC, and may be the presenting finding or the first sign of recurrence. In addition, paraneoplastic syndromes may mimic metastatic disease and, unless detected, lead to inappropriate palliative rather than curative treatment. Often the paraneoplastic syndrome may be relieved with successful treatment of the tumor. In some cases, the pathophysiology of the paraneoplastic syndrome is known, particularly when a hormone with biological activity is secreted by a tumor. However, in many cases, the pathophysiology is unknown. Systemic symptoms of anorexia, cachexia, weight loss (seen in 30% of patients), fever, and suppressed immunity are paraneoplastic syndromes of unknown etiology or at least not well defined. Weight loss greater than 10% of total body weight is considered a bad prognostic sign. Endocrine syndromes are seen in 12% of patients; hypercalcemia resulting from ectopic production of parathyroid hormone (PTH), or more commonly, PTH-related peptide, is the most common life-threatening metabolic complication of malignancy, primarily occurring with squamous cell carcinomas of the lung. Clinical symptoms include nausea, vomiting, abdominal pain, constipation, polyuria, thirst, and altered mental status.

Hyponatremia may be caused by the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) or possibly atrial natriuretic peptide (ANP) ([Chap. 89](#)). SIADH resolves within 1–4 weeks of initiating chemotherapy in the vast majority of cases. During this period, serum sodium can usually be managed and maintained above 128 mEq/L via fluid restriction. Demeocycline can be a useful adjunctive measure when fluid restriction alone is insufficient. Vasopressin receptor antagonists like tolvaptan also have been used in the management of SIADH. However, there are significant limitations to the use of tolvaptan including liver injury and overly rapid correction of the hyponatremia, which can lead to irreversible neurologic injury. Likewise, the cost of tolvaptan may be prohibitive (as high as \$300 per tablet in some areas). Of note, patients with ectopic ANP may have worsening hyponatremia if sodium intake is not concomitantly increased. Accordingly, if hyponatremia fails to improve or worsens after 3–4 days of adequate fluid restriction, plasma levels of ANP should be measured to determine the causative syndrome.

Ectopic secretion of ACTH by SCLC and pulmonary carcinoids usually results in additional electrolyte disturbances, especially hypokalemia, rather than the changes in body habitus that occur in Cushing's syndrome from a pituitary adenoma ([Chap. 89](#)). Treatment with standard medications, such as metyrapone and ketoconazole, is largely ineffective due to extremely high cortisol levels. The most effective strategy for management of the Cushing's syndrome is effective treatment of the underlying SCLC. Bilateral adrenalectomy may be considered in extreme cases.

Skeletal-connective tissue syndromes include clubbing in 30% of cases (usually NSCLCs) and hypertrophic primary osteoarthropathy in 1–10% of cases (usually adenocarcinomas). Patients may develop periostitis, causing pain, tenderness, and swelling over the affected bones and a positive bone scan. Neurologic-myopathic syndromes are seen in only 1% of patients but are dramatic and include the myasthenic Eaton-Lambert syndrome and retinal blindness with SCLC, whereas peripheral neuropathies, subacute cerebellar degeneration, cortical degeneration, and polymyositis are seen with all lung cancer types. Many of these are caused by autoimmune responses such as the development of anti-voltage-gated calcium channel antibodies in Eaton-Lambert syndrome. Patients with this disorder present with proximal muscle weakness, usually in the lower extremities, occasional autonomic dysfunction, and rarely, cranial nerve symptoms or involvement of the bulbar or respiratory muscles. Depressed deep tendon reflexes are frequently present. In contrast to patients with myasthenia gravis, strength improves with serial effort. Some patients who respond to chemotherapy will have resolution of the neurologic abnormalities. Thus, chemotherapy is the initial treatment of choice. Paraneoplastic encephalomyelitis and sensory neuropathies, cerebellar

degeneration, limbic encephalitis, and brainstem encephalitis occur in SCLC in association with a variety of antineuronal antibodies such as anti-Hu, anti-CRMP5, and ANNA-3. Paraneoplastic cerebellar degeneration may be associated with anti-Hu, anti-Yo, or P/Q calcium channel autoantibodies. Coagulation or thrombotic or other hematologic manifestations occur in 1–8% of patients and include migratory venous thrombophlebitis (Trousseau's syndrome), nonbacterial thrombotic (marantic) endocarditis with arterial emboli, and disseminated intravascular coagulation with hemorrhage, anemia, granulocytosis, and leukoerythroblastosis. Thrombotic disease complicating cancer is usually a poor prognostic sign. Cutaneous manifestations such as dermatomyositis and acanthosis nigricans are uncommon (1%), as are the renal manifestations of nephrotic syndrome and glomerulonephritis (<1%).

DIAGNOSING LUNG CANCER

Tissue sampling is required to confirm a diagnosis in all patients with suspected lung cancer. In patients with suspected metastatic disease, a biopsy of a distant site of disease is preferred for tissue confirmation. Given the greater emphasis placed on molecular testing for NSCLC patients, a core biopsy is preferred to ensure adequate tissue for analysis. Tumor tissue may be obtained via minimally invasive techniques such as bronchial or transbronchial biopsy during fiberoptic bronchoscopy, by fine-needle aspiration (FNA) or percutaneous biopsy using image guidance, or via endobronchial ultrasound (EBUS)-guided biopsy. Depending on the location, lymph node sampling may occur via transesophageal endoscopic ultrasound-guided biopsy (EUS), EBUS, or blind biopsy. In patients with clinically palpable disease such as a lymph node or skin metastasis, a biopsy may be obtained. In patients with suspected metastatic disease, a diagnosis may be confirmed by percutaneous biopsy of a soft tissue mass, lytic bone lesion, bone marrow, pleural or liver lesion, or an adequate cell block obtained from a malignant pleural effusion. In patients with a suspected malignant pleural effusion, if the initial thoracentesis is negative, a repeat thoracentesis is warranted. Although the majority of pleural effusions are due to malignant disease, particularly if they are exudative or bloody, some may be parapneumonic. In the absence of distant disease, such patients should be considered for possible curative treatment.

The diagnostic yield of any biopsy depends on several factors including location (accessibility) of the tumor, tumor size, tumor type, and technical aspects of the diagnostic procedure including the experience level of the bronchoscopist and pathologist. In general, central lesions such as squamous cell carcinomas, small-cell carcinomas, or endobronchial lesions such as carcinoid tumors are more readily diagnosed by bronchoscopic examination, whereas peripheral lesions such as adenocarcinomas and large-cell carcinomas are more amenable to transthoracic biopsy. Diagnostic accuracy for SCLC versus NSCLC for most specimens is excellent, with lesser accuracy for subtypes of NSCLC.

Bronchoscopic specimens include bronchial brush, bronchial wash, bronchioloalveolar lavage, transbronchial FNA, and core biopsy. For more accurate histologic classification, mutation analysis, or investigational purposes, reasonable efforts (e.g., a core needle biopsy) should be made to obtain more tissue than what is contained in a routine cytology specimen obtained by FNA. Overall sensitivity for combined use of bronchoscopic methods is ~80%, and together with tissue biopsy, the yield increases to 85–90%. Like transbronchial core biopsy specimens, transthoracic core biopsy specimens are also preferred. Sensitivity is highest for larger lesions and peripheral tumors. In general, core biopsy specimens, whether transbronchial, transthoracic, or EUS-guided, are superior to other specimen types. This is primarily due to the higher percentage of tumor cells with fewer confounding factors such as obscuring inflammation and reactive nonneoplastic cells.

Sputum cytology is inexpensive and noninvasive but has a lower yield than other specimen types due to poor preservation of the cells and more variability in acquiring a good-quality specimen. The yield for sputum cytology is highest for larger and centrally located tumors such as squamous cell carcinoma and small-cell carcinoma histology. The specificity for sputum cytology averages close to 100%, although sensitivity is generally <70%. The accuracy of sputum cytology

improves with increased numbers of specimens analyzed. Consequently, analysis of at least three sputum specimens is recommended.

STAGING LUNG CANCER

Lung cancer staging consists of two parts: first, a determination of the location of the tumor and possible metastatic sites (anatomic staging), and second, an assessment of a patient's ability to withstand various antitumor treatments (physiologic staging). All patients with lung cancer should have a complete history and physical examination, with evaluation of all other medical problems, determination of performance status, and history of weight loss. The most significant dividing line is between those patients who are candidates for surgical resection and those who are inoperable but will benefit from chemotherapy, radiation therapy, or both. Staging with regard to a patient's potential for surgical resection is principally applicable to NSCLC.

■ ANATOMIC STAGING OF PATIENTS WITH LUNG CANCER

The accurate staging of patients with NSCLC is essential for determining the appropriate treatment in patients with resectable disease and for avoiding unnecessary surgical procedures in patients with advanced disease (Fig. 74-3). All patients with NSCLC should undergo initial radiographic imaging with CT scan, positron emission tomography (PET), or preferably CT-PET. PET scanning attempts to identify sites of malignancy based on glucose metabolism by measuring the uptake of ¹⁸F-fluorodeoxyglucose (FDG). Rapidly dividing cells, presumably in the lung tumors, will preferentially take up ¹⁸F-FDG and appear as a "hot spot." To date, PET has been mostly used for staging and detection of metastases in lung cancer and in the detection of nodules >15 mm in diameter. Combined ¹⁸F-FDG PET-CT imaging has been shown to improve the accuracy of staging in NSCLC compared to visual correlation of PET and CT or either study alone. CT-PET has been found to be superior in identifying pathologically enlarged mediastinal lymph nodes and extrathoracic metastases. A standardized uptake value (SUV) of >2.5 on PET is highly suspicious for malignancy. False negatives can be seen in diabetes, in lesions <8 mm, and in slow-growing tumors (e.g., carcinoid tumors or well-differentiated adenocarcinoma).

False positives can be seen in certain infections and granulomatous disease (e.g., tuberculosis). Thus, PET should never be used alone to diagnose lung cancer, mediastinal involvement, or metastases. Confirmation with tissue biopsy is required. For brain metastases, magnetic resonance imaging (MRI) is the most effective method. MRI can also be useful in selected circumstances, such as superior sulcus tumors to rule out brachial plexus involvement, but in general, MRI does not play a major role in NSCLC staging.

In patients with NSCLC, the following are contraindications to potential curative resection: extrathoracic metastases, superior vena cava syndrome, vocal cord and, in most cases, phrenic nerve paralysis, malignant pleural effusion, cardiac tamponade, tumor within 2 cm of the carina (potentially curable with combined chemoradiotherapy), metastasis to the contralateral lung, metastases to supraclavicular lymph nodes, contralateral mediastinal node metastases (potentially curable with combined chemoradiotherapy), and involvement of the main pulmonary artery. In situations where it will make a difference in treatment, abnormal scan findings require tissue confirmation of malignancy so that patients are not precluded from having potentially curative therapy.

The best predictor of metastatic disease remains a careful history and physical examination. If signs, symptoms, or findings from the physical examination suggest the presence of malignancy, then sequential imaging starting with the most appropriate study should be performed. If the findings from the clinical evaluation are negative, then imaging studies beyond CT-PET are unnecessary and the search for metastatic disease is complete. More controversial is how one should assess patients with known stage III disease. Because these patients are more likely to have asymptomatic occult metastatic disease, current guidelines recommend a more extensive imaging evaluation including imaging of the brain with either CT scan or MRI. In patients in whom distant metastatic disease has been ruled out, lymph node status needs to be assessed via a combination of radiographic imaging and/or minimally invasive techniques such as those mentioned above and/or invasive techniques such as mediastinoscopy, mediastinotomy, thoracoscopy, or thoracotomy. Approximately one-quarter to one-half of patients diagnosed with NSCLC will have mediastinal lymph node

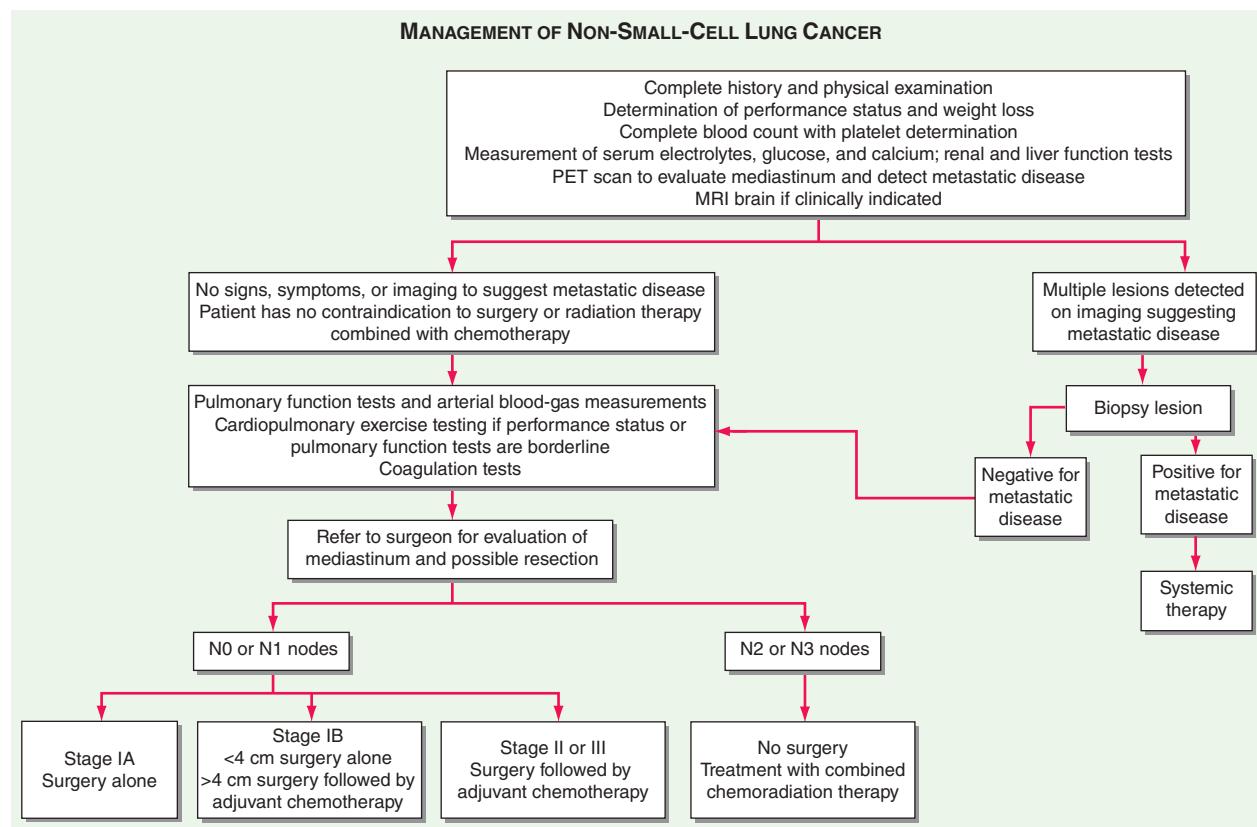


FIGURE 74-3 Algorithm for management of non-small-cell lung cancer. MRI, magnetic resonance imaging; PET, positron emission tomography.

metastases at the time of diagnosis. Lymph node sampling is recommended in all patients with enlarged nodes detected by CT or PET scan and in patients with large tumors or tumors occupying the inner third of the lung. The extent of mediastinal lymph node involvement is important in determining the appropriate treatment strategy: surgical resection followed by adjuvant chemotherapy versus combined chemoradiation alone (see below). A standard nomenclature for referring to the location of lymph nodes involved with lung cancer has evolved (Fig. 74-4).

In SCLC patients, current staging recommendations include a PET-CT scan and MRI of the brain (positive in 10% of asymptomatic patients) (Fig. 74-5). Bone marrow biopsies and aspirations are rarely performed now given the low incidence of isolated bone marrow metastases. Confirmation of metastatic disease, ipsilateral or contralateral lung nodules, or metastases beyond the mediastinum may be achieved by the same modalities recommended earlier for patients with NSCLC.

If a patient has signs or symptoms of spinal cord compression (pain, weakness, paralysis, urinary retention), a spinal CT or MRI scan and examination of the cerebrospinal fluid cytology should be performed. If metastases are evident on imaging, a neurosurgeon should be consulted for possible palliative surgical resection and/or a radiation oncologist should be consulted for palliative radiotherapy to the site of compression. If signs or symptoms of leptomeningitis develop at any time in a patient with lung cancer, an MRI of the brain and spinal cord should be performed, as well as a spinal tap, for detection of malignant cells. If the spinal tap is negative, a repeat spinal tap should be considered. There is currently no approved therapy for the treatment of leptomeningeal disease.

■ STAGING SYSTEM FOR NON-SMALL-CELL LUNG CANCER

The tumor-node-metastasis (TNM) international staging system provides useful prognostic information and is used to stage all patients

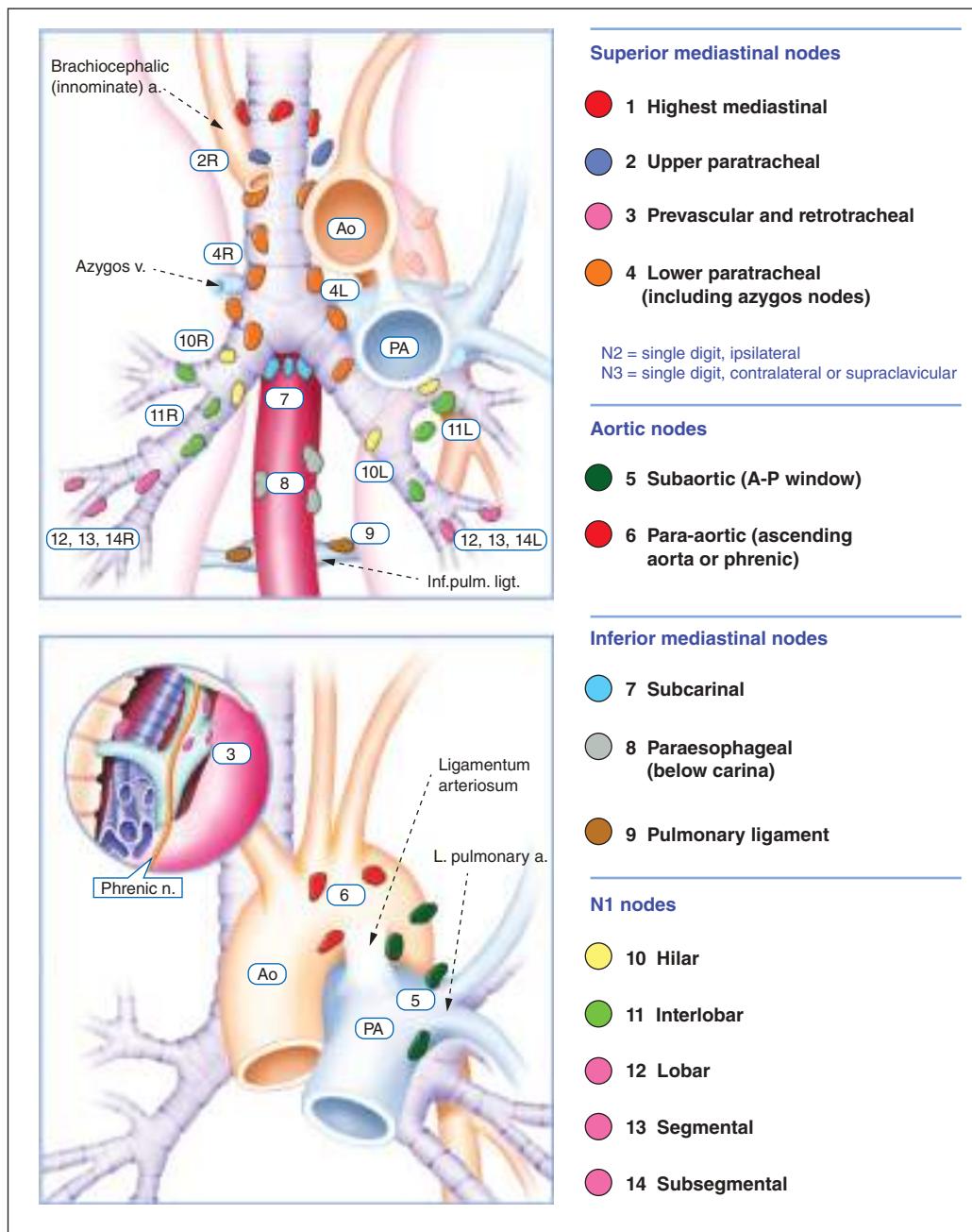


FIGURE 74-4 Lymph node stations in staging non-small-cell lung cancer. The International Association for the Study of Lung Cancer (IASLC) lymph node map, including the proposed grouping of lymph node stations into “zones” for the purposes of prognostic analyses. a., artery; Ao, aorta; Inf. pulm. ligt., inferior pulmonary ligament; n., nerve; PA, pulmonary artery; v., vein.

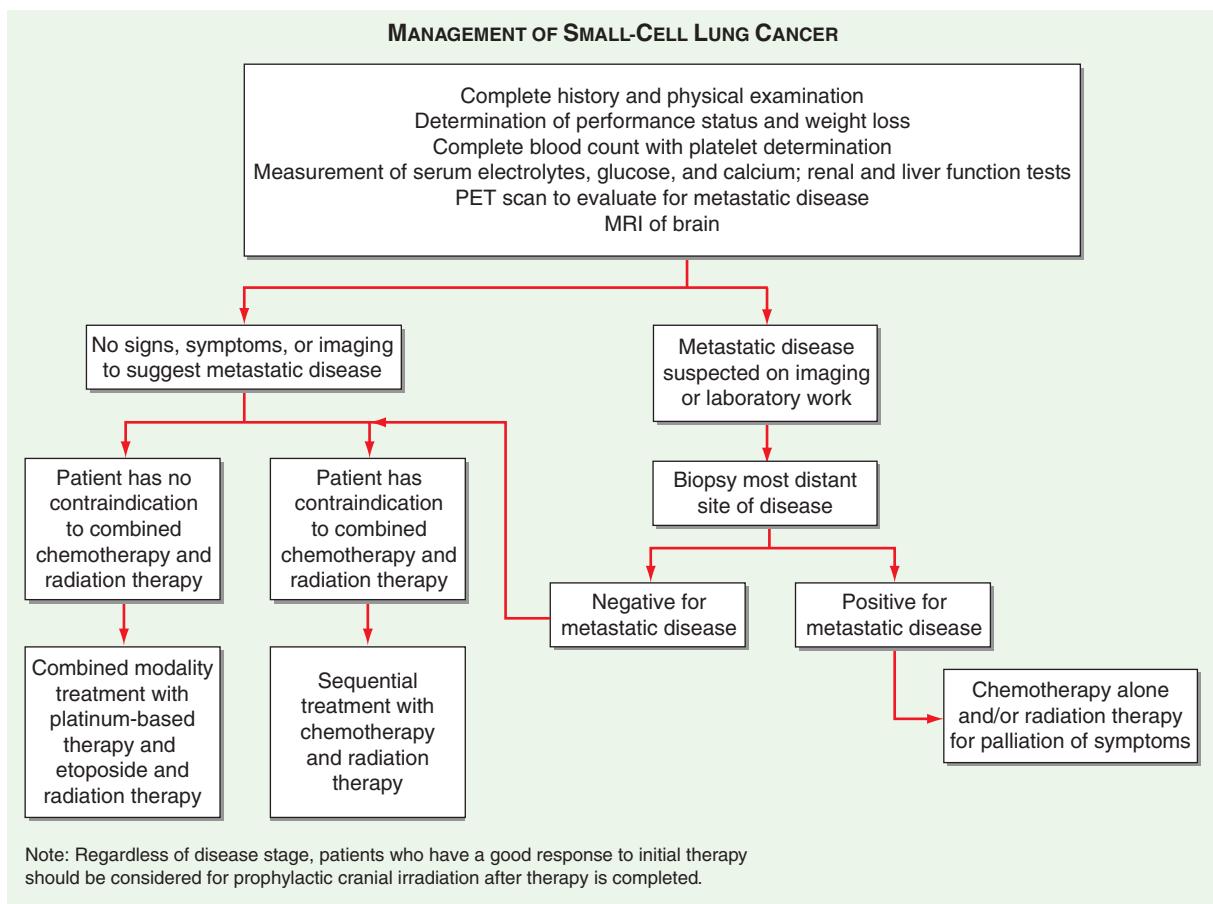


FIGURE 74-5 Algorithm for management of small-cell lung cancer. MRI, magnetic resonance imaging; PET, positron emission tomography.

with NSCLC. The various T (tumor size), N (regional node involvement), and M (presence or absence of distant metastasis) are combined to form different stage groups (**Tables 74-6 and 74-7**). The seventh edition of the TNM staging system went into effect in 2010 and developed using a much more robust database of more than 100,000 patients with lung cancer who were treated in multiple countries between 1990 and 2000. Data from 67,725 patients with NSCLC were then used to reevaluate the prognostic value of the TNM descriptors. In the current edition T1 tumors are divided into tumors ≤ 2 cm in size, as these patients were found to have a better prognosis compared to patients with tumors >2 cm but ≤ 3 cm. T2 tumors are divided into those that are >3 cm but ≤ 5 cm and those that are >5 cm but ≤ 7 cm. Tumors that are >7 cm are considered T3 tumors. T3 tumors also include tumors with invasion into local structures such as chest wall and diaphragm and additional nodules in the same lobe. T4 tumors include tumors of any size with invasion into mediastinum, heart, great vessels, trachea, or esophagus or multiple nodules in the ipsilateral lung. The eighth edition of the TNM has been proposed and differences are outlined in Tables 74-6 and 74-7. The major changes are in the T and M staging while no changes have been made to the current classification of lymph node involvement (N). Patients with metastasis may be classified as M1a (malignant pleural or pericardial effusion, pleural nodules, or nodules in the contralateral lung) or M1b (distant metastasis; e.g., bone, liver, adrenal, or brain metastasis), M1b single metastasis to a single organ or M1c multiple metastases to a single organ or metastases to multiple organs. Based on these data, approximately one-third of patients have localized disease that can be treated with curative attempt (surgery or radiotherapy), one-third have local or regional disease that may or may not be amenable to a curative attempt, and one-third have metastatic disease at the time of diagnosis.

■ STAGING SYSTEM FOR SMALL-CELL LUNG CANCER

In patients with SCLC, it is now recommended that both the Veterans Administration system and the American Joint Committee on Cancer/

International Union Against Cancer seventh edition system (TNM) be used to classify the tumor stage. The Veterans Administration system is a distinct two-stage system dividing patients into those with limited- or extensive-stage disease. Patients with limited-stage disease (LD) have cancer that is confined to the ipsilateral hemithorax and can be encompassed within a tolerable radiation port. Thus, contralateral supraclavicular nodes, recurrent laryngeal nerve involvement, and superior vena caval obstruction can all be part of LD. Patients with extensive-stage disease (ED) have overt metastatic disease by imaging or physical examination. Cardiac tamponade, malignant pleural effusion, and bilateral pulmonary parenchymal involvement generally qualify disease as ED, because the involved organs cannot be encompassed safely or effectively within a single radiation therapy port. Sixty to 70% of patients are diagnosed with ED at presentation. The TNM staging system is preferred in the rare SCLC patient presenting with what appears to be clinical stage I disease (see above).

■ PHYSIOLOGIC STAGING

Patients with lung cancer often have other comorbid conditions related to smoking including cardiovascular disease and COPD. To improve their preoperative condition, correctable problems (e.g., anemia, electrolyte and fluid disorders, infections, cardiac disease, and arrhythmias) should be addressed, appropriate chest physical therapy should be instituted, and patients should be encouraged to stop smoking. Patients with a forced expiratory volume in 1 s (FEV₁) of greater than 2 L or greater than 80% of predicted can tolerate a pneumonectomy, and those with an FEV₁ greater than 1.5 L have adequate reserve for a lobectomy. In patients with borderline lung function but a resectable tumor, cardiopulmonary exercise testing could be performed as part of the physiologic evaluation. This test allows an estimate of the maximal oxygen consumption (VO_{2max}). A VO_{2max} <15 mL/(kg·min) predicts for a higher risk of postoperative complications. Patients deemed unable to tolerate lobectomy or pneumonectomy from a pulmonary functional standpoint may be candidates for more limited resections, such as wedge

TABLE 74-6 Comparison of Seventh and Eighth Edition TNM Staging Systems for Non-Small-Cell Lung Cancer

	TNM STAGING SYSTEM FOR LUNG CANCER (7TH EDITION)	TNM STAGING SYSTEM FOR LUNG CANCER (8TH EDITION)
Primary Tumor (T)		
T1	Tumor ≤3 cm diameter, surrounded by lung or visceral pleura, without invasion more proximal than lobar bronchus	T1 tumor ≤3 cm in diameter surrounds by lung or visceral pleural without evidence of main bronchus
T1a	Tumor ≤2 cm in diameter	Tumor <1 cm
T1b	Tumor >2 cm but ≤3 cm in diameter	Tumor ≥1 cm but ≤2 cm
T1c	N/A	Tumor >2 cm but ≤3 cm
T2	Tumor >3 cm but ≤7 cm, or tumor with any of the following features: Involves main bronchus ≥2 cm distal to carina Invades visceral pleura Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung	T2 tumor >3 cm but ≤5 cm or tumor with any of the following features that does not involve the entire lung Involves main bronchus ≥2 cm distal to carina Invades visceral pleura Associate with atelectasis or obstructive pneumonitis that extends to the hilar region
T2a	Tumor >3 cm but ≤5 cm	Tumor >3 cm but ≤4 cm
T2b	Tumor >5 cm but ≤7 cm	Tumor >4 cm but ≤5 cm
T3	Tumor >7 cm or any of the following: Directly invades any of the following: chest wall, diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium, main bronchus <2 cm from carina (without involvement of carina) Atelectasis or obstructive pneumonitis of the entire lung Separate tumor nodules in the same lobe	>5 cm but ≤7 cm or any of the following: Directly invades any of the following chest wall, diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium, main bronchus <2 cm from carina (without involvement of carina) Atelectasis or obstructive pneumonitis of the entire lung
T4	Tumor of any size that invades the mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina, or with separate tumor nodules in a different ipsilateral lobe	7 cm or any of the following invades the mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina, or with separate tumor nodules in a different ipsilateral lobe
Regional Lymph Nodes (N)		
N0	No regional lymph node metastases	N0 No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension	N1 Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)	N2 Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)	N3 Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
Distant Metastasis (M)		
M0	No distant metastasis	M0 No distant metastasis
M1	Distant metastasis	Distant metastasis
M1a	Separate tumor nodule(s) in a contralateral lobe; tumor with pleural nodules or malignant pleural or pericardial effusion	M1 a separate nodule(s) in a contralateral tumor with pleural nodules or malignant pleural or pericardial effusion
M1b	Distant metastasis (in extrathoracic organs)	Single metastasis in a single organ
M1c		multiple metastases in a single organ or in several organs

Abbreviation: TNM, tumor-node-metastasis.

Source: Reproduced with permission from P Goldstraw et al: J Thorac Oncol 2:706, 2007.

TABLE 74-7 Comparison of Seventh and Eighth Edition TNM Staging Systems for Non-Small-Cell Lung Cancer

STAGE GROUPINGS SEVENTH EDITION				STAGE GROUPINGS EIGHTH EDITION			
Stage IA	T1a-T1b	N0	M0	Stage IA1	T1a	N0	M0
				Stage IA2	T1b	N0	M0
				Stage IA3	T1c	N0	M0
Stage IB	T2a	N0	M0	Stage IB	T2a	N0	M0
Stage IIA	T1a,T1b,T2a T2b	N1 N0	M0 M0	Stage IIA T2bNOM0	T2b	N0	M0
Stage IIB	T2b T3	N1 N0	M0 M0	Stage IIB	T1a-T2b T3	N1 N0	M0 M0
Stage IIIA	T1a,T1b,T2a,T2b T3 T4	N2 N1,N2 N0,N1	M0 M0 M0	Stage IIIA	T1-2b T3 T4	N2 N1 N0/N1	M0 M0 M0
Stage IIIB	T4 Any T	N2 N3	M0 M0	Stage IIIB	T1-2b T3/T4	N3 N0/N1	M0 M0
Stage IIIC	N/A				T3/T4	N3	M0
Stage IV	Any T	Any N	M1a or M1b	Stage IVA Stage IV B	Any T Any T	Any N Any N	M1a/M1b M1c

or anatomic segmental resection, although such procedures are associated with significantly higher rates of local recurrence and a trend toward decreased overall survival. All patients should be assessed for cardiovascular risk using American College of Cardiology and American Heart Association guidelines. A myocardial infarction within the past 3 months is a contraindication to thoracic surgery because 20% of patients will die of reinfarction. An infarction in the past 6 months is a relative contraindication. Other major contraindications include uncontrolled arrhythmias, an FEV₁ of less than 1 L, CO₂ retention (resting PCO₂ >45 mmHg), DLCO <40%, and severe pulmonary hypertension.

TREATMENT

Non-Small-Cell Lung Cancer

The overall treatment approach to patients with NSCLC is shown in Fig. 74-3.

OCCULT AND STAGE 0 CARCINOMAS

Patients with severe atypia on sputum cytology have an increased risk of developing lung cancer compared to those without atypia. In the uncommon circumstance where malignant cells are identified in a sputum or bronchial washing specimen but the chest imaging appears normal (TX tumor stage), the lesion must be localized. More than 90% of tumors can be localized by meticulous examination of the bronchial tree with a fiberoptic bronchoscope under general anesthesia and collection of a series of differential brushings and biopsies. Surgical resection following bronchoscopic localization has been shown to improve survival compared to no treatment. Close follow-up of these patients is indicated because of the high incidence of second primary lung cancers (5% per patient per year).

SOLITARY PULMONARY NODULE AND "GROUND-GLASS" OPACITIES

A solitary pulmonary nodule is defined as an x-ray density completely surrounded by normal aerated lung with circumscribed margins, of any shape, usually 1–6 cm in greatest diameter. The approach to a patient with a solitary pulmonary nodule is based on an estimate of the probability of cancer, determined according to the patient's smoking history, age, and characteristics on imaging (Table 74-8). Prior CXRs and CT scans should be obtained if available for comparison. A PET scan may be useful if the lesion is greater than 7–8 mm in diameter. If no diagnosis is apparent, Mayo investigators reported that clinical characteristics (age, cigarette smoking status, and prior cancer diagnosis) and three radiologic characteristics (nodule diameter, spiculation, and upper lobe location) were independent predictors of malignancy. At present, only two radiographic criteria are thought to predict the benign nature of a solitary pulmonary nodule: lack of growth over a period >2 years and certain characteristic patterns of calcification. Calcification alone, however, does not exclude malignancy; a dense central nidus, multiple punctate foci, and "bulls eye" (granuloma) and "popcorn ball" (hamartoma) calcifications are highly suggestive of a benign lesion. In contrast,

a relatively large lesion, lack of or asymmetric calcification, chest symptoms, associated atelectasis, pneumonitis, or growth of the lesion revealed by comparison with an old x-ray or CT scan or a positive PET scan may be suggestive of a malignant process and warrant further attempts to establish a histologic diagnosis. An algorithm for assessing these lesions is shown in Fig. 74-6.

Since the advent of screening CTs, small "ground-glass" opacities (GGOs) have often been observed, particularly as the increased sensitivity of CTs enables detection of smaller lesions. Many of these GGOs, when biopsied, are found to be atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), or minimally invasive adenocarcinoma (MIA). AAH is usually a nodule of <5 mm and is minimally hazy, also called nonsolid or ground glass (i.e., hazy slightly increased attenuation, no solid component, and preservation of bronchial and vascular margins). On thin-section CT, AIS is usually a nonsolid nodule and tends to be slightly more opaque than AAH. MIA is mainly solid, usually with a small (<5 mm) central solid component. However, overlap exists among the imaging features of the preinvasive and minimally invasive lesions in the lung adenocarcinoma spectrum. Lepidic adenocarcinomas are usually solid but may be nonsolid. Likewise, the small invasive adenocarcinomas also are usually solid but may exhibit a small nonsolid component.

MANAGEMENT OF STAGES I AND II NSCLC

Surgical Resection of Stage I and II NSCLC Surgical resection, ideally by an experienced thoracic surgeon, is the treatment of choice for patients with clinical stage I and II NSCLC who are able to tolerate the procedure. Operative mortality rates for patients resected by thoracic or cardiothoracic surgeons are lower compared to general surgeons. Moreover, survival rates are higher in patients who undergo resection in facilities with a high surgical volume compared to those performing fewer than 70 procedures per year, even though the higher-volume facilities often serve older and less socioeconomic advantaged populations. The improvement in survival is most evident in the immediate postoperative period. The extent of resection is a matter of surgical judgment based on findings at exploration. In patients with stage IA NSCLC, lobectomy is superior to wedge resection with respect to rates of local recurrence. There is also a trend toward improvement in overall survival. In patients with comorbidities, compromised pulmonary reserve, and small peripheral lesions, a limited resection, wedge resection, and segmentectomy (potentially by video-assisted thoracoscopic surgery) may be reasonable surgical option. Pneumonectomy is reserved for patients with central tumors and should be performed only in patients with excellent pulmonary reserve. The 5-year survival rates are 60–80% for patients with stage I NSCLC and 40–50% for patients with stage II NSCLC.

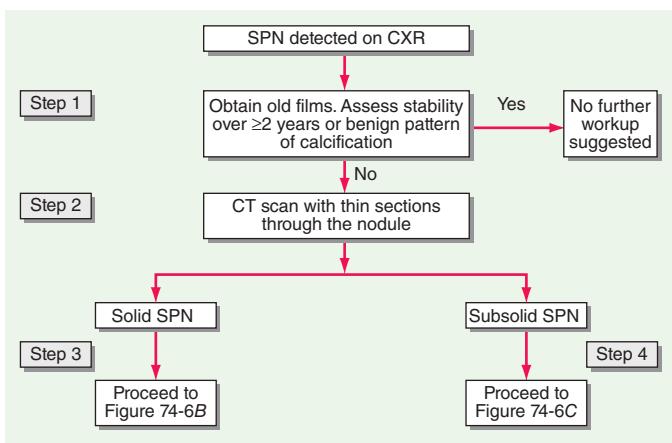
Accurate pathologic staging requires adequate segmental, hilar, and mediastinal lymph node sampling. Ideally this includes a mediastinal lymph node dissection. On the right side, mediastinal stations 2R, 4R, 7, 8R, and 9R should be dissected; on the left side, stations 5, 6, 7, 8L, and 9L should be dissected. Hilar lymph nodes are typically resected and sent for pathologic review, although it is helpful to specifically dissect and label level 10 lymph nodes when possible. On the left side, level 2 and sometimes level 4 lymph nodes are generally obscured by the aorta. Although the therapeutic benefit of nodal dissection versus nodal sampling is controversial, a pooled analysis of three trials involving patients with stages I to IIIA NSCLC demonstrated a superior 4-year survival in patients undergoing resection and a complete mediastinal lymph node dissection compared with lymph node sampling. Moreover, complete mediastinal lymphadenectomy added little morbidity to a pulmonary resection for lung cancer when carried out by an experienced thoracic surgeon.

Radiation Therapy in Stages I and II NSCLC There is currently no role for postoperative radiation therapy in patients following resection of stage I or II NSCLC with negative margins. However, patients with stage I and II disease who either refuse or are not

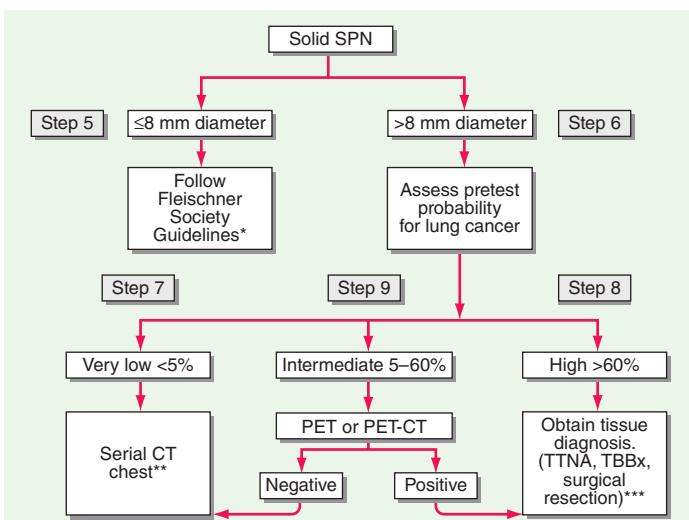
TABLE 74-8 Assessment of Risk of Cancer in Patients with Solitary Pulmonary Nodules

VARIABLE	RISK		
	LOW	INTERMEDIATE	HIGH
Diameter (cm)	<1.5	1.5–2.2	≥2.3
Age (years)	<45	45–60	>60
Smoking status	Never smoker	Current smoker (<20 cigarettes/d)	Current smoker (>20 cigarettes/d)
Smoking cessation status	Quit ≥7 years ago or quit	Quit <7 years ago	Never quit
Characteristics of nodule margins	Smooth	Scalloped	Corona radiata or spiculated

Source: Reproduced with permission from D Ost et al: N Engl J Med 348:2535, 2003.



A



*Fleischner society guidelines; modified from: H. MacMahon, et al: Radiology 2005; 237:395–400

Nodule size (a)	Low-risk patient (b):	High-risk patient (c):
≤4 mm	No follow-up needed (d)	Follow-up at 12 months; if unchanged, no further follow-up
>4–≤6 mm	Follow-up CT at 12 months; if unchanged, no further follow-up	Follow-up CT at 6–12 months; then 18–24 months if no change
>6–≤8 mm	Follow-up CT at 6–12 months; then 18–24 months if no change	Follow-up CT at 3–6 months; then 9–12 and 24 months if no change
>8 mm	Follow-up CT at 3, 9, and 24 months; dynamic contrast-enhanced CT, PET, and/or biopsy	Same as low-risk patient

(a) Average of largest and smallest axial diameters of the nodule

(b) No smoking history and absence of other risk factors

(c) Previous or current smoking history or other risk factors

(d) Risk of malignancy (<0.1%) is substantially lower than for an asymptomatic smoker

**ACCP guidelines (see MK Gould et al: Chest 132(suppl 3):108s, 2007).

***Consider patient preference, severity of medical comorbidities, center specific expertise prior to tissue diagnosis.

C

FIGURE 74-6 **A.** Algorithm for evaluation of solitary pulmonary nodule (SPN). **B.** Algorithm for evaluation of solid SPN. **C.** Algorithm for evaluation of semisolid SPN. CT, computed tomography; CXR, chest radiograph; GGN, ground-glass nodule; PET, positron emission tomography; TTNA, transthoracic needle biopsy; TBBx, transbronchial biopsy. (Adapted from VK Patel et al: Chest 143:840, 2013.)

suitable candidates for surgery should be considered for radiation therapy with *curative* intent. Stereotactic body radiation therapy (SBRT) is a technique used to treat patients with isolated pulmonary nodules (≤ 5 cm) who are not candidates for or refuse surgical resection. Treatment is typically administered in three to five fractions delivered over 1–2 weeks. In uncontrolled studies, disease control rates are $>90\%$, and 5-year survival rates of up to 60% have been reported with SBRT. By comparison, survival rates typically range from 13 to 39% in patients with stage I or II NSCLC treated with standard external-beam radiotherapy. Cryoablation is another technique occasionally used to treat small, isolated tumors (i.e., ≤ 3 cm). However, very little data exist on long-term outcomes with this technique.

Chemotherapy in Stages I and II NSCLC Although a landmark meta-analysis of cisplatin-based adjuvant chemotherapy trials in patients with resected stages I to IIIA NSCLC (the Lung Adjuvant Cisplatin Evaluation [LACE] Study) demonstrated a 5.4% improvement in 5-year survival for adjuvant chemotherapy compared to surgery alone, the survival benefit was seemingly confined to patients with stage II or III disease (Table 74-9). By contrast, survival was actually worsened in stage IA patients with the application of adjuvant therapy. In stage IB, there was a modest improvement in survival of questionable clinical significance.

TABLE 74-9 Adjuvant Chemotherapy Trials in Non-Small-Cell Lung Cancer

TRIAL	STAGE	TREATMENT	NO. OF PATIENTS	5-YEAR SURVIVAL (%)	P
IALT	I–III	Cisplatin-based	932	44.5	<.03
		Control	835	40.4	
BR10	IB–II	Cisplatin + vinorelbine	242	69	.03
		Control	240	54	
ANITA	IB–IIIA	Cisplatin + vinorelbine	407	60	.017
		Control	433	58	
ALPI	I–III	MVP	548	50	.49
		Control	540	45	
BLT	I–III	Cisplatin-based	192	60	.90
		Control	189	58	
CALGB	IB	Carboplatin + paclitaxel	173	59	.10
			171	57	

Abbreviations: ALPI, Adjuvant Lung Cancer Project Italy; ANITA, Adjuvant Navelbine International Trialist Association; BLT, Big Lung Trial; CALGB, Cancer and Lung Cancer Group B; IALT, International Adjuvant Lung Cancer Trial; MVP, mitomycin, vindesine, and cisplatin.

Adjuvant chemotherapy was also detrimental in patients with poor performance status (Eastern Cooperative Oncology Group [ECOG] performance status = 2). These data suggest that adjuvant chemotherapy is best applied in patients with resected stage II or III NSCLC. There is no apparent role for adjuvant chemotherapy in patients with resected stage IA or IB NSCLC. A possible exception to the prohibition of adjuvant therapy in this setting is the stage IB patient with a resected lesion ≥ 4 cm. At present, targeted therapies and immunotherapies are not used in the adjuvant setting, unless given as part of a clinical trial.

As with any treatment recommendation, the risks and benefits of adjuvant chemotherapy should be considered on an individual patient basis. If a decision is made to proceed with adjuvant chemotherapy, in general, treatment should be initiated 6–12 weeks after surgery, assuming the patient has fully recovered, and should be administered for no more than four cycles. Although a cisplatin-based chemotherapy is the preferred treatment regimen, carboplatin can be substituted for cisplatin in patients who are unlikely to tolerate cisplatin for reasons such as reduced renal function, presence of neuropathy, or hearing impairment. No specific chemotherapy regimen is considered optimal in this setting, although platinum plus vinorelbine is most commonly used.

Neoadjuvant chemotherapy, which is the application of chemotherapy administered *before* an attempted surgical resection, has been advocated by some experts on the assumption that such an approach will more effectively extinguish occult micrometastases compared to postoperative chemotherapy. In addition, it is thought that preoperative chemotherapy might render an inoperable lesion resectable. With the exception of superior sulcus tumors, however, the role of neoadjuvant chemotherapy in stage I to III disease is not well defined. However, a meta-analysis of 15 randomized controlled trials involving more than 2300 patients with stage I to III NSCLC suggested there may be a modest 5-year survival benefit (i.e., ~5%) that is virtually identical to the survival benefit achieved with postoperative chemotherapy. Accordingly, neoadjuvant therapy may prove useful in selected cases (see below). A decision to use neoadjuvant chemotherapy should always be made in consultation with an experienced surgeon.

It should be noted that all patients with resected NSCLC are at high risk of recurrence, most of which occurs within 18–24 months of surgery, or developing a second primary lung cancer. Thus, it is reasonable to follow these patients with periodic imaging studies. Given the results of the NLST, periodic CT scans appear to be the most appropriate screening modality. Based on the timing of most recurrences, some guidelines recommend a contrasted chest CT scan every 6 months for the first 3 years after surgery, followed by yearly CT scans of the chest without contrast thereafter.

MANAGEMENT OF STAGE III NSCLC

Management of patients with stage III NSCLC usually requires a combined-modality approach. Patients with stage IIIA disease commonly are stratified into those with “nonbulky” or “bulky” mediastinal lymph node (N2) disease. Although the definition of “bulky” N2 disease varies somewhat in the literature, the usual criteria include the size of a dominant lymph node (i.e., >2 –3 cm in short-axis diameter as measured by CT), groupings of multiple smaller lymph nodes, evidence of extracapsular nodal involvement, or involvement of more than two lymph node stations. The distinction between nonbulky and bulky stage IIIA disease is mainly used to select potential candidates for *upfront* surgical resection or for resection after neoadjuvant therapy. Many aspects of therapy of patients with stage III NSCLC remain controversial, and the optimal treatment strategy has not been clearly defined. Moreover, although there are many potential treatment options, none yields a very high probability of cure. Furthermore, because stage III disease is highly heterogeneous, no single treatment approach can be recommended for all patients. Key factors guiding treatment choices include the particular combination of tumor (T) and nodal (N) disease, the ability to achieve a complete surgical resection if indicated, and the patient’s

overall physical condition and preferences. For example, in carefully selected patients with limited stage IIIA disease where involved mediastinal lymph nodes can be completely resected, initial surgery followed by postoperative chemotherapy (with or without radiation therapy) may be indicated. By contrast, for patients with clinically evident bulky mediastinal lymph node involvement, the standard approach to treatment is concurrent chemoradiotherapy. Nevertheless, in some cases, the latter group of patients may be candidates for surgery following chemoradiotherapy.

Absent and Nonbulky Mediastinal (N2, N3) Lymph Node Disease For the subset of stage IIIA patients initially thought to have clinical stage I or II disease (i.e., pathologic involvement of mediastinal [N2] lymph nodes is *not* detected preoperatively), surgical resection is often the treatment of choice. This is followed by adjuvant chemotherapy in patients with microscopic lymph node involvement in a resection specimen. Postoperative radiation therapy (PORT) may also have a role for those with close or positive surgical margins. Patients with tumors involving the chest wall or proximal airways within 2 cm of the carina with hilar lymph node involvement (but not N2 disease) are classified as having T3N1 stage IIIA disease. They too are best managed with surgical resection, if technically feasible, followed by adjuvant chemotherapy if completely resected. Patients with tumors exceeding 7 cm in size also are now classified as T3 and are considered stage IIIA if tumor has spread to N1 nodes. The appropriate initial management of these patients involves surgical resection when feasible, provided the mediastinal staging is negative, followed by adjuvant chemotherapy for those who achieve complete tumor resection. Patients with T3N0 or T3N1 disease due to the presence of satellite nodules within the same lobe as the primary tumor also are candidates for surgery, as are patients with ipsilateral nodules in another lobe and negative mediastinal nodes (IIIA, T4N0 or T4N1). Although data regarding adjuvant chemotherapy in the latter subsets of patients are limited, it is often recommended.

Patients with T4N0-1 were reclassified as having stage IIIA tumors in the seventh edition of the TNM system. These patients may have involvement of the carina, superior vena cava, or a vertebral body and yet still be candidates for surgical resection in selected circumstances. The decision to proceed with an attempted resection must be made in consultation with an experienced thoracic surgeon often in association with a vascular or cardiac surgeon and an orthopedic surgeon depending on tumor location. However, if an incomplete resection is inevitable or if there is evidence of N2 involvement (stage IIIB), surgery for T4 disease is contraindicated. Most T4 lesions are best treated with chemoradiotherapy.

The role of PORT in patients with completely resected stage III NSCLC is controversial. To a large extent, the use of PORT is dictated by the presence or absence of N2 involvement and, to a lesser degree, by the biases of the treating physician. Using the Surveillance, Epidemiology, and End Results (SEER) database, a recent meta-analysis of PORT identified a significant increase in survival in patients with N2 disease but not in patients with N0 or N1 disease. An earlier analysis by the PORT Meta-analysis Trialist Group using an older database produced similar results.

Known Mediastinal (N2, N3) Lymph Node Disease When pathologic involvement of mediastinal lymph nodes is documented preoperatively, a combined-modality approach is recommended assuming the patient is a candidate for treatment with curative intent. These patients are at high risk for both local and distant recurrence if managed with resection alone. For patients with stage III disease who are not candidates for initial surgical resection, *concurrent* chemoradiotherapy is most commonly used as the initial treatment. Concurrent chemoradiotherapy has been shown to produce superior survival compared to *sequential* chemoradiotherapy; however, it also is associated with greater host toxicities (including fatigue, esophagitis, and neutropenia). Therefore, for patients with a good performance status, concurrent chemoradiotherapy is the preferred treatment approach, whereas sequential chemoradiotherapy

may be more appropriate for patients with a performance status that is not as good. For patients who are *not* candidates for a combined-modality treatment approach, typically due to a poor performance status or a comorbidity that makes chemotherapy untenable, radiotherapy alone may provide a modest survival benefit in addition to symptom palliation.

For patients with potentially resectable N2 disease, it remains uncertain whether surgery after neoadjuvant chemoradiotherapy improves survival. In an NCI-sponsored Intergroup randomized trial comparing concurrent chemoradiotherapy alone to concurrent chemoradiotherapy followed by attempted surgical resection, no survival benefit was observed in the trimodality arm compared to the bimodality therapy. In fact, patients subjected to a pneumonectomy had a worse survival outcome. By contrast, those treated with a lobectomy appeared to have a survival advantage based on a retrospective subset analysis. Thus, in carefully selected, otherwise healthy patients with nonbulky mediastinal lymph node involvement, surgery may be a reasonable option if the primary tumor can be fully resected with a lobectomy. This is not the case if a pneumonectomy is required to achieve complete resection.

Superior Sulcus Tumors (Pancoast Tumors) Superior sulcus tumors represent a distinctive subset of stage III disease. These tumors arise in the apex of the lung and may invade the second and third ribs, the brachial plexus, the subclavian vessels, the stellate ganglion, and adjacent vertebral bodies. They also may be associated with Pancoast syndrome, characterized by pain that may arise in the shoulder or chest wall or radiate to the neck. Pain characteristically radiates to the ulnar surface of the hand. Horner's syndrome (enophthalmos, ptosis, miosis, and anhydrosis) due to invasion of the paravertebral sympathetic chain may be present as well. Patients with these tumors should undergo the same staging procedures as all patients with stage II and III NSCLC. Neoadjuvant chemotherapy or combined chemoradiotherapy followed by surgery is reserved for those without N2 involvement. This approach yields excellent survival outcomes (>50% 5-year survival in patients with an R0 resection). Patients with N2 disease are less likely to benefit from surgery and can be managed with chemoradiotherapy alone. Patients presenting with metastatic disease can be treated with radiation therapy (with or without chemotherapy) for symptom palliation.

MANAGEMENT OF METASTATIC NSCLC

Approximately 40% of NSCLC patients present with advanced, stage IV disease at the time of diagnosis. In addition, a significant number of patients who first presented with early-stage NSCLC will eventually relapse with distant disease. Patients who have recurrent disease have a better prognosis than those presenting with metastatic disease at the time of diagnosis. Standard medical management, the judicious use of pain medications, and the appropriate use of radiotherapy and systemic therapy—which may compromise of traditional cytotoxic chemotherapy, targeted therapy, and immunotherapy depending on the specific diagnosis and molecular subtype—form the cornerstone of management. Systemic therapy palliates symptoms, improves the quality of life, and improves survival in patients with stage advanced NSCLC, particularly in patients with good performance status. Of note, the early application of palliative care in conjunction with chemotherapy is associated with improved survival and a better quality of life.

Cytotoxic Chemotherapy for Metastatic or Recurrent NSCLC A landmark meta-analysis published in 1995 provided the earliest meaningful indication that chemotherapy could provide a survival benefit in metastatic NSCLC as opposed to supportive care alone. However, the survival benefit was seemingly confined to cisplatin-based chemotherapy regimens (hazard ratio 0.73; 27% reduction in the risk of death; 10% improvement in survival at 1 year). To date, platinum-based regimens remain the cornerstone of the cytotoxic chemotherapy regimens used for patients with metastatic NSCLC (Table 74-10). Several different platinum “doublet” regimens have

TABLE 74-10 First-Line Chemotherapy Trials for Metastatic Non-Small-Cell Lung Cancer

TRIAL	REGIMEN	NO. OF PATIENTS	RR (%)	MEDIAN SURVIVAL (MONTHS)
ECOG1594	Cisplatin + paclitaxel	288	21	7.8
	Cisplatin + gemcitabine	288	22	8.1
	Cisplatin + docetaxel	289	17	7.4
	Carboplatin + paclitaxel	290	17	8.1
TAX-326	Cisplatin + docetaxel	406	32	11.3
	Cisplatin + vinorelbine	394	25	10.1
	Carboplatin + docetaxel	404	24	9.4
EORTC	Cisplatin + paclitaxel	159	32	8.1
	Cisplatin + gemcitabine	160	37	8.9
	Paclitaxel + gemcitabine	161	28	6.7
ILCP	Cisplatin + gemcitabine	205	30	9.8
	Carboplatin + paclitaxel	204	32	9.9
	Cisplatin + vinorelbine	203	30	9.5
SWOG	Cisplatin + vinorelbine	202	28	8.0
	Carboplatin + paclitaxel	206	25	8.0
FACS	Cisplatin + irinotecan	145	31	13.9
	Carboplatin + paclitaxel	145	32	12.3
	Cisplatin + gemcitabine	146	30	14.0
	Cisplatin + vinorelbine	145	33	11.4
Scagliotti	Cisplatin + gemcitabine	863	28	10.3
	Cisplatin + pemetrexed	862	31	10.3

Abbreviations: ECOG, Eastern Cooperative Oncology Group; EORTC, European Organization for Research and Treatment of Cancer; ILCP, Italian Lung Cancer Project; SWOG, Southwest Oncology Group; FACS, Follow-up After Colorectal Surgery.

been used—combining platinum (cisplatin or carboplatin) with another type of chemotherapy (for example, paclitaxel, docetaxel, pemetrexed, gemcitabine, or vinorelbine). Although specific tumor histology was once considered irrelevant to treatment choice in NSCLC, with the recent recognition that selected chemotherapy agents perform quite differently in squamous versus adenocarcinomas, accurate determination of histology has become essential. Specifically, in a landmark randomized phase III trial, patients with nonsquamous NSCLC were found to have an improved survival when treated with cisplatin and pemetrexed compared to cisplatin and gemcitabine. By contrast, patients with squamous carcinoma had an improved survival when treated with cisplatin and gemcitabine. This survival difference is thought to be related to the differential expression of thymidylate synthase (TS), one of the targets of pemetrexed, between tumor types. Squamous cancers have a much higher expression of TS compared to adenocarcinomas, accounting for their lower responsiveness to pemetrexed. By contrast, the activity of gemcitabine is not impacted by the levels of TS.

Maintenance Therapy for Metastatic NSCLC Several large phase III randomized trials have failed to show a meaningful benefit for increasing the duration of platinum-doublet chemotherapy beyond four to six cycles. In fact, longer duration of platinum-doublet chemotherapy has been associated with increased toxicities and impaired quality of life. Maintenance chemotherapy in nonprogressing patients (patients with a complete response, partial response, or stable disease) is divided into two types of maintenance strategies: (1) switch maintenance therapy, where patients receive four to six cycles of platinum-based chemotherapy and are switched to an entirely different regimen; and (2) continuation maintenance therapy, where patients receive four to six cycles of platinum-based chemotherapy and then the platinum agent is discontinued but the agent it is paired with is continued (Table 74-11). Two studies investigated switch maintenance single-agent chemotherapy with docetaxel or pemetrexed in nonprogressing patients following treatment with first-line platinum-based chemotherapy. Both trials randomized patients to immediate single-agent therapy

TABLE 74-11 Maintenance Therapy Trials

GROUP	CT	NO. OF PATIENTS	SURVIVAL	
			OS (MONTHS)	PFS (MONTHS)
Switch Maintenance				
Fidias	Immediate docetaxel	153	12.3	5.7
	Delayed docetaxel	156	9.7	2.7
Ciuleanu	Pemetrexed	444	13.4	4.3
	BSC	222	10.6	2.6
Paramount	Pemetrexed	472	13.9	4.1
	BSC	297	11.0	2.8
ATLAS	Bev + erlotinib	384	15.9	4.8
	Bev + placebo	384	13.9	3.8
SATURN	Erlotinib	437	12.3	2.9
	Placebo	447	11.1	2.6
Continuation Maintenance				
ECOG4599	Bev 15 mg/kg	444	12.3	6.2
	BSC	434	10.3	4.5
AVAIL	Bev 15 mg/kg	351	13.4	6.5
	Bev 7.5 mg/kg	345	13.6	6.7
	Placebo	347	13.1	6.1
POINTBREAK	Pemetrexed + Bev 15 mg/kg			8.6
	Bev 15 mg/kg			6.9

Abbreviations: Bev, bevacizumab; BSC, best supportive care; CT, chemotherapy; OS, overall survival; PFS, progression-free survival.

versus observation and reported improvements in progression-free and overall survival. In both trials, a significant portion of patients in the observation arm did not receive therapy with the agent under investigation upon disease progression; 37% of study patients never received docetaxel in the docetaxel study and 81% of patients never received pemetrexed in the pemetrexed study. In the trial of maintenance docetaxel versus observation, survival was identical to the treatment group in the subset of patients who received docetaxel on progression, indicating this is an active agent in NSCLC. These data are not available for the pemetrexed study. Two additional trials evaluated switch maintenance therapy with erlotinib after platinum-based chemotherapy in patients with advanced NSCLC and reported an improvement in progression-free survival and overall survival in the erlotinib treatment group; however, erlotinib is not recommended in patients with EGFR wild type tumors. Bevacizumab, a monoclonal antibody against VEGF, has been shown to improve response rate, progression-free survival, and overall survival in patients with advanced disease when combined with chemotherapy. However, bevacizumab cannot be given to patients with squamous cell histology NSCLC because of their tendency to experience serious hemorrhagic effects. Currently, carboplatin/paclitaxel and bevacizumab or carboplatin/pemetrexed and bevacizumab are appropriate regimens for first-line treatment for stage IV nonsquamous NSCLC patients followed by maintenance bevacizumab or maintenance pemetrexed/bevacizumab respectively. Currently, maintenance pemetrexed following platinum-based chemotherapy in patients with advanced NSCLC is also approved by the U.S. FDA. Maintenance erlotinib is only approved in patients with EGFR mutations (see below). It should be noted that there are no approved maintenance regimens for patients with squamous cell histology. Moreover, maintenance therapy is not without toxicity and, at this time, should be considered on an individual patient basis.

Targeted Therapies for Select Molecular Cohorts of NSCLC As the efficacy of traditional cytotoxic chemotherapeutic agents plateaued in NSCLC, there was a critical need to define novel therapeutic treatment strategies. For a cohort of NSCLC patients, the presence of an oncogenic driver allows the use of oral therapies with significant tumor regression. These driver mutations occur in genes

encoding signaling proteins that, when aberrant, drive initiation and maintenance of tumor cells. Importantly, driver mutations can serve as Achilles' heels for tumors, if their gene products can be targeted therapeutically with small-molecule inhibitors. For example, EGFR mutations have been detected in 10–15% of North American patients diagnosed with NSCLC. EGFR mutations are associated with younger age, light (<10 pack-year) and nonsmokers, and adenocarcinoma histology. Approximately 90% of these mutations are exon 19 deletions or exon 21 L858R point mutations within the EGFR TK domain, resulting in hyperactivation of both EGFR kinase activity and downstream signaling. Lung tumors that harbor activating mutations within the EGFR kinase domain display high sensitivity to small-molecule EGFR TKIs. Erlotinib, gefitinib, and afatinib are FDA-approved oral small-molecule TKIs that inhibit EGFR. Several large, international, phase III studies have demonstrated improved response rates, progression-free survival, and overall survival in patients with EGFR mutation-positive NSCLC patients treated with an EGFR TKI as compared with standard first-line chemotherapy regimens (Table 74-12). A phase III trial also compared gefitinib to afatinib as first-line therapy in patients with EGFR mutation-positive NSCLC and demonstrated superior efficacy, with increasing toxicity for patients treated with afatinib. Unfortunately, all patients with EGFR mutation-positive NSCLC treated with EGFR TKIs eventually developed acquired resistance. Disease progression occurs usually around 12 months. Approximately 50% of patients have tumors that harbor a second site mutation, most commonly the T790M mutation occurring within exon 20. Osimertinib, a third generation mutant-selective EGFR TKI received approval in 2015 for patients who progress on erlotinib, gefitinib, or afatinib and whose tumors harbor the T790M mutation.

With the rapid pace of scientific discovery, additional driver mutations in lung cancer have been identified and targeted therapeutically with impressive clinical results. For example, chromosomal rearrangements involving the anaplastic lymphoma kinase (ALK) gene on chromosome 2 have been found in ~3–7% of NSCLC. The result of these ALK rearrangements is hyperactivation of the ALK TK domain. Similar to EGFR, ALK rearrangements are typically (but not exclusively) associated with younger age, light (<10 pack-year) and nonsmokers, and adenocarcinoma histology. Remarkably, ALK rearrangements were initially described in lung cancer in 2007, and by 2011, the first ALK inhibitor, crizotinib, received FDA approval for patients with lung tumors harboring ALK rearrangements. Two additional ALK inhibitors, ceritinib and alectinib, are currently approved in patients who progress on crizotinib. ALK testing may be performed via fluorescence in situ hybridization (FISH),

TABLE 74-12 Results of Phase III Trials Comparing Chemotherapy and First-Line EGFR TKI in EGFR Mutation-Positive Patients

STUDY	THERAPY	NO. OF PATIENTS	ORR (%)	PFS (MONTHS)
IPASS	CbP	129	47	6.3
	Gefitinib	132	71	9.3
EURTAC	CG	87	15	5.2
	Erlotinib	86	58	9.7
OPTIMAL	CG	72	36	4.6
	Erlotinib	82	83	13.1
NEJ002	CG	114	31	5.4
	Gefitinib	114	74	10.8
WJTOG3405	CD	89	31	6.3
	Gefitinib	88	62	9.2
LUX LUNG 3	CP	115	23	6.9
	Afatinib	230	56	11.1
LUX LUNG 6	CG	122	23	5.6
	Afatinib	242	67	11.0

Abbreviations: CbP carboplatin and paclitaxel; CD, cisplatin and docetaxel; CG, cisplatin and gemcitabine; CP cisplatin and paclitaxel; CG, cisplatin and gemcitabine; ORR, overall response rate; PFS, progression-free survival.

immunohistochemistry (IHC), or next generation sequencing. *ROS1* fusions have been identified in ~1% of patients with NSCLC and similar to EGFR and ALK, *ROS* rearrangements are typically associated with younger age and light or never smoking status. Crizotinib, which inhibits both ALK and *ROS1* kinases, was recently FDA approved for patients whose tumors harbor a *ROS1* fusion.

In addition to *EGFR*, *ALK*, and *ROS1* other driver mutations have been discovered with varying frequencies in NSCLC, including *KRAS*, *BRAF*, *PIK3CA*, *NRAS*, *AKT1*, *MET*, *MEK1* (*MAP2K1*), *NTRK*, and *RET*. Mutations within the *KRAS* GTPase are found in ~20% of lung adenocarcinomas. To date, however, no small-molecule inhibitors are available to specifically target mutant *KRAS*. Each of the other driver mutations occurs in less than 1–3% of lung adenocarcinomas. The great majority of the driver mutations are mutually exclusive, and there are ongoing clinical studies for their specific inhibitors. For example, the *BRAF* inhibitors dabrafenib and vemurafenib and the *RET* inhibitors cabozantinib and vandetanib have already demonstrated efficacy in patients with lung cancer harboring *BRAF* mutations or *RET* gene fusions, respectively. Most of these mutations are present in adenocarcinoma; however, mutations that may be linked to future targeted therapies in squamous cell carcinomas are emerging. In addition, there are active research efforts aimed at defining novel targetable mutations in lung cancer as well as defining mechanisms of acquired resistance to small-molecule inhibitors used in the treatment of patients with NSCLC.

Second-Line Therapy and Beyond Second-line therapy for advanced NSCLC was almost never recommended until a seminal study in 2000 showed that docetaxel improved survival compared to supportive care alone. Little progress had been made in the second-line setting for NSCLC patients until the introduction of immunotherapy agents (see below) with only pemetrexed and docetaxel available as FDA-approved agents, and erlotinib recommended in patients with EGFR mutation-positive NSCLC who did not receive a first line EGFR TKI. Ramucirumab is a recombinant human IgG1 monoclonal antibody that targets VEGFR-2 and blocks the interaction of VEGF ligands and VEGFR-2. A phase III trial demonstrated a significant improvement in progression-free survival and overall survival when ramucirumab was combined with docetaxel as second-line therapy in patients who had progressed on platinum-based chemotherapy. Contrary to bevacizumab, ramucirumab was safe in patients with both squamous and nonsquamous NSCLC and is approved regardless of histology.

Immunotherapy Immune checkpoint inhibitors are a novel class of agents that have significantly improved the quality of life and survival for a group of patients with advanced NSCLC. Immune checkpoint inhibitors work by blocking interactions between T cells and antigen presenting cells (APCs) or tumor cells that lead to T-cell inactivation. By inhibiting this interaction, the immune system is effectively upregulated and T cells become activated against tumor cells. Several large randomized phase III trials demonstrated superior overall survival for both the anti-PD1 antibodies, nivolumab and pembrolizumab and the anti-PD-L1 antibody atezolizumab compared to second-line docetaxel in patients with NSCLC who have progressed on platinum-based chemotherapy (Table 74-13). Nivolumab and atezolizumab are approved as second-line therapy in patients who have progressed following platinum-based chemotherapy regardless of the presence of PD-L1 while pembrolizumab is approved in patients with tumors positive for PD-L1 expression in ≥1% of tumor cells. Pembrolizumab demonstrated superior efficacy to first-line platinum-based chemotherapy in patients with tumors expressing PD-L1 in greater than 50% of tumor cells, as assessed with immunohistochemistry. A similarly designed study did not show efficacy when nivolumab was compared to chemotherapy; however, in this study patients with tumors expressing PD-L1 in greater than 1% of tumor cells were enrolled. Pembrolizumab is approved as first-line therapy in patients with tumors that are positive for PD-L1 expression in ≥50% of tumor cells. While PD-L1 has been identified as a biomarker that can predict

TABLE 74-13 Results of Phase III Trials Comparing Chemotherapy and Immunotherapy in Patients with NSCLC

STUDY	THERAPY	NO. OF PATIENTS	OS (MONTHS)	PFS (MONTHS)
Checkmate 017	Docetaxel	137	6.0	2.8
Squamous	Nivolumab	135	9.2	3.5
Checkmate 057	Docetaxel	290	9.4	4.2
Nonsquamous	Nivolumab	292	12.2	2.3
Keynote 10	Docetaxel	212	8.5	4.0
PD-L1 ≥1%	Pembrolizumab 2 mg/kg	259	10.4	2.9
	Pembrolizumab 10 mg/kg	255	12.7	2.9
OAK	Docetaxel	425	10.3	2.8
	Atezolizumab	425	12.6	4.0
Keynote 24	Platinum-chemotherapy	116	NR	6.0
PD-L1 ≥50%	Pembrolizumab	73	NR	10.3
Checkmate 26	Platinum-chemotherapy	212	13.2	5.9
PD-L1 ≥1%	Nivolumab	211	14.4	4.2

Abbreviations: OS, overall survival; PFS, progression-free survival; Platinum-chemotherapy refers to first-line platinum-doublet chemotherapy.

response to immune checkpoint inhibitors, responses are observed in patients who do not appear to express the biomarker and not all PD-L1 positive patients respond to checkpoint inhibition. Complicating matter is that each checkpoint inhibitor is being developed in conjunction with its own antibody to assess PD-L1 expression and a large effort is underway to compare these tests. Further evaluation of these agents in both NSCLC and SCLC is ongoing in combination with already approved chemotherapy and targeted agents as well as other checkpoint inhibitors.

Supportive Care No discussion of the treatment strategies for patients with advanced lung cancer would be complete without a mention of supportive care. Coincident with advances in chemotherapy and targeted therapy was a pivotal study that demonstrated that the early integration of palliative care with standard treatment strategies improved both quality of life and mood for patients with advanced lung cancer (Chaps. 9 and 65). Aggressive pain and symptom control is an important component for optimal treatment of these patients.

TREATMENT

Small-Cell Lung Cancer

The overall treatment approach to patients with SCLC is shown in Fig. 74-5.

SURGERY FOR LIMITED-DISEASE SMALL-CELL LUNG CANCER

SCLC is a highly aggressive disease characterized by its rapid doubling time, high growth fraction, early development of disseminated disease, and dramatic response to first-line chemotherapy and radiation. In general, surgical resection is *not* routinely recommended for patients because even patients with LD-SCLC still have occult micrometastases. However, the most recent American College of Chest Physicians Evidence-Based Clinical Practice Guidelines recommend surgical resection over nonsurgical treatment in SCLC patients with clinical stage I disease after a thorough evaluation for distant metastases and invasive mediastinal stage evaluation (grade 2C). After resection, these patients should receive platinum-based adjuvant chemotherapy (grade 1C). If the histologic diagnosis of SCLC is made in patients on review of a resected surgical specimen, such patients should receive standard SCLC chemotherapy as well.

CHEMOTHERAPY

Chemotherapy significantly prolongs survival in patients with SCLC. Four to six cycles of platinum-based chemotherapy with either cisplatin or carboplatin plus either etoposide or irinotecan has been the mainstay of treatment for nearly three decades and is recommended over other chemotherapy regimens irrespective of initial stage. Cyclophosphamide, doxorubicin (Adriamycin), and vincristine (CAV) may be an alternative for patients who are unable to tolerate a platinum-based regimen. Despite response rates to first-line therapy as high as 80%, the median survival ranges from 12 to 20 months for patients with LD and from 7 to 11 months for patients with ED. Regardless of disease extent, the majority of patients relapse and develop chemotherapy-resistant disease. Only 6–12% of patients with LD-SCLC and 2% of patients with ED-SCLC live beyond 5 years. The prognosis is especially poor for patients who relapse within the first 3 months of therapy; these patients are said to have *chemotherapy-resistant disease*. Patients are said to have *sensitive disease* if they relapse more than 3 months after their initial therapy and are thought to have a somewhat better overall survival. These patients also are thought to have the greatest potential benefit from second-line chemotherapy (Fig. 74-7). Topotecan is the only FDA-approved agent for second-line therapy in patients with SCLC. Topotecan has only modest activity and can be given either intravenously or orally. In one randomized trial, 141 patients who were not considered candidates for further IV chemotherapy were randomized to receive either oral topotecan or best supportive care. Although the response rate to oral topotecan was only 7%, overall survival was significantly better in patients receiving chemotherapy (median survival time, 26 weeks vs 14 weeks; $p = 0.01$). Moreover, patients given topotecan had a slower decline in quality of life than did those not receiving chemotherapy. Other agents with similar low levels of activity in the second-line setting include irinotecan, paclitaxel, docetaxel, vinorelbine, oral etoposide, and gemcitabine. Clearly novel treatments for this all too common disease are desperately needed.

THORACIC RADIATION THERAPY

Thoracic radiation therapy (TRT) is a standard component of induction therapy for good performance status and limited-stage SCLC patients. Meta-analyses indicate that chemotherapy combined with

chest irradiation improves 3-year survival by ~5% as compared with chemotherapy alone. The 5-year survival rate, however, remains disappointingly low at ~10–15%. Most commonly, TRT is combined with cisplatin and etoposide chemotherapy due to a superior toxicity profile as compared to anthracycline-containing chemotherapy regimens. As observed in locally advanced NSCLC, *concurrent* chemoradiotherapy is more effective than *sequential* chemoradiation but is associated with significantly more esophagitis and hematologic toxicity. Ideally TRT should be administered with the first two cycles of chemotherapy because later application appears slightly less effective. If for reasons of fitness or availability, this regimen cannot be offered, TRT should follow induction chemotherapy. With respect to fractionation of TRT, twice-daily 1.5-Gy fractionated radiation therapy has been shown to improve survival in LD-SCLC patients but is associated with higher rates of grade 3 esophagitis and pulmonary toxicity. Although it is feasible to deliver once-daily radiation therapy doses up to 70 Gy concurrently with cisplatin-based chemotherapy, there are no data to support equivalency of this approach compared with the 45-Gy twice-daily radiotherapy dose. Therefore, the current standard regimen of a 45-Gy dose administered in 1.5-Gy fractions twice daily for 30 days is being compared with higher-dose regimens in two phase III trials, one in the United States and one in Europe. Patients should be carefully selected for concurrent chemoradiation therapy based on good performance status and adequate pulmonary reserve. The role of radiotherapy in ED-SCLC is largely restricted to palliation of tumor-related symptoms such as bone pain and bronchial obstruction.

PROPHYLACTIC CRANIAL IRRADIATION

Prophylactic cranial irradiation (PCI) should be considered in all patients with either LD-SCLC or ED-SCLC who have responded well to initial therapy. A meta-analysis including seven trials and 987 patients with LD-SCLC who had achieved a complete remission after upfront chemotherapy yielded a 5.4% improvement in overall survival for patients treated with PCI. In patients with ED-SCLC who have responded to first-line chemotherapy, a prospective randomized phase III trial showed that PCI reduced the occurrence of symptomatic brain metastases and prolonged disease-free and overall survival compared to no radiation therapy. Long-term toxicities, including deficits in cognition, have been reported after PCI but are difficult to sort out from the effects of chemotherapy or normal aging.

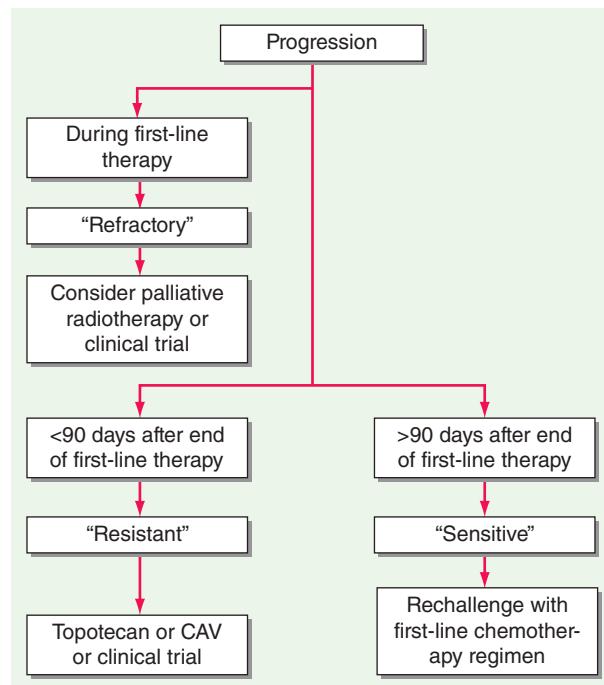


FIGURE 74-7 Management of recurrent small-cell lung cancer (SCLC). CAV, cyclophosphamide, doxorubicin, and vincristine. (Adapted with permission from JP van Meerbeek et al: Lancet 378:1741, 2011.)

THYMIC TUMORS

Thymic tumors are rare malignancies accounting for 0.5–1.5% of all malignancies in the United States with a higher incidence among Asian populations. They are particularly rare among children and young adults with incidence peaking in the fifth decade of life. There is no difference between sexes and no clear risk factors have been identified.

CLINICAL MANIFESTATIONS

The majority of thymic tumors occur in the anterior mediastinum. Approximately 40% of patients with mediastinal masses will be asymptomatic with an incidental finding on chest imaging. In patients presenting with an anterior mediastinal mass, if appropriate, serum beta-HCG (human chorionic gonadotropin) and α fetoprotein (AFP) should be sent to rule out a germ cell tumor. A patient with a sign or symptom of thymoma or thymic carcinoma may present with chest pain, dyspnea, cough or superior vena cava syndrome secondary to effects on adjacent organs or a paraneoplastic syndrome, most commonly myasthenia gravis, pure red cell aplasia or hypogammaglobulinemia. More rare paraneoplastic syndromes include limbic encephalitis, aplastic anemia, hemolytic anemia, and autoimmune disease such as Sjogren's syndrome, polymyositis, rheumatoid arthritis, ulcerative colitis among others.

STAGING

Given the rarity of the tumor, patients with suspected thymoma should be evaluated by a multidisciplinary team including a surgeon, medical

TABLE 74-14 Staging Thymic Tumors

MASAOKA STAGE	DEFINITION
I	Grossly and microscopically encapsulated
IIA	Microscopic transcapsular invasion
IIB	Macroscopic invasion into surrounding tissue excluding pericardium, lung, and great vessels
III	Macroscopic invasion into neighboring organs of the lower neck or upper chest
IVA	Pleural or pericardial dissemination
IVB	Hematogenous or lymphatic dissemination to distal organs
WHO	
A	Tumor with few lymphocytes
AB	Tumor with features of type A and foci rich in lymphocytes
B1	Tumor with features of normal epithelial cells with vesicular nuclei and distinct nucleoli and an abundant population of lymphocytes. Also known as cortical thymoma, lymphocyte-rich thymoma
B2	Thymoma with no or mild atypia with round or polygonal shaped cells with small component of lymphocytes
B3	Well differentiated thymic carcinoma with mild atypia
C	Thymic carcinoma with high atypia

and radiation oncologist as well as pathologist with experience in treating the disease. A CT scan of the chest with contrast is recommended to determine if the mass is resectable based on relationship to surrounding structures. An MRI with contrast may be performed if clinically indicated. A PET scan may be useful in the evaluation of a patient with thymic tumors although may be less useful in the staging of thymoma compared to thymic carcinoma. A core needle biopsy is considered standard of care for obtaining a histological diagnosis of an anterior mediastinal tumor. This may be obtained via CT or ultrasound imaging. However, in some circumstances a mediastinoscopy or open biopsy may be required.

Thymomas are commonly staged using the Masaoka system or the World Health Organization (WHO) staging system as described in Table 74-14. WHO type A, AB, and B1 tend to be more well-differentiated,

type B2 and B3 are moderately differentiated, and C are poorly differentiated.

TREATMENT

Surgical resection is the mainstay of treatment for patients with Masaoka type I and II thymic tumors. In patients with type III and IV who are potentially resectable thymic tumors, neoadjuvant chemotherapy may be given to decrease the tumor size and allow for a resection with negative margins. Surgery remains controversial and provides a limited role in the treatment of stage III and IV disease. No additional therapy may be required in patients with type I who have a resection with negative margins. Postoperative radiation therapy may be recommended based on extracapsular extension and the presence of positive margins in patients with type II or III thymic tumors or histological evaluation WHO B3 and C. Radiation therapy may be beneficial in patients with locally advanced disease (type III or IV) or in patients with symptoms secondary to compression of surrounding structures. Chemotherapy with cisplatin, doxorubicin, and cyclophosphamide (CAP) remains the mainstay of therapy in the neoadjuvant and adjuvant setting as well as first-line therapy in patients with metastatic thymoma, while carboplatin and paclitaxel are often employed in patients with thymic carcinoma. Limited additional agents are recommended based on small phase II trials as second-line therapy and beyond.

SUMMARY

The management of NSCLC has undergone major change in the past decade. To a lesser extent, the same is true for SCLC and thymic tumors. For patients with early-stage disease, advances in radiotherapy and surgical procedures as well as new systemic therapies have greatly improved prognosis in all diseases. For patients with advanced lung cancer, major progress in understanding tumor genetics and tumor immunology has led to the development of rational targets and specific inhibitors which have documented efficacy in specific subsets of NSCLC. Furthermore, increased understanding of how to activate the immune system to drive antitumor immunity has proven to be a successful therapeutic strategy for a subset of patients with advanced lung cancer. In Fig. 74-8, we propose an algorithm of the treatment approach for patient with stage IV NSCLC. However, the reality is that only a small subset of patients responds to immune checkpoint

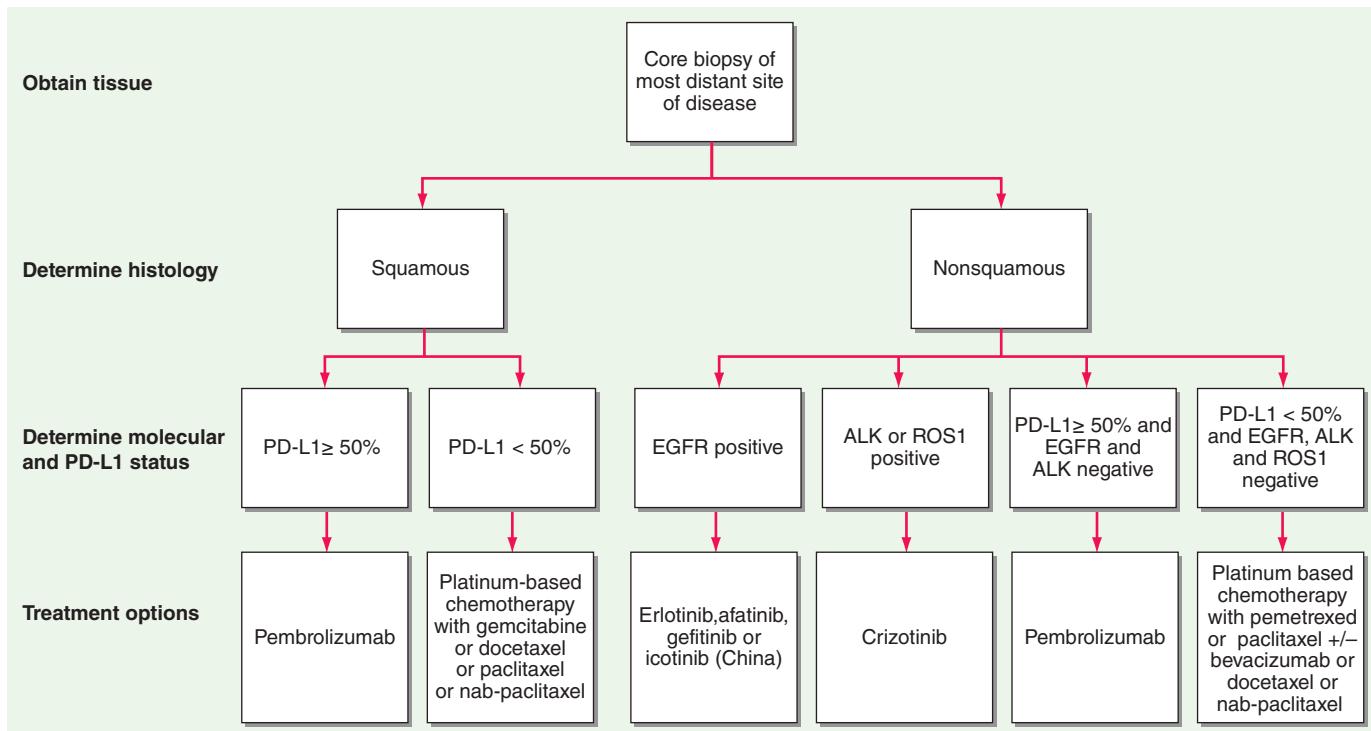


FIGURE 74-8 Approach to first-line therapy in a patient with stage IV non-small-cell lung cancer (NSCLC).

inhibitors and the majority of patients treated with targeted therapies or chemotherapy eventually develop resistance, which provides strong motivation for further research and enrollment of patients onto clinical trials in this rapidly evolving area.

ACKNOWLEDGMENT

David Johnson contributed to this chapter in the prior edition and material from that chapter has been retained here.

FURTHER READING

- ANTONIO SJ et al: Nivolumab alone and nivolumab plus ipilimumab in recurrent small cell lung cancer (CheckMate 032): A multicenter, open-label, phase 1/2 trial. *Lancet Oncol* 17:883, 2016.
- BORGHAEI H et al: Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med* 373:1627, 2015.
- HERBST RS et al: Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): A randomised controlled trial. *Lancet* 387:1540, 2016.
- JÄNNE PA et al: AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* 372:1689, 2015.
- KRIS M et al: Using multiplexed assays of oncogenic drivers in lung cancers to selected targeted drugs. *JAMA* 311:1998, 2014.
- MOK TS et al: Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 361:947, 2009.
- SHAW A et al: Ceritinib in ALK-rearranged non-small cell lung cancer. *N Engl J Med* 370:1189, 2014.
- SHAW A et al: Crizotinib in ROS-1 rearranged non-small cell lung cancer. *N Engl J Med* 371:1963, 2014.
- SOLOMON B et al: First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med* 371:2167, 2014.
- TRAVIS WD et al: The 2015 World Health Organization classification of lung tumors. *J Thorac Oncol* 10:1243, 2015.

early menarche, late first full-term pregnancy, and late menopause. These three factors account for 70–80% of the variation in breast cancer frequency in different countries. Also, duration of maternal nursing correlates with substantial risk reduction independent of either parity or age at first full-term pregnancy.

International variation and immigration statistics of incidence provide insight into hormonal carcinogenesis. A woman living to age 80 years in North America has one chance in nine of developing invasive breast cancer. Asian women have traditionally had only 1/5th to 1/10th the risk of breast cancer of women in North America or Western Europe. However, with shifts from agrarian to industrialized economic systems, and in immigrant populations, Asian women living in modern, Western-style environments have risks identical to those of their Western counterparts.

Presumably, these differences are secondary to menstrual, and associated intrinsic estrogen exposure, histories. However, differences in diets have also been implicated, although the role of diet in breast cancer etiology is controversial. While there are associative links between total caloric and fat intake and breast cancer risk, the exact role of fat in the diet is unproven and may actually intersect with menstrual history and estrogenic exposure.

Central obesity is both a risk factor for occurrence and recurrence of breast cancer. Moderate alcohol intake also increases the risk by an unknown mechanism. Folic acid supplementation appears to modify risk in women who use alcohol but is not additionally protective in abstainers. Recommendations favoring abstinence from alcohol must be weighed against other social pressures and the possible cardioprotective effect of moderate alcohol intake. Chronic low-dose aspirin use is associated with a decreased incidence of breast cancer. Depression is also associated with both occurrence and recurrence of breast cancer.

Exogenous use of female hormones also plays a role in breast cancer incidence. Oral contraceptive use causes a small increased risk of breast cancer. However, this risk is more than balanced by avoidance of an undesired pregnancy and a substantial protective effect against ovarian epithelial and endometrial cancers.

Hormone replacement therapy (HRT) with conjugated equine estrogens plus progestins increases the risk of breast cancer and adverse cardiovascular events, but decreases the risk of bone fractures and colorectal cancer. On balance, there appear to be more negative events with HRT; 6–7 years of HRT nearly doubled the risk of breast cancer. Of note, administration of conjugated estrogens alone (estrogen replacement therapy in women who have had hysterectomies) produces no significant increase in breast cancer incidence. Thus, there are serious concerns about long-term HRT, especially in combination with progestins, in terms of cardiovascular disease and breast cancer. No comparable safety data are available for other less potent forms of estrogen replacement, such as bioequivalent estrogen found in soy, and they should not be routinely used as substitutes. Rapid decrease in the number of women on HRT has already led to a coincident decrease in breast cancer incidence. HRT in women previously diagnosed with breast cancer, especially of the subtype that expresses estrogen receptors, increases recurrence rates.

In addition to the other factors, radiation is a risk factor in younger women. Women who have been exposed before age 30 years to radiation in the form of multiple fluoroscopies (200–300 cGy) or treatment for Hodgkin's disease (>3600 cGy) have a substantial increase in risk of breast cancer, whereas radiation exposure after age 30 years appears to have a minimal carcinogenic effect on the breast.

FURTHER READING

- The genetics of breast cancer require an understanding of the distinction between inherited, germline genetic differences among individuals and acquired, somatic genetic changes within cancers. The former, often called single nucleotide polymorphisms (SNPs), if deleterious, may lead to higher susceptibility to developing cancer and/or to a patient's response to or toxicity from a given treatment (pharmacogenetics). Somatic genetic changes that are not inherited, including mutations, amplifications, deletions, translocations, and others, are responsible for the malignant behavior of a cancer, including

75

Breast Cancer

Daniel F. Hayes, Marc E. Lippman



Breast cancer is a malignant proliferation of epithelial cells lining the ducts or lobules of the breast. In the year 2017, ~247,000 cases of invasive and 61,000 cases of in situ breast cancer and 41,000 deaths will occur in the United States. In addition, ~2000 men will be diagnosed with breast cancer. Epithelial malignancies of the breast are the most common cause of cancer in women (excluding skin cancer), accounting for about one-third of all cancer in women. As a result of improved treatment and earlier detection, the mortality rate from breast cancer has begun to decrease very substantially in the United States. This chapter does not consider rare malignancies presenting in the breast, such as sarcomas and lymphomas, but focuses on the epithelial cancers.

EPIDEMIOLOGY AND RISK FACTORS

Breast cancer is principally a disease of older women. Seventy-five percent of all breast cancers occur in women aged >50 years. The female-to-male ratio is ~150:1. It is also a hormone-dependent disease. Women without functioning ovaries, or who experience an early menopause, and who never receive combination estrogen/progesterone replacement therapy, are much less likely to develop breast cancer than those who have a normal menstrual history. A log-log plot of incidence versus age for breast cancer shows two components: a straight-line increase with age but with a decrease in slope beginning at the age of menopause. Length of menstrual life—particularly the fraction occurring before first full-term pregnancy—is a substantial component of the total risk of breast cancer. Breast cancer risk is increased in women with

unrestrained proliferation, as well as extravasation from one site and migration and establishment of metastases into another.

In this regard, human breast cancer is a clonal disease. One or more transformed cells, which arise due to a combination of inherited germline susceptibility and environmentally driven somatic changes, are eventually able to express full malignant potential. Thus, breast cancer may exist for a long period as either a noninvasive disease or an invasive but nonmetastatic disease. These facts have significant clinical ramifications, including overdiagnosis of biologically nonmalignant but anatomically apparent cancers.

Germline Genetic Susceptibility Although family history is an important risk factor, for most women the increased risk associated with a family member who has had breast cancer appears to be related to both a weak, and probably multi-gene germline susceptibility and/or similar exposure to environmental/life style risk factors. Not >10% of human breast cancers can be linked directly to single germline SNPs. However, when they are present, the relative and absolute risk for that individual's developing breast, and other, cancers in her lifetime are extraordinary.

Of these, the *BRCA1* and 2 genes are the best characterized and have the greatest clinical importance. *BRCA1* has been identified at the chromosomal locus 17q21; this gene encodes a zinc finger protein, and the protein product functions as a transcription factor and is involved in gene repair. Women who inherit a mutated allele of this gene from either parent have at least a 60–80% lifetime chance of developing breast cancer and about a 33% chance of developing ovarian cancer. The cancers that arise within a *BRCA1*-mutated patient are almost exclusively negative for estrogen and progesterone receptors (ER, PgR) and for human epidermal receptor 2 (HER2) (so-called “triple negative” breast cancers), and ~20% of women with triple negative breast cancers will be positive for deleterious germline *BRCA1* SNPs. Nonetheless, the risk of breast cancer penetrance is variable within the *BRCA1*-affected population and is higher among women born after 1940, presumably due to promotional effects of hormonal factors. Men who carry a mutant allele of the gene have an increased incidence of prostate cancer and breast cancer.

BRCA2, which has been localized to chromosome 13q12, is also associated with an increased incidence of breast cancer in women. It should be noted that cancers that arise in *BRCA2* contexts are more likely to be ER positive, compared to those in *BRCA1* families, in which they are almost universally negative for ER, PgR, and HER2 expression. Of interest, men with *BRCA2* deleterious SNPs also have a higher risk of breast cancer, although most male breast cancer cases do not occur in *BRCA2*-mutated men, and the risk of breast cancer in men who do carry the *BRCA2* mutation is lower than that in women with this genetic abnormality.

Germline mutations in *BRCA1* and *BRCA2* can be readily detected in blood tests of normal circulating leucocytes. However, most experts do not recommend testing all women, since the rate of germline SNPs in this gene in the general population is quite low (well below 1%) and the tests are not 100% accurate. Further, it is not infrequent to identify variants of unknown significance (VUS) that may increase patient anxiety without a clear-cut set of recommendations about management. Consensus guidelines on who should be tested include any patient with a triple negative breast cancer and any patient with contralateral breast cancer or who has a first-degree relative (mother, father, or sister) with breast cancer. Further, any man with breast cancer should also be tested. Some guidelines suggest testing any patient with breast cancer who is of Ashkenazi descent, since the incidence in this population of a specific founder *BRCA1* mutation (substitution of adenine for guanine at position 185) is ~2%. Patients with these mutations should be counseled appropriately.

Over the last 5 years, panels of germline genes have been offered in addition to *BRCA1* and 2. These include genes that are known to be risk factors for breast cancer if the individual harbors deleterious SNPs, including *p53*, *PTEN*, and *PALB1*. However, several of the other genes included in these panels are less well-studied, and therefore it is less clear how to counsel affected individuals.

Somatic Genetic Changes in Breast Cancer Abnormalities in these and other genes can also be acquired, leading to breast cancer and its specific behavior. The specific causes of these mutations in breast cancer are generally unknown. A *p53* mutation is present in ~40% of human breast cancers as an acquired defect. Acquired mutations in *PTEN* occur in ~10% of the cases. *BRCA1* mutation in sporadic primary breast cancer has not been reported. However, decreased expression of *BRCA1* messenger RNA (mRNA) (possibly via gene methylation) and abnormal cellular location of the *BRCA1* protein have been found in some breast cancers. Loss of heterozygosity of *BRCA1* and *BRCA2* suggests that tumor-suppressor activity may be inactivated in sporadic cases of human breast cancer.

Approximately 80% of all breast cancers overexpress ER. Many of these cancers respond to antiestrogen treatments. Likewise, increased expression of the dominant oncogene *erbB2*, often due to amplification, occurs in approximately one-quarter of human breast cancer cases. The product of this gene, HER2, contributes to transformation of human breast epithelium. HER2 is the target of effective systemic therapy in adjuvant and metastatic disease settings.

A series of other acquired “driver” mutations has been identified in sporadic breast cancer by major sequencing consortia. Of interest, activating mutations in the gene that encodes for ER (*ESR1*) have been reported in ~20% of metastatic breast cancers after prior endocrine treatment, but almost never in untreated primary cancers. Similarly, activating mutations in *erb2* are reported in 3–5% of breast cancers. Both these findings may have therapeutic implications. Multiple academic and commercial entities are offering exon sequencing for these and many other possible mutations on either tumor biopsies or on circulating DNA shed from tumors. Unfortunately, most occur in no more than 5% of cases. Further, they are either not associated with any known targeted therapeutic agents, or the abnormalities are associated with response to an agent in another disease, but at present not in breast cancer. Therefore, while appealing, “personalized medicine” is for now more of a dream than a reality.

PREVENTION OF BREAST CANCER

One major reason to determine risk would be to develop and apply effective prevention strategies. These might either be lifestyle changes or surgical or pharmacologic interventions. At present, although diet and exercise are certainly recommended approaches to healthy living, none has been proven to specifically decrease a woman’s risk of breast cancer. Avoidance of combined estrogen/progestin HRT avoids their associated increased risk of breast cancer.

Prophylactic removal of the breasts is an effective, albeit usually unacceptable, preventive strategy. Retrospective and prospective registries have demonstrated that bilateral prophylactic mastectomies reduce the risk of breast cancer incidence and mortality by more than 95%. Because breasts are not encapsulated organs, some normal breast tissue is always left behind, and therefore women who elect to have prophylactic mastectomies should be counseled that they still have some risk of developing a new breast cancer. Because of its obvious adverse effect on sexuality, cosmesis, and breast-feeding, this approach is not considered appropriate for a woman of average risk.

As noted, cessation of menses and/or other means of reducing estrogen exposure, such as aromatase inhibition in postmenopausal women, and use of the selective estrogen receptor modulators (SERMs) tamoxifen and raloxifene are effective methods to lower breast cancer risk. So-called chemoprevention with SERMs or aromatase inhibition lowers risk of ER-positive breast cancer by approximately one-third to one-half, although it has no effect on the more lethal ER-negative breast cancers. Of interest, prophylactic bilateral oophorectomy and salpingo-oophorectomy, which is often performed in women with high genetic risk (such as those with inherited *BRCA1/2* deleterious SNPs), also reduces breast cancer risk.

SCREENING FOR BREAST CANCER

A recent review by the American Cancer Society (ACS) supports the perception that screening mammography reduces breast cancer mortality by one-quarter to one-third in women aged ≥50 years. The data for

a relative reduction in breast cancer mortality for women between ages 40 and 50 years are almost as positive; however, since the incidence of breast cancer is much lower in younger women, the number of women whose lives are saved is much lower than in older women, and because they have denser breasts, and therefore there are more false-positive findings and the positive predictive factor is lower.

Further, screening mammography and early detection are more likely to identify tumors at a stage more appropriate for conservative local therapy. Better technology, including digitized mammography, routine use of magnified views, and greater skill in mammographic interpretation, have all improved the accuracy of mammography. Newer diagnostic techniques (magnetic resonance spectroscopy [MRI], positron emission tomography [PET], etc.) appear to have higher sensitivity, but their specificity is often lower. Further, many authors have raised concern about diagnosis of anatomically defined cancers that may be biologically insignificant, raising the specter of overdiagnosis and treatment. Since none of these newer technologies has been shown to be superior to mammography in regards to mortality reduction, screening of women with standard risk by any technique other than mammography is not recommended.

Screening with more sensitive but less specific techniques, in particular MRI, is recommended for women with genetic risk, such as *BRCA1* or *BRCA2* carriers or those with Li-Fraumeni, Cowden's, or Bannayan-Riley-Ruvalcaba syndromes; untested first-degree relatives of women with cancer; women with a history of radiation therapy to the chest between ages 10 and 30 years; or women with a lifetime risk of breast cancer of at least 20%. In these women, the positive predictive value of MRI is higher because of the higher incidence of cancer, and, furthermore, many of them are considering prophylactic mastectomy as an alternative, and therefore the lower specificity and risk of a false positive finding has been considered more acceptable.

Research does not show a clear benefit of individual self-examination or by physical breast examinations done by a health professional. Because of this lack of evidence, regular clinical breast examination and breast self-examination are not recommended. Still, all women should be familiar with how their breasts normally look and feel and report any changes to a health care provider right away. Moreover, because the breasts are a common site of potentially fatal malignancy in women, examination of the breast is an essential part of the physical examination. Although breast cancer in men is unusual, unilateral lesions should be evaluated in the same manner as in women, with the recognition that gynecomastia in men can sometimes begin unilaterally and is often asymmetric.

EVALUATION OF BREAST MASSES IN MEN AND WOMEN

Virtually all breast cancer is diagnosed by biopsy of a nodule detected either on a mammogram or by palpation. Algorithms have been developed to enhance the likelihood of diagnosing breast cancer and reduce the frequency of unnecessary biopsy.

THE PALPABLE BREAST MASS

If a patient brings a breast abnormality to the attention of a health care giver, or if a lesion is appreciated during routine examination, proper attention needs to be given to ensure appropriate evaluation and treatment. Lesions with certain features are more likely to be cancerous. These include enigmatically, painless masses, and, more importantly, hard, irregular masses, especially if tethered or fixed to the underlying chest wall. In contrast, those that are cystic appearing on physical examination or are associated with pain, are less likely malignant. However, none of these is a terribly accurate positive or negative finding. Likewise, a negative mammogram in the presence of a persistent lump in the breast does not exclude malignancy. Any concerning, and persistent, breast finding should be referred to an experienced breast diagnostician.

In premenopausal women, lesions that are either equivocal or non-suspicious on physical examination should be reexamined in 2–4 weeks, during the follicular phase of the menstrual cycle. Days 5–7 of the cycle are the best time for breast examination. A dominant mass in

a postmenopausal woman or a dominant mass that persists through a menstrual cycle in a premenopausal woman should be referred to an experienced breast diagnostician for further evaluation, including biopsy if appropriate.

Several points are essential in pursuing these management decision trees. First, risk-factor analysis is not part of the decision structure. No constellation of risk factors, by their presence or absence, can be used to exclude biopsy. Second, fine-needle aspiration should be used only in centers that have proven skill in obtaining such specimens and analyzing them. The patient and physician must be aware of a 1% risk of false negatives. Third, additional technologies such as MRI, ultrasound, and sestamibi imaging cannot be used to exclude the need for biopsy; although in unusual circumstances, they may provoke a biopsy.

THE ABNORMAL MAMMOGRAM

Diagnostic mammography, which is performed after a palpable abnormality has been detected, should not be confused with *screening mammography*, which is performed in an asymptomatic woman with no prediscovered abnormalities. Diagnostic mammography is aimed at evaluating the rest of the breast before biopsy is performed or occasionally is part of the triple-test strategy to exclude immediate biopsy.

Subtle abnormalities that are first detected by screening mammography should be evaluated carefully by compression or magnified views. These abnormalities include clustered, heterogeneous, linear, and branching microcalcifications; densities (especially if spiculated); and new or enlarging architectural distortion. For some nonpalpable lesions, ultrasound may be helpful either to identify cysts or to guide biopsy. If there is no palpable lesion and detailed mammographic studies are unequivocally benign, the patient should have routine follow-up appropriate to the patient's age. If a nonpalpable mammographic lesion has a low index of suspicion, mammographic follow-up in 3–6 months is reasonable. However, it cannot be stressed too strongly that in the presence of a breast lump a negative mammogram does not rule out cancer, and if it persists or enlarges during follow-up, the patient should be referred to an experienced breast diagnostician.

BREAST MASSES IN THE PREGNANT OR LACTATING WOMAN

During pregnancy, the breast grows under the influence of estrogen, progesterone, prolactin, and human placental lactogen. Lactation is suppressed by progesterone, which blocks the effects of prolactin. After delivery, lactation is promoted by the fall in progesterone levels, which leaves the effects of prolactin unopposed. The development of a dominant mass during pregnancy or lactation should never be attributed to hormonal changes. A dominant mass must be treated with the same concern in a pregnant woman as any other. Breast cancer develops in 1 in every 3000–4000 pregnancies. Stage for stage, breast cancer in pregnant patients is no different from premenopausal breast cancer in nonpregnant patients. However, pregnant women often have more advanced disease because the significance of a breast mass was not fully considered and/or because of endogenous hormone stimulation. Persistent lumps in the breast of pregnant or lactating women *cannot* be attributed to benign changes based on physical findings; such patients should be promptly referred for diagnostic evaluation.

BENIGN BREAST MASSES

Only ~1 in every 5–10 breast biopsies leads to a diagnosis of cancer, although the rate of positive biopsies varies in different countries and clinical settings. These differences may be related to interpretation, medico-legal considerations, and availability of mammograms. The vast majority of benign breast masses are due to "fibrocystic" changes, a descriptive term for small fluid-filled cysts and modest epithelial cell and fibrous tissue hyperplasia. The subset of women with ductal or lobular cell proliferation (~30% of patients), particularly the small fraction (3%) with atypical hyperplasia, have a fourfold greater risk of developing breast cancer than those women who have not had a biopsy. The increase in the risk is about nine-fold for women in this category who also have an affected first-degree relative. Thus, careful follow-up of these patients is required. By contrast, patients with a

benign biopsy without atypical hyperplasia are at little risk and may be followed routinely.

STAGING

Correct staging of breast cancer patients is of extraordinary importance. Not only does it permit an accurate prognosis, but in many cases, therapeutic decision making is based largely on the TNM (primary tumor, regional nodes, metastasis) classification. Comparison with historic series should be undertaken with caution, as the staging has changed several times in the past 20 years. The current staging is complex and results in significant changes in outcome by stage as compared with prior staging systems.

NONINVASIVE BREAST CANCER

Breast cancer develops as a series of molecular changes in the epithelial cells that lead to ever more malignant behavior. Increased use of mammography has led to more frequent diagnoses of noninvasive breast cancer. These lesions fall into two groups: ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (lobular neoplasia or LCIS). The management of both entities is controversial.

Ductal Carcinoma In Situ Proliferation of cytologically malignant breast epithelial cells within the ducts is termed *ductal carcinoma in situ* (DCIS). Atypical hyperplasia may be difficult to differentiate from DCIS. At least one-third of patients with untreated DCIS develop invasive breast cancer within 5 years. However, many low-grade DCIS lesions do not appear to progress over many years; therefore, many patients are overtreated. Unfortunately, there is no reliable means of distinguishing patients who require treatment from those who may be safely observed.

For many years, the standard treatment for DCIS was mastectomy. Although no studies have compared breast-preserving therapy to mastectomy, the ~100% ten year survival rates with the former suggest that it is a satisfactory strategy. Breast-preserving surgery alone may also be acceptable. However, although survival was identical in the two arms of a randomized trial comparing wide excision plus or minus irradiation, the latter caused a substantial reduction in the local recurrence rate as compared with wide excision alone. Addition of tamoxifen or an aromatase inhibitor (AI) to any DCIS surgical/radiation therapy regimen further improves local control. However, in the largest trial comparing the two in DCIS, anastrozole did not improve distant disease-free or overall survival compared to tamoxifen.

Several prognostic features may help to identify patients at high risk for local recurrence after either lumpectomy alone or lumpectomy with radiation therapy, and therefore might provide an indication for mastectomy. These include extensive disease; age <40; and cytologic features such as necrosis, poor nuclear grade, and comedo subtype with overexpression of *erbB2*. In summary, it is reasonable to recommend breast-preserving surgery for patients who have a localized focus of DCIS with clear margins followed by breast irradiation and tamoxifen or anastrozole. For patients with localized DCIS, axillary lymph node dissection is unnecessary.

More controversial is the question of what management is optimal when there is any degree of invasion. Because of a significant likelihood (10–15%) of axillary lymph node involvement even when the primary lesion shows only microscopic invasion, it is prudent to do at least a sentinel lymph node sampling for all patients with any degree of invasion. Further management is dictated by the presence of nodal spread.

Lobular Neoplasia Proliferation of cytological malignant cells within the lobules is termed *lobular neoplasia* (LCIS). Nearly 30% of patients who have had adequate local excision, or incidentally discovered LCIS, or just a biopsy a needle biopsy of a suspicious area develop a subsequent breast cancer (usually infiltrating ductal carcinoma) over the next 15–20 years. Ipsilateral and contralateral cancers are equally common. Therefore, LCIS may be considered a premalignant condition with associated elevated risk of subsequent breast cancer, rather than a form of malignancy itself, and aggressive local management seems unreasonable. Management options include careful observation with

routine mammography or chemoprevention with either a SERM or an AI (for postmenopausal women) for 5 years as well as concurrent and subsequent annual mammography and semiannual physical examinations. A third option, although no more effective and associated with substantial cosmetic, and perhaps emotional, morbidity is bilateral prophylactic mastectomy.

TREATMENT

Breast Cancer

BIOLOGICAL CONSIDERATIONS

One of the most important advances in our understanding of breast cancer has been the appreciation that it can be classified by gene expression patterns into a series of subtypes.

- Luminal:** Luminal breast cancers are almost always positive for ER and negative for HER2 amplification. They are divided into two groups:
 - Luminal A:** Luminal A tumors have the highest levels of ER expression as well as of downstream ER-dependent genes, such as PgR. They are almost universally negative or low in HER2, and they have low proliferative thrust. They are usually low grade, are most likely to respond to endocrine therapy, and have a favorable prognosis. They appear to be less responsive to chemotherapy.
 - Luminal B:** Luminal B breast cancers are also of luminal epithelial origin, but with a gene expression pattern distinct from luminal A. They tend to be PgR negative and have evidence of higher proliferative activity. They also tend to express HER2, but not to the level of the so-called "HER2 amplified" cancers. Their grade is more often higher than luminal A cancers. Prognosis is somewhat worse. They may be more sensitive to chemotherapy.
- HER2 amplified:** These tumors have amplification of the *HER2* gene on chromosome 17q and frequently exhibit coamplification and overexpression of other genes adjacent to *HER2*. Historically the clinical prognosis of such tumors was poor. However, with the advent of trastuzumab and other targeted therapies, the clinical outcome of *HER2* positive patients is markedly improved compared to 20 or more years ago.
- Basal:** These ER/PgR-negative and *HER2*-negative tumors (so-called triple negative) are characterized by markers of basal/myoepithelial cells. They tend to be high grade, and express cytokeratins 5/6 and 17 as well as vimentin, p63, CD10, α -smooth muscle actin, and epidermal growth factor receptor (EGFR). Patients with *BRCA1* mutations also fall within this molecular subtype. They also have stem cell characteristics.
- Normal breast-like:** These tumors have a gene expression profile reminiscent of nonmalignant "normal" breast epithelium. Prognosis is similar to the luminal B group. This subtype is somewhat controversial and may represent contamination of the sample by normal mammary epithelium.
- Claudin-low:** These cancers are often triple negative but they have low expression of cell-cell junction proteins including E-cadherin. They are frequently associated with lymphocytic infiltration.

GENERAL TREATMENT CONSIDERATIONS

Treatment of breast cancer depends on whether the patient does or does not have evidence of distant (meaning outside the breast, chest wall, and regional lymph nodes) metastases, as detected by scintigraphic or radiologic imaging and biopsy. For patients with no evidence of detectable distant metastases, the goal of therapy is cure, or at least substantial survival prolongation, and is divided into primary and systemic considerations. Primary therapies consist of surgical and radiation treatments directed toward the breast and locoregional lymph nodes. These approaches are designed to excise and eliminate the cancer and sterilize unaffected breast tissue as appropriate. Adjuvant systemic

treatments, consisting of antiestrogen (or endocrine), anti-HER2, and/or chemotherapies, are given to treat micrometastases that may have already escaped to distant sites but are not yet detectable.

All treatments for breast cancer are based on prognostic and predictive factors. Prognostic factors provide an indication of how likely a cancer will recur, either locally or in distant organs, in the future if a patient is not treated with the respective treatments. Predictive factors are used to determine if a given treatment is likely to work or not, assuming the patient's prognosis justifies treatment (or further treatment assuming the patient has been treated in some manner already).

Prognostic features guide both whether and what type of primary and adjuvant systemic treatments should be pursued. Anatomic prognostic features include visual and physical examination findings of locally advanced breast cancer (T4 lesions: skin erythema [“inflammatory”] or edema [“peau d’orange”], nodules, or ulceration or tumor fixation to the chest wall). In patients without any of these findings, the most important prognostic features are tumor size and lymph node status (TN in the staging system). As discussed below, biologic features, such as histologic tumor grade as well as ER, PgR, and HER2, are also prognostic. Over the last decade, several multiparameter tests based on gene expression have been developed to determine prognosis in patients who have node-negative, ER-positive, and HER2-negative disease.

Predictive features are usually used to guide systemic therapies. These include ER for endocrine treatments and HER2 for anti-HER2 therapies, such as trastuzumab. There are no established predictive factors to predict response to radiation treatment. The issue of chemoresistance in luminal A cancers is under large-scale investigations.

■ EARLY-STAGE BREAST CANCER

Primary Therapies Prior to 1980, the Halsted radical mastectomy, in which the breast, chest wall muscles, and complete axillary nodal contents were removed, was the standard treatment of choice for women with newly diagnosed breast cancer. In the 1980s, prospective randomized trials demonstrated that recurrence and survival rates were the same with the less disfiguring modified radical mastectomy, in which the chest wall muscles were preserved and only a sampling of axillary lymph nodes were removed.

In the same decade, breast-conserving treatments, consisting of the removal of the primary tumor by some form of surgical excision (designated as lumpectomy, quadrantectomy, or partial mastectomy), were shown to result in equal, if not slightly superior, to that associated with mastectomy. Several of these trials also demonstrated that the in-breast recurrence rate was quite high in the absence of breast radiation, while it was reduced substantially if radiation was provided. Therefore, for women undergoing breast conservation, postlumpectomy radiation is usually indicated, although it may be less necessary and withheld in older women with ER-positive, node-negative breast cancer, since their risk of subsequent in-breast recurrence is quite low with surgery and endocrine therapy only. When lumpectomy with negative tumor margins is achieved and radiation is delivered appropriately, breast conservation is associated with a recurrence rate in the breast of <5%.

Not all patients are candidates for breast-conserving therapy. Contraindications include large tumor to breast ratio, inability to achieve clear margins with adequate cosmesis after extensive surgery, multifocal cancers, extensive four-quadrant DCIS, and inability to receive radiation. The latter issue arises in women with dermal autoimmune disease (such as lupus erythematosus), prior radiation to the site, and/or lack of available radiation treatment facilities. Further, although not contraindicated, breast-conserving therapy may be less cosmetically acceptable than mastectomy with reconstruction if the nipple-areolar complex is involved with cancer and must be sacrificed. This is a personal choice, and some women prefer mastectomy, especially those with high genetic risks for second breast cancers.

For patients who do undergo mastectomy, postoperative chest wall and regional nodal radiation is also associated with an improvement in survival if they have a high risk of local-regional recurrence, such as tumors ≥5 cm, four or more positive axillary lymph nodes,

or postoperative positive margins. Postmastectomy radiation is not indicated in women with cancers <2 cm, negative lymph nodes, and negative margins. It is considered for women who fall into the areas between these (2–5 cm, one to three positive nodes, or close margins), and is usually recommended if a patient has one to three involved axillary lymph nodes.

At present, nearly one-third of women in the United States are managed by lumpectomy, and recent data suggest that the fraction of women treated with breast-conserving therapy is decreasing. It appears that many women still undergo mastectomy who could safely avoid this procedure and probably would if appropriately counseled.

Axillary node sampling or dissection is unnecessary in many cases. Sentinel lymph node mapping and biopsy (SLNB) is generally the standard of care for women with localized breast cancer and clinically negative axilla. If SLNB is negative, more extensive axillary surgery is not required, avoiding much of the risk of lymphedema following more extensive axillary dissections. Even in the presence of sentinel lymph node involvement, further axillary surgery may not be required for selected patients, such as older women and those with ER-positive cancers.

The survival of patients who have recurrence in the breast after proper treatment (adequate surgery and radiation if indicated) is somewhat worse than that of women who do not, but it is not worse than those who suffer local-regional recurrence after mastectomy. Thus, local-regional recurrence is a negative prognostic variable for long-term survival but not the *cause* of distant metastasis. Most patients should consult with a radiation oncologist before making a final decision concerning local therapy. However, a multimodality clinic in which the surgeon, radiation oncologist, medical oncologist, and other caregivers cooperate to evaluate the patient and develop a treatment plan is usually considered a major advantage by patients.

Adjuvant Systemic Therapies The concept of adjuvant systemic therapy is based on the observation that cancer is a condition of genetic instability, and with increasing generations of cellular replication, genetic abnormalities accumulate. Although these occur randomly, and therefore may lead to sensitivity or resistance to therapies, the latter is of greater concern. Thus, as a consequence of accumulation of mutations to resistance, almost all patients with metastatic breast cancer are destined to die with, if not of their cancer.

However, treatment with the same therapies administered earlier, in the setting of micrometastatic disease only, has been repeatedly shown to be more effective than waiting until symptomatic, documented metastases occur. Put simply, the use of systemic therapy as an adjuvant to local management of breast cancer substantially improves survival. More than half of the women who would otherwise die of metastatic breast cancer remain disease-free and experience considerable survival advantaged when treated with the appropriate adjuvant systemic regimen. These data have grown more and more impressive with longer follow-up and more effective regimens.

PROGNOSTIC VARIABLES As noted, prognostic factors help define who most likely needs, or perhaps more importantly does not need, adjuvant systemic therapy. The most important prognostic variables are provided by tumor staging: tumor size (T), lymph node status (N) and detectable distant metastases (M) (**Table 75-1**). Histologic classification of the tumor has also been used as a prognostic factor. Tumors with a

TABLE 75-1 5-Year Survival Rate for Breast Cancer by Stage

STAGE	5-YEAR SURVIVAL, %
0	99
I	92
IIA	82
IIB	65
IIIA	47
IIIB	44
IV	14

Source: Modified from data of the National Cancer Institute: Surveillance, Epidemiology, and End Results (SEER).

poor nuclear grade have a higher risk of recurrence than tumors with a good nuclear grade. Semiquantitative measures such as the Elston score improve the reproducibility of this measurement. Importantly, there is no need to perform imaging for distant metastases in a patient with no signs or symptoms of widespread disease and who has a T3 or less tumor and fewer than four involved axillary lymph nodes.

Adjuvant systemic therapy may not be needed at all for patients with very small (<1 cm) tumors and negative lymph nodes. However, there is no patient with invasive breast cancer who does not have some risk of subsequent distant metastases, and therefore who might not benefit at all. This consideration raises two issues: (1) the differences in odds of benefit and odds of toxicities of the various types of therapies and (2) the judgment between the patient and her caregiver regarding the calculated absolute benefit-risk ratio for specific types of adjuvant systemic treatments.

There are three types of adjuvant systemic therapies: (1) chemotherapy; (2) endocrine; and (3) anti-HER2 therapies. The decision whether to apply each of these depends on prognostic and predictive features as well as the combined judgment of the patient and caregiver. For example, a patient might be much more likely to accept endocrine therapy for a very small potential benefit than she would accept chemotherapy for the same calculated advantage, since the former is much less often associated with either life-taking, life-threatening, or permanently life-changing toxicities than the latter. Thus, one has to consider prognostic and predictive factors for each type of therapy, separately.

The greatest controversy concerns the recommendation for adjuvant *chemotherapy*, since there is no good predictive factor for this class of treatments, and the decision must be made on prognosis alone. Large overview analyses suggest that chemotherapy reduces the risk of recurrence over the 10 years subsequent to primary diagnosis by approximately one-third. For patients with positive lymph nodes and/or features that render the cancer T4, the risk of distant recurrence (and thus not being cured) over that decade is 50% or higher. Therefore, a one-third reduction of at least 50% means that 15–20%, or more, women will be cured who would not have been in the absence of adjuvant chemotherapy. The life-taking, life-threatening, or permanently life-changing toxicities of adjuvant chemotherapy are ~1–2%, and therefore almost all medical oncologists would recommend adjuvant chemotherapy in this setting.

In contrast, there is rarely justification for adjuvant chemotherapy in most women with tumors <1 cm in size whose axillary lymph nodes are negative. However, this decision is very much weighed by the expression of ER and HER2. For example, the risk of recurrence of such a patient whose tumor is negative for ER, PgR, and HER2 (so-called triple-negative breast cancer) over the succeeding 10 years without any adjuvant is ~15%. If chemotherapy reduces this risk by approximately one-third or more, which is what large overview analyses suggest, then 5%, or perhaps even higher, of patients will be cured who would otherwise be destined to die of their disease. Likewise, a patient with ER and PgR-negative, but HER2-positive, disease has a slightly worse prognosis (risk of recurrence over 10 years is ~20%), and will benefit not only from the adjuvant chemotherapy but from anti-HER2 therapy as well, so that her potential absolute benefit is even higher. Many, but not all, clinicians would recommend adjuvant chemotherapy for such patients.

On the other hand, patients with ER-positive disease have a better prognosis than those with ER-negative breast cancer, and adjuvant endocrine therapy will further reduce the odds of recurrence by approximately one-half. Therefore, the same patient in the example above (<1 cm, node negative) but who has an ER-positive and HER2-negative cancer has a lower initial risk of recurrence (~10% over 10 years). Given the relatively low life-taking, life-threatening, or permanently life-changing toxicities, she is very likely to accept adjuvant endocrine therapy, further lowering her estimated risk of recurrence to ~5%. If chemotherapy reduces this risk by approximately one-third, no more than 1–2% of patients will benefit. This potential benefit is approximately the same as the number of patients who will suffer life-taking, life-threatening, or permanently life-changing toxicities. Thus, in this case, most clinicians would recommend adjuvant endocrine, but not chemotherapy.

These examples represent extremes. In the screening era, up to 30% of newly diagnosed patients have T2-3, node-negative, ER-positive cancers. These patients have an intermediate risk between the two extremes, and the calculated absolute benefit of adjuvant chemotherapy is ~3–5%. It is unclear if this small but real benefit is sufficient to justify adjuvant chemotherapy. Detection of breast cancer cells either in the circulation or bone marrow is associated with an increased relapse rate. However, the finding of bone marrow micrometastases only portends a slightly worse prognosis, especially in node negative patients, and bone marrow biopsies are not recommended in patients with early stage disease.

The most exciting development in this area is the use of gene expression arrays to analyze patterns of tumor gene expression, especially for node-negative, ER-positive cancers. Several groups have independently defined gene sets that reliably predict disease-free and overall survival far more accurately than any single prognostic variable. The Oncotype DX® Recurrence Score (RS) analysis of 21 genes was the first such assay to be adopted. A number of retrospective and more recently prospective studies have documented its utility in identifying patients with node-negative, ER-positive breast cancer whose prognosis, assuming adequate adjuvant endocrine therapy, is so good that they can forego adjuvant chemotherapy. Basically, the 30–50% of patients with ER positive, node negative, but low RS, appear to have luminal A breast cancers, and they do not need chemotherapy, whereas those with high RS appear to have luminal B cancers and the benefits of adjuvant chemotherapy clearly outweigh the risks. For those with intermediate RS, the answer is still unclear and has been the focus of now completed, but as yet unreported, prospective trials.

More recently, other assays, including the Prosigna®, EndoPredict®, and Breast Cancer Index® have also been shown to have clinical utility in this setting. Only one of these tests should be ordered for a single patient, since they do not always give the same results and there are no data to determine which, in the case of discordance, might be "correct." Also, the use of such standardized risk assessment tools such as Adjuvant! Online (www.adjuvantonline.com) is very helpful. These tools are highly recommended in otherwise ambiguous circumstances.

Several *measures of tumor growth rate* correlate with early relapse, but their use is problematic due to analytical variability. Of these, assessment using immunochemical assays for the proliferation marker, Ki67, is the most widespread. However, there is substantial lab-to-lab variability and disagreement regarding optimal cut points. At present, in standard practice outside of a highly skilled laboratory, use of Ki67 is not recommended to make clinical decisions.

Molecular changes in the tumor are also useful. Tumors that overexpress erbB2 (HER2/neu) have a worse prognosis, but expression of this gene for prognosis is most important in patients with ER-positive, node-negative disease. Indeed, patients with HER2-positive breast cancer are so likely to have a high RS that it is not recommended that the Oncotype DX®, or for that matter any of the other multiparameter assays, be ordered. HER2 should be performed on every breast cancer biopsy, however, because of its predictive role for anti-HER2 therapies.

Predictive Factors to Choose Adjuvant Systemic Therapy

The decision to recommend AST is also based on predictive factors; those that provide a prediction of the likelihood that a given class, or even specific drug within a class, will have activity or not. The two important predictive factors, which should be ordered in all breast cancer biopsies (primary or metastatic), are ER and HER2.

There is no detectable benefit in patients with ER-poor, or -negative, cancers, whereas adjuvant endocrine therapy reduces the risk of recurrence by one-half or more in patients with ER-rich cancers. ER is most commonly measured by counting the percent of positive cells within the cancer after immunohistochemical (IHC) staining. Endocrine therapy is recommended for any patient with $\geq 10\%$ positive cells, whereas it is not for those whose cancers only have 0–1% staining. The evidence supporting benefit in 1–9% cases is weak, but given the potential benefit and relatively low toxicities of endocrine therapy, it is recommended for patients in this circumstance, with a low threshold for discontinuation if side effects are intolerable.

HER2 is the target for anti-HER2 therapies. Adjuvant trastuzumab therapy reduces the risk of distant recurrence by one-third or more, with associated substantial risk of dying of breast cancer. Most, if not all, of the large adjuvant trastuzumab trials have been performed in patients with HER2-“positive” breast cancer. HER2 status is determined using either IHC staining for protein overexpression, or fluorescent *in situ* hybridization (FISH) for gene amplification. IHC staining of 3+ (on a scale of 0–3+) is considered positive, whereas 0–1+ is considered negative. For cases with 2+ staining, reflex FISH analysis is recommended. FISH can either be used as the initial evaluation, or for additional evaluation in IHC 2+ cases. FISH results are considered positive if the ratio of HER2 to centromere signal on chromosome 17 is ≥2.0. There is no reason to do FISH if IHC is 3+ or 0–1+, nor is there reason to order IHC testing if FISH is ≥2.0. If note, preclinical studies and retrospective analyses of a few selected cases from the prospective randomized trials have suggested that perhaps trastuzumab might be effective in cases with IHC 1–2+ results. A large prospective randomized clinical trial addressing this issue is completed but not yet reported.

There are no reliable predictive factors for chemotherapy, in general or for specific types of chemotherapies. It has been hypothesized that chemotherapy may be more active in ER-negative and/or HER2-positive cancers. More recently, this issue has evolved to imply that luminal B cancers may be more chemosensitive, whereas luminal A cancers are perceived to be relatively chemoresistant. At present, none of the tests for intrinsic subtype should be used to determine whether to give chemotherapy or not, based on *prediction* of resistance in patients with poor *prognosis*, such as those with T4 or node-positive disease. Attempts to identify reliable predictive factors for individual classes of chemotherapeutic agents (such as anthracyclines, alkylating agents, or taxanes) have been unsuccessful. The platin salts (carbo-, cis-platin) may have higher activity in patients with triple-negative breast cancer and perhaps in patients with HER2-positive disease.

Adjuvant Regimens If chemotherapy is indicated, it should include multiple agents, either in combination or as sequential single agents. If indicated, anti-HER2 therapy should include at least 1 year of trastuzumab, and preliminary data have supported addition of pertuzumab for at least three months. Endocrine therapy should be administered to patients with ER-positive breast cancer following completion of chemotherapy and administered for at least 5 years, and probably longer.

Endocrine Therapy There are two proven endocrine therapy strategies: the SERM, tamoxifen, or estrogen ablation. In addition to being effective in preventing new cancers and reducing the risk of local-regional recurrences in patients with DCIS, tamoxifen reduces the risk of distant recurrence and death due to invasive breast cancer by ~40% over the decade following diagnosis. It is equally effective in pre- and postmenopausal women, although it may be slightly less effective in very young (<40 years) patients. Because tamoxifen is a SERM, it has mixed ER antagonism (in the breast and brain) and agonism (in the bone, liver, and uterus). Therefore, it is active against breast cancer in the prevention, adjuvant, and metastatic settings, but frequently causes hot flashes. The agonistic effect results in reduction of osteopenia/osteoporosis, especially in postmenopausal women, but it increases thrombosis and endometrial cancers due to this effect in the liver and uterus, respectively.

Estrogen depletion can be achieved surgically in premenopausal women by oophorectomy or ovarian suppression with a gonadotropin-releasing hormone super-agonist (GnRH agonist), such as goserelin, that results in tachyphylaxis of the pituitary. However, women with nonfunctioning ovaries, whether induced or by natural menopause, still produce small amounts of estrogen. Estrogen production in these women occurs by adrenal synthesis of estrogen precursors (testosterone, dehydroepiandrosterone [DHEA]) that are converted to estradiol and estrone by aromatase activity in peripheral fat and possible cancer cells. In postmenopausal women, circulating estrogen can be reduced to nearly imperceptible levels with the use of oral AIs. There are three such agents available (anastrozole, letrozole, and exemestane).

Although there is no perceptible difference in activity or toxicity among the three AIs, they are all slightly more effective than tamoxifen.

It is recommended that all postmenopausal women with ER-positive breast cancer be treated for at least 3–5 years with an AI, unless there is a contraindication. The most common concern is the presence of severe osteoporosis, since this is the most frequent life-taking or life-threatening toxicity of the AIs. Likewise, ~15–20% of patients cannot tolerate the AIs due to musculoskeletal symptoms mimicking osteoarthritis and arthralgias. For both these groups of women, tamoxifen is a reasonable therapy, again assuming no contraindications exist. The most important of these is a past history of thrombosis, or high risk of cerebrovascular disease.

For premenopausal women, the decision of optimal endocrine therapy depends on prognosis and patient choice. Complete estrogen depletion is slightly more effective than tamoxifen alone, but it may also be associated with more bothersome side effects, such as hot flashes, vaginal dryness, and sexual dysfunction. Recent studies have suggested that complete estrogen depletion, consisting of either oophorectomy or chemical suppression of gonadotropins coupled with an AI, is indicated for women with worse prognosis, in particular node positivity. For those with more favorable prognosis, tamoxifen alone may be preferable. The AIs should not be administered to women with functioning, or dormant, ovaries, since the negative hypothalamic-pituitary feedback can result in a rebound hyperestrogenic production effect.

The duration of adjuvant endocrine treatment is unclear. Until recently, the standard recommendation was at least 5 years of therapy. Several studies have now demonstrated that although 5 years of adjuvant endocrine treatment clearly reduces the risk of recurrence during that time and for a few years after discontinuation, the annual risk of distant recurrence during the subsequent 15 years is 0.5–3%, depending on the initial T and N status. Further, so-called extended adjuvant endocrine therapy with either tamoxifen or an AI, for at least years 6–10, continues to reduce this late risk of relapse. The decision of whether to continue adjuvant endocrine therapy or not after 5 years must therefore take into consideration initial risk (T, N, grade), current side effects and potential cumulative toxicities, and the patient’s perception of the relative and absolute benefits and risks.

Chemotherapy If adjuvant chemotherapy is indicated, as discussed above, one must consider the optimal regimen. Several studies, and a combined overview analysis, have demonstrated that multiple-agent chemotherapy is more effective than single agent. However, at least two studies have shown that sequential single-agent chemotherapy is as effective, and may be slightly less toxic, than simultaneous combination chemotherapy although it requires longer total duration to deliver. Administration of four to six cycles of chemotherapy appears to be optimal; one cycle is less effective than six, but more than six have generally increased toxicity without further efficacy.

Several chemotherapeutic agents have activity in the adjuvant setting. These include alkylating agents, (principally cyclophosphamide), anthracyclines (doxorubicin, epirubicin), antimetabolites (5-fluorouracil [5FU], capecitabine, methotrexate), and the taxanes (paclitaxel, docetaxel). Within classes, randomized trials have failed to demonstrate superiority of one agent versus another (e.g., doxorubicin vs epirubicin, or paclitaxel vs docetaxel). Dose escalation above an optimal dose is not more effective. The advantage of more frequent scheduling for most individual agents is unclear, but weekly or every other week paclitaxel is superior to every 3-week infusion, while, enigmatically, the opposite is true for its cousin, docetaxel. However, one benefit of a “dose dense” regimen (e.g., every 2 weeks with cytokine support vs every 3 weeks) is earlier completion of therapy.

These agents are usually combined within a single regimen. The oldest of these is cyclophosphamide, methotrexate, and 5FU (CMF). Addition of an anthracycline, or substitution of an anthracycline for the antimetabolite, improves outcomes slightly, albeit with slightly increased risk of heart failure and secondary leukemia. Addition of a taxane to an anthracycline-based regimen further reduces the chances of distant recurrence and death, albeit only modestly. Recent studies have suggested that addition of an anthracycline to a taxane-based

regimen is also modestly more effective than a taxane plus cyclophosphamide alone.

Which regimen is appropriate for a patient must be individualized based on prognosis, comorbid conditions, and the perspective of the patient. For example, the modest relative improvement of giving an anthracycline, cyclophosphamide, and a taxane (AC-T) may not transfer to a sufficiently large absolute improvement in survival in a patient with a relative small (T2) tumor and negative nodes, whereas that same relative reduction in death may translate to a sufficiently large absolute benefit in a patient with a worse prognosis. Therefore, the former patient might best be served with a taxane/cyclophosphamide (TC) regimen alone, while the latter might wish to accept the added risk of congestive heart failure and leukemia associated with the anthracyclines.

Neoadjuvant treatment involves the administration of adjuvant systemic therapy, most commonly chemotherapy, before definitive surgery and radiation therapy. The objective partial and complete response rates of patients with breast cancer to neoadjuvant chemotherapy exceed 75%. Thus, many patients will be “downstaged” by neoadjuvant chemotherapy. In this circumstance, patients with locally advanced, inoperable cancers may become candidates for surgery, and a small fraction of patients who are not considered eligible for breast-conserving surgery may become so due to shrinkage of their cancer. However, overall survival has not been improved using this approach as compared with the same drugs given postoperatively.

Patients who achieve a pathologic complete remission after neoadjuvant chemotherapy have a substantially improved survival compared to those who do not. It is unknown if this observation implies that the latter group did not benefit, or just had a worse initial prognosis, yet still gained some benefit. Although it is appealing to consider treating patients who have not had a pathologic complete response with even more chemotherapy, no studies have demonstrated that doing so improves overall survival. It is possible that these patients have chemoresistant disease, and therefore more chemotherapy will not be of value. However, it is essential that all patients, regardless of response to neoadjuvant chemotherapy, receive adjuvant endocrine therapy if they have an ER-positive breast cancer and adjuvant anti-HER2 therapy if their cancer is HER2 positive.

The neoadjuvant setting also provides an appealing opportunity for the evaluation of new agents. For example, a second HER2-targeting antibody, pertuzumab, has been shown to provide increased rates of pathologic complete response when combined with trastuzumab in the neoadjuvant setting. However, this approach is controversial; it is not clear that demonstration of higher response rates in the neoadjuvant setting will translate into better overall survival. For example, neoadjuvant trials demonstrated that combination trastuzumab and lapatinib resulted in higher pathologic complete responses than trastuzumab alone, yet a classically performed adjuvant trial failed to demonstrate improved survival for this regimen.

Chemotherapy is associated with nausea, vomiting, and alopecia in ~100% of patients, although the former two are well controlled with modern antiemetics. More importantly, chemotherapy causes neutropenia and fever, with a risk of infection of ~1%. The neutropenia can be prevented in most patients with appropriate use of the growth factor filgrastim. Secondary myelodysplasia and leukemia occur in ~0.5–1% of patients treated with anthracyclines as well as with high cumulative doses of cyclophosphamide, usually occurring within 2–5 years of treatment. The anthracyclines cause cumulative dose-related congestive heart failure, which occurs in ~1% of patients treated with standard four to five cycles at 60 mg/m². Peripheral neuropathy is the major dose-limiting and life-changing toxicity of the taxanes. Neuropathy occurs during treatment in ~15–20% of patients, and permanent, chronic neuropathy persists in 3–5%.

Anti-HER2 Therapy The emergence of therapies directed toward HER2 has been one of the great success stories of all oncology. Several trials have demonstrated that the humanized monoclonal antibody, trastuzumab, decreases both risk of recurrence and mortality in early-stage breast cancer. While trastuzumab administered after

chemotherapy is effective, the accumulated evidence suggests that it is optimally delivered concurrently with chemotherapy, particularly in association with a taxane. However, concurrent treatment with an anthracycline is generally avoided, since the main toxicity of trastuzumab is cardiac dysfunction, which appears more often when the agent is delivered simultaneously with doxorubicin. Therefore, if an anthracycline is to be used, it is most commonly given prior to administration of trastuzumab—for example as AC for four cycles followed by a taxane plus trastuzumab. In patients with reasonably favorable prognosis (T1 or 2, node negative), single-agent paclitaxel plus trastuzumab appears to be an adequate regimen.

Twelve months of trastuzumab therapy is optimal. Randomized trials have demonstrated no additional benefit beyond 12 months, whereas 6 months has been shown to be inferior to 12. Trastuzumab is administered intravenously weekly or every 3 weeks.

Other, anti-HER2 treatments that are effective in the metastatic setting are appealing candidates for adjuvant therapies. As noted, neoadjuvant studies have demonstrated that chemotherapy with the combination of trastuzumab and pertuzumab results in higher pathologic complete responses than trastuzumab alone. The U.S. Food and Drug Administration (FDA) has granted this combination with accelerated approval, but final approval for the combination is pending more clinically meaningful results (disease-free, overall survival) from now-completed, classic adjuvant trials. Although lapatinib did not add to trastuzumab therapy and single-agent adjuvant lapatinib is inferior to single agent trastuzumab, another anti-HER2 tyrosine kinase inhibitor, neratinib, is superior to no anti-HER2 therapy. Neratinib has not been compared to trastuzumab, either as a single agent or in combination. Ado-trastuzumab emtansine, an antibody-drug conjugate, has activity in the metastatic setting even in patients who have progressed on trastuzumab and is now being tested in the adjuvant setting.

Skeletal Strengthening Agents Bone-strengthening agents that are commonly used to treat osteoporosis appear to have some, but limited, activity in preventing recurrent breast cancer, particularly in postmenopausal women. In an overview analysis of all trials addressing bisphosphonate therapy, improvement in overall survival was not significantly associated with any specific bisphosphonate class, treatment schedule, ER status, nodal status, tumor grade, or concomitant chemotherapy. No differences were seen in nonbreast cancer mortality. Bone fractures were reduced (relative risk [RR] 0.85, 95% confidence interval [CI] 0.75–0.97; $p = 0.02$). At present, there is no clear consensus regarding routine use of bisphosphonates as an adjuvant therapy, although patients with advancing osteopenia or confirmed osteoporosis should be treated accordingly.

Novel Adjuvant Systemic Agents Other exciting adjuvant strategies are being tested, such as poly-ADP ribose polymerase (PARP) inhibitors in patients with known germline *BRCA1* or *BRCA2* mutations or those with triple-negative cancers that share similar defects in DNA repair in their etiology. The remarkable results of immune checkpoint inhibitors in other cancers have led to studies of this approach in both metastatic and post-neoadjuvant chemotherapy settings but are still considered highly investigational.

Recommendations for adjuvant therapy are found in Table 75-2.

■ STAGE III BREAST CANCER

Between 10 and 25% of patients present with so-called locally advanced, or stage III, breast cancer at diagnosis. Many of these cancers are technically operable, whereas others, particularly cancers with chest wall involvement, inflammatory breast cancers, or cancers with large matted axillary lymph nodes, cannot be managed with surgery initially. As noted, neoadjuvant compared may be no more effective than postsurgical adjuvant chemotherapy in prolonging survival, but the advantages of downstaging and therefore facilitating local therapy are accepted. Radiotherapy either to the chest wall after mastectomy or to the breast after tumor excision is almost always recommended, as is regional lymph node treatment. Adjuvant anti-HER2 and endocrine therapies are also used, as appropriate. These patients should be managed in multimodality clinics to coordinate surgery, radiation

TABLE 75-2 Suggested Approaches to Adjuvant Systemic Therapy^a

NODAL STATUS	TUMOR SIZE	ER	HER2	MULTI-PARAMETER ASSAY	MENSTRUAL STATUS	CHEMOTHERAPY	ENDOCRINE THERAPY	ANTI-HER2 THERAPY
Positive	Any	Neg	Neg	Not indicated	Any	Multidrug	None	None
		Pos			Prem		Ovarian ablation + AI	
					Post		AI	
		Any	Pos		Any			Trastuzumab X12 mos; pertuzumab X12 weeks
Negative	<1 cm	Neg	Neg	Not indicated	Any	Consider multidrug	None	None
	≥1 cm	Neg	Neg		Any	Multidrug	None	None
	1–5 cm	Pos	Neg	Low RS	Prem	None	Tam	None
					Post		AI	None
				Intermed	Any	Consider multidrug	Tam (pre) or AI (post)	None
				Hi	Any	Multidrug	As for node pos	None
		Any	Pos	Not indicated	Any	Single-agent paclitaxel	As for node pos	Trastuzumab X12 mos

^aMeant for guidance only. Each patient should be considered independently based on tumor and comorbidity status.

therapy, and systemic chemo-, endocrine, and anti-HER2 therapies, as indicated. Such approaches produce long-term disease-free survival in ~30–50% of patients.

BREAST CANCER SURVIVORSHIP ISSUES

The odds of surviving breast cancer have increased dramatically over the last 35 years due to a combination of early detection and more effective therapies. Although detection bias improves case fatality rates, age-adjusted mortality rates (mortality/100,000 women in society/year) have declined by >30%. Therefore, while ~40,000 American women will die of metastatic breast cancer in 2016, >60,000 would have suffered breast cancer mortality without these advances. Thus, all clinicians, not just oncologists, need to be aware of survivorship issues in patients with previously diagnosed and treated breast cancer.

No special follow-up procedures, such as serial circulating tumor biomarkers or systemic radiographic/scintigraphic imaging, are indicated in an asymptomatic patient with no physical findings of recurrence. Although randomized trials have demonstrated slightly higher incidence of detection of metastases with lead times of 3–12 months by screening asymptomatic patients compared to no special follow-up, there is no evidence of improved overall survival. If anything, one of these studies suggested a worse quality of life due to higher anxiety levels associated with the testing, and toxicities associated with earlier treatment in patients who were otherwise doing well at that time. These recommendations are summarized in **Table 75-3**.

It is important to carefully assess and evaluate new symptoms, considering whether they might be due to the cancer, the treatment, or an unassociated condition. Judgment needs to be used to decide if blood tests or imaging are required, in order to avoid missing a lesion for which appropriate treatment would improve the patient's quality of life but to diminish overtesting, with associated inconvenience, anxieties, false positives, and cost. Serial echocardiography should be performed every 3 months for patients on adjuvant trastuzumab, but not after it is discontinued.

Likewise, there is no role for serial monitoring for long-term, life-threatening toxicities associated with chemotherapy, such as myelodysplastic syndromes or congestive heart failure, since these are quite uncommon and likely to cause obvious symptoms requiring proper evaluation if they occur.

For patients on endocrine therapy, quality-of-life issues may be critical, including hot flashes, sexual difficulties, musculoskeletal complaints, and risk of osteoporosis. Although estrogen therapy, given orally, transdermally, or transvaginally, effectively reduces these side effects, it should not be given to these patients, since it may counteract the efficacy of the endocrine therapy. Nonhormonal treatments, such as selected antidepressants for hot flashes and musculo-skeletal symptoms, and counseling and water-based lubricants for sexual issues

can be quite helpful. It is important to screen bone density in patients on an AI more frequently than is recommended for the average postmenopausal woman, since total estrogen depletion results in enhanced risk of osteoporosis and fracture. All women should be counseled to take daily calcium and vitamin D replacement, and if osteoporosis is present or osteopenia is worsening, bone strengthening agents should be administered.

THERAPY OF METASTATIC DISEASE

About 15–20% of patients treated for localized breast cancer develop metastatic disease in the subsequent decade after diagnosis.

TABLE 75-3 Surveillance Guidelines for Breast Cancer Patients after Primary and Adjuvant Therapy during Routine Follow-up

TEST	FREQUENCY
Recommended	
History; eliciting symptoms; physical examination	q3–6 months × 3 years; q6–12 months × 2 years; then annually
Breast self-examination	Monthly
Mammography	Annually
Pelvic examination	Annually as per age-appropriate guidelines (particularly for patients on SERMs)
Patient education about symptoms of recurrence	Ongoing
Coordination of care	Ongoing
Assessment of side effects if on endocrine therapy	Ongoing
Echocardiography if on trastuzumab	Every 3 months; discontinue when trastuzumab therapy complete
Not Recommended (if asymptomatic)	
Complete blood count	
Serum chemistry studies	
Chest radiographs	
Bone scans	
Ultrasound examination of the liver	
Computed tomography of chest, abdomen, or pelvis	
Tumor markers CA 15-3, CA 27-29, CEA, CTC	
Transvaginal endometrial ultrasonography	

Abbreviations: CA, cancer antigen; CEA, carcinoembryonic antigen; CTC, circulating tumor cell; SERM, selective estrogen receptor modulator.

Source: Recommended Breast Cancer Surveillance Guidelines, ASCO Education Book, Fall, 1997. From J Clin Oncol 15:2149, 1997; with permission.

Soft tissue, bony, and visceral (lung and liver) metastases all account for approximately one-third of sites of initial relapses. However, by the time of death, most patients will have bony involvement. Recurrences can appear at any time after primary therapy, but at least half occur >5 years after initial therapy. This observation is particularly true in patients with ER-positive disease, for whom the risk of distant recurrence remains constant for as long as 20 years and is the basis for recommendation of extended adjuvant endocrine therapy. It is now clear that a variety of host factors can influence recurrence rates, including depression and central obesity, and these diseases should be managed as aggressively as possible.

For patients with no prior history of metastases, a biopsy of suspicious physical or radiographic lesions should be performed, both for confirmation that the lesion does, indeed, represent recurrent cancer and to reevaluate ER and HER2, which can differ between the primary and metastatic lesions in up to 15% of cases. One should not assume that an apparent abnormality is a breast cancer metastasis. Many benign conditions, such as tuberculosis, gallstones, sarcoidosis, or other nonmalignant diseases, can mimic a recurrent breast cancer and are of course treated much differently.

Although treatable, metastatic disease is rarely if ever cured. The median survival for all patients diagnosed with metastatic breast cancer is <3 years, but with remarkable variability depending on intrinsic subtype and effective treatments. Patients with triple-negative metastatic breast cancer have the shortest expected survival, while those with ER-positive disease can expect to live the longest. HER2 positivity was initially found to be a very poor prognostic factor in metastatic breast cancer, but the availability of several effective treatments has improved the expected survival rates to at least those of ER-positive patients, if not better.

In the absence of cure, the overall goal of treatment of metastatic disease is palliation, or, put simply, to "keep the patient feeling as well as she can for as long as she can." A secondary goal is improved survival. It is important to point out that survival has not been improved by advocating more aggressive, or toxic, therapies, such as high-dose or combination chemotherapy, but rather by more selective and biologically based therapy, such as use of endocrine or anti-HER2 therapies in patients with ER- or HER2-positive breast cancers, respectively. Generally, a new treatment is continued until either progression or unacceptable toxicities are evident. These are both evaluated by serial history and physical examinations and periodic serologic evaluation for hematologic or hepatic abnormalities, as well as circulating tumor biomarker tests (assays for MUC1, such as CA15-3 or CA27.29, and for carcinoembryonic antigen or occasionally CA125). If all these evaluations fail to suggest progression, it is unlikely that imaging will contribute. However, if one or more of these suggest progression, whole-body imaging with either a PET/CT or a scintigraphic bone scan and dedicated CT are indicated. Brain imaging is not recommended unless the patient has some sort of central nervous system (CNS) symptom or finding.

The choice of therapy requires consideration of local therapy needs, specifically surgical approaches to particularly worrisome long-bone lytic lesions or isolated CNS metastases. New back pain in patients with breast cancer should be explored aggressively on an emergent basis; to wait for neurologic symptoms is a potentially catastrophic error. Metastatic involvement of endocrine organs can occasionally cause profound dysfunction, including adrenal insufficiency and hypopituitarism. Similarly, obstruction of the biliary tree or other impaired organ function may be better managed with a local therapy than with a systemic approach. Radiation as an adjunct to or instead of surgery is an important consideration for particularly symptomatic disease in long or vertebral bones, local-regional recurrences, and CNS metastases. In many cases, systemic therapy can be withheld while the patient is managed with appropriate local therapy.

There is no evidence that aggressive local treatment, such as excision; radiation; radiofrequency ablation; or cryotherapy of metastases to the lung, liver, or other distant sites, improves survival. Although appealing, these strategies are associated with increased toxicity and cost and should be reserved for palliation.

Selection of the systemic therapy strategy depends on the overall medical condition of the patient, the hormone receptor and HER2 status of the tumor, and clinical judgment. Because therapy of systemic disease is palliative, the potential toxicities of therapies should be balanced against expected response rates. Several variables influence the response to systemic therapy. For example, the presence of ER and PgR is a strong indication for endocrine therapy, even for patients with limited visceral (lung/liver) disease. On the other hand, patients with short disease-free intervals or rapidly progressive visceral disease (liver and lung) with end-organ dysfunction, such as lymphangitic pulmonary disease, are unlikely to respond to endocrine therapy.

Many patients with bone-only or bone-dominant disease have a relatively indolent course. Because the goal of therapy is to maintain well-being for as long as possible, emphasis should be placed on avoiding the most hazardous complications of metastatic disease, including pathologic fracture of the axial skeleton and spinal cord compression. Under such circumstances, systemic chemotherapy has a modest effect, whereas radiation therapy may be effective for long periods. Other systemic treatments, such as strontium-89, may provide a palliative benefit without inducing objective responses. Patients with bone involvement should receive concurrent bone strengthening agents, such as bisphosphonates or the humanized monoclonal anti-RANK ligand antibody, denosumab.

Many patients are inappropriately treated with toxic regimens into their last days of life. Often, oncologists are unwilling to have the difficult conversations that are required with patients nearing the end of life, and not uncommonly, patients and families can pressure physicians into treatments with very little survival value. Palliative care consultation and realistic assessment of treatment expectations need to be reviewed with patients and families. We urge consideration of palliative care consultations for patients who have received at least two lines of therapy for metastatic disease.

Endocrine Therapy ER-positive breast cancer will respond to endocrine therapy ~30–70% of the time. Potential endocrine therapies are summarized in Table 75-4. As in the adjuvant setting, one can choose among the SERM, tamoxifen, the AIs (anastrozole, letrozole, exemestane), or other strategies. Among the latter, the selective estrogen receptor downregulator (SERD), fulvestrant, has substantial activity. Early clinical studies with this drug were unexciting, but more recent studies have proven a very steep dose-response curve, and at higher levels (500 mg/month), it is as or more active than either tamoxifen or the AIs. Additive endocrine therapies, including treatment with progestins, and androgens, and enigmatically, pharmacologic doses of estrogens, are all active, but they may be associated with unacceptable side effects in many women. The mechanism of action of these latter therapies is unknown. Cases in which tumors shrink in response to tamoxifen withdrawal (as well as withdrawal of pharmacologic doses of estrogens) have been reported, but with the advent of so many other therapies for metastatic disease, this strategy is rarely used in modern oncology.

TABLE 75-4 Endocrine Therapies for Breast Cancer

THERAPY	COMMENTS
Castration Surgical LHRH agonists	For premenopausal women
Antiestrogens	
Tamoxifen	Useful in pre- and postmenopausal women ^a
Fulvestrant	Responses in tamoxifen-resistant and aromatase inhibitor-resistant patients ^a
Aromatase inhibitors	Low toxicity; now first choice for metastatic disease ^a
High-dose progestogens	Common fourth-line choice after aromatase inhibitors, tamoxifen, and fulvestrant
Additive androgens or estrogens	Plausible fourth-line therapies; potentially toxic

^aConsider retreatment with everolimus in combination for disease progression.

Abbreviation: LHRH, luteinizing hormone-releasing hormone.

TABLE 75-5 Common Agents Added to Endocrine Therapies for Metastatic Breast Cancer

CLASS	HOW ADMINISTERED	AGENTS	COMMON TOXICITIES
Anti-mTOR	Oral	Everolimus	Mucositis, diarrhea, rash
CDK4/6 inhibitors	Oral	Palbociclib, ribociclib, abemaciclib (not FDA approved)	Neutropenia; uncommon leukopenia, fatigue, and nausea

Abbreviations: CDK4/6, cyclin D kinase 4/6; FDA, Food and Drug Administration; mTOR, mammalian target of rapamycin.

The sequence of endocrine therapy is variable. Patients who respond to one endocrine therapy have at least a 50% chance of responding to a second endocrine therapy. It is not uncommon for patients to respond to two or three sequential endocrine therapies. In most postmenopausal patients, the initial endocrine therapy should be an AI rather than tamoxifen. As noted, AIs are not used in premenopausal women because their hypothalamus can respond to estrogen deprivation by producing gonadotropins that promote estrogen synthesis. Tamoxifen and fulvestrant are usually used in sequence after AI therapy. Combination endocrine therapies increase the chances of response initially, but they do not appear to increase the ultimate time to chemotherapy use or overall survival. Combinations of chemotherapy with endocrine therapy are not useful, as summarized in Table 75-4.

At least two different targeted agents have been shown to enhance outcomes of patients with ER-positive metastatic breast cancer when combined with endocrine therapy. Addition of an inhibitor of the mammalian target of rapamycin (mTOR), everolimus, to the hormonal treatment can lead to AIs, tamoxifen, or fulvestrant improves time to progression, and this agent is now being explored as front-line therapy and in the adjuvant setting. Likewise, inhibitors of cyclin D kinase 4/6 (CDK4/6) (palbociclib, ribociclib, abemaciclib) have also been shown to substantially improve progression-free survival when combined either with an AI or fulvestrant. These agents are also being tested in the adjuvant setting. Data regarding overall survival benefits from the mTOR or CDK4/6 inhibitors are still pending, but addition of one or the other in combination with ET for women with ER-positive metastatic breast cancer is becoming the standard of care. These should not be given simultaneously but rather in sequence as appropriate, as summarized in Table 75-5.

Chemotherapy Unlike many other epithelial malignancies, breast cancer responds to multiple chemotherapeutic agents, including anthracyclines, alkylating agents, taxanes, and antimetabolites. Multiple combinations of these agents have been found to improve response rates somewhat, but they have had little effect on duration of response or survival. Unless patients have rapidly progressive visceral (lung, liver) metastases with end-organ dysfunction, single-agent chemotherapy, used in sequence as one drug fails going on the next, is preferable. Given the significant toxicity of most drugs, the use of a single effective agent will minimize toxicity by sparing the patient exposure to drugs that would be of little value. No method to select the drugs most efficacious for a given patient has been demonstrated to be useful.

Most oncologists use either capecitabine or an anthracycline or a taxane for first-line chemotherapy, either in a patient with ER-positive disease that is refractory to endocrine therapy or for a patient with ER-negative breast cancer. Within these general classes, it is not clear that one particular agent (such as doxorubicin vs epirubicin or paclitaxel vs docetaxel) is preferable, and the choice has to be balanced with individual needs. Objective responses in previously treated patients may also be seen with gemcitabine, vinorelbine, and oral etoposide, as well as a new class of agents, epothilones. Platinum-based agents have become far more widely used in both the adjuvant and advanced disease settings for some breast cancers, particularly those of the “triple-negative” subtype.

Anti-HER2 therapy Treatment of patients with anti-HER2 metastatic breast cancer is one of the great success stories in the last

TABLE 75-6 Common Anti-HER2 Agents for Breast Cancer

CLASS	HOW ADMINISTERED	AGENTS	COMMON TOXICITIES
Humanized monoclonal antibodies	IV	Trastuzumab, pertuzumab	Cardiac dysfunction GI (diarrhea)
Tyrosine Kinase Inhibitors	Oral	Lapatinib neratinib (not FDA approved)	Diarrhea, mucositis, rash
Antibody-drug conjugate	IV	Ado-trastuzumab emtansine	Peripheral neuropathy, thrombocytopenia

30 years of oncology. Initial use of a trastuzumab, either alone or with chemotherapy, was shown to improve response rate and survival for women with HER2-positive disease. Indeed, anecdotal reports of a few patients with remarkably sustained complete responses suggest that, on occasion, a few may be cured. Chronologically, the tyrosine kinase, lapatinib, was subsequently shown to be effective when added to chemotherapy after patients progressed on prior trastuzumab. Further, both continuation of trastuzumab after progression, in combination with the next chemotherapeutic regimen and combination of trastuzumab and lapatinib in patients who had progressed on trastuzumab are both superior to discontinuing the trastuzumab.

For patients who have become refractory to trastuzumab-based therapy, and more recently even in the upfront setting, other therapies have remarkably high activity. A novel antibody drug conjugate (ADC) that links trastuzumab to a cytotoxic agent, ado-trastuzumab emtansine, is active even in patients who have progressed on trastuzumab. More recently, the combination of chemotherapy and trastuzumab and pertuzumab has been shown to result in prolonged overall survival compared to trastuzumab alone. These recommendations are summarized in Table 75-6.

Other Therapies Bevacizumab is an agent that targets the vascular endothelial growth factor (VEGF). Bevacizumab with paclitaxel or other chemotherapeutic agents modestly increases the response rate and response duration to paclitaxel, but without improvement in overall survival and with occasional major toxicities. After initial excitement and FDA approval, its use has been mostly abandoned in breast cancer. As in the metastatic setting, trials are ongoing testing the value of PARP (poly-ADP ribose polymerase) inhibitors in patients with known germline *BRCA1/2* mutations or cancers that have BRCA-like biologies. The excitement over immune check-point inhibitors has spread to metastatic breast cancer, especially of the triple-negative subtype, but at present there are no agents approved for it.

■ MALE BREAST CANCER

Breast cancer is ~1/150th as frequent in men as in women; ~2000 men developed breast cancer annually in the United States. Risk factors include inherited, deleterious SNPs in *BRCA2*, as well as Klinefelter's syndrome. Men with Klinefelter's syndrome have two or more copies of the X chromosome and have lower levels of and higher levels of estrogen. Other conditions of hyperestrogenism, such as in hepatic failure, are also associated with higher risk of male breast cancers. However, the vast majority of men who present with breast cancer have none of these conditions.

Breast cancer usually presents in men as a unilateral lump in the breast and is frequently not diagnosed promptly. Given the small amount of soft tissue and the unexpected nature of the problem, locally advanced presentations are somewhat more common. Although gynecomastia may initially be unilateral or asymmetric, any unilateral mass in a man aged >40 years should receive a careful workup including biopsy. On the other hand, bilateral symmetric breast development rarely represents breast cancer and is almost invariably due to endocrine disease or a drug effect. It should be kept in mind, nevertheless, that the risk of cancer is much greater in men with gynecomastia; in such men, gross asymmetry of the breasts should arouse suspicion of cancer.

Approximately 90% of male breast cancers contain ERs, and it behaves similarly to that in a postmenopausal woman. When matched to female breast cancer by age and stage, its overall prognosis is identical. Male breast cancer is best managed by mastectomy and axillary lymph node dissection or SLNB, although some men prefer breast-conserving therapy. Patients with locally advanced disease or positive nodes should also be treated with irradiation, and ~60% of cases with metastatic disease respond to endocrine therapy. Tamoxifen is usually the agent of choice, and it is unknown if the AIs are effective in men. No randomized studies have evaluated adjuvant therapy for male breast cancer. Two historic experiences suggest that the disease responds well to adjuvant systemic therapy, and, if not medically contraindicated, the same criteria for the use of adjuvant therapy in women should be applied to men.

The sites of relapse and spectrum of response to chemotherapeutic drugs are virtually identical for breast cancers in either sex.

FURTHER READING

- COLLEONI M et al: Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: Results from the International Breast Cancer Study Group Trials I to V. *J Clin Oncol* 34:927, 2016.
- EARLY BREAST CANCER TRIALISTS' COLLABORATIVE (EBCTCG) et al: Adjuvant bisphosphonate treatment in early breast cancer: Meta-analyses of individual patient data from randomised trials. *Lancet* 386:1353, 2015.
- EARLY BREAST CANCER TRIALISTS' COLLABORATIVE (EBCTCG) et al: Aromatase inhibitors versus tamoxifen in early breast cancer: Patient-level meta-analysis of the randomised trials. *Lancet* 386:1341, 2015.
- EASTON DF et al: Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 372:2243, 2015.
- LE TOURNEAU C et al: Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): A multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol* 16:1324, 2015.
- OEFFINGER KC et al: Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. *JAMA* 314:1599, 2015.
- PEREZ EA et al: Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: Planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *J Clin Oncol* 32:3744, 2014.
- PETO R et al: Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* 379:432, 2012.
- RUNOWICZ CD et al: American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. *J Clin Oncol* 34:611–35, 2016.
- SPARANO JA et al: Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med* 373:2005, 2015.

adenocarcinomas; the two histologic subtypes have a similar clinical presentation but different causative factors.

Worldwide, squamous cell carcinoma is the more common cell type, having an incidence that rises strikingly in association with geographic location. It occurs frequently within a region extending from the southern shore of the Caspian Sea on the west to northern China on the east, encompassing parts of Iran, central Asia, Afghanistan, Siberia, and Mongolia. Familial increased risk has been observed in regions with high incidence, although gene associations are not yet defined. High-incidence "pockets" of the disease are also present in such disparate locations as Finland, Iceland, Curaçao, southeastern Africa, and northwestern France. In North America and western Europe, the disease is more common in blacks than whites and in males than females; it appears most often after age 50 and seems to be associated with a lower socioeconomic status. Such cancers generally arise in the cervical and thoracic portions of the esophagus.

A variety of causative factors have been implicated in the development of squamous cell cancers of the esophagus (Table 76-1). In the United States, the etiology of such cancers is primarily related to excess alcohol consumption and/or cigarette smoking. The relative risk increases with the amount of tobacco smoked or alcohol consumed, with these factors acting synergistically. The consumption of whiskey is linked to a higher incidence than the consumption of wine or beer. Squamous cell esophageal carcinoma has also been associated with the ingestion of nitrates, smoked opiates, and fungal toxins in pickled vegetables, as well as mucosal damage caused by such physical insults as long-term exposure to extremely hot tea, the ingestion of lye, radiation-induced strictures, and chronic achalasia. The presence of an esophageal web in association with glossitis and iron deficiency (i.e., Plummer-Vinson or Paterson-Kelly syndrome) and congenital hyperkeratosis and pitting of the palms and soles (i.e., tylosis palmaris et plantaris) have each been linked with squamous cell esophageal cancer, as have dietary deficiencies of molybdenum, zinc, selenium, and vitamin A. Patients with head and neck cancer are at increased risk of squamous cell cancer of the esophagus.

For unclear reasons, the incidence of squamous cell esophageal cancer has decreased somewhat in both the black and white populations in the United States over the past 40 years, whereas the rate of adenocarcinoma has risen sevenfold, particularly in white males (male-to-female ratio of 6:1). Whereas squamous cell cancers comprised the vast majority of esophageal cancers in the United States as recently as 40–50 years ago, >75% of esophageal tumors are now adenocarcinomas, with the incidence of this histologic subtype continuing to increase rapidly. Understanding the cause for this increase is the focus of current investigation.

Several strong etiologic associations have been observed to account for the development of adenocarcinoma of the esophagus (Table 76-2).

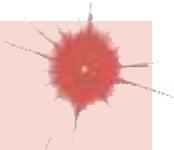
TABLE 76-1 Some Etiologic Factors Associated with Squamous Cell Cancer of the Esophagus

Excess alcohol consumption
Cigarette smoking
Other ingested carcinogens
Nitrates (converted to nitrites)
Smoked opiates
Fungal toxins in pickled vegetables
Mucosal damage from physical agents
Hot tea
Lye ingestion
Radiation-induced strictures
Chronic achalasia
Host susceptibility
Esophageal web with glossitis and iron deficiency (i.e., Plummer-Vinson or Paterson-Kelly syndrome)
Congenital hyperkeratosis and pitting of the palms and soles (i.e., tylosis palmaris et plantaris)
? Dietary deficiencies of selenium, molybdenum, zinc, and vitamin A

76

Upper Gastrointestinal Tract Cancers

Robert J. Mayer



Upper gastrointestinal cancers include malignancies arising in the esophagus, stomach, and small intestine.

ESOPHAGEAL CANCER

INCIDENCE AND ETIOLOGY

 Cancer of the esophagus is an increasingly common and extremely lethal malignancy. The diagnosis was made in 16,940 Americans in 2017 and led to 15,690 deaths. Almost all esophageal cancers are either squamous cell carcinomas or

TABLE 76-2 Some Etiologic Factors Associated with Adenocarcinoma of the Esophagus

Chronic gastroesophageal reflux
Obesity
Barrett's esophagus
Male sex
Cigarette smoking

Such tumors arise in the distal esophagus in association with chronic gastric reflux, often in the presence of Barrett's esophagus (replacement of the normal squamous epithelium of the distal esophagus by columnar mucosa), which occurs more commonly in obese individuals. Adenocarcinomas arise within dysplastic columnar epithelium in the distal esophagus. Even before frank neoplasia is detectable, aneuploidy and *p53* mutations are found in the dysplastic epithelium. The value of proton-pump inhibitors in reducing the risk of cancer in individuals with chronic gastric reflux or Barrett's esophagus is uncertain. These adenocarcinomas behave clinically like gastric adenocarcinomas, although they are not associated with *Helicobacter pylori* infections. Approximately 15% of esophageal adenocarcinomas overexpress the *HER2/neu* gene.

■ CLINICAL FEATURES

About 5% of esophageal cancers occur in the upper third of the esophagus (cervical esophagus), 20% in the middle third, and 75% in the lower third. Squamous cell carcinomas and adenocarcinomas cannot be distinguished radiographically or endoscopically.

Progressive dysphagia and weight loss of short duration are the initial symptoms in the vast majority of patients. Dysphagia initially occurs with solid foods and gradually progresses to include semisolids and liquids. By the time these symptoms develop, the disease is already very advanced, because difficulty in swallowing does not occur until >60% of the esophageal circumference is infiltrated with cancer. Dysphagia may be associated with pain on swallowing (odynophagia), pain radiating to the chest and/or back, regurgitation or vomiting, and aspiration pneumonia. The disease most commonly spreads to adjacent and supraclavicular lymph nodes, liver, lungs, pleura, and bone. Tracheoesophageal fistulas may develop, primarily in patients with upper and mid-esophageal tumors. As with other squamous cell carcinomas, hypercalcemia may occur in the absence of osseous metastases, probably from parathormone-related peptide secreted by tumor cells (Chap. 89).

■ DIAGNOSIS

Attempts at endoscopic and cytologic screening for carcinoma in patients with Barrett's esophagus, while effective as a means of detecting high-grade dysplasia, have not yet been shown to reduce the likelihood of death from esophageal adenocarcinoma. Esophagoscopy should be performed in all patients suspected of having an esophageal abnormality, to both visualize and identify a tumor and also to obtain histopathologic confirmation of the diagnosis. Because the population of persons at risk for squamous cell carcinoma of the esophagus (i.e., smokers and drinkers) also has a high rate of cancers of the lung and the head and neck region, endoscopic inspection of the larynx, trachea, and bronchi should also be carried out. A thorough examination of the fundus of the stomach (by retroflexing the endoscope) is imperative as well. The extent of tumor spread to the mediastinum and para-aortic lymph nodes should be assessed by computed tomography (CT) scans of the chest and abdomen and by endoscopic ultrasound. Positron emission tomography scanning provides a useful assessment of the presence of distant metastatic disease, offering accurate information regarding spread to mediastinal lymph nodes, which can be helpful in defining radiation therapy fields. Such scans, when performed sequentially, appear to provide a means of making an early assessment of responsiveness to preoperative chemotherapy and have increasingly been used in guiding a change in clinical management.

TREATMENT

Esophageal Cancer

The prognosis for patients with esophageal carcinoma is poor. Approximately 10% of patients survive 5 years after the diagnosis; thus, management focuses on symptom control. Surgical resection of all gross tumor (i.e., total resection) is feasible in only 45% of cases, with residual tumor cells frequently present at the resection margins. Such esophagectomies have been associated with a postoperative mortality rate of ~5% due to anastomotic fistulas, subphrenic abscesses, and cardiopulmonary complications. Although debate regarding the comparative benefits of transthoracic versus transhiatal resections has continued, experienced thoracic surgeons are now favoring minimally invasive transthoracic esophagectomies. Endoscopic resections of superficial squamous cell cancers or adenocarcinomas are being examined but have not yet been shown to result in a similar likelihood of survival as observed with conventional surgical procedures. Similarly, the value of endoscopic ablation of dysplastic lesions in an area of Barrett's esophagus on reducing subsequent mortality from esophageal carcinoma is uncertain. Some experts have advocated fundoplication surgery (i.e., the removal of the gastroesophageal junction) as a means of cancer prevention in patients with Barrett's esophagus; again, objective data are not yet available to fully assess the risks versus benefits of this invasive procedure. About 20% of patients who survive a total surgical resection live for 5 years. The evaluation of chemotherapeutic agents in patients with esophageal carcinoma has been hampered by ambiguity in the definition of "response" and the debilitated physical condition of many treated individuals, particularly those with squamous cell cancers. Nonetheless, significant reductions in the size of measurable tumor masses have been reported in 15–25% of patients given single-agent treatment and in 30–60% of patients treated with drug combinations that include a platinum form of chemotherapy. In the small subset of patients whose tumors overexpress the *HER2/neu* gene, the addition of the monoclonal antibody trastuzumab (Herceptin) appears to further enhance the likelihood of benefit, particularly in patients with gastroesophageal lesions. The use of the antiangiogenic agent bevacizumab (Avastin) seems to be of limited value in the setting of esophageal cancer. Combination chemotherapy and radiation therapy as the initial therapeutic approach, either alone or followed by an attempt at operative resection, seems to be beneficial. When administered along with radiation therapy, chemotherapy produces a better survival outcome than radiation therapy alone. The use of preoperative chemotherapy and radiation therapy followed by esophageal resection appears to prolong survival compared with surgery alone according to several randomized trials and a meta-analysis; some reports suggest that no additional benefit accrues when surgery is added if significant shrinkage of tumor has been achieved by the chemoradiation combination.

For the incurable, surgically unresectable patient with esophageal cancer, dysphagia, malnutrition, and the management of tracheoesophageal fistulas are major issues. Approaches to palliation include repeated endoscopic dilatation, the surgical placement of a gastrostomy or jejunostomy for hydration and feeding, endoscopic placement of an expansive metal stent to bypass the tumor, and radiation therapy.

TUMORS OF THE STOMACH

■ GASTRIC ADENOCARCINOMA

 **Incidence and Epidemiology** For unclear reasons, the incidence and mortality rates for gastric cancer have decreased in the United States during the past 80 years, although the disease remains the third most frequent cause of worldwide cancer-related death. The mortality rate from gastric cancer in the United States has dropped in men from 28 to 7.4 per 100,000 persons, whereas in women, the rate has decreased from 27 to 2.4 per 100,000. Nonetheless,

in 2017, 28,000 new cases of stomach cancer were diagnosed in the United States, and 10,960 Americans died of the disease. Although the incidence of gastric cancer has decreased worldwide, it remains high in such disparate geographic regions as Japan, China, Chile, and Ireland.

The risk of gastric cancer is greater among lower socioeconomic classes. Migrants from high- to low-incidence nations maintain their susceptibility to gastric cancer, whereas the risk for their offspring approximates that of the new homeland. These findings suggest that an environmental exposure, probably beginning early in life, is related to the development of gastric cancer, with dietary carcinogens considered the most likely factor(s).

Pathology About 85% of stomach cancers are adenocarcinomas, with 15% due to lymphomas, gastrointestinal stromal tumors (GISTs), and leiomyosarcomas. Gastric adenocarcinomas may be subdivided into two pathologically defined categories: a *diffuse type*, in which cell cohesion is absent, so that individual cells infiltrate and thicken the stomach wall without forming a discrete mass; and an *intestinal type*, characterized by cohesive neoplastic cells that form glandlike tubular structures. The diffuse carcinomas occur more often in younger patients, develop throughout the stomach (including the cardia), result in a loss of distensibility of the gastric wall (so-called *linitis plastica*, or “leather bottle” appearance), and carry a poorer prognosis. Diffuse cancers have defective intercellular adhesion, mainly as a consequence of loss of expression of E-cadherin. Intestinal-type lesions are frequently ulcerative, more commonly appear in the antrum and lesser curvature of the stomach, and are often preceded by a prolonged precancerous process, often initiated by *H. pylori* infection. Although the incidence of diffuse carcinomas is similar in most populations, the intestinal type tends to predominate in the high-risk geographic regions and is less likely to be found in areas where the frequency of gastric cancer is declining. Thus, different etiologic factor(s) are likely involved in these two subtypes. In the United States, ~30% of gastric cancers originate in the distal stomach, ~20% arise in the midportion of the stomach, and ~40% originate in the proximal third of the stomach. The remaining 10% involve the entire stomach.

Genomic profiling of gastric adenocarcinomas has led to subdividing the disease into four molecularly defined subgroups: chromosomally unstable tumors (50% of cases correlating with intestinal type histology), genetically stable tumors (20% of cases correlating with diffuse type histology), microsatellite unstable tumors (22% of cases), and Epstein-Barr virus (EBV) positive tumors (9% of cases) (Fig. 76-1). Efforts to incorporate these molecular subtypes into clinical management are underway.

Etiology The long-term ingestion of high concentrations of nitrates found in dried, smoked, and salted foods appears to be associated with a higher risk. The nitrates are thought to be converted to carcinogenic nitrites by bacteria (Table 76-3). Such bacteria may be introduced exogenously through the ingestion of partially decayed foods, which are consumed in abundance worldwide by the lower socioeconomic classes. Bacteria such as *H. pylori* may also contribute to this effect by causing chronic inflammatory atrophic gastritis, loss of gastric acidity, and bacterial growth in the stomach. Although the risk for developing gastric cancer is thought to be sixfold higher in people infected with *H. pylori*, it remains uncertain whether eradicating the bacteria after infection has already occurred actually reduces this risk. Loss of acidity may occur when acid-producing cells of the gastric antrum have been removed surgically to control benign peptic ulcer disease or when achlorhydria, atrophic gastritis,

TABLE 76-3 Nitrate-Converting Bacteria as a Factor in the Causation of Gastric Carcinoma

Exogenous sources of nitrate-converting bacteria:

Bacterially contaminated food (common in lower socioeconomic classes, who have a higher incidence of the disease; diminished by improved food preservation and refrigeration)

Helicobacter pylori infection

Endogenous factors favoring growth of nitrate-converting bacteria in the stomach:

Decreased gastric acidity

Prior gastric surgery (antrectomy) (15- to 20-year latency period)

Atrophic gastritis and/or pernicious anemia

? Prolonged exposure to histamine H₂-receptor antagonists

^aHypothesis: Dietary nitrates are converted to carcinogenic nitrites by bacteria.

and even pernicious anemia develop in the elderly. Serial endoscopic examinations of the stomach in patients with atrophic gastritis have documented replacement of the usual gastric mucosa by intestinal-type cells. This process of intestinal metaplasia may lead to cellular atypia and eventual neoplasia. Because the declining incidence of gastric cancer in the United States primarily reflects a decline in distal, ulcerating, intestinal-type lesions, it is conceivable that better food preservation and the availability of refrigeration for all socioeconomic classes have decreased the dietary ingestion of exogenous bacteria. *H. pylori* has not been associated with the diffuse, more proximal form of gastric carcinoma or with cancers arising at the gastroesophageal junction or in the distal esophagus. Approximately 10–15% of adenocarcinomas appearing in the proximal stomach, the gastroesophageal junction, and the distal esophagus overexpress the *HER2/neu* gene; individuals whose tumors demonstrate this overexpression benefit from treatment directed against this target (i.e., trastuzumab [Herceptin]).

Several additional etiologic factors have been associated with gastric carcinoma. Gastric ulcers and adenomatous polyps have occasionally been linked, but data on a cause-and-effect relationship are unconvincing. The inadequate clinical distinction between benign gastric ulcers and small ulcerating carcinomas may, in part, account for this presumed association. The presence of extreme hypertrophy of gastric rugal folds (i.e., Ménétrier’s disease), giving the impression of polypoid

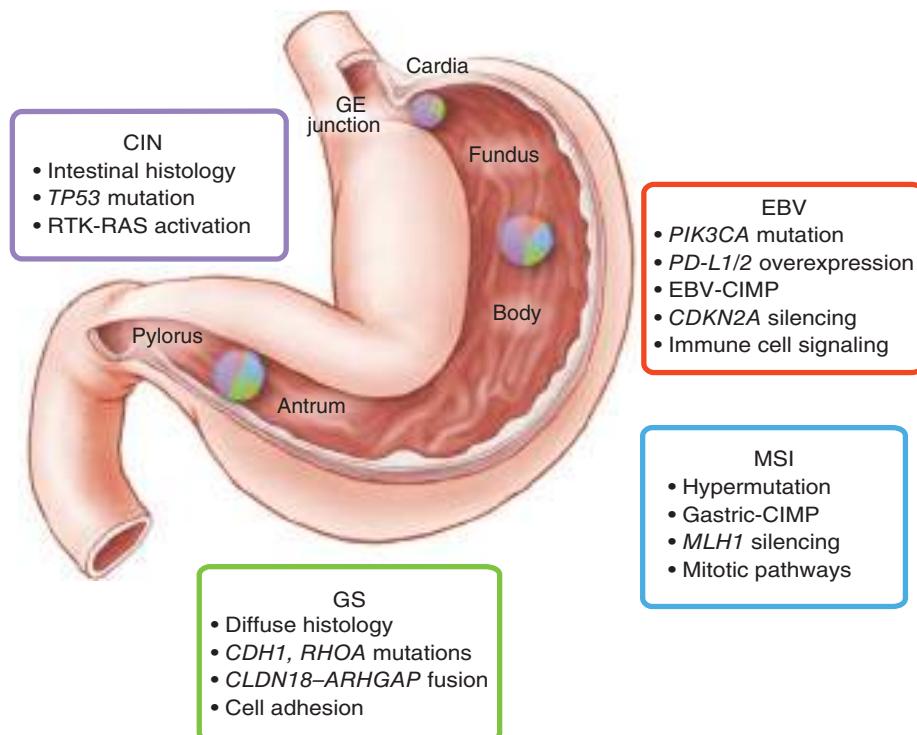


FIGURE 76-1 Molecular/genomic characterization of subtypes of gastric carcinomas. CIMP, CpG-island methylator phenotype; CIN, chromosomally unstable; EBV, Epstein-Barr virus-associated; GS, genetically stable; MSI, microsatellite instability-associated.

lesions, has been associated with a striking frequency of malignant transformation; such hypertrophy, however, does not represent the presence of true adenomatous polyps. Individuals with blood group A have a higher incidence of gastric cancer than persons with blood group O; this observation may be related to differences in the mucous secretion, leading to altered mucosal protection from carcinogens. A germline mutation in the E-cadherin gene (*CDH1*), inherited in an autosomal dominant pattern and coding for a cell adhesion protein, has been linked to a high incidence of occult diffuse-type gastric cancers in young asymptomatic carriers in whom the endoscopic appearance of the gastric mucosa appears normal but foci of tumor are frequently present deeper in the stomach wall; this observation has led to a recommendation that they undergo a prophylactic gastrectomy. Carriers of this mutation are also at greater risk for the development of lobular breast cancer. Duodenal ulcers are not associated with gastric cancer.

Clinical Features Gastric cancers, when superficial and surgically curable, usually produce no symptoms. As the tumor becomes more extensive, patients may complain of an insidious upper abdominal discomfort varying in intensity from a vague, postprandial fullness to a severe, steady pain. Anorexia, often with slight nausea, is very common but is not the usual presenting complaint. Weight loss may eventually be observed, and nausea and vomiting are particularly prominent in patients whose tumors involve the pylorus; dysphagia and early satiety may be the major symptoms caused by diffuse lesions originating in the cardia. There may be no early physical signs. A palpable abdominal mass indicates long-standing growth and predicts regional extension. Gastric carcinomas spread by direct extension through the gastric wall to the perigastric tissues, occasionally adhering to adjacent organs such as the pancreas, colon, or liver. The disease also spreads via lymphatics or by seeding of peritoneal surfaces. Metastases to intraabdominal and supraclavicular lymph nodes occur frequently, as do metastatic nodules to the ovary (Krukenberg's tumor), periumbilical region ("Sister Mary Joseph node"), or peritoneal cul-de-sac (Blumer's shelf palpable on rectal or vaginal examination); malignant ascites may also develop. The liver is the most common site for hematogenous spread of tumor.

The presence of iron-deficiency anemia in men and of occult blood in the stool in both sexes mandates a search for an occult gastrointestinal tract lesion. A careful assessment is of particular importance in patients with atrophic gastritis or pernicious anemia. Unusual clinical features associated with gastric adenocarcinomas include migratory thrombophlebitis, microangiopathic hemolytic anemia, diffuse seborrheic keratoses (so-called Leser-Trélat sign), and acanthosis nigricans.

Diagnosis The use of double-contrast radiographic examinations has been supplanted by esophagogastroduodenoscopy and CT scanning for the evaluation of patients with epigastric complaints.

Gastric ulcers identified at the time of such endoscopic procedure may appear benign but merit biopsy in order to exclude a malignancy. Malignant gastric ulcers must be recognized before they penetrate into surrounding tissues, because the rate of cure of early lesions limited to the mucosa or submucosa is >80%. Because gastric carcinomas are difficult to distinguish clinically or endoscopically from gastric lymphomas, endoscopic biopsies should be made as deeply as possible, due to the submucosal location of lymphoid tumors.

The clinical staging system for gastric carcinoma is shown in Table 76-4.

TREATMENT

Gastric Adenocarcinoma

Complete surgical removal of the tumor with resection of adjacent lymph nodes offers the only chance for cure. However, this is possible in less than a third of patients. A subtotal gastrectomy is the treatment of choice for patients with distal carcinomas, whereas total or near-total gastrectomies are required for more proximal tumors. The inclusion of extended lymph node dissection in these procedures appears to confer an added risk for complications without providing a meaningful enhancement in survival. The prognosis following complete surgical resection depends on the degree of

TABLE 76-4 Staging System for Gastric Carcinoma

STAGE	TNM	FEATURES	DATA FROM ACS IN THE UNITED STATES	
			NO. OF CASES, %	5-YEAR SURVIVAL, %
0	T _{is} NOMO	Node negative; limited to mucosa	1	90
IA	T1NOMO	Node negative; invasion of lamina propria or submucosa	7	59
IB	T2NOMO T1N1MO	Node negative; invasion of muscularis propria	10	44
II	T1N2MO T2N1MO T3N0MO	Node positive; invasion beyond mucosa but within wall or Node negative; extension through wall	17	29
IIIA	T2N2MO T3N1-2MO	Node positive; invasion of muscularis propria or through wall	21	15
IIIB	T4N0-1MO	Node negative; adherence to surrounding tissue	14	9
IIIC	T4N2-3MO T3N3MO	>3 nodes positive; invasion of serosa or adjacent structures 7 or more positive nodes; penetrates wall without invading serosa or adjacent structures		
IV	T4N2MO T1-4N0-2-M1	Node positive; adherence to surrounding tissue or Distant metastases	30	3

Abbreviations: ACS, American Cancer Society; TNM, tumor, node, metastasis.

tumor penetration into the stomach wall and is adversely influenced by regional lymph node involvement and vascular invasion, characteristics found in the vast majority of American patients. As a result, the probability of survival after 5 years for the 25–30% of patients able to undergo complete resection is ~20% for distal tumors and <10% for proximal tumors, with recurrences continuing for at least 8 years after surgery. In the absence of ascites or extensive hepatic or peritoneal metastases, even patients whose disease is believed to be incurable by surgery should be offered resection of the primary lesion. Reduction of tumor bulk is the best form of palliation and may enhance the probability of benefit from subsequent therapy. In high-incidence regions such as Japan and Korea, where the use of endoscopic screening programs has identified patients with superficial tumors, the use of laparoscopic gastrectomy has gained popularity. In the United States and western Europe, the use of this less invasive surgical approach remains investigational.

Gastric adenocarcinoma is a relatively radioresistant tumor, and the adequate control of the primary tumor requires doses of external-beam irradiation that exceed the tolerance of surrounding structures, such as bowel mucosa and spinal cord. As a result, the major role of radiation therapy in patients has been palliation of pain. Radiation therapy alone after a complete resection does not prolong survival. In the setting of surgically unresectable disease limited to the epigastrium, patients treated with 3500–4000 cGy did not live longer than similar patients not receiving radiotherapy; however, survival was prolonged slightly when 5-fluorouracil (5-FU) plus leucovorin was given in combination with radiation therapy (3-year survival 50% vs 41% for radiation therapy alone). In this clinical setting, the 5-FU likely functions as a radiosensitizer.

The administration of combinations of cytotoxic drugs to patients with advanced gastric carcinoma has been associated with partial

responses in 30–50% of cases; responders appear to benefit from treatment. Such drug combinations have generally included cisplatin combined with epirubicin or docetaxel and infusional 5-FU or capecitabine, or with either irinotecan or oxaliplatin. Despite the encouraging response rates, complete remissions are uncommon, the partial responses are transient, and the overall impact of multidrug therapy on survival has been limited; the median survival time for patients treated in this manner remains less than 12 months. As with adenocarcinomas arising in the esophagus, the addition of bevacizumab (Avastin) to chemotherapy regimens in treating gastric cancer appears to provide limited benefit. However, preliminary results utilizing another antiangiogenic compound—ramucirumab (Cyranza)—in the treatment of gastric cancer are encouraging, particularly when combined with paclitaxel. Additionally, initial experiences with checkpoint inhibitors (PD-1 and PD-2) have shown such immunotherapy to provide benefit to some patients. The administration of adjuvant chemotherapy alone following the complete resection of a gastric cancer has only minimally improved survival. However, combination chemotherapy administered before and after surgery (*perioperative treatment*) as well as postoperative chemotherapy combined with radiation therapy reduces the recurrence rate and prolongs survival.

PRIMARY GASTRIC LYMPHOMA

Primary lymphoma of the stomach is relatively uncommon, accounting for <15% of gastric malignancies and ~2% of all lymphomas. The stomach is, however, the most frequent extranodal site for lymphoma, and gastric lymphoma has increased in frequency during the past 35 years. The disease is difficult to distinguish clinically from gastric adenocarcinoma; both tumors are most often detected during the sixth decade of life; present with epigastric pain, early satiety, and generalized fatigue; and are usually characterized by ulcerations with a ragged, thickened mucosal pattern demonstrated by contrast radiographs or endoscopic appearance. The diagnosis of lymphoma of the stomach may occasionally be made through cytologic brushings of the gastric mucosa but usually requires a biopsy at gastroscopy or laparotomy. Failure of gastroscopic biopsies to detect lymphoma in a given case should not be interpreted as being conclusive, because superficial biopsies may miss the deeper lymphoid infiltrate. The macroscopic pathology of gastric lymphoma may also mimic adenocarcinoma, consisting of either a bulky ulcerated lesion localized in the corpus or antrum or a diffuse process spreading throughout the entire gastric submucosa and even extending into the duodenum. Microscopically, the vast majority of gastric lymphoid tumors are lymphomas of B-cell origin. Histologically, these tumors may range from well-differentiated, superficial processes (mucosa-associated lymphoid tissue [MALT]) to high-grade, large-cell lymphomas. Like gastric adenocarcinoma, infection with *H. pylori* increases the risk for gastric lymphoma in general and MALT lymphomas in particular. Large-cell lymphomas of the stomach spread initially to regional lymph nodes (often to Waldeyer's ring) and may then disseminate.

TREATMENT

Primary Gastric Lymphoma

Primary gastric lymphoma is a far more treatable disease than adenocarcinoma of the stomach, a fact that underscores the need for making the correct diagnosis. Antibiotic treatment to eradicate *H. pylori* infection has led to regression of about 75% of gastric MALT lymphomas and should be considered before surgery, radiation therapy, or chemotherapy is undertaken in patients having such tumors. A lack of response to such antimicrobial treatment has been linked to a specific chromosomal abnormality, i.e., t(11;18). Responding patients should undergo periodic endoscopic surveillance because it remains unclear whether the neoplastic clone is eliminated or merely suppressed, although the response to antimicrobial treatment is quite durable. Subtotal gastrectomy, usually followed by combination chemotherapy, has led to 5-year survival

rates of 40–60% in patients with localized high-grade lymphomas. The need for a major surgical procedure has been questioned, particularly in patients with preoperative radiographic evidence of nodal involvement, for whom chemotherapy (CHOP [cyclophosphamide, doxorubicin, vincristine, and prednisone]) plus rituximab is highly effective therapy. A role for radiation therapy is not defined because most recurrences develop at distant sites.

GASTRIC (NONLYMPHOID) SARCOMA

Leiomyosarcomas and GISTs make up 1–3% of gastric neoplasms. They most frequently involve the anterior and posterior walls of the gastric fundus and often ulcerate and bleed. Even those lesions that appear benign on histologic examination may behave in a malignant fashion. These tumors rarely invade adjacent viscera and characteristically do not metastasize to lymph nodes, but they may spread to the liver and lungs. The treatment of choice is surgical resection with 3 years of postoperative therapy to be considered following the removal of a GIST if the primary tumor demonstrates high-risk features. All such tumors should be analyzed for a mutation in the c-kit receptor. GISTs are unresponsive to conventional chemotherapy; yet ~50% of patients experience objective response and prolonged survival when treated with imatinib mesylate (Gleevec) (400–800 mg PO daily), a selective inhibitor of the c-kit tyrosine kinase. Many patients with GIST whose tumors have become refractory to imatinib subsequently benefit from sunitinib (Sutent) or regorafenib (Stivarga), other inhibitors of the c-kit tyrosine kinase.

TUMORS OF THE SMALL INTESTINE

Small-bowel tumors comprise <3% of gastrointestinal neoplasms. Because of their rarity and inaccessibility, a correct diagnosis is often delayed. Abdominal symptoms are usually vague and poorly defined, and conventional radiographic studies of the upper and lower intestinal tract often appear normal. Small-bowel tumors should be considered in the differential diagnosis in the following situations: (1) recurrent, unexplained episodes of crampy abdominal pain; (2) intermittent bouts of intestinal obstruction, especially in the absence of inflammatory bowel disease (IBD) or prior abdominal surgery; (3) intussusception in the adult; and (4) evidence of chronic intestinal bleeding in the presence of negative conventional and endoscopic examination. A careful small-bowel barium study should be considered in such a circumstance; the diagnostic accuracy may be improved by infusing barium through a nasogastric tube placed into the duodenum (enteroclysis). Alternatively, capsule endoscopic procedures have been used.

BENIGN TUMORS

The histology of benign small-bowel tumors is difficult to predict on clinical and radiologic grounds alone. The symptomatology of benign tumors is not distinctive, with pain, obstruction, and hemorrhage being the most frequent symptoms. These tumors are usually discovered during the fifth and sixth decades of life, more often in the distal rather than the proximal small intestine. The most common benign tumors are adenomas, leiomyomas, lipomas, and angiomas.

Adenomas These tumors include those of the islet cells and Brunner's glands as well as polypoid adenomas. *Islet cell adenomas* are occasionally located outside the pancreas; the associated syndromes are discussed in Chap. 80. *Brunner's gland adenomas* are not truly neoplastic but represent a hypertrophy or hyperplasia of submucosal duodenal glands. These appear as small nodules in the duodenal mucosa that secrete a highly viscous alkaline mucus. Most often, this is an incidental radiographic finding not associated with any specific clinical disorder.

Polypoid Adenomas About 25% of benign small-bowel tumors are polypoid adenomas (see Table 77-2). They may present as single polypoid lesions or, less commonly, as papillary villous adenomas. As in the colon, the sessile or papillary form of the tumor is sometimes associated with a coexisting carcinoma. Occasionally, patients with Gardner's syndrome develop premalignant adenomas in the small bowel; such lesions are generally in the duodenum. Multiple polypoid tumors may occur throughout the small bowel (and occasionally the

stomach and colorectum) in the Peutz-Jeghers syndrome. The polyps are usually hamartomas (juvenile polyps) having a low potential for malignant degeneration. Mucocutaneous melanin deposits as well as tumors of the ovary, breast, pancreas, and endometrium are also associated with this autosomal dominant condition.

Leiomyomas These neoplasms arise from smooth-muscle components of the intestine and are usually intramural, affecting the overlying mucosa. Ulceration of the mucosa may cause gastrointestinal hemorrhage of varying severity. Cramping or intermittent abdominal pain is frequently encountered.

Lipomas These tumors occur with greatest frequency in the distal ileum and at the ileocecal valve. They have a characteristic radiolucent appearance and are usually intramural and asymptomatic, but on occasion cause bleeding.

Angiomas While not true neoplasms, these lesions are important because they frequently cause intestinal bleeding. They may take the form of telangiectasia or hemangiomas. Multiple intestinal telangiectasias occur in a nonhereditary form confined to the gastrointestinal tract or as part of the hereditary Osler-Rendu-Weber syndrome. Vascular tumors may also take the form of isolated hemangiomas, most commonly in the jejunum. Angiography, especially during bleeding, is the best procedure for evaluating these lesions.

MALIGNANT TUMORS

While rare, small-bowel malignancies occur in patients with long-standing regional enteritis and celiac sprue as well as in individuals with AIDS. Malignant tumors of the small bowel are frequently associated with fever, weight loss, anorexia, bleeding, and a palpable abdominal mass. After ampullary carcinomas (many of which arise from biliary or pancreatic ducts), the most frequently occurring small-bowel malignancies are adenocarcinomas, lymphomas, carcinoid tumors, and leiomyosarcomas.

■ ADENOCARCINOMAS

The most common primary cancers of the small bowel are adenocarcinomas, accounting for ~50% of malignant tumors. These cancers occur most often in the distal duodenum and proximal jejunum, where they tend to ulcerate and cause hemorrhage or obstruction. Radiologically, they may be confused with chronic duodenal ulcer disease or with Crohn's disease if the patient has long-standing regional enteritis. The diagnosis is best made by endoscopy and biopsy under direct vision. Surgical resection is the treatment of choice with suggested postoperative adjuvant chemotherapy options generally following treatment patterns used in the management of colon cancer.

■ LYMPHOMAS

Lymphoma in the small bowel may be primary or secondary. A diagnosis of a primary intestinal lymphoma requires histologic confirmation in a clinical setting in which palpable adenopathy and hepatosplenomegaly are absent and no evidence of lymphoma is seen on chest radiograph, CT scan, or peripheral blood smear or on bone marrow aspiration and biopsy. Symptoms referable to the small bowel are present, usually accompanied by an anatomically discernible lesion. Secondary lymphoma of the small bowel consists of involvement of the intestine by a lymphoid malignancy extending from involved retroperitoneal or mesenteric lymph nodes ([Chap. 104](#)).

Primary intestinal lymphoma accounts for ~20% of malignancies of the small bowel. These neoplasms are non-Hodgkin's lymphomas; they usually have a diffuse, large-cell histology and are of T cell origin. Intestinal lymphoma involves the ileum, jejunum, and duodenum, in decreasing frequency—a pattern that mirrors the relative amount of normal lymphoid cells in these anatomic areas. The risk of small-bowel lymphoma is increased in patients with a prior history of malabsorptive conditions (e.g., celiac sprue), regional enteritis, and depressed immune function due to congenital immunodeficiency syndromes, prior organ transplantation, autoimmune disorders, or AIDS.

The development of localized or nodular masses that narrow the lumen results in periumbilical pain (made worse by eating) as well as

weight loss, vomiting, and occasional intestinal obstruction. The diagnosis of small-bowel lymphoma may be suspected from the appearance on contrast radiographs of patterns such as infiltration and thickening of mucosal folds, mucosal nodules, areas of irregular ulceration, or stasis of contrast material. The diagnosis can be confirmed by surgical exploration and resection of involved segments. Intestinal lymphoma can occasionally be diagnosed by peroral intestinal mucosal biopsy, but because the disease mainly involves the lamina propria, full-thickness surgical biopsies are usually required.

Resection of the tumor constitutes the initial treatment modality. While postoperative radiation therapy has been given to some patients following a total resection, most authorities favor short-term (three cycles) systemic treatment with combination chemotherapy. The frequent presence of widespread intraabdominal disease at the time of diagnosis and the occasional multicentricity of the tumor often make a total resection impossible. The probability of sustained remission or cure is ~75% in patients with localized disease but is ~25% in individuals with unresectable lymphoma. In patients whose tumors are not resected, chemotherapy may lead to bowel perforation.

A unique form of small-bowel lymphoma, diffusely involving the entire intestine, was first described in oriental Jews and Arabs and is referred to as *immunoproliferative small intestinal disease* (IPSID), *Mediterranean lymphoma*, or *α heavy chain disease*. This is a B-cell tumor. The typical presentation includes chronic diarrhea and steatorrhea associated with vomiting and abdominal cramps; clubbing of the digits may be observed. A curious feature in many patients with IPSID is the presence in the blood and intestinal secretions of an abnormal IgA that contains a shortened α heavy chain and is devoid of light chains. It is suspected that the abnormal α chains are produced by plasma cells infiltrating the small bowel. The clinical course of patients with IPSID is generally one of exacerbations and remissions, with death frequently resulting from either progressive malnutrition and wasting or the development of an aggressive lymphoma. The use of oral antibiotics such as tetracycline appears to be beneficial in the early phases of the disorder, suggesting a possible infectious etiology. Combination chemotherapy has been administered during later stages of the disease, with variable results. Results are better when antibiotics and chemotherapy are combined.

■ CARCINOID TUMORS

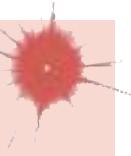
Carcinoid tumors arise from argentaffin cells of the crypts of Lieberkühn and are found from the distal duodenum to the ascending colon, areas embryologically derived from the midgut. More than 50% of intestinal carcinoids are found in the distal ileum, with most congregating close to the ileocecal valve. Most intestinal carcinoids are asymptomatic and of low malignant potential, but invasion and metastases may occur, leading to the carcinoid syndrome ([Chap. 80](#)).

■ LEIOMYOSARCOMAS

Leiomyosarcomas often are >5 cm in diameter and may be palpable on abdominal examination. Bleeding, obstruction, and perforation are common. Such tumors should be analyzed for the expression of mutant c-kit receptor (defining GIST), and in the presence of metastatic disease, justifying treatment with imatinib mesylate (Gleevec) or, in imatinib-refractory patients, sunitinib (Sutent) or regorafenib (Stivarga).

■ FURTHER READING

- BASS AJ et al: The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513:202, 2014.
- ENZINGER PC et al: CALGB 80403 (Alliance)/E1206: A randomized phase II study of three chemotherapy regimens plus cetuximab in metastatic esophageal and gastroesophageal junction cancers. *J Clin Oncol* 34:2736, 2016.
- PENNATHUR A et al: Oesophageal carcinoma. *Lancet* 381:400, 2013.
- RUSTGI AK, EL-SERAG HB: Esophageal carcinoma. *N Engl J Med* 371:2499, 2014.
- VAN CUTSEM E et al: Gastric cancer. *Lancet* 388:2654, 2016.
- VAN HAGEN P et al: Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med* 366:2074, 2012.



Lower gastrointestinal cancers include malignant tumors of the colon, rectum, and anus.

COLORECTAL CANCER

INCIDENCE

 Cancer of the large bowel is second only to lung cancer as a cause of cancer death in the United States: 135,430 new cases occurred in 2017, and 50,260 deaths were due to colorectal cancer. The incidence rate has decreased significantly during the past 25 years, likely due in large part to enhanced and more compliantly followed screening practices. Similarly, mortality rates in the United States have decreased by ~25%, resulting largely from earlier detection and improved treatment.

POLYPS AND MOLECULAR PATHOGENESIS

Most colorectal cancers, regardless of etiology, arise from adenomatous polyps. A polyp is a grossly visible protrusion from the mucosal surface and may be classified pathologically as a nonneoplastic hamartoma (e.g., *juvenile polyp*), a hyperplastic mucosal proliferation (*hyperplastic polyp*), or an adenomatous polyp. Only adenomas are clearly premalignant, and only a minority of adenomatous polyps evolve into cancer. Adenomatous polyps may be found in the colons of ~30% of middle-aged and ~50% of elderly people; however, <1% of polyps ever become malignant. Most polyps produce no symptoms and remain clinically undetected. Occult blood in the stool is found in <5% of patients with polyps.

A number of molecular changes are noted in adenomatous polyps and colorectal cancers that are thought to reflect a multistep process in the evolution of normal colonic mucosa to life-threatening invasive carcinoma. These developmental steps toward carcinogenesis include, but are not restricted to, point mutations in the *K-ras* protooncogene; hypomethylation of DNA, leading to gene activation; loss of DNA (*allelic loss*) at the site of a tumor-suppressor gene (the adenomatous polyposis coli [*APC*] gene) on the long arm of chromosome 5 (5q21); allelic loss at the site of a tumor-suppressor gene located on chromosome 18q (the deleted in colorectal cancer [*DCC*] gene); and allelic loss at chromosome 17p, associated with mutations in the *p53* tumor-suppressor gene (see Fig. 67-2). Thus, the altered proliferative pattern of the colonic mucosa, which results in progression to a polyp and then to carcinoma, may involve the mutational activation of an oncogene followed by and coupled with the loss of genes that normally suppress tumorigenesis. It remains uncertain whether the genetic aberrations always occur in a defined order. Based on this model, however, cancer is believed to develop only in those polyps in which most (if not all) of these mutational events take place.

Clinically, the probability of an adenomatous polyp becoming a cancer depends on the gross appearance of the lesion, its histologic features, and its size. Polyps may be pedunculated (stalked) or sessile (flat-based), adenomatous or serrated. Invasive cancers develop more frequently in sessile, serrated (i.e. "flat") polyps. Histologically, adenomatous polyps may be tubular, villous (i.e., papillary), or tubulovillous. Villous adenomas, most of which are sessile, become malignant more than three times as often as tubular adenomas. The likelihood that any polypoid lesion in the large bowel contains invasive cancer is related to the size of the polyp, being negligible (<2%) in lesions <1.5 cm, intermediate (2–10%) in lesions 1.5–2.5 cm, and substantial (10%) in lesions >2.5 cm in size.

Following the detection of an adenomatous polyp, the entire large bowel should be visualized endoscopically because synchronous lesions are noted in about one-third of cases. Colonoscopy should

then be repeated periodically, even in the absence of a previously documented malignancy, because such patients have a 30–50% probability of developing another adenoma and are at a higher-than-average risk for developing a colorectal carcinoma. Adenomatous polyps are thought to require >5 years of growth before becoming clinically significant; colonoscopy need not be carried out more frequently than every 3 years for the vast majority of patients.

ETIOLOGY AND RISK FACTORS

 Risk factors for the development of colorectal cancer are listed in Table 77-1.

Diet The etiology for most cases of large-bowel cancer appears to be related to environmental factors. The disease occurs more often in upper socioeconomic populations who live in urban areas. Mortality from colorectal cancer is directly correlated with per capita consumption of calories, meat protein, and dietary fat and oil as well as elevations in the serum cholesterol concentration and mortality from coronary artery disease. Geographic variations in incidence largely are unrelated to genetic differences, since migrant groups tend to assume the large-bowel cancer incidence rates of their adopted countries. Furthermore, population groups such as Mormons and Seventh Day Adventists, whose lifestyle and dietary habits differ somewhat from those of their neighbors, have significantly lower-than-expected incidence and mortality rates for colorectal cancer. The incidence of colorectal cancer has increased in Japan since that nation has adopted a more "Western" diet. At least three hypotheses have been proposed to explain the relationship to diet, none of which is fully satisfactory.

ANIMAL FATS One hypothesis is that the ingestion of animal fats found in red meats and processed meat leads to an increased proportion of anaerobes in the gut microflora, resulting in the conversion of normal bile acids into carcinogens. This provocative hypothesis is supported by several reports of increased amounts of fecal anaerobes in the stools of patients with colorectal cancer. Diets high in animal (but not vegetable) fats are also associated with high serum cholesterol, which is also associated with enhanced risk for the development of colorectal adenomas and carcinomas.

INSULIN RESISTANCE The large number of calories in Western diets coupled with physical inactivity has been associated with a higher prevalence of obesity. Obese persons develop insulin resistance with increased circulating levels of insulin, leading to higher circulating concentrations of insulin-like growth factor type I (IGF-I). This growth factor appears to stimulate proliferation of the intestinal mucosa.

FIBER Contrary to prior beliefs, the results of randomized trials and case-controlled studies have *failed* to show any value for dietary fiber or diets high in fruits and vegetables in preventing the recurrence of colorectal adenomas or the development of colorectal cancer.

The weight of epidemiologic evidence, however, implicates diet as being the major etiologic factor for colorectal cancer, particularly diets high in animal fat and in calories.

HEREDITARY FACTORS AND SYNDROMES

Up to 25% of patients with colorectal cancer have a family history of the disease, suggesting a hereditary predisposition. Inherited large-bowel cancers can be divided into two main groups: the well-studied but uncommon polyposis syndromes and the more common nonpolyposis syndromes (Table 77-2).

TABLE 77-1 Risk Factors for the Development of Colorectal Cancer

Diet: Animal fat
Hereditary syndromes
Polyposis coli
MYH-associated polyposis
Nonpolyposis syndrome (Lynch's syndrome)
Inflammatory bowel disease
Streptococcus bovis bacteremia
? Tobacco use

TABLE 77-2 Heritable (Autosomal Dominant) Gastrointestinal Polyposis Syndromes

SYNDROME	DISTRIBUTION OF POLYPS	HISTOLOGIC TYPE	MALIGNANT POTENTIAL	ASSOCIATED LESIONS
Familial adenomatous polyposis	Large intestine	Adenoma	Common	None
Gardner's syndrome	Large and small intestines	Adenoma	Common	Osteomas, fibromas, lipomas, epidermoid cysts, ampullary cancers, congenital hypertrophy of retinal pigment epithelium
Turcot's syndrome	Large intestine	Adenoma	Common	Brain tumors
MYH-associated polyposis	Large intestine	Adenoma	Common	None
Nonpolyposis syndrome (Lynch's syndrome)	Large intestine (often proximal)	Adenoma	Common	Endometrial and ovarian tumors (most frequently) gastric, genitourinary, pancreatic, biliary cancers (less frequently)
Peutz-Jeghers syndrome	Small and large intestines, stomach	Hamartoma	Rare	Mucocutaneous pigmentation; tumors of the ovary, breast, pancreas, endometrium
Juvenile polyposis	Large and small intestines, stomach	Hamartoma, rarely progressing to adenoma	Rare	Various congenital abnormalities

Polyposis Coli Polyposis coli (familial polyposis of the colon) is a rare condition characterized by the appearance of thousands of adenomatous polyps throughout the large bowel. It is transmitted as an autosomal dominant trait; the occasional patient with no family history probably developed the condition due to a spontaneous mutation. Polyposis coli is associated with a deletion in the long arm of chromosome 5 (including the *APC* gene) in both neoplastic (somatic mutation) and normal (germline mutation) cells. The loss of this genetic material (i.e., allelic loss) results in the absence of tumor-suppressor genes whose protein products would normally inhibit neoplastic growth. The presence of soft tissue and bony tumors, congenital hypertrophy of the retinal pigment epithelium, mesenteric desmoid tumors, and ampullary cancers in addition to the colonic polyps characterizes a subset of polyposis coli known as *Gardner's syndrome*. The appearance of malignant tumors of the central nervous system accompanying polyposis coli defines *Turcot's syndrome*. The colonic polyps in all these conditions are rarely present before puberty but are generally evident in affected individuals by age 25. If the polyposis is not treated surgically, colorectal cancer will develop in almost all patients aged <40. Polyposis coli results from a defect in the colonic mucosa, leading to an abnormal proliferative pattern and impaired DNA repair mechanisms. Once the multiple polyps are detected, patients should undergo a total colectomy. Medical therapy with nonsteroidal anti-inflammatory drugs (NSAIDs) such as sulindac and selective cyclooxygenase-2 inhibitors such as celecoxib can decrease the number and size of polyps in patients with polyposis coli; however, this effect on polyps is only temporary, and the use of NSAIDs has not been shown to reduce the risk of cancer. Colectomy remains the primary therapy/prevention. The offspring of patients with polyposis coli, who often are prepubertal when the diagnosis is made in the parent, have a 50% risk for developing this premalignant disorder and should be carefully screened by annual flexible sigmoidoscopy until age 35. Proctosigmoidoscopy is a sufficient screening procedure because polyps tend to be evenly distributed from cecum to anus, making more invasive and expensive techniques such as colonoscopy or barium enema unnecessary. Testing for occult blood in the stool is an inadequate screening maneuver. If a causative germline *APC* mutation has been identified in an affected family member, an alternative method for identifying carriers is testing DNA from peripheral blood mononuclear cells for the presence of the specific *APC* mutation. The detection of such a germline mutation can lead to a definitive diagnosis before the development of polyps.

MYH-Associated Polyposis MYH-associated polyposis (MAP) is a rare autosomal recessive syndrome caused by a biallelic mutation in the *MUT4H* gene. This hereditary condition may have a variable clinical presentation, resembling polyposis coli or colorectal cancer occurring in younger individuals without polyposis. Screening and colectomy guidelines for this syndrome are less clear than for polyposis coli, but annual to biennial colonoscopic surveillance is generally recommended starting at age 25–30.

Hereditary Nonpolyposis Colon Cancer Hereditary nonpolyposis colon cancer (HNPCC), also known as *Lynch's syndrome*, is another autosomal dominant trait. It is characterized by the presence of three or more relatives with histologically documented colorectal cancer, one of whom is a first-degree relative of the other two; one or more cases of colorectal cancer diagnosed before age 50 in the family; and colorectal cancer involving at least two generations. In contrast to polyposis coli, HNPCC is associated with an unusually high frequency of cancer arising in the proximal large bowel. The median age for the appearance of an adenocarcinoma is <50 years, 10–15 years younger than the median age for the general population. Despite having a poorly differentiated, mucinous histologic appearance, the proximal colon tumors that characterize HNPCC have a better prognosis than sporadic tumors from patients of similar age. Families with HNPCC often include individuals with multiple primary cancers; the association of colorectal cancer with either ovarian or endometrial carcinomas is especially strong in women, and an increased appearance of gastric, small-bowel, genitourinary, pancreaticobiliary, and sebaceous skin tumors has been reported as well. It has been recommended that members of such families undergo annual or biennial colonoscopy beginning at age 25 years, with intermittent pelvic ultrasonography and endometrial biopsy for afflicted women; such a screening strategy has not yet been validated. HNPCC is associated with germline mutations of several genes, particularly *hMSH2* on chromosome 2 and *hMLH3* on chromosome 3. These mutations lead to errors in DNA replication and are thought to result in DNA instability because of defective repair of DNA mismatches resulting in abnormal cell growth and tumor development. Testing tumor cells through molecular analysis of DNA for "microsatellite instability" or immunohistochemical staining for deficiency in mismatch repair proteins in patients with colorectal cancer and a positive family history for colorectal or endometrial cancer may identify probands with HNPCC.

■ INFLAMMATORY BOWEL DISEASE

(Chap. 319) Large-bowel cancer is increased in incidence in patients with long-standing inflammatory bowel disease (IBD). Cancers develop more commonly in patients with ulcerative colitis than in those with granulomatous (i.e., Crohn's) colitis, but this impression may result in part from the occasional difficulty of differentiating these two conditions. The risk of colorectal cancer in a patient with IBD is relatively small during the initial 10 years of the disease, but then appears to increase at a rate of ~0.5–1% per year. Cancer may develop in 8–30% of patients after 25 years. The risk is higher in younger patients with pancolitis.

Cancer surveillance strategies in patients with IBD are unsatisfactory. Symptoms such as bloody diarrhea, abdominal cramping, and obstruction, which may signal the appearance of a tumor, are similar to the complaints caused by a flare-up of the underlying disease. In patients with a history of IBD lasting ≥15 years who continue to experience exacerbations, the surgical removal of the colon can significantly

reduce the risk for cancer and also eliminate the target organ for the underlying chronic gastrointestinal disorder. The value of such surveillance techniques as colonoscopy with mucosal biopsies and brushings for less symptomatic individuals with chronic IBD is uncertain. The lack of uniformity regarding the pathologic criteria that characterize dysplasia and the absence of data that such surveillance reduces the development of lethal cancers have made this costly practice an area of controversy.

■ OTHER HIGH-RISK CONDITIONS

Streptococcus bovis Bacteremia For unknown reasons, individuals who develop endocarditis or septicemia from this fecal bacterium have a high incidence of occult colorectal tumors and, possibly, upper gastrointestinal cancers as well. Endoscopic or radiographic screening appears advisable.

Tobacco Use Cigarette smoking is linked to the development of colorectal adenomas, particularly after >35 years of tobacco use. No biologic explanation for this association has yet been proposed.

■ PRIMARY PREVENTION

Several orally administered compounds have been assessed as possible inhibitors of colon cancer. The most effective class of chemopreventive agents is aspirin and other NSAIDs, which are thought to suppress cell proliferation by inhibiting prostaglandin synthesis. Regular aspirin use reduces the risk of colon adenomas and carcinomas as well as death from large-bowel cancer; such use also appears to diminish the likelihood for developing additional premalignant adenomas following successful treatment for a prior colon carcinoma. This effect of aspirin on colon carcinogenesis increases with the duration and dosage of drug use. Emerging data linking adequate plasma levels of vitamin D with reduced risk of adenomatous polyps and colorectal cancer appear promising. The value of vitamin D as a form of chemoprevention is under study. Antioxidant vitamins such as ascorbic acid, tocopherols, and β-carotene are ineffective at reducing the incidence of subsequent adenomas in patients who have undergone the removal of a colon adenoma. Estrogen replacement therapy has been associated with a reduction in the risk of colorectal cancer in women, conceivably by an effect on bile acid synthesis and composition or by decreasing synthesis of IGF-I.

■ SCREENING

The rationale for colorectal cancer screening programs is that the removal of adenomatous polyps will prevent colorectal cancer, and that earlier detection of localized, superficial cancers in asymptomatic individuals will increase the surgical cure rate. Such screening programs are particularly important for individuals with a family history of the disease in first-degree relatives. The relative risk for developing colorectal cancer increases to 1.75 in such individuals and may be even higher if the relative was afflicted before age 60. The prior use of rigid proctosigmoidoscopy as a screening tool was based on the observation that 60% of early lesions are located in the rectosigmoid. For unexplained reasons, however, the proportion of large-bowel cancers arising in the rectum has been decreasing during the past several decades, with a corresponding increase in the proportion of cancers in the more proximal descending colon. As such, the potential for rigid proctosigmoidoscopy to detect a sufficient number of occult neoplasms to make the procedure cost-effective has been questioned.

Screening strategies for colorectal cancer that have been examined during the past several decades are listed in **Table 77-3**.

Many programs directed at the early detection of colorectal cancers have focused on digital rectal examinations and fecal occult blood (i.e., stool guaiac) testing. The digital examination should be part of any routine physical evaluation in adults aged >40 years, serving as a screening test for prostate cancer in men, a component of the pelvic examination in women, and an inexpensive maneuver for the detection of masses in the rectum. However, because of the proximal migration of colorectal tumors, its value as an overall screening modality for colorectal cancer has become limited. The development of the fecal

TABLE 77-3 Screening Strategies for Colorectal Cancer

Digital rectal examination
Stool testing
• Occult blood
• Fecal DNA
Imaging
• Contrast barium enema
• Virtual (i.e., computed tomography colonography)
Endoscopy
• Flexible sigmoidoscopy
• Colonoscopy

occult blood test has greatly facilitated the detection of occult fecal blood. Unfortunately, even when performed optimally, the fecal occult blood test has major limitations as a screening technique. About 50% of patients with documented colorectal cancers have a negative fecal occult blood test, consistent with the intermittent bleeding pattern of these tumors. When random cohorts of asymptomatic persons have been tested, 2–4% have fecal occult blood-positive stools. Colorectal cancers have been found in <10% of these “test-positive” cases, with benign polyps being detected in an additional 20–30%. Thus, a colorectal neoplasm will not be found in most asymptomatic individuals with occult blood in their stool. Nonetheless, persons found to have fecal occult blood-positive stool routinely undergo further medical evaluation, including sigmoidoscopy and/or colonoscopy—procedures that are not only uncomfortable and expensive but also associated with a small risk for significant complications. The added cost of these studies would appear justifiable if the small number of patients found to have occult neoplasms because of fecal occult blood screening could be shown to have an improved prognosis and prolonged survival. Prospective controlled trials have shown a statistically significant reduction in mortality rate from colorectal cancer for individuals undergoing annual stool guaiac screening. However, this benefit only emerged after >13 years of follow-up and was extremely expensive to achieve because all positive tests (most of which were falsely positive) were followed by colonoscopy. Moreover, these colonoscopic examinations quite likely provided the opportunity for cancer prevention through the removal of potentially premalignant adenomatous polyps because the eventual development of cancer was reduced by 20% in the cohort undergoing annual screening.

With the appreciation that the carcinogenic process leading to the progression of the normal bowel mucosa to an adenomatous polyp and then to a cancer is the result of a series of molecular changes, investigators have examined fecal DNA for evidence of mutations associated with such molecular changes as evidence of the occult presence of precancerous lesions or actual malignancies. Such a strategy has been tested in >4000 asymptomatic individuals whose stool was assessed for occult blood and for 21 possible mutations in fecal DNA; these study subjects also underwent colonoscopy. Although the fecal DNA strategy suggested the presence of more advanced adenomas and cancers than did the fecal occult blood testing approach, the overall sensitivity, using colonoscopic findings as the standard, was <50%, diminishing enthusiasm for further pursuit of the fecal DNA screening strategy.

The use of imaging studies to screen for colorectal cancers has also been explored. Air contrast barium enemas had been used to identify sources of occult blood in the stool prior to the advent of fiberoptic endoscopy; the cumbersome nature of the procedure and inconvenience to patients limited its widespread adoption. The introduction of computed tomography (CT) scanning led to the development of virtual (i.e., CT) colonography as an alternative to the growing use of endoscopic screening techniques. Virtual colonography was proposed as being equivalent in sensitivity to colonoscopy and being available in a more widespread manner because it did not require the same degree of operator expertise as fiberoptic endoscopy. However, virtual colonography requires the same cathartic preparation that has limited widespread acceptance in association with endoscopic colonoscopy, is diagnostic but not therapeutic (i.e., patients with suspicious findings

must undergo a subsequent endoscopic procedure for polypectomy or biopsy), and, in the setting of general radiology practices, appears to be less sensitive as a screening technique when compared with endoscopic procedures.

With the appreciation of the inadequacy of fecal occult blood testing alone, concerns about the practicality of imaging approaches, and the wider adoption of endoscopic examinations by the primary care community, screening strategies in asymptomatic persons have changed. At present, both the American Cancer Society and the National Comprehensive Cancer Network recommend either fecal occult blood testing annually coupled with flexible sigmoidoscopy every 5 years or colonoscopy every 10 years beginning at age 50 in asymptomatic individuals with no personal or family history of polyps or colorectal cancer. The recommendation for the inclusion of flexible sigmoidoscopy is strongly supported by the recently published results of three randomized trials performed in the United States, the United Kingdom, and Italy, involving >350,000 individuals, which consistently showed that periodic (even single) sigmoidoscopic examinations, after more than a decade of median follow-up, lead to an ~21% reduction in the development of colorectal cancer and a >25% reduction in mortality from the malignant disease. Less than 20% of participants in these studies underwent a subsequent colonoscopy. In contrast to the cathartic preparation required before colonoscopic procedures, which is only performed by highly trained specialists, flexible sigmoidoscopy requires only an enema as preparation and can be accurately performed by nonspecialty physicians or physician-extenders. The randomized screening studies using flexible sigmoidoscopy led to the estimate that ~650 individuals needed to be screened to prevent one colorectal cancer death; this contrasts with the data for mammography where the number of women needing to be screened to prevent one breast cancer death is 2500, reinforcing the efficacy of endoscopic surveillance for colorectal cancer screening. Presumably the benefit from the sigmoidoscopic screening is the result of the identification and removal of adenomatous polyps; it is intriguing that this benefit has been achieved using a technique that leaves the proximal half of the large bowel unvisualized.

It remains to be seen whether surveillance colonoscopy, which has gained increasing popularity in the United States for colorectal cancer screening, will prove to be more effective than flexible sigmoidoscopy. Ongoing randomized trials being conducted in Europe are addressing this issue. Although flexible sigmoidoscopy only visualizes the distal half of the large bowel, leading to the assumption that colonoscopy represents a more informative approach, colonoscopy has been reported as being less accurate for screening the proximal rather than the distal colon, perhaps due to technical considerations but also possibly because of a greater frequency of serrated (i.e., "flat") polyps in the right colon, which are more difficult to identify. At present, colonoscopy performed every 10 years has been offered as an alternative to annual fecal occult blood testing with periodic (every 5 years) flexible sigmoidoscopy. Colonoscopy has been shown to be superior to double-contrast barium enema and also to have a higher sensitivity for detecting villous or dysplastic adenomas or cancers than the strategy using occult fecal blood testing and flexible sigmoidoscopy. Whether colonoscopy performed every 10 years beginning at age 50 is medically superior and economically equivalent to flexible sigmoidoscopy remains to be determined.

■ CLINICAL FEATURES

Presenting Symptoms Symptoms vary with the anatomic location of the tumor. Because stool is relatively liquid as it passes through the ileocecal valve into the right colon, cancers arising in the cecum and ascending colon may become quite large without resulting in any obstructive symptoms or noticeable alterations in bowel habits. Lesions of the right colon commonly ulcerate, leading to chronic, insidious blood loss without a change in the appearance of the stool. Consequently, patients with tumors of the ascending colon often present with symptoms such as fatigue, palpitations, and even angina pectoris and are found to have a hypochromic, microcytic anemia indicative of iron



FIGURE 77-1 Double-contrast air-barium enema revealing a sessile tumor of the cecum in a patient with iron-deficiency anemia and guaiac-positive stool. The lesion at surgery was a stage II adenocarcinoma.

deficiency. Because the cancers may bleed intermittently, a random fecal occult blood test may be negative. As a result, the unexplained presence of iron-deficiency anemia in any adult (with the possible exception of a premenopausal, multiparous woman) mandates a thorough endoscopic and/or radiographic visualization of the entire large bowel (Fig. 77-1).

Because stool becomes more formed as it passes into the transverse and descending colon, tumors arising there tend to impede the passage of stool, resulting in the development of abdominal cramping, occasional obstruction, and even perforation. Radiographs of the abdomen often reveal characteristic annular, constricting lesions ("apple-core" or "napkin-ring") (Fig. 77-2).

Cancers arising in the rectosigmoid are often associated with hematochezia, tenesmus, and narrowing of the caliber of stool; anemia is an infrequent finding. While these symptoms may lead patients and their

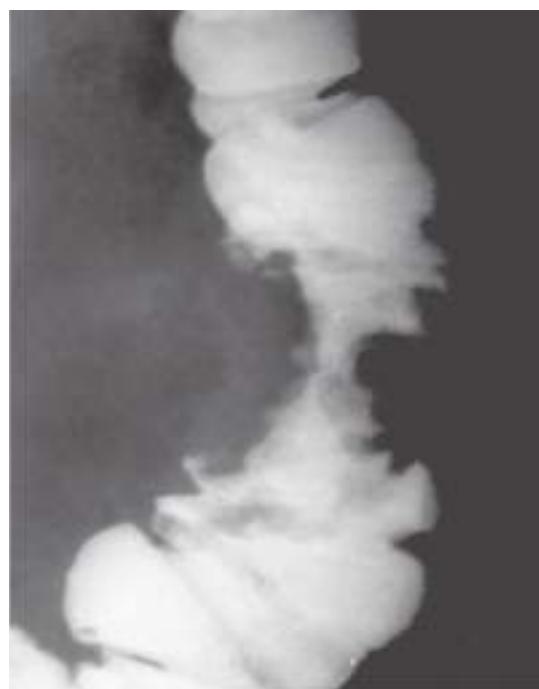


FIGURE 77-2 Annular, constricting adenocarcinoma of the descending colon. This radiographic appearance is referred to as an "apple-core" lesion and is always highly suggestive of malignancy.

physicians to suspect the presence of hemorrhoids, the development of rectal bleeding and/or altered bowel habits demands a prompt digital rectal examination and proctosigmoidoscopy.

Staging, Prognostic Factors, and Patterns of Spread The prognosis for individuals having colorectal cancer is related to the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases. These variables are incorporated into a TNM classification method, in which T represents the depth of tumor penetration, N the presence of lymph node involvement, and M the presence or absence of distant metastases (Fig. 77-3). Superficial lesions that do not involve regional lymph nodes and do not penetrate through the submucosa (T1) or the muscularis (T2) are designated as *stage I* (T1–N0M0) disease; tumors that penetrate through the muscularis but have not spread to lymph nodes are *stage II* disease (T3–N0M0); regional lymph node involvement defines *stage III* (TXN1–M0) disease; and metastatic spread to sites such as liver, lung, or bone indicates *stage IV* (TXNM1) disease. Unless gross evidence of metastatic disease is present, disease stage cannot be determined accurately before surgical resection and pathologic analysis of the operative specimens.

Most recurrences after a surgical resection of a large-bowel cancer occur within the first 4 years, making 5-year survival a fairly reliable indicator of cure. The likelihood for 5-year survival in patients with colorectal cancer is stage-related (Fig. 77-3). That likelihood has improved during the past several decades when similar surgical stages have been compared. The most plausible explanation for this improvement is more thorough intraoperative and pathologic staging. In particular, more exacting attention to pathologic detail has revealed that the prognosis following the resection of a colorectal cancer is not related merely to the presence or absence of regional lymph node involvement; rather, prognosis may be more precisely gauged by the number of involved lymph nodes (one to three lymph nodes ["N1"] vs four or more lymph nodes ["N2"]) and the number of nodes examined. A minimum of 12 sampled lymph nodes is thought necessary to accurately define tumor stage, and the more nodes examined, the better. Other predictors of a poor prognosis after a total surgical resection include tumor penetration through the bowel wall into pericolic fat, poorly differentiated histology, perforation and/or tumor adherence to adjacent organs (increasing the risk for an anatomically adjacent recurrence), and venous invasion by tumor (Table 77-4). Regardless

TABLE 77-4 Predictors of Poorer Outcomes Following Total Surgical Resection of Colorectal Cancer

Tumor spread to regional lymph nodes
Number of regional lymph nodes involved
Tumor penetration through the bowel wall
Poorly differentiated histology
Perforation
Tumor adherence to adjacent organs
Venous invasion
Preoperative elevation of CEA titer (>5 ng/mL)
Specific chromosomal deletion (e.g., mutation in the <i>b-raf</i> gene)
Right-sided location of primary tumor

Abbreviation: CEA, carcinoembryonic antigen.

of the clinicopathologic stage, a preoperative elevation of the plasma carcinoembryonic antigen (CEA) level predicts eventual tumor recurrence. The presence of specific chromosomal aberrations, particularly a mutation in the *b-raf* gene in tumor cells, appears to predict for a higher risk for metastatic spread. Conversely, the detection of microsatellite instability in tumor tissue indicates a more favorable outcome. Tumors arising in the left colon are associated with a better prognosis than those appearing in the right colon, likely due to differences in molecular patterns. In contrast to most other cancers, the prognosis in colorectal cancer is not influenced by the size of the primary lesion when adjusted for nodal involvement and histologic differentiation.

Cancers of the large bowel generally spread to regional lymph nodes or to the liver via the portal venous circulation. The liver represents the most frequent visceral site of metastasis; it is the initial site of distant spread in one-third of recurring colorectal cancers and is involved in more than two-thirds of such patients at the time of death. In general, colorectal cancer rarely spreads to the lungs, supraclavicular lymph nodes, bone, or brain without prior spread to the liver. A major exception to this rule occurs in patients having primary tumors in the distal rectum, from which tumor cells may spread through the paravertebral venous plexus, escaping the portal venous system and thereby reaching the lungs or supraclavicular lymph nodes without hepatic involvement. The median survival after the detection of distant metastases has increased during the last 30 years from 6–9 months (hepatomegaly, abnormal liver chemistries) to 27–30 months (small

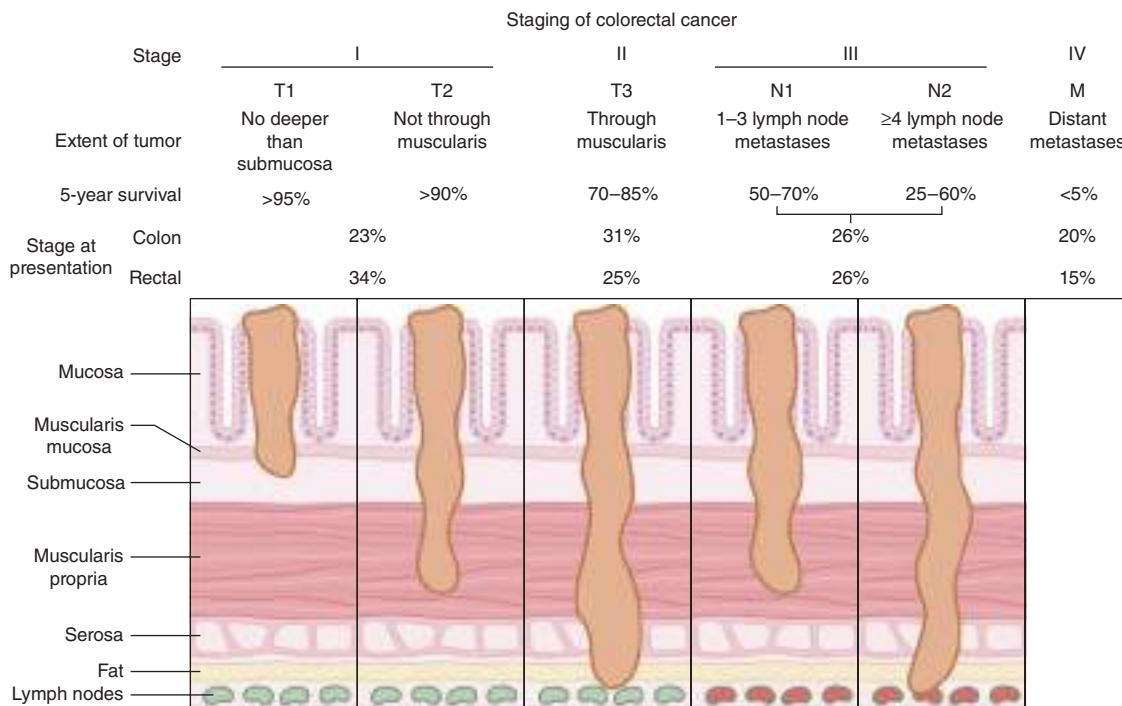


FIGURE 77-3 Staging and prognosis for patients with colorectal cancer.

liver nodule initially identified by elevated CEA level and subsequent CT scan) with increasingly effective systemic therapy improving this prognosis further.

Efforts to use gene expression profiles to identify patients at risk of recurrence or those particularly likely to benefit from adjuvant therapy have not yet yielded practice-changing results. Despite a burgeoning literature examining a host of prognostic factors, pathologic stage at diagnosis remains the best predictor of long-term prognosis. Patients with lymphovascular invasion and high preoperative CEA levels are likely to have a more aggressive clinical course.

TREATMENT

Colorectal Cancer

Total resection of tumor is the optimal treatment when a malignant lesion is detected in the large bowel. An evaluation for the presence of metastatic disease, including a thorough physical examination, biochemical assessment of liver function, measurement of the plasma CEA level, and a CT scan of the chest, abdomen, and pelvis, should be performed before surgery. When possible, a colonoscopy of the entire large bowel should be performed to identify synchronous neoplasms and/or polyps. The detection of metastases should not preclude surgery in patients with tumor-related symptoms such as gastrointestinal bleeding or obstruction, but it often prompts the use of a less radical operative procedure. The necessity for a primary tumor resection in asymptomatic individuals with metastatic disease is an area of controversy. At the time of laparotomy, the entire peritoneal cavity should be examined, with thorough inspection of the liver, pelvis, and hemidiaphragm and careful palpation of the full length of the large bowel. Following recovery from a complete resection, patients should be observed carefully for 5 years by semiannual physical examinations and blood chemistry measurements. If a complete colonoscopy was not performed preoperatively, it should be carried out within the first several postoperative months. Some authorities favor measuring plasma CEA levels at 3-month intervals because of the sensitivity of this test as a marker for otherwise undetectable tumor recurrence. Subsequent endoscopic surveillance of the large bowel, probably at triennial intervals, is indicated, because patients who have been cured of one colorectal cancer have a 3–5% probability of developing an additional bowel cancer during their lifetime and a >15% risk for the development of adenomatous polyps. Anastomotic (“suture-line”) recurrences are infrequent in colorectal cancer patients, provided the surgical resection margins were adequate and free of tumor. The value of periodic CT scans of the abdomen, assessing for an early, asymptomatic indication of tumor recurrence, while uncertain, has been recommended annually for the first 3 postoperative years.

Radiation therapy to the pelvis is recommended for patients with rectal cancer because it reduces the 20–25% probability of regional recurrences following complete surgical resection of stage II or III tumors, especially if they have penetrated through the serosa. This alarmingly high rate of local disease recurrence is believed to be due to the fact that the contained anatomic space within the pelvis limits the extent of the resection and because the rich lymphatic network of the pelvic side wall immediately adjacent to the rectum facilitates the early spread of malignant cells into surgically inaccessible tissue. The use of sharp rather than blunt dissection of rectal cancers (*total mesorectal excision*) appears to reduce the likelihood of local disease recurrence to ~10%. Radiation therapy, either pre- or postoperatively, further reduces the likelihood of pelvic recurrences but does not appear to prolong survival. Combining radiation therapy with 5-fluorouracil (5-FU)-based chemotherapy, preferably prior to surgical resection, lowers local recurrence rates and improves overall survival. Preoperative radiotherapy is indicated for patients with large, potentially unresectable rectal cancers; such lesions may shrink enough to permit subsequent surgical removal. Radiation therapy alone is not effective as the primary treatment of colon cancer.

Systemic therapy for patients with colorectal cancer has become more effective. 5-FU remains the backbone of treatment for this disease. Partial responses are obtained in 15–20% of patients. The probability of tumor response appears to be somewhat greater for patients with liver metastases when chemotherapy is infused directly into the hepatic artery, but intraarterial treatment is costly and toxic and does not appear to appreciably prolong survival. The concomitant administration of folinic acid (leucovorin [LV]) improves the efficacy of 5-FU in patients with advanced colorectal cancer, presumably by enhancing the binding of 5-FU to its target enzyme, thymidylate synthase. 5-FU is generally administered intravenously but may also be given orally in the form of capecitabine (Xeloda) with seemingly similar efficacy.

Irinotecan (CPT-11), a topoisomerase 1 inhibitor, has been added to 5-FU and LV (e.g., FOLFIRI) with resultant improvement in response rates and survival of patients with metastatic disease. The *FOLFIRI regimen* is as follows: irinotecan, 180 mg/m² as a 90-min infusion on day 1; LV, 400 mg/m² as a 2-h infusion during irinotecan administration; immediately followed by 5-FU bolus, 400 mg/m², and 46-h continuous infusion of 2.4–3 g/m² every 2 weeks. Diarrhea is the major side effect from irinotecan. Oxaliplatin, a platinum analogue, also improves the response rate when added to 5-FU and LV (FOLFOX) as initial treatment of patients with metastatic disease. The *FOLFOX regimen* is as follows: 2-h infusion of LV (400 mg/m² per day) followed by a 5-FU bolus (400 mg/m² per day) and 22-h infusion (1200 mg/m²) every 2 weeks, together with oxaliplatin, 85 mg/m² as a 2-h infusion on day 1. Oxaliplatin frequently causes a dose-dependent sensory neuropathy that often but not always resolves following the cessation of therapy. FOLFIRI and FOLFOX are equal in efficacy. In metastatic disease, these regimens may produce median survivals of 2 years.

Monoclonal antibodies are also effective in patients with advanced colorectal cancer. Cetuximab (Erbitux) and panitumumab (Vectibix) are directed against the epidermal growth factor receptor (EGFR), a transmembrane glycoprotein involved in signaling pathways affecting growth and proliferation of tumor cells. Both cetuximab and panitumumab, when given alone, have been shown to benefit a small proportion of previously treated patients, and cetuximab appears to have therapeutic synergy with such chemotherapeutic agents as irinotecan, even in patients previously resistant to this drug; this suggests that cetuximab can reverse cellular resistance to cytotoxic chemotherapy. The antibodies are not effective in the ~65% subset of colon tumors that contain mutations in *ras* or *b-raf* genes. The use of both cetuximab and panitumumab can lead to an acne-like rash, with the development and severity of the rash being correlated with the likelihood of antitumor efficacy. Inhibitors of the EGFR tyrosine kinase such as erlotinib (Tarccea) or sunitinib (Sutent) do not appear to be effective in colorectal cancer.

Bevacizumab (Avastin) is a monoclonal antibody directed against the vascular endothelial growth factor (VEGF) and is thought to act as an antiangiogenesis agent. The addition of bevacizumab to irinotecan-containing combinations and to FOLFOX initially appeared to significantly improve the outcome observed with chemotherapy alone, but subsequent studies have suggested a more modest degree of benefit. The use of bevacizumab can lead to hypertension, proteinuria, and an increased likelihood of thromboembolic events.

Preliminary data suggest that the use of checkpoint inhibitors (i.e., PD-1 and PD-2) as immunotherapy is effective in the small subset of patients with metastatic colorectal cancer whose tumors are mismatch repair protein deficient (i.e., microsatellite unstable). Patients with solitary hepatic metastases without clinical or radiographic evidence of additional tumor involvement should be considered for partial liver resection, because such procedures are associated with 5-year survival rates of 25–30% when performed on selected individuals by experienced surgeons.

The administration of 5-FU and LV for 6 months after resection of tumor in patients with stage III disease leads to a 40% decrease in recurrence rates and 30% improvement in survival. The likelihood of recurrence has been further reduced when oxaliplatin has been

combined with 5-FU and LV (e.g., FOLFOX), particularly in patients whose tumor has spread to 4 or more regional lymph nodes (N2). Unexpectedly, the addition of irinotecan to 5-FU and LV as well as the addition of either bevacizumab or cetuximab to FOLFOX did not significantly enhance outcome. Patients with stage II tumors do not appear to benefit appreciably from adjuvant therapy, with the use of such treatment generally restricted to those patients having biologic characteristics (e.g., perforated tumors, T4 lesions, lymphovascular invasion) that place them at higher likelihood for recurrence. The addition of oxaliplatin to adjuvant treatment for patients aged >70 and those with stage II disease does not appear to provide any therapeutic benefit.

In rectal cancer, the delivery of preoperative or postoperative combined-modality therapy (5-FU or capecitabine plus radiation therapy) reduces the risk of recurrence and increases the chance of cure for patients with stage II and III tumors, with the preoperative approach being better tolerated.

CANCERS OF THE ANUS

Cancers of the anus account for 1–2% of the malignant tumors of the large bowel. Most such lesions arise in the anal canal, the anatomic area extending from the anorectal ring to a zone approximately halfway between the pectinate (or dentate) line and the anal verge. Carcinomas arising proximal to the pectinate line (i.e., in the transitional zone between the glandular mucosa of the rectum and the squamous epithelium of the distal anus) are known as *basaloid*, *cuboidal*, or *cloacogenic* tumors; about one-third of anal cancers have this histologic pattern. Malignancies arising distal to the pectinate line have squamous histology, ulcerate more frequently, and constitute ~55% of anal cancers. The prognosis for patients with basaloid and squamous cell cancers of the anus is identical when corrected for tumor size and the presence or absence of nodal spread.

The development of anal cancer is associated with infection by human papillomavirus, the same organism etiologically linked to cervical cancer. The virus is sexually transmitted. The infection may lead to anal warts (*condyloma acuminata*), which may progress to anal intraepithelial neoplasia and on to squamous cell carcinoma. The risk for anal cancer is increased among homosexual males, presumably related to anal intercourse. Anal cancer risk is increased in both men and women with AIDS, possibly because their immunosuppressed state permits more severe papillomavirus infection. Vaccination against human papilloma viruses appears to reduce the eventual risk for anal cancer. Anal cancers occur most commonly in middle-aged persons and are more frequent in women than men. At diagnosis, patients may experience bleeding, pain, sensation of a perianal mass, and pruritus.

Radical surgery (abdominal-perineal resection with lymph node sampling and a permanent colostomy) was once the treatment of choice for this tumor type. The 5-year survival rate after such a procedure was 55–70% in the absence of spread to regional lymph nodes and <20% if nodal involvement was present. An alternative therapeutic approach combining external beam radiation therapy with concomitant chemotherapy (5-FU and mitomycin C) has resulted in biopsy-proven disappearance of all tumor in >80% of patients whose initial lesion was <3 cm in size. Tumor recurrences develop in <10% of these patients, meaning that ~70% of patients with anal cancers can be cured with nonoperative treatment and without the need for a colostomy. Surgery should be reserved for the minority of individuals who are found to have residual tumor after being managed initially with radiation therapy combined with chemotherapy.

FURTHER READING

- BRENNER H et al: Colorectal cancer. Lancet 383:1490, 2014.
- CAO Y et al: Population-wide impact of long-term use of aspirin and the risk of cancer. JAMA Oncol 2:762, 2016.
- LIEBERMAN D et al: Screening for colorectal cancer and evolving issues for physicians and patients. A review. JAMA 316:2135, 2016.
- MEYERHARDT JA et al: Follow-up care, surveillance protocol, and secondary prevention measures for survivors of colorectal cancer:

American Society of Clinical Oncology Clinical Practice Guidelines Endorsement. J Clin Oncol 31:4465, 2013.

PETRELLI F et al: Prognostic survival associated with left-sided vs right-sided colon cancer. A systemic review and meta-analysis. JAMA Oncol 3:211, 2017.

SHRIDHAR R et al: Anal cancer: Current standards in care and recent changes in practice. CA Cancer J Clin 65:139, 2015.

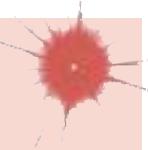
STRUM WB: Colorectal adenomas. N Engl J Med 374:1065, 2016.

VAN CUTSEM E et al: Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and *RAS* mutations in colorectal cancer. J Clin Oncol 33:692, 2015.

78

Tumors of the Liver and Biliary Tree

Josep M. Llovet



The burden of cancer is increasing worldwide. Lung, breast, and colorectal cancers are the most commonly diagnosed while lung and liver cancers are the most common causes of cancer death. Liver cancer is the sixth most common cancer worldwide, the second leading cause of cancer-related deaths and one of the few neoplasms whose incidence and mortality rates have been steadily increasing. Liver cancer comprises a heterogeneous group of malignant tumors with different histologic features and unfavorable prognosis that range from hepatocellular carcinoma (HCC; 85–90% cases), intrahepatic cholangiocarcinoma (iCCA; 10%), and other malignancies accounting for <1% of tumors, such as fibrolamellar HCC, mixed HCC-iCCA, epithelioid hemangioendothelioma, and the pediatric cancer hepatoblastoma. The burden of liver cancer is increasing globally in almost all countries, and it is estimated to reach one million cases by 2030.

HEPATOCELLULAR CARCINOMA

EPIDEMIOLOGY AND RISK FACTORS

 Overall, liver cancer accounts for 7% of all cancers (~850,000 new cases each year), and HCC represents 90% of primary liver cancers. The highest incidence rates of HCC occur in Asia and sub-Saharan Africa due to the high prevalence of hepatitis B virus (HBV) infection, with 20–35 cases per 100,000 inhabitants. Southern Europe, and now North America have intermediate incidence rates (10 cases per 100,000), whereas Northern and Western Europe have low incidence rates of less than 5 cases per 100,000 inhabitants. In the United States, liver cancer is ranked number one in terms of increased mortality during the past two decades (Fig. 78-1), with an incidence of 35,000 cases per year. HCC has a strong male preponderance with a male to female ratio estimated to be 2.5. The incidence increases with age, reaching a peak at 65–70 years old. In Chinese and in black African populations (where vertical transmission of HBV occurs), the mean age is 40–50 years. By contrast, in Japan mean age in men is now around 75 years.

The risk factors for HCC are well established (Fig. 78-2). The main risk factor is cirrhosis—and associated chronic liver damage caused by inflammation and fibrosis—of any etiology, which underlies 80% of HCC cases worldwide and results from chronic infection by HBV or hepatitis C virus (HCV) infection, alcohol abuse, metabolic syndrome, and hemochromatosis (associated to *HFE1* gene germ-line mutations). Cirrhotic patients represent 1% of the human population and one-third of them will develop HCC during their lifetime. Long-term follow-up studies have established an annual risk of HCC development of 2% in HBV-infected cirrhotic patients and 3–7% in HCV-infected cirrhotic patients. HCC is less common in cirrhosis associated with alpha-1 antitrypsin deficiency, autoimmune hepatitis, Wilson's disease, and cholestatic liver disorders. Predictors of liver cancer development

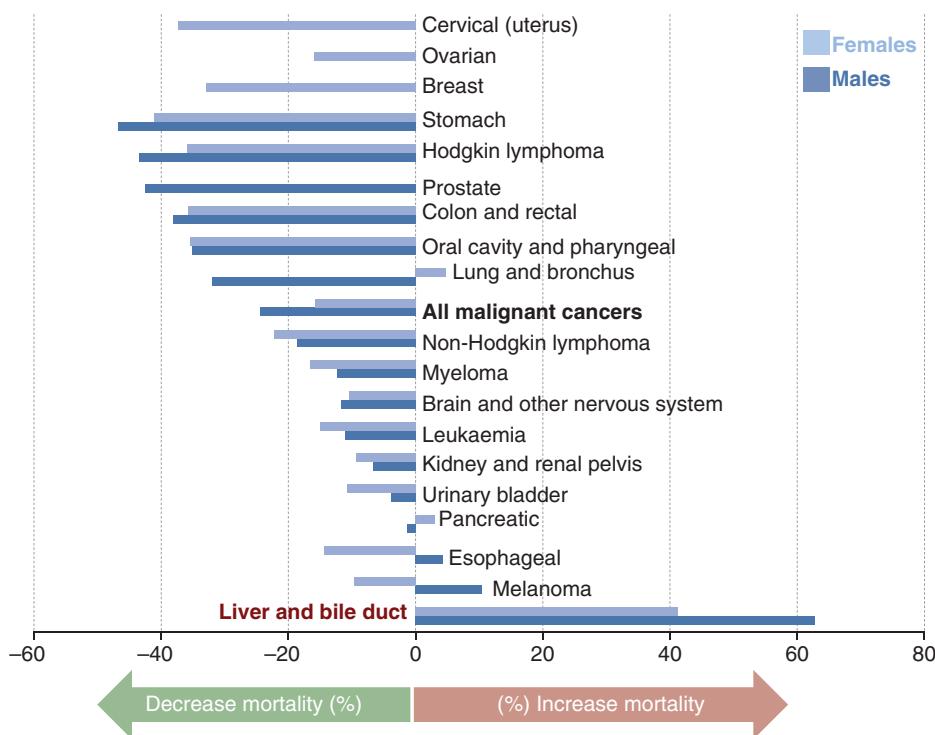


FIGURE 78-1 Mortality trends of patients with different malignancies in the United States between 1990 and 2009. Changes in cancer mortality across tumor types in the United States. Liver and bile duct cancer rank first in terms of increase mortality for both men and women. (Reprinted with permission from JM Llovet et al: *Nat Rev Clin Oncol* 12:408, 2015.)

among cirrhotic patients have been associated with liver disease severity (platelet count of $<100,000/\text{mm}^3$, presence of portal hypertension), the degree of liver stiffness as measured by transient elastography, and liver gene signatures capturing the *cancer field effect*.

has not yet been established.

Alcohol consumption and metabolic syndrome due to diabetes and obesity are responsible for ~20% of cases. Non-alcoholic steatohepatitis (NASH), related to metabolic syndrome, is now an emerging cause of

In terms of attributable risk fraction, HBV infection—a DNA virus that can cause insertional mutagenesis and affects 400 million people globally—accounts for 50% of HCC cases, and is the predominant cause in Asia and Africa. Among patients with HBV infection, a family history of HCC, HBeAg seropositivity, high viral load and genotype C are independent predictors of HCC development. Chronic treatments with effective antiviral HBV therapies are able to significantly decrease the risk of cancer. HCV infection—an RNA virus that affects 170 million people—is responsible for 30% of cases, and is the main cause of HCC in Europe and North America. Among patients with HCV infection, HCC occurs almost exclusively when relevant liver damage is present (either advanced fibrosis—Metavir F3 [Metavir is a scoring system for hepatic histology that grades fibrosis from 0 to 4 with higher numbers indicating more fibrosis]—or cirrhosis), particularly if associated with HCV genotype 1b. In addition, a polymorphism that activates EGFR, the EGF receptor, has been established as associated with HCV-HCC in several studies. Antiviral therapies with interferon regimes are able to prevent cirrhosis development and HCC occurrence. The impact of new direct-acting antiviral (DAA) regimes on HCC incidence

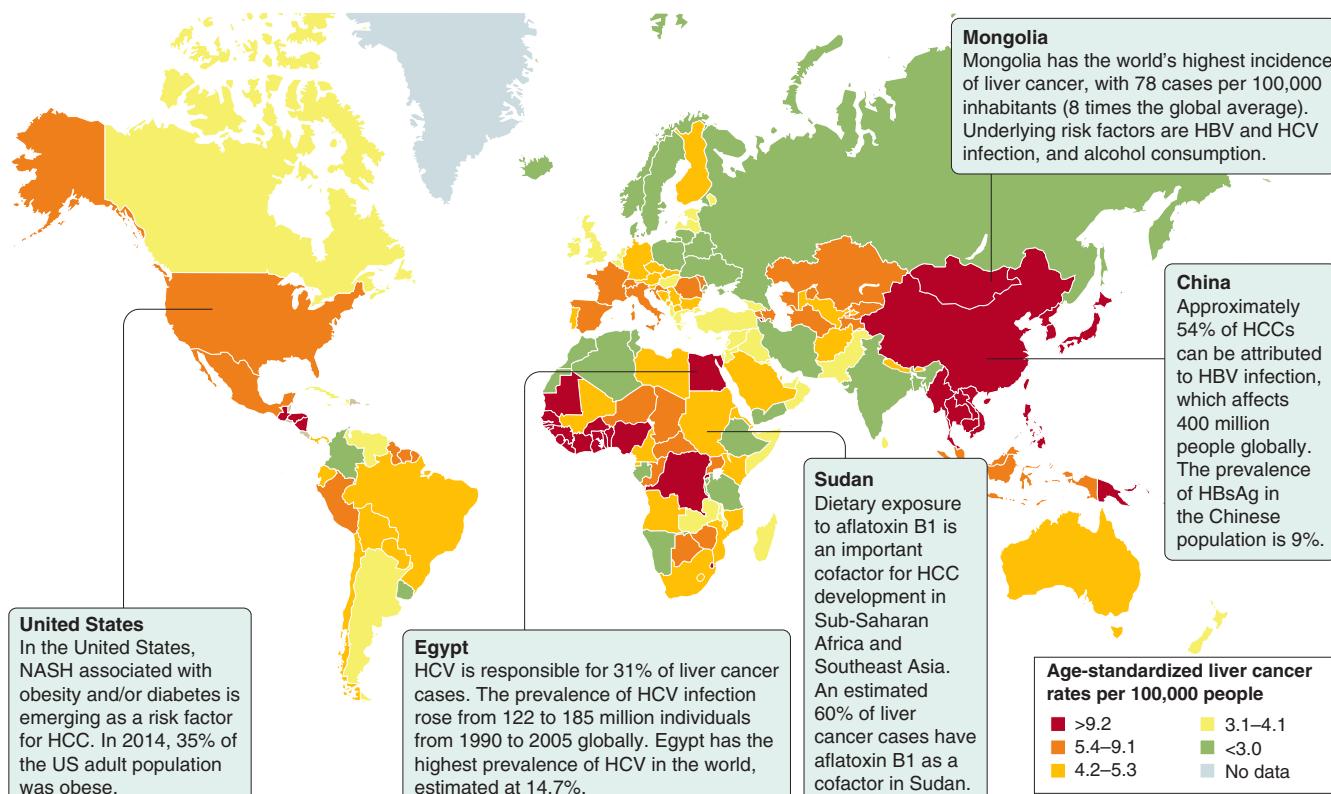


FIGURE 78-2 The global burden of hepatocellular carcinoma. Incidence of hepatocellular carcinoma (HCC) and risk factors. ASR, Age-standardized rate per 100,000 inhabitants. The main risk factors for HCC development are HBV infection (China), HCV infection (example: Egypt), alcohol intake, non-alcoholic steatohepatitis (United States), and aflatoxin B1 (Sudan). Mongolia has the highest incidence of HCC globally, with 78 cases per 100,000 inhabitants. (Reprinted with permission from JM Llovet et al: *Nat Rev Dis Primers* 2:16018, 2016.)

HCC in developed countries. A *PNPLA3* polymorphism is strongly associated with fatty and alcoholic chronic liver diseases and HCC occurrence. Other co-factors contributing to HCC development are tobacco and aflatoxin B1, a fungal carcinogen present in food supplies that induces *TP53* mutations. Finally, infection with adeno-associated virus 2 is associated with HCC in individuals without cirrhosis. Aside from the associations described above, genome-wide association studies have not yet confirmed polymorphisms predisposing to HCC development.

There is a growing incidence of HCC world-wide. The growth in U.S. incidence results from the emergence of end-stage liver disease due to hepatitis C, the increase in HBV-related HCC among immigrants from endemic countries, and the accelerating prevalence of obesity and fatty liver disease. In Europe and Asia, the growth has been less prominent. In some countries the incidence is declining, like Italy and Japan, a country where the impact of HCV-related HCC was first noticed after World War II. Finally, the impact of universal infant vaccination against HBV has decreased the rate of HBV-related HCC in endemic countries, such as Taiwan.

MOLECULAR PATHOGENESIS

HCC development is a complex multistep process that starts with pre-cancerous cirrhotic nodules, so-called low-grade dysplastic nodules (LGDN) that evolve to high-grade dysplastic nodules (HGDN) that can transform into early-stage HCC. Molecular studies support the pivotal role of adult hepatocytes as the cell of origin, either by directly transforming to HCC or by de-differentiating into hepatocyte precursor cells. Alternatively, progenitor cells also give rise to HCC with progenitor markers.

Genomic analysis has provided a clear picture of the main drivers responsible for HCC initiation and progression. This tumor results from the accumulation of around 35–40 somatic genomic alterations per tumor, among which 4–8 are considered driver cancer genes. HCC is a prototypical inflammation-associated cancer, where immune microenvironment and oxidative stress present in chronically damaged livers play pivotal roles in inducing mutations. In pre-neoplastic HGDN, mutations in telomere reverse transcriptase (*TERT*) gene (20% cases) and gains in 8q have been described. Oncogenic transformation occurs upon additional genomic hits including Wnt/β-catenin pathway activation, re-expression of fetal genes, deregulation of protein folding machinery and the response to oxidative stress. Genomic studies and next-generation sequencing conducted during the past decade enables a description of the landscape of mutations, signaling pathways and a molecular classification of the disease. Nonetheless, none of these data have yet translated into actual clinical benefits for molecularly defined tumor subgroups.

Molecular Drivers The landscape of mutational drivers in HCC identified by deep-genome sequencing of ~1,000 samples is detailed in **Table 78-1**. The most common mutations are in the telomerase reverse transcriptase (*TERT*) promoter (56%), *TP53* (27%), *CTNNB1* (26%), *ARID2* (7%), *ARID1A* (6%), and *AXIN1* (5%) genes. These mutated genes participate in cell-cycle control and senescence -*TERT* or *TP53*-, in cell differentiation—*CTNNB1* and *AXIN1*, and chromatin remodeling -*ARID2* and *ARID1A*. Genes commonly mutated in other solid tumors such as *EGFR*, *HER2*, *PIK3CA*, *BRAF*, or *KRAS* are rarely mutated in HCC (<5%). Thus, the most prominent drivers are not currently targetable. Some risk factors have been associated with specific molecular aberrations. HBV integrates into the genome of driver genes, such as the *TERT* promoter, *MLL4*, and cyclin E1 (*CCNE1*). HCV infection and alcohol abuse have been significantly associated with *CTNNB1* mutations. *TP53* mutations are the most frequent alterations with a specific hotspot of mutation (R249S) in patients with aflatoxin B1 exposure.

Studies assessing copy-number alterations in HCCs have consistently identified: (i) high level amplifications at 5–10% prevalence containing oncogenes in 11q13 (*CCND1* and *FGF19*) and 6p21 (*VEGFA*), *TERT* focal amplification and homozygous deletion of *CDKN2A*; and (ii) common amplifications containing *MYC* (8q gain) and *MET* genes (focal gains 7q31). High-level amplifications of 11q13 include *CCND1* and *FGF19*, which have been demonstrated as prominent oncogenes in HCC and are potential therapeutic targets. Similarly, high-level gains

TABLE 78-1 Molecular Aberrations Common in HCC^a

PATHWAY	TARGET	PREVALENCE (%)
Mutations		
Telomere stability	<i>TERT</i> promoter	56
p53/cell cycle control	<i>TP53</i>	27
	<i>ATM</i>	3
	<i>RB1</i>	3
Wnt/β-catenin signaling	<i>CTNNB1</i>	26
	<i>AXIN1</i>	5
Chromatin remodeling	<i>ARID1A</i>	6
	<i>ARID2</i>	7
	<i>KMT2A</i>	3
	<i>KMT2C</i>	3
Ras/PI3K/mTOR pathway	<i>RPS6KA3</i>	3
	<i>TSC1/TSC2</i>	3
Oxidative stress	<i>NFE2L2</i>	3
	<i>KEAP1</i>	3
High-Level Focal Amplifications		
VEGF signaling	<i>VEGFA</i>	3
FGF signaling	<i>FGF19</i>	6
Cell-cycle control	<i>CCND1</i>	7
Target with Homozygous Deletion		
TP53/cell-cycle control	<i>CDKN2A</i>	5
	<i>TP53</i>	4
	<i>RB1</i>	4
Wnt/β-catenin signaling	<i>AXIN1</i>	3

^aRecurrent mutations, focal amplifications or homozygous deletions in HCC based on next-generation sequencing analyses.

of 6p21 containing more than four copies of *VEGFA* were identified in 4–8% of HCCs. *VEGFA* amplification can induce tumor proliferation by unleashing macrophage-mediated hepatocyte growth factor secretion.

Signaling Pathways Several signaling pathways have been implicated in HCC progression and dissemination. Activation of these pathways can result from structural alterations (mutations and amplifications/losses), or epigenetic modifications. In brief, (a) *TERT* overexpression occurs in 90% of cases, particularly related to promoter *TERT* mutations or amplifications; (b) inactivation of p53 and alterations of cell cycle are major defects in HCC, particularly in cases related to HBV infection; (c) Wnt/β-Catenin pathway activation occurs in 50% of cases, either as a result of β-catenin or *AXIN1* mutation, or overexpression of Frizzled receptors or inactivation of E-cadherin; (d) PI3K/PTEN/Akt/mTOR pathway is activated in 40–50% of HCCs due to mutation and focal deletion of the tuberous sclerosis complex (TSC1/TSC2) genes, *PTEN* or ligand overexpression of EGF or IGF upstream signals; (e) Ras MAPK signaling is activated in half of early and almost all advanced HCCs, activation results from up-stream signaling by EGF, IGF, and MET activation, and from the epigenetic silencing of tumor suppressors such as *NORE1A* and *RASSF1A15*; (f) insulin-like growth factor receptor (IGFR) signaling is activated in 20% of cases through overexpression of the oncogenic ligand IGF2 or allelic loss affecting the tumor suppressor IGF2R; (g) dysregulation of the c-MET receptor and its ligand HGF, critical for hepatocyte regeneration after liver injury, are common events in advanced HCC (50%); (h) vascular endothelial growth factor (VEGF) signaling is the cornerstone of angiogenesis in HCC, along with activated angiogenic pathways such as Ang2 and FGF signaling; and (i) chromatin remodelling complexes and epigenetic regulators are frequently altered in HCC due to *ARID1A* and *ARID2* mutations. Several agents that target these different processes are currently being tested in Phase I-III trials.

Molecular Classes and Prognostic Gene Signatures

Genomic studies have revealed two molecular subclasses of HCC, each representing ~50% of patients. The proliferative subclass is enriched by activation of Ras, mTOR, and insulin-like growth factor (IGF) signaling

and FGF19 amplification, and is associated with HBV-related etiologies, overexpression of α -fetoprotein and poor outcomes (particularly those tumors enriched in progenitor cell markers). By contrast, the so called non-proliferative subclass contains a subtype characterized by *CTNNB1* mutations and better outcome. Another classification based upon immune status has been proposed. It defines an immune HCC class in ~25% of cases characterized by immune infiltrate with expression of PD1/PDL1, enrichment of T-cell activation, and better outcome. Direct translation of molecular subclasses into clinical decision-making is yet to be achieved.

■ PREVENTION AND EARLY DETECTION

Prevention Primary prevention of HCC can be achieved by vaccination against HBV and effective treatment of HBV and HCV infection. Studies assessing the impact of universal vaccination against HBV infection started in Taiwan in 1984 have reported a significant decrease of the incidence of HCC. Nowadays, HBV vaccination is recommended to all newborns and high risk groups, following World Health Organization guidelines. Vaccination is also recommended in people with risk factors for acquiring HBV infection, such as health workers, travelers to areas where HBV-infection is prevalent, injecting drug users, and people with multiple sex partners.

Effective antiviral treatments for patients with chronic HBV infection—achieving undetectable viral titres (circulating HBV-DNA)—reduce the risk of HCC development. Evidence of this effect is supported by one randomized trial and several cohort studies. Regarding HCV infection, eradication of hepatitis C results in decreased HCC incidence. Anti-viral therapies achieving a sustained virological response (SVR) in patients with chronic hepatitis prevent the development of advanced stage disease and cirrhosis, hence resulting in a decreased risk of HCC development. However, once cirrhosis is established no high-level evidence suggests that SVR leads to HCC prevention. A meta-analysis of observational studies concluded that interferon-based regimens achieving SVR in patients with cirrhosis were associated with a substantially reduced risk of HCC development. Treatment of HCV has dramatically advanced with the new DAAs (drug antiviral agents) that yield >90% SVR rates after 12 weeks of treatment. A few observational studies with a short follow-up reported an HCC annual incidence of 3–5% in patients with cirrhosis following successful DAA therapy, an incidence similar to that of untreated patients, and higher than those observed with interferon-based therapies. Similarly, there is controversy on the effect of DAA-based SVR on HCC recurrence after curative therapies. Some studies suggest a 6-month recurrence rate higher than historical controls, thus emphasizing the need for large prospective studies. It is too early to estimate the effect of DAA therapy on the burden of HCC. Due to all these circumstances, surveillance remains recommended in patients with cirrhosis achieving SVR.

Additional putative chemopreventive agents have been proposed to reduce HCC incidence in at-risk populations, including statins and metformin. Nonetheless, the evidence is not strong enough to recommend using these therapies in at-risk patients. Finally, coffee consumption is associated with a reduced risk of HCC in population studies.

Surveillance The aim of surveillance is to obtain a reduction in disease-related mortality. This is usually achieved through early detection that enhances the applicability and cost-effectiveness of curative therapies. United States and European guidelines recommend surveillance for patients at high risk for HCC on the basis of cost-effectiveness analyses. As a general rule, high-risk populations are considered those presenting an incidence cut-off > 1.5% for patients with cirrhosis and 0.2% for patients with chronic hepatitis B. However, the strength of evidence supporting surveillance is modest, and is based upon two randomized studies conducted in China and meta-analysis of observational studies. Overall, these studies conclude that surveillance identifies patients with smaller tumors who are more likely to undergo curative procedures. Because of lead time bias and length time bias it cannot be concluded that surveillance ultimately reduces HCC-related mortality.

Surveillance is recommended for cirrhotic patients owing to any cause, those with HCV-related advanced fibrosis (Metavir score of F3), and for patients with chronic HBV infection if Asian aged >40 years,

African aged >20 years or family history of HCC. In terms of liver dysfunction, the presence of advanced cirrhosis (Child-Pugh class C) prevents potentially curative therapies from being employed, and thus surveillance is not recommended. As an exception, patients on the waiting list for liver transplantation, regardless of liver functional status, should be screened for HCC in order to detect tumors exceeding conventional criteria and to define priority policies for transplantation. Complex scoring systems to identify at-risk populations are not yet recommended by guidelines.

Ultrasonography every 6 months is the recommended method of surveillance. It has a sensitivity of 65–80% and a specificity of >90% for early detection. A 3-month interval does not enhance outcomes, and survival is lower with 12 month compared with 6 month intervals. A shorter follow-up interval (every 3–4 months) is recommended when a nodule of <1 cm has been detected. Computed tomography (CT) and magnetic resonance imaging (MRI) are not recommended as screening tools due to lack of data on accuracy, high cost and possible harm (i.e., radiation with CT). Exceptionally, these techniques can be considered in patients with obesity and fatty liver, where visualization with ultrasound is difficult. Accurate tumor biomarkers for early detection need to be developed. Use of alpha-fetoprotein (AFP) levels identifies patients with HCC with 60% sensitivity, but high false-positive results. One main limitation of AFP is that only a small proportion of early tumors (~20%) present with abnormal AFP serum levels. Combining AFP with ultrasound performed by experienced personnel only increase 6–8% the HCC detection rate. Nonetheless, testing AFP is still widely used and this remains an area of controversy. Particularly, testing AFP might be considered in special populations or health care environments when ultrasound is not available. The accuracy of other serum biomarkers proposed, such as des- γ carboxyprothrombin (DCP) and the L3 fraction of AFP (AFP-L3), in early detection is not known.

Despite the fact that surveillance is cost-effective in HCC, the global implementation of such programs is estimated to engage ~50% of the target population in Europe and ~30% in the United States. Public health policies encouraging the implementation of such programs should lead to an increase in early tumor detection.

Diagnosis HCC is generally diagnosed at early or intermediate stages in Western countries, but at advanced stages in most Asian (except Japan) and African countries. A surveillance program yields early diagnosis in 70–80% of cases. At these stages the tumor is asymptomatic, and diagnosis can be made by non-invasive (radiological) or invasive (biopsy) approaches. Without surveillance, HCC is discovered either as a radiological finding or due to cancer-related symptoms. If symptoms are present the disease is already at an advanced stage with a median life expectancy <1 year. Symptoms include malaise, weight loss, anorexia, abdominal discomfort, or signs related to advanced liver dysfunction.

NON-INVASIVE (RADIOLOGICAL) DIAGNOSIS Patients enrolled in a surveillance program are diagnosed by identification of a new liver nodule on abdominal ultrasound. Non-invasive criteria can only be applied to cirrhotic patients and are based on imaging techniques obtained by 4-phase multidetector CT scan (four phases are unenhanced, arterial, venous, and delayed) or dynamic contrast-enhanced MRI. A flow-chart of diagnosis and recall policy recommended by U.S. and European guidelines is summarized in Fig. 78-3. Radiological diagnosis is achieved with a high degree of confidence if the lesion is ≥2 cm in diameter and shows the *radiological hallmarks of HCC* by one imaging technique. Using contrast-enhanced imaging techniques, the typical hallmark of HCC consists of vascular uptake of the nodule in the arterial phase with washout in the portal venous or delayed phases. This radiological pattern captures the hypervascular nature characteristic of HCC. In these scenarios the diagnostic specificity is ~95–100% and a biopsy is not necessary. For lesions 1–2 cm in diameter, the radiological hallmarks of HCC define diagnosis, but need to be confirmed by two imaging techniques in non-specialized centers. Nodules <1 cm in size are unlikely to be HCC and would be very difficult to diagnose, and thus ultrasound follow-up at 3–4 months is recommended. MRI with liver specific contrast agents might help in the diagnosis of HCC, but

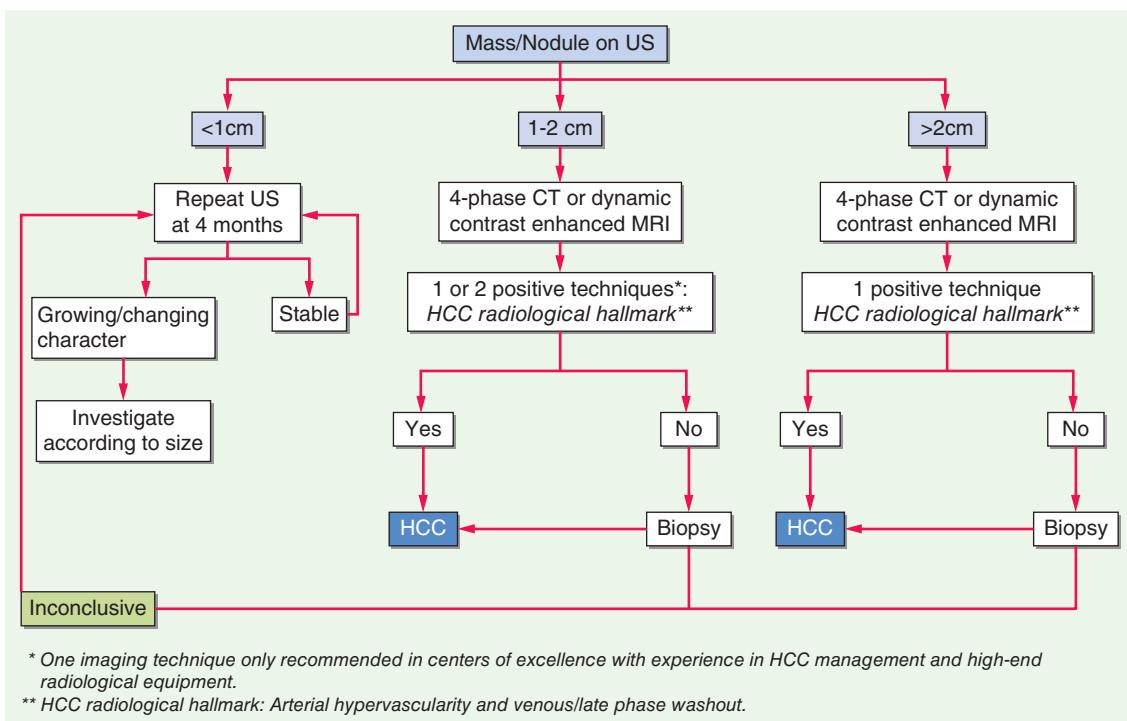


FIGURE 78-3 Recall diagnosis schedule for HCC (EASL). EASL, European Association for the Study of Liver Disease; HCC, hepatocellular carcinoma. (Reprinted with permission from EASL-EORTC guidelines. *J Hepatology* 56: 908, 2012.)

the specificity of these agents is still suboptimal. Contrast-enhanced ultrasound (CEUS) and angiography are less accurate for HCC diagnosis. Positron emission tomography (PET)-scan performs poorly for early diagnosis. AFP levels ≥ 400 ng/dL are highly suspicious, but not diagnostic of HCC according to guidelines.

PATHOLOGICAL DIAGNOSIS Pathological diagnosis is required in two scenarios: (a) in patients without cirrhosis, and (b) if radiology is not typical in at least one of two imaging techniques (CT and MRI). This occurs mainly with early-stage HCC lesions. Biopsy is not an ideal gold standard, because of variation introduced by sampling and complications. Sensitivity of liver biopsies ranges between 70 and 90% for all tumor sizes, but decrease to <50% in tumors 1–2 cm in size. The risk of complications such as tumor seeding and bleeding after liver biopsy is ~3%. Biopsies should be assessed by an expert hepatopathologist. The use of special stains may help to resolve diagnostic uncertainties. Positive staining in two of four markers [glypican 3 (GPC3), glutamine synthetase, heat shock protein 70 (HSP70), and clathrin heavy chain] is highly specific for HCC. Gene expression blueprints (glypican 3, LYVE1, and survivin) are also able to differentiate high grade dysplastic nodules from early HCC. Additional staining can be considered to detect progenitor cell features (K19 and EpCAM) or assess neovascularization (CD34). A negative biopsy does not eliminate the diagnosis of HCC. A second biopsy is recommended in case of inconclusive findings, or growth or change in enhancement pattern identified during follow up. European guidelines advocate obtaining tissue samples in the setting of all research studies in HCC, even if radiological criteria are met.

TREATMENT

Staging Systems and Treatment Allocation Staging systems are aimed at stratifying patients according to prognostic factors and outcome, and to allocate the best available therapies according to evidence. The most accepted staging system is the Barcelona-Clinic-Liver Cancer (BCLC) Classification, which is endorsed by U.S. and European clinical practice guidelines (Fig. 78-4). This staging system defines five prognostic subclasses and allocates specific treatments for each stage. The BCLC staging system has been externally validated by numerous studies. It is an evolving system that allows incorporation of new therapies and treatment-dependent variables as new evidence emerges. Six

treatments have been demonstrated to improve survival in HCC, five are adopted by guidelines and BCLC classification: surgical resection, liver transplantation, radiofrequency (RF) ablation, chemoembolization, and systemic therapies (sorafenib, regorafenib, lenvatinib, cabozantinib, ramucirumab). The BCLC system will also incorporate lenvatinib in first line and regorafenib as standard of care in patients with advanced HCC progressing on sorafenib as a consequence of a positive randomized controlled trial (RCT). The BCLC assigns each patient with a given treatment allocation. Treatment stage migration is also applied by this scheme, meaning that if patients are not candidates for the selected therapy, the next effective therapy at more advanced stages can be given.

In HCC, three parameters are relevant for defining treatment strategy: tumor status, cancer-related symptoms, and liver dysfunction. The BCLC staging captures all three variables and allocates patients to treatments according to evidence. Since >80% of patients have two diseases, HCC and cirrhosis, a clear measurement of liver dysfunction should be in place. The prognosis of chronic liver disease is commonly assessed using the Child-Pugh score, which uses five clinical measures—total bilirubin, serum albumin, prothrombin time, ascites severity, and hepatic encephalopathy grade—to classify patients into one of three groups (A–C) of predicted survival rates. In brief, Child-Pugh's A reflects well-preserved liver function, Child's B moderate liver dysfunction with a median life expectancy of ~3 years and Child C severe liver dysfunction with life expectancy of ~1 year. At early BCLC stages more granular criteria to define patients with very-well preserved liver function (Child-Pugh's hyper-A class without portal hypertension) needs to be in place to select candidates for resection. Modifications of Child-Pugh scoring or model for end-stage liver disease (MELD) score have not been adopted for treatment allocation, except for prioritization on the waiting list for liver transplantation (MELD score). More sophisticated measures of liver dysfunction (i.e., assessment of portal hypertension) are recommended for preoperative assessment of candidates for resection. Performance status is assessed by Eastern Cooperative Oncology Group (ECOG) and presence of cancer-related symptoms (ECOG 1–2) is considered a sign of advanced stage. Patients with severe liver dysfunction (Child-Pugh's C class) or performance status impairment (ECOG 3–4) are offered supportive care management.

Considering all these prognostic/predictive variables and evidence-based treatment efficacy, five BCLC stages have been defined

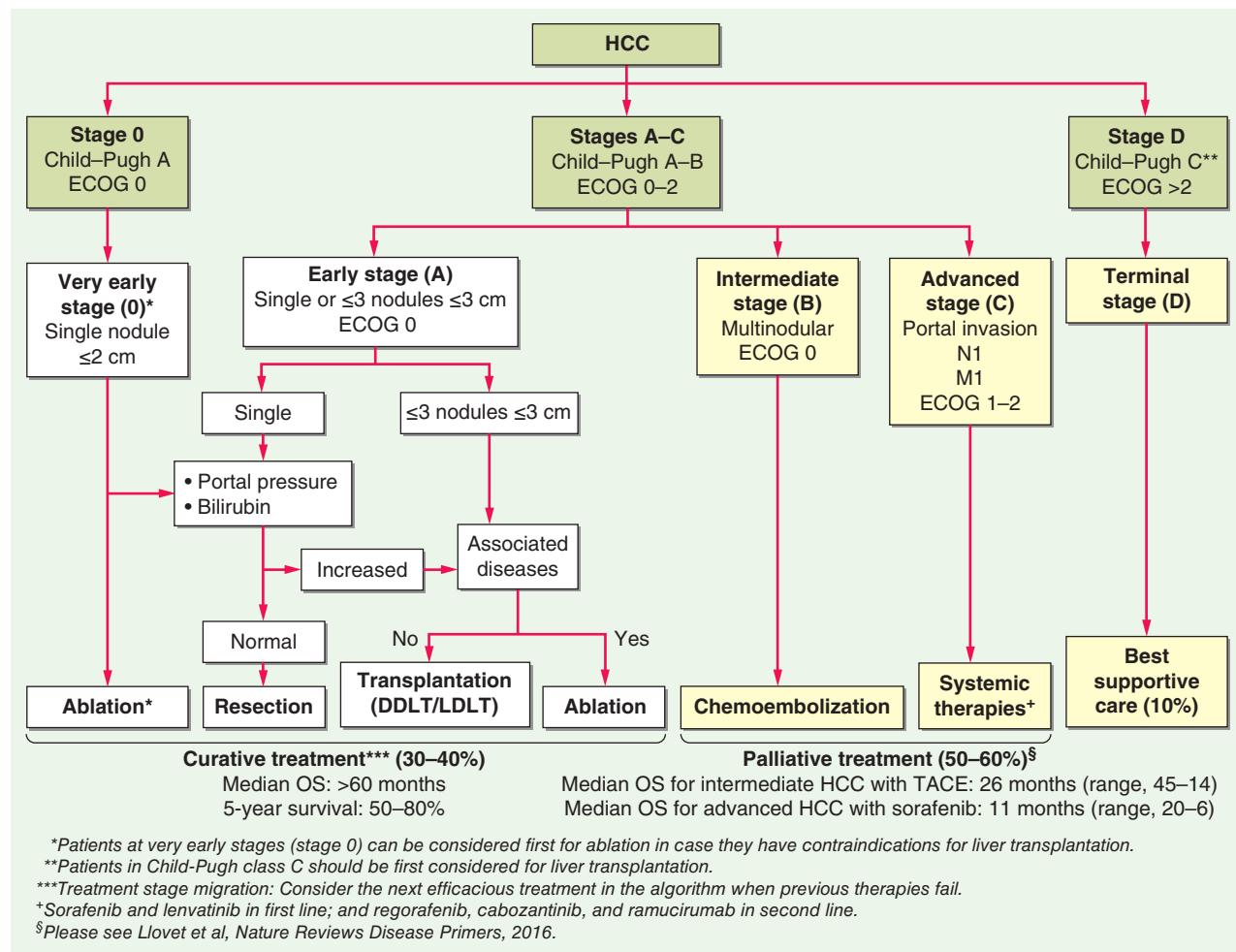


FIGURE 78-4 BCLC staging system and therapeutic strategy. BCLC classification comprises five stages that select the best candidates for therapies according to evidence-based data. Patients with asymptomatic early tumors (stages 0-A) are candidates for radical therapies (resection, transplantation, or local ablation). Asymptomatic patients with multinodular HCC (stage B) are suitable for transcatheter arterial chemoembolization (TACE), whereas patients with advanced symptomatic tumors and/or an invasive tumoral pattern (stage C) are candidates to receive sorafenib. End-stage disease (stage D) includes patients with poor prognosis that should be treated by best supportive care. BCLC, Barcelona Clinic Liver Cancer; DDLT, deceased donor liver transplantation; EASL, European Association for the Study of Liver Disease; ECOG, Eastern Cooperative Oncology Group Performance Status; EORTC, European Organisation for Research and Treatment of Cancer; GRADE, grading of recommendations assessment, development, and evaluation; HCC, hepatocellular carcinoma; LDLT, living donor liver transplantation; OS, overall survival; PEI, percutaneous ethanol injection; RF, radiofrequency ablation; TACE, transcatheter arterial chemoembolization. (Reprinted with permission from JM Llovet et al: *Nat Rev Dis Primers* 2:16018, 2016; A Forner, JM Llovet, J Bruix: *Hepatocellular carcinoma*. *Lancet* 379:1245, 2012.)

(Fig. 78-4). Patients with liver only neoplastic disease, no symptoms (ECOG 0) and with mild to moderate liver dysfunction (Child-Pugh A-B) can be classified as very early (Stage 0) or early (Stage A) or intermediate (stage B) stages depending upon tumor size and number. Very early HCC (BCLC 0) is defined by single tumors ≤2 cm (if pathology is available they should be well-differentiated with absence of microvascular invasion or satellites). Early HCC (BCLC A) includes either single tumors or a maximum of three nodules of ≤3 cm in diameter. Intermediate stage (BCLC B) is defined by all other liver-only tumors. Conversely, HCC is considered at advanced stages (BCLC C) when patients present cancer-related symptoms (ECOG 1-2) or tumors with macrovascular invasion (of any type, including branch, hepatic, or portal vein), lymph node involvement, or extrahepatic spread. Finally, end-stage disease (BCLC D) is considered in cases of several impairment of quality of life/cancer-related symptoms (ECOG 3-4) or severe liver dysfunction (Child-Pugh C).

Around 40% of patients are diagnosed at Stages 0 and A, and hence are eligible for potentially curative therapies, resection, transplantation, or local ablation. These treatments provide median survival rates of 60 months and beyond, which are in sharp contrast with outcomes of 36 months reported in historical controls (Fig. 78-5). No adjuvant therapy is recommended. Patients at intermediate stage (Stage B) with preserved liver function have a documented natural history of around 16 months. These patients benefit from transarterial chemoembolization (TACE) as

reported in two randomized studies and one meta-analysis, and achieve an estimated survival of 25–30 months. None of the combination therapies with TACE have shown outcome advantages. Patients progressing on TACE or at advanced stage (Stage C) benefit from systemic sorafenib, which extends survival by ~3 months (from 7.9 to 10.7 months). Lenvatinib showed non inferiority results compared to sorafenib (13.6 months vs. 12.3 month, respectively). Regorafenib improves survival from 7.8 to 10.6 months in patients progressing on sorafenib (second-line advanced HCC). Therefore, these treatments have been adopted by guidelines and incorporated to the BCLC classification. Patients with end-stage disease (BCLC D) should be considered for nutritional and psychological support and proper management of pain.

Although the BCLC establishes validated stages and treatment assignment according to evidence, clinical practice is not always aligned with this classification. In large cohort studies and surveys, only half of patients, or even less in Asia, are treated accordingly. Alternative staging or scoring systems have been proposed, but none of them has acquired global consensus. In contrast to BCLC, some proposed systems capture the standard of practice in Asia, such as the Hong Kong classification or the Japan Integrated Staging score. These systems capture extended indications for resection and TACE applied in clinical practice in Asia. Other systems define prognostic stratification, such as the Cancer of the Liver Italian Program (CLIP) score, although they do not incorporate treatment allocation to distinct stages.

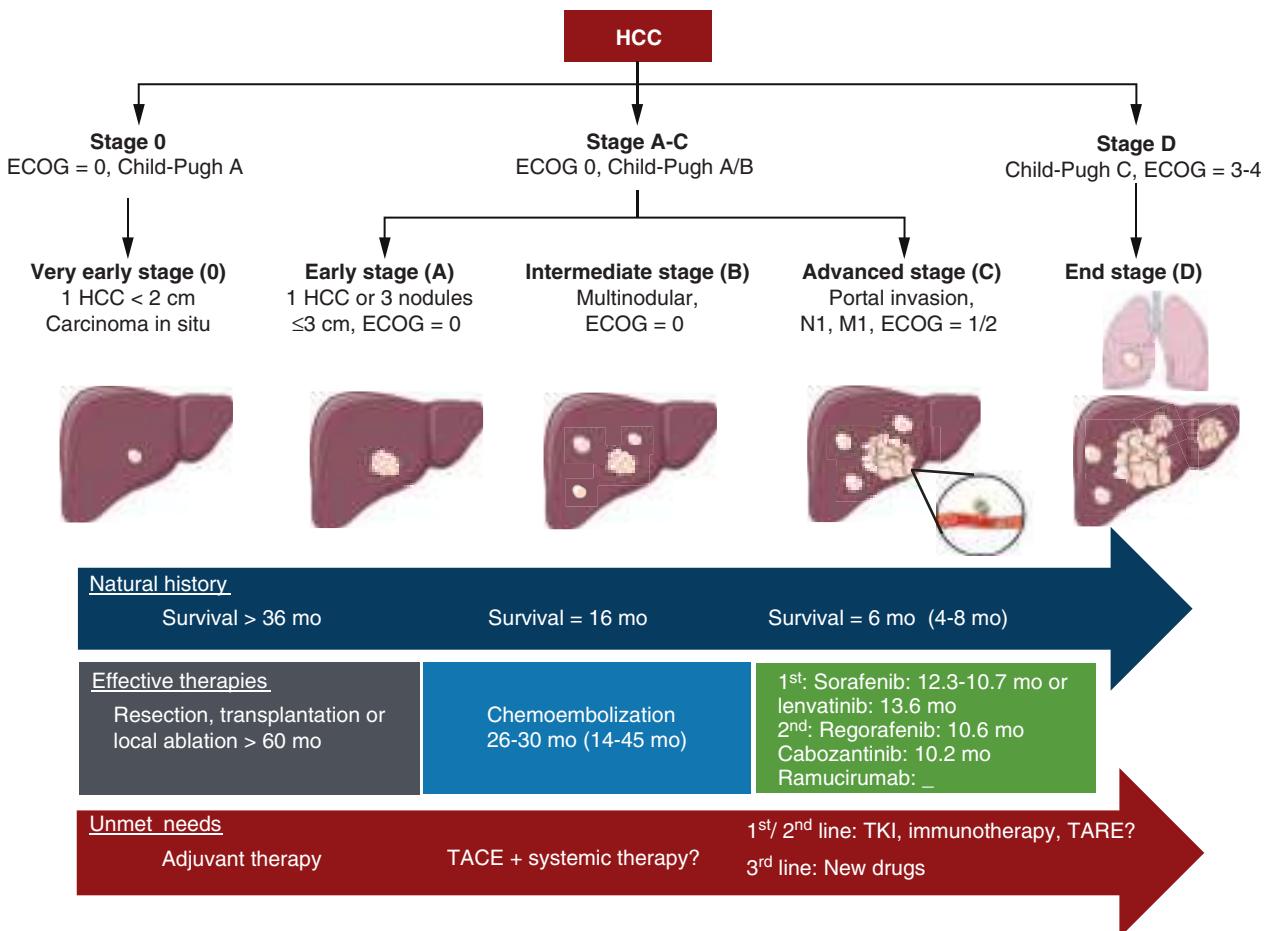


FIGURE 78-5 Natural history, impact of therapies, and unmet needs in HCC.

Finally, the tumor-node-metastasis (TNM) staging system is not used in HCC since it does not incorporate the main prognostic variables related to liver function and performance status.

The BCLC system does not incorporate molecular classes or biomarkers. Some biomarkers (i.e., AFP at a cut-off of >200 or 400 ng/mL) or molecular classes/signatures have prognostic or biological significance. However, they are not ready for clinical application due to the lack of data on biomarker-based response to therapies. A few Phase III studies are currently conducted based upon biomarker-enriched populations (i.e., ramucirumab in AFP >400 ng/mL) or as proof-of-concept studies (i.e., FGFR4 inhibitors in patients with FGF19 amplification/overexpression). Ramucirumab trial has been positive, and thus biomarkers will be incorporated in the treatment allocation system.

Due to the complexities of HCC diagnosis and management, it is recommended to send patients to a referral center where all the armamentarium of therapies can be offered. In principle, patient management and outcome benefit from liver cancer multidisciplinary programs that include hepatologist, oncologist, hepatobiliary and transplant surgeons, interventional and body imaging radiologist, hepatopathologist, and specialized nurses.

SURGICAL THERAPIES

Resection Surgical resection is the first-line option for non-cirrhotic patients at early-stage HCC (BCLC 0 or A) with solitary tumors (Fig. 78-4). In cirrhotic patients, ablation competes with resection for BCLC 0 tumors (<2 cm in diameter). Which is better is not defined. Cost-effectiveness approaches report a benefit for local ablation with RF. For single tumors >2 cm (BCLC A), resection remains the mainstay of treatment in patients with Child-Pugh's hyper-A class, those patients with normal bilirubin and absence of portal hypertension (portal hypertension is defined by hepatic venous pressure gradient ≥10 mmHg). Surrogate measures of portal hypertension are presence of esophageal varices or platelet count <100,000/mm³ associated with

splenomegaly. Anatomic resections following the functional segments of the liver are recommended to spare uninvolved liver parenchyma and to remove satellite tumors. Predictors of recurrence are tumor size, number, presence of microsatellites, or microvascular invasion at the specimen analysis. Macrovascular invasion, extrahepatic involvement, and liver dysfunction (Child-Pugh B-C) are major contraindications for resection.

ADJUVANT TREATMENTS Tumor recurrence represents the major complication of resection (and local ablation) and occurs in 70% of cases at 5 years. Most of recurrences are intrahepatic metastases, but at least one third are considered *de novo* tumor, new clones developing in the cirrhotic carcinogenic field. The type of recurrence can only be defined by molecular studies. So far, no adjuvant therapies have proven to improve outcome or prevent recurrence after resection/ablation. Randomized trials testing adjuvant sorafenib, retinoids, chemotherapies or chemoembolization have been negative. Some trials testing adoptive immunotherapy or interferon showed positive results, but these treatments have not been adopted by guidelines of management due to weakness of the evidence or small magnitude of benefit. Therefore, the current recommendations are that in case of recurrence patients will be re-assessed by BCLC staging, and re-treated accordingly.

Liver Transplantation Liver transplantation is the first treatment choice for cirrhotic patients with single tumors ≤5 cm and portal hypertension (including Child-Pugh's B and C) or with small multinodular tumors (≤3 nodules, each ≤3 cm) (Fig. 78-4). These so-called Milan criteria have been validated over the years and a meta-analysis reported 5-year and 10-year survival rates of ~70 and ~50%, respectively, similar to outcomes achieved in non-HCC transplantation indications. Perioperative mortality rates have been reduced to <3%. Transplantation simultaneously cures the tumor and the underlying cirrhosis, and it is associated with a low risk of recurrence, around 10–15% at 5 years. No immunosuppressive regimens or anti-tumor therapies

after transplantation have demonstrated any preventive effect on recurrence. Milan criteria are integrated in the BCLC treatment strategy (BCLC 0 and A) and have also been adopted by the United Network for Organ Sharing (UNOS) pre-transplant staging for organ allocation in the United States (Stage T2). Aside from size and number, conventional contraindications for organ transplantation procedures (ABO incompatibility, co-morbidities, etc.) are applied in this setting.

Liver transplantation has a couple of important limitations, such as cost and donor availability, which limit this procedure to <5% of HCC cases. The scarcity of donors represents a major drawback of liver transplantation. Donor scarcity varies geographically, and deceased liver donation is almost zero in some Asian countries. Due to the shortage of donors, median waiting times in Western programs is ~6–12 months leading to 20% of candidates dropping off the list due to tumor progression before receiving the procedure. Predictors of drop-out are treatment failure, baseline AFP >400 ng/mL or steady increase of AFP level >15 ng/mL per month. Several strategies have been proposed to overcome this limitation. First, apply neo-adjuvant therapies in patients on the waiting list. Neo-adjuvant treatments testing TACE or RF ablation have been assessed in the setting of cohort and cost-effectiveness studies. In principle, the use of these therapies is recommended when the waiting time exceeds 6 months, even though impact on long-term outcome is uncertain. Second, a priority policy has been established for patients enlisted. UNOS has implemented a scoring system based upon the dropout risk, giving priority to tumors 2–5 cm in size and multinodular tumors.

The Milan criteria are universally used as the basis for transplant eligibility, and adherence to them yields good post-transplant survival. Modest expansion of Milan criteria applying the “up-to-seven” criteria (i.e., those HCCs having the number 7 as the sum of the size of the largest tumor and the number of tumors) in patients without microvascular invasion achieves competitive outcomes. These pathologically-defined criteria are being used in clinical practice to predict the expected outcome after transplantation. Similarly, *down-staging to Milan criteria* is currently explored by several groups. Down-staging is defined as the reduction of HCC burden by loco-regional treatments to achieve Milan staging before transplantation. A few studies claim that down-staging lasting for >3 months achieves competitive outcomes, but robust long-term survival data is scarce, and thus it cannot yet be recommended. Down-staging policy is only endorsed by guidelines for patients outgrowing the Milan criteria while on the waiting list.

Since policies for enhancing organ donation have reached a ceiling during the past several years, alternatives to donation have emerged. Living donor liver transplantation represents a plausible alternative that accounts of ~5% of total transplants performed globally. Outcomes reported are similar to those with deceased liver donors, and it is recommended as an alternative option in patients on a waiting list exceeding 6–7 months. The risks and benefits of this procedure should take into account both donor (death is estimated in 0.3%) and recipient, a concept known as *double equipoise*. Due to the complexity of this treatment, it must be restricted to centers of excellence in hepatobiliary surgery and transplantation.

■ LOCO-REGIONAL THERAPIES

Local Ablation RF ablation is recommended as the primary ablative technique (Fig. 78-4). The energy generated by RF ablation (heating of tissue at 80°–100°C) induces coagulative necrosis of the tumor producing a *safety ring* in the peritumoral tissue, which might eliminate small-undetected satellites. Treatment consists of 1 or 2 sessions performed using a percutaneous approach, although in some instances ablation with laparoscopy is needed. RF ablation is more effective in response rate and time-to-recurrence compared with the once-conventional percutaneous ethanol injection. Long-term outcome of HCC patients treated by RF ablation have 5-year survival rates of ~60%. In tumors <2 cm, BCLC 0, RF ablation achieves complete responses in >90% of cases with good long-term outcome and is competitive with resection in terms of cost-effectiveness. For BCLC A cases, RF ablation is the first-line treatment for single tumors 2–5 cm or multinodular up to three nodules, each ≤3 cm in diameter, unsuitable for surgery.

The main limitation of RF ablation is that its failure rate increases in tumors >3 cm because of the heat loss due to perfusion-mediated tissue cooling within the area ablated. In tumors 3–5 cm in diameter, complete pathological tumor necrosis of <50% has been reported. Particularly, ~10–15% of tumors with difficult-to-treat locations, such as a subcapsular location or adjacent to the gallbladder, have a higher risk of incomplete ablation or major complications and can be approached by ethanol injection. Several approaches have been proposed to enhance the anti-tumor activity of RF ablation. The combination of RF ablation with either chemoembolization or with a heat-activated formulation of liposomal doxorubicin yielded good results in cohort studies. Other treatments, such as microwave ablation, high-intensity focused ultrasound or stereotactic body radiotherapy for small tumors are under investigation.

Chemoembolization TACE is the most widely used primary treatment for unresectable HCC worldwide, and the first-line indication for patients with intermediate BCLC B stage (Fig. 78-4). Conventional chemoembolization (c-TACE) consists of the local hepatic artery administration of chemotherapy (either doxorubicin 50 mg/m² or cisplatin) mixed with an emulsion of lipiodol followed by obstruction of the feeding artery with sponge particles. c-TACE mainly benefits patients with liver-only disease, Child-Pugh A Class, or B without ascites, good performance status (ECOG 0), and absence of branch or trunk vascular invasion. Median survival is ~20 months (compared to 16 months for pooled control arms). The best randomized trial and subsequent Phase II studies have provided median survivals for TACE of 25–30 months in properly selected populations. Median objective response rates are of 50–70%. In randomized studies, the treatment is performed at a regular schedule of 0, 2, and 6 months (median number of sessions: 3), although no consensus has been established. TACE procedures should be stopped upon tumor progression or any other contraindication. Exceptionally, occurrence of a new small untreated nodule as the only progression feature can be considered for treatment. Around ~50% of patients present a limited postembolization syndrome of fever and abdominal pain related to ischemic injury and release of cytokines. Less than 5% of patients present major complications (liver abscess, ischemic cholecystitis, or liver failure) and in <2% of cases treatment-related death occurs.

Applicability of c-TACE in BCLC B patients is limited to half of cases, mostly as a result of the presence of liver failure (Child B, or ascites or encephalopathy), technical contraindications to the procedure (i.e., impaired portal-vein blood flow), or infiltrative/massive tumor burden (i.e., generally main tumor size >10 cm). Super-selective TACE minimize the ischemic insult to non-tumor tissue. According to guidelines, treatment-stage migration allows performing TACE on patients at early stages not suitable for surgical or ablative therapies. In selective studies, median survival rates of 5 year have been reported in patients with single HCC treated by supra-selective TACE. On the other hand, TACE performed beyond guidelines as a conventional practice to patients with formal contraindications (generally BCLC C) yields poor outcomes.

Drug-eluting beads chemoembolization (DEB-TACE) differs from c-TACE in the use of more standardized embolic spheres of regular size embedded with chemotherapy. This strategy ensures drug release over a 1-week period resulting in an enhancement of drug concentration within the tumor. DEB-TACE achieves similar anti-tumor activity (objective responses of ~60%) as c-TACE associated with significantly less systemic cytotoxic effects and better tolerance, but with no clear differences in clinical outcomes. Phase II and III studies have compared DEB-TACE with the combination of DEB-TACE with sorafenib or brivanib, a VEGF receptor inhibitor. Median survival in both arms of these international trials was 25–30 months.

■ Radioembolization and Other Intraarterial Therapies

Radioembolization using beads coated with yttrium-90 (Y-90)—an isotope that emits short-range β radiation—is the most promising alternative to TACE. Several Phase II studies reported objective responses and overall outcome with a safe profile. Due to the lack of Phase III trials, this treatment is currently not recommended in guidelines. Whether radioembolization might be effective in patients at an intermediate-stage not eligible for TACE needs to be studied. Radioembolization

requires prevention of severe lung shunting and intestinal radiation before the procedure. Around 20% of patients present liver-related toxicity and 3% treatment-related death. Due to the minimally embolic effect of Y-90 microspheres, treatment can be safely used in patients with portal vein thrombosis, a setting where survival results in Phase II were encouraging and Phase III investigations in combination with sorafenib are ongoing. Head to head comparison of Y-90 vs sorafenib did not hit the primary end-point of overall survival.

TACE should be distinguished from other intraarterial therapies, such as chemo-lipiodolization, which involves the delivery of an emulsion of chemotherapy mixed with lipiodol, bland transcatheter embolization (TAE), where no chemotherapeutic agent is delivered, and intra-arterial chemotherapy, where no embolization is performed. None of these approaches is recommended due to the lack of survival benefit.

■ SYSTEMIC THERAPIES

Conventional systemic chemotherapy and radiotherapy have not produced survival advantages. Randomized studies also failed with anti-estrogen therapies and vitamin D derivatives. External beam liver-directed radiotherapy (stereotactic body radiotherapy) efficacy is currently being tested with and without sorafenib in Phase III trials. In 2007 a Phase III trial demonstrated survival benefits for patients with advanced stage disease treated with sorafenib, and more recently lenvatinib showed similar effects to sorafenib in first line treatment. A second multikinase inhibitor, regorafenib, has been shown to benefit patients progressing to sorafenib.

Molecular Targeted Therapies Sorafenib is the standard of care systemic therapy for HCC. It is indicated for patients with well-preserved liver function (Child-Pugh A class) and with advanced tumors (i.e., BCLC C) or those tumors at intermediate stage (i.e., BCLC B) progressing upon loco-regional therapies (Fig. 78-4). A Phase III study comparing sorafenib vs placebo showed increased survival from 7.9 months to 10.7 months (HR 0.69; 31% reduction of risk of death). In

this trial, 80% of patients were BCLC C and 20% BCLC B progressed to TACE. Overall, 35% of patients presented with macrovascular invasion and 50% with extrahepatic spread. A similar magnitude of benefit was observed in another positive Phase III study conducted in parallel in Asian patients, mostly with HBV-related HCC. Interestingly, objective responses account for 2% of patients assessed by RECIST criteria and ~10% assessed by the more refined modified RECIST (mRECIST) criteria. Patients with HCV-related HCC achieve significantly better outcomes with sorafenib, with a median survival of 14 months. No predictive biomarkers of responsiveness to sorafenib have been identified.

The recommended daily dose of sorafenib is 800 mg. Median treatment duration is about six months. Treatment is associated with manageable adverse events, such as diarrhea, hand-foot skin reactions, fatigue, and hypertension. Treatment-related liver failure or life-threatening complications are unusual. These toxicities lead to treatment discontinuation in 20% of patients and dose-reduction in up to half. Not all patients at advanced stages can receive sorafenib. It has been estimated that this therapy cannot be administered to around one-third of the targeted patients due to primary intolerance, advanced age, or liver failure (ascites or encephalopathy). Active vascular disease, either coronary or peripheral, is considered a formal contraindication.

The efficacy of sorafenib probably results from a balance between targeting cancer cells and the microenvironment by blocking up to 40 kinases, including anti-angiogenic (vascular endothelial growth factor receptor [VEGFR], platelet-derived growth factor receptor [PDGFR]), and anti-proliferative drivers (serine/threonine-protein kinase B-raf [BRAF] and mast/stem cell growth factor receptor [c-Kit]). Median time to progression on sorafenib is of 4–5 months in Phase III trials. Activation of MAPK14 signaling, IGF signaling, and enrichment in tumor-initiating cells is the main mechanisms of acquired resistance.

Several other agents have been tested with negative results in most of the cases (Table 78-2). Recently, a phase III study comparing lenvatinib (an inhibitor of VEGFR, fibroblast growth factor receptor [FGFR],

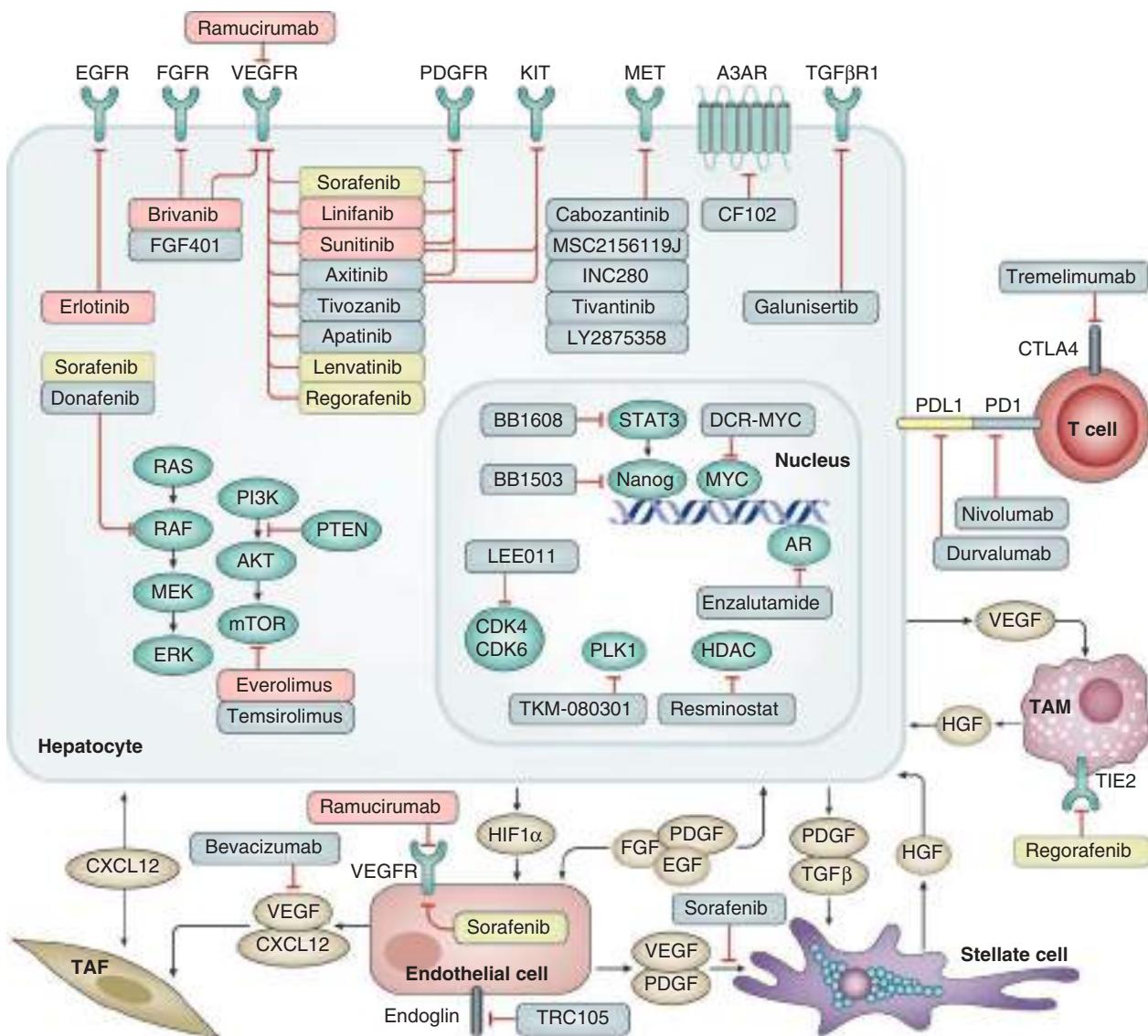
TABLE 78-2 Phase III Trials Testing Molecular Therapies in Advanced HCC the Past 10 Years

	DRUGS**	n	MEDIAN OS (month)	HAZARD RATIO (p-value)	MEDIAN TTP (month)	HAZARD RATIO (p-value)	OBJECTIVE RESPONSE (%)
First-Line							
SHARP	Sorafenib	299	10.7	0.69	5.5	0.58	2.3
	Placebo	303	7.9	(<0.001)	2.8	(<0.001)	0.7
Asian-Pacific	Sorafenib	150	6.5	0.68	2.8	0.57	3.3
	Placebo	76	4.2	(0.01)	1.4	(<0.001)	1.3
Sunitinib	Sunitinib	530	7.9	1.3	4.1	1.13	6.6
	Sorafenib	544	10.2	(0.001)	3.8	(0.308)	6.1
BRISK-FL	Brivanib	577	9.5	1.06	4.2	1.01	12*
	Sorafenib	578	9.9	(0.31)	4.1	(0.853)	8.8*
LIGHT	Linifanib	514	9.1	1.04	5.4	0.76	10.1
	Sorafenib	521	9.8	(0.52)	4	(0.001)	6.1
SEARCH	Sorafenib + Erlotinib	362	9.5	0.92	3.2	1.13	6.6
	Sorafenib	358	8.5	(0.2)	4	(0.18)	3.9
Lenvatinib	Lenvatinib	478	13.6	0.92	8.9	0.63	24*
	Sorafenib	476	12.3	(-)	3.7	(<0.001)	9*
Second-Line							
BRISK-PS	Brivanib	263	9.4	0.89	4.2	0.56	9.9*
	Placebo	132	8.2	(0.33)	2.7	(<0.001)	1.5*
EVOLVE-1	Everolimus	362	7.6	1.05	3	0.93 (NS)	2.2
	Placebo	184	7.3	(0.68)	2.6		1.6
REACH	Ramucirumab	283	9.2	0.86	3.5	0.59	7.1
	Placebo	282	7.6	(0.13)	2.6	(<0.001)	0.7
RESOURCE	Regorafenib	379	10.6	0.63	3.2	0.44	10.6*
	Placebo	194	7.8	(<0.001)	1.5	(<0.001)	4.1*

Abbreviations: NS, not significant; OS, Overall Survival; TTP, Time-to-progression.

*Refers to Objective response by mRECIST criteria. Celestial Cabozantinib: n 470, Median OS (month) 10.2, Hazard ratio (p-value) 0.76, Median TTP (month) 5.5, Hazard ratio (p-value) 0.40, Objective response (%) 4 Placebo: n 237, Median OS (month) 8.0, Hazard ratio (p-value) 0.0049, Median TTP (month) 1.9, Hazard ratio (p-value) p<0.001, Objective response (%) 0.4.

**Ramucirumab vs placebo (positive phase III in second line in patients with advanced HCC and AFP>400 ng/mL).



Nature Reviews | Disease Primers

FIGURE 78-6 Effective and emerging therapies in HCC. Summary of treatments tested in Phase II–III clinical trials. Yellow boxes indicate drugs with positive Phase III studies, red boxes indicate drugs with negative results from Phase III trials and drugs in grey boxes have been tested in Phase II studies. (Reprinted with permission from JM Llovet et al: Nat Rev Dis Primers 2:16018, 2016.)

PDGFR, RET, and c-Kit) with sorafenib has shown non-inferiority results in terms of overall survival (HR = 0.92). It is estimated that only half of the patients progressing on sorafenib can be considered for second-line therapies, and their median survival with no treatment is 7–8 months (obtained from patients allocated to the placebo arm).

Recently, lenvatinib showed similar efficacy compared to sorafenib in a phase III trial (median survival 13.6 months vs 12.3 months, respectively). A Phase III study comparing regorafenib (a more potent multi-kinase inhibitor than sorafenib, but targeting similar kinases) vs placebo in patients progressing to sorafenib has reported a benefit in survival from 7.8 to 10.6 months (HR: 0.62; 38% reduction of risk of death) (Fig. 78-5). Treatment improved survival in all patient subgroups. In this trial, 88% of patients were BCLC C and 12% BCLC B, and all had progressed on sorafenib. Around 30% of patients presented with macrovascular invasion, 70% with extrahepatic spread, and 45% with AFP >400 ng/dL. Response rate was 10% based upon mRECIST. Treatment was started at 160 mg/day (3 weeks on /1 week off). Median time on treatment was 3.5 months. Prevalence of toxicity (hand-foot reaction, fatigue, and hypertension) was higher compared with reported toxicity from sorafenib, but adverse events only led to treatment discontinuation in 10% of cases. Patients progressing after second-line therapy, along with those at a BCLC D stage should receive best supportive palliative care

including management of pain, nutrition, and psychological support. Emerging therapies are shown in Fig. 78-6.

CHOLANGIOPRIMARY CARCINOMA

Cholangiocarcinoma (CCA) is classified according to its anatomic location as intrahepatic (iCCA; ~30%), perihilar (pCCA; ~50%), and distal (dCCA; ~20%). The latter two are also known as extrahepatic cholangiocarcinomas (eCCA), with the second-order bile ducts acting as the separation point (Fig. 78-7). This classification is endorsed by the 7th edition of the American Joint Committee on Cancer (AJCC) Staging Manual. In addition, iCCA has been recognized as a distinct entity with specific *ad hoc* clinical practice guidelines. Treatment options beyond surgery are limited, and unlike most solid tumors, no molecular targeted therapies have been approved for its treatment. The three subtypes of CCA differ in their anatomic location, epidemiology and risk factors, cell of origin, pathogenesis, and treatment. iCCA originates from adult cholangiocytes, trans-differentiation of adult hepatocytes and hepatic progenitor cells, whereas eCCA arises from the biliary epithelium and peribiliary glands. Moreover, their mutational profile also differs. FGFR2 fusions and IDH1/2 mutations only occur in iCCA, whereas ERBB2 amplifications, APOBEC-associated mutation signatures, and PKA fusions occur in eCCA. Thus, clinical management



FIGURE 78-7 Classification of cholangiocarcinoma subtypes. The 7th edition AJCC/UICC TNM staging classification includes intrahepatic (iCCA, **A**), perihilar (pCCA, **B**) and distal (dCCA, **C**) tumors. (Reprinted with permission from S Rizvi, GJ Gores: *Hepatology* 63:1356, 2016.)

and trials testing molecular therapies should be tailored according to each biological/anatomical subtype of CCA, as opposed to a common approach for all biliary tract cancers.

■ EPIDEMIOLOGY, RISK FACTORS, AND MOLECULAR TRAITS

CCA is the second most common liver cancer following HCC, with a 5-year survival of 10%. iCCA has globally increasing incidence and mortality rates. The incidence of iCCA varies according to exposure to risk factors, ranging from 1–2 cases per 100,000 inhabitants in Europe and North America to the highest incidence in some areas of Southeast Asia, particularly in Thailand (>80 cases/100,000 inhabitants). The male/female ratio is 1.2. Overall, most cases occur with unknown risk factors. The classical risk factors for CCA development include primary sclerosing cholangitis (PSC), biliary duct cysts, hepatolithiasis, Caroli's disease. Parasitic biliary infestation with flukes (i.e., most common is *Opisthorchis viverrini* and *Clonorchis sinensis*), is a prevalent etiology in Asia that can be prevented with an antihelminth therapy, praziquantel. PSC is a clear risk factor for iCCA and pCCA development, with a lifetime incidence ranging from 5 to 10%. Surveillance in PSC patients is recommended with annual imaging techniques and CA 19.9 serum determination. Common risk factors for HCC, such as HBV and HCV infection and cirrhosis, have been associated to iCCA development.

Molecular classification and drivers. There is no established molecular classification of CCA. Genomic studies have provided insight on two subclasses of iCCA, a proliferation subclass—characterized by activation of oncogenic signaling pathways (including RAS and MET)—and an inflammation subclass, characterized by activation of inflammatory pathways, overexpression of cytokines, and STAT3 activation. The landscape of mutations discovered by whole exome sequencing techniques defines a distinct mutational fingerprint depending on etiology and CCA subtype (Fig. 78-7). iCCA mutation portrait is characterized by ~50–60% of tumors having at least one targetable driver including FGFR2 fusion events (~25%), mutations in IDH1-2 (15%), KRAS (15%), BRAF (5%) and EGFR (3%), and amplifications in FGF19/CCDN1 (4%). While mutations in P53 (~30%) and KRAS (~25%) are more common in eCCA than in iCCA, some molecular drivers are specific for subtypes, such as fusion of PRKACA or PRKACB for eCCA or ERBB2 amplifications (~20%) for gallbladder cancer. Liver flukes-associated CCA have higher incidence of TP53 and SMAD4 mutations. Host genetic polymorphisms predisposing to CCA have not been established.

■ INTRAHEPATIC CHOLANGIOPANCREATIC CANCER

Diagnosis and Staging Diagnosis of iCCA requires pathological confirmation. Guidelines are currently not recommending surveillance

for early diagnosis, when patients are asymptomatic, since at-risk populations are ill-defined. Cirrhotic patients at risk of HCC development are enrolled in surveillance programs, and can benefit for early detection of iCCA. Otherwise, incidental diagnosis occurs due to cross-sectional imaging performed for other reasons. In most cases, iCCA is diagnosed at advanced stages where symptoms such as weight loss, malaise, abdominal discomfort, or jaundice are present. Pathological diagnosis of iCCA is based on the WHO criteria. Differential diagnosis should be established with metastatic adenocarcinoma and mixed iCCA-HCC tumors, which may require evaluation of markers such as Hep-Par-1, GPC3, HSP70, and glutamine synthetase markers. Imaging studies with CT/MRI are not accurate enough to establish iCCA non-invasive diagnosis. Dynamic CT scanning characterizes 80% of iCCA as liver mass-forming tumors with progressive contrast uptake from the arterial to the venous/delayed phase. MRI dynamic images also show peripheral enhancement in the arterial phase followed by progressive filling-in of the tumor. Atypical radiological behavior with arterial enhancement recapitulating HCC occurs in 10% of cases. MRI with cholangiopancreatography (MRI/MRCP) is useful to visualize the ductal system and vascular structures. Guidelines do not recommend PET scan for diagnosis. Tumor biomarker carbohydrate antigen (CA) 19-9 at a cut-off level of 100 U/mL has prognostic significance, but lacks accuracy (sensitivity and specificity of ~60%) for early diagnosis.

Radiological criteria are inadequate for iCCA diagnosis in cirrhotic patients. However, in non-cirrhotic patients, guidelines endorse a presumed diagnosis of iCCA (i.e., venous phase contrast enhancement on dynamic CT/MRI) if resection is considered. Assessment of disease extent (venous or arterial invasion and extrahepatic disease) and resectability is best accomplished with CT and/or MRI studies. Doppler ultrasound is accurate in defining vascular invasion. Before surgery, PET scanning may be considered to rule out an occult primary or metastatic site.

Staging system. The staging system for iCCA resected cases is based on the TNM staging as per the 7th edition of the AJCC/UICC staging, which is a new system that has already been validated. T1 tumors are solitary without vascular invasion; T2 disease includes multiple tumors (e.g., multi-focal disease, satellitosis, intrahepatic metastasis), or with vascular invasion (microvascular or major vascular invasion); T3 tumors directly invade adjacent structures; and T4 disease includes tumors with any periductal-infiltrating component. Regional lymph node metastasis in the hilar, peridiaphragmatic, and peripancreatic nodes are considered N1 disease, while distant spread is considered M1 disease. TNM stages I, II, and III overlap with T status, whereas stage IV includes either periductal invasion or N1/M1 disease.

TREATMENT

After adopting the TNM staging system, the International Liver Cancer Association (ILCA) guidelines for management of iCCA proposed the treatment algorithm depicted in Fig. 78-8. Overall, most of the treatments endorsed have a modest level of evidence and, thus guidelines are providing physicians with recommendations as standards of practice rather than standards of care supported by robust evidence-based data. Surgical resection represents the sole curative treatment option in 30–40% of patients with a 5-year survival of 30%. The largest systematic review including ~4500 iCCA patients undergoing resection reported a median survival of 28 months. In non-cirrhotic individuals, the best candidates for resection are patients at TNM stage I-II, whereas in patients with cirrhosis liver function should be assessed as previously described for HCC. Preoperative disease assessment should discard vascular invasion, N1 and M1. Lymphadenectomy of regional nodes is recommended given its prognostic value. The main predictors of recurrence (~50–60% at 3 years) and survival are identified at the pathological examination, including presence of vascular invasion, lymph node metastases, and poor differentiation. There is no established adjuvant therapy. Liver transplantation remains controversial, and few studies have reported good outcomes for single tumors ≤2 cm.

Non-surgical candidates have a dismal life expectancy. Overall, patients at stage III might be considered for loco-regional therapies, such as chemoembolization or radioembolization, but the level of evidence is low, mostly based on cohort studies. A meta-analysis of 14 trials testing loco-regional therapies reported median survival times of 15 months. External beam radiation therapy is not recommended as standard therapy. At more advanced stages (stage IV) in patients ECOG 0-1, systemic chemotherapy with the combination of gemcitabine and cisplatin is considered the standard of practice yielding median survival rates of 11.7 months compared to 8 months for gemcitabine alone. This recommendation for first-line treatment of advanced tumors is based on a subgroup analysis of 80 iCCA patients included in a large randomized Phase III trial ($n = 410$, ABC trial-02) of patients with advanced biliary tract tumors.

No molecular targeted therapy has been proven effective for iCCA. Patient stratification based on molecular biomarkers is ongoing with FGFR2 aberrations and IDH1/2 mutations. Preliminary data of a Phase II trial testing BGJ398 in advanced iCCA harboring FGFR2 gene fusions reported ~20% objective response.

Mixed HCC-iCCA is a rare neoplasm accounting for <0.5% of all primary liver cancers. Diagnosis is based on pathology. The 2010 WHO classification defined two subtypes: the classical and the stem cell feature type. Molecular data has also characterized a third unique entity, cholangiolocellular carcinoma, with distinct molecular traits and better outcome. Due to its low incidence, the demographic features and clinical behavior of these tumors remain ill-defined. Survival rates are similar to iCCA, and until specific guidelines are available, they should be managed following the treatment algorithm of iCCA.

EXTRAHEPATIC CHOLANGIOPAPILLARY CARCINOMA

Perihilar (pCCA) and Distal Cholangiocarcinoma (dCCA)

The 7th edition AJCC/UICC TNM staging classification has established pCCA as tumors that arise between the second-order bile ducts up to the insertion of the cystic duct, whereas dCCA arise from this point to the ampulla of Vater (Fig. 78-7). Thus, dCCA can be difficult to distinguish from early pancreatic cancer. Both entities have a similar diagnostic approach. Acute onset of painless jaundice occurs in 90% of patients with pCCA, and 10% present with cholangitis. Primary biliary cholangitis with a cut-off for CA19.9 >129 U/mL is suspicious for CCA. Imaging assessment starts with CT and MRI; they have a good sensitivity and specificity (>85%) for detecting the degree of bile duct involvement, and hepatic and portal vein invasion. MRI-cholangiography is optimal for defining the extent of the bile duct lesion. Ruling out IgG-4 cholangiopathy by assessing serum IgG4 is mandatory. As a second step, endoscopic retrograde cholangiography with brushing to explore cytology and fluorescence in situ hybridization (FISH)—for exploring polysomy—is recommended. FISH enhances the sensitivity of cytology from 20 to ~40%.

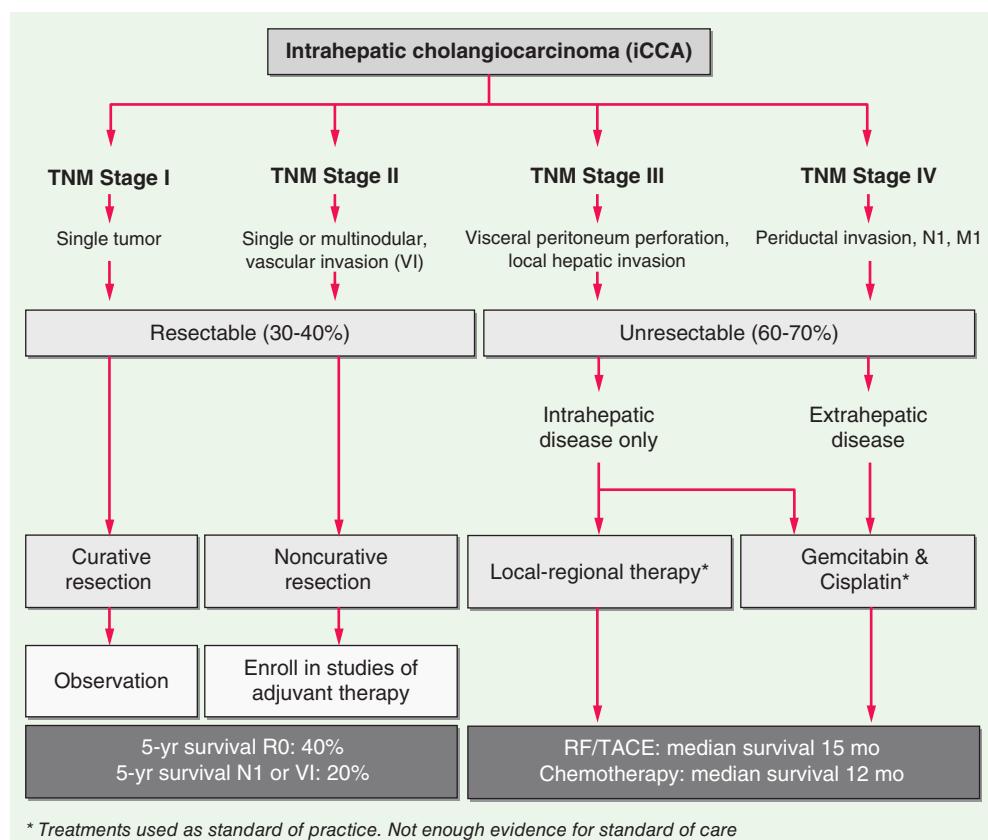


FIGURE 78-8 Staging and treatment schedule for iCCA proposed by the International Liver Cancer Association. (Reprinted with permission from J Bridgewater: *J Hepatology* 60:1268–1289, 2014.)

Diagnosis is based upon pathology. The treatment algorithm for pCCA indicates that in cases of a dominant stricture with positive cytology/biopsy or polysomy, a lymph node biopsy through endoscopic ultrasound should be obtained. pCCA with negative lymph node involvement is best treated by surgery, resection, or transplantation, the sole curative options. Resection entails hepatic and bile duct removal, Roux-en-Y-hepaticojejunostomy with regional lymphadenectomy. Bilobular involvement is considered a surgical contraindication. In few referral centers, unresectable single pCCA <3 cm without dissemination can be considered for liver transplantation with neoadjuvant chemoradiation. This procedure is associated with 5-year survival rates of ~70%. If lymph node involvement is present, systemic chemotherapy can be considered along with biliary tract stenting. Of note, the subgroup analysis of the Phase III ABC trial-02 did not identify differences between gemcitabine alone or in combination with cisplatin for pCCA. Surgical resection (Whipple procedure) is the primary option for management of dCCA, a procedure that achieves a median survival of 2 years and 5-year survival rates of ~25%. Main contraindications for resection are presence of distant lymph node involvement, metastases, or major vascular invasion. At the pathological examination, perineural invasion, lymph node metastasis, R0 resection (absence of residual tumor at pathological examination), and tumor differentiation are predictors of survival. Adjuvant therapy has not shown outcome benefits. There is no evidence of benefit of chemotherapy for unresectable cases. No molecular targeted therapies are available for these entities.

GALLBLADDER CANCER

Gallbladder cancer is the most common cancer of the biliary tract worldwide. The estimated cases of gallbladder cancer in the United States in 2016 are 11,400, more than CCA. The female:male ratio is 3:1. Cholelithiasis is the major risk factor, but <1% of patients with cholelithiasis develop this cancer. Gallbladder polyps at risk of transformation are those with ≥ 10 mm in diameter. Early cases are discovered incidentally at routine cholecystectomy. Clinical symptoms, such as jaundice, pain, and weight loss, are associated with advanced stages. Staging of gallbladder cancer follows the TNM classification. The most accurate technique to define staging and vascular and biliary tract invasion is the magnetic resonance cholangiopancreatography. CT and PET scan can be also useful for preoperative staging.

The mainstay of treatment is surgical, either simple or radical cholecystectomy (partial hepatectomy and regional lymph node dissection) for stage I or II disease, respectively. Only ~20% of patients are candidates for surgery with a curative intent. Survival rates are near 80–90% at 5 years for stage I, and range from 60 to 90% at 5 years for stage II. Regional nodal status and the depth of tumor invasion (T status) are the two most important prognostic factors. Adjuvant therapy has not proven effective. Gallbladder cancers at stage III and IV are considered unresectable. For patients with ECOG 0–1, chemotherapy with gemcitabine and cisplatin is the standard of practice based on data from the subgroup analysis including 181 patients with gallbladder cancer in the setting of two clinical trials. Overall, median survival is 10–12 months in advanced cases. Percutaneous transhepatic drainage is indicated in case of biliary obstruction. Radiotherapy is not effective.

OTHER MALIGNANT LIVER TUMORS

Fibrolamellar Hepatocellular Carcinoma (FLC) FLC is a rare form of primary liver cancer that typically affects children and young adults (10–30 years of age) without background liver disease. FLC accounts for 0.85% of all primary hepatic malignancies in the United States, and its incidence rate is 0.02 cases per 100,000 inhabitants. FLC is considered a unique entity with a specific fusion oncogene *PRKACA-DNAJB1* present in 80–100% of cases. A few mutations have been described, all at a level of <10%. FLC has a better prognosis than HCC, probably due to the absence of cirrhosis and the earlier age of presentation. Surgical resection is the mainstay of treatment and indications are less restrictive than for HCC. A retrospective series of 575 FLC cases reported a median survival of 70 months after resection. At

advanced stages, the expected outcome is <20 months. Chemotherapy is not effective and there is no standard of care.

Hepatoblastoma (HB) HB is the most frequent primary liver tumor in children. The incidence of the disease is 1.5 cases per 1,000,000. Background liver disease is rare in these patients. WNT signaling plays a major role, with *CTNNB1* mutations (70%) as the most frequently reported molecular event. A gene signature is able to discriminate two molecular classes with distinct outcome. Resection followed by chemotherapy with doxorubicin is the mainstay treatment strategy. A study including 1605 patients randomized in eight clinical trials reported better outcome for patients with stage I-II of the PRETEXT classification (out of four stages), age <3 years, AFP $>1,000$ ng/mL, and absence of metastases. As opposed to HCC, low AFP indicates poor prognosis. Outcomes for best candidates after resection (stages I/II with small tumors, age <3 years and AFP >100 ng/mL) achieve 5-year disease-free survival of 90%, compared with worst candidates (metastatic disease and AFP <100 ng/mL) with 5-year disease-free survival of 20–30%.

BENIGN LIVER TUMORS

The most common benign liver tumors are hemangiomas, focal nodular hyperplasia (FNH), and hepatocellular adenomas (HCA). Most benign tumors are identified incidentally by abdominal ultrasound or other imaging techniques. *Hemangiomas* are present in ~5% of the general population, are diagnosed by ultrasound except in cirrhotic patients or oncology patients where contrast enhanced imaging (contrast enhanced ultrasound, CT, or MRI) is required. Conservative management is appropriate and follow-up is not recommended. Exceptionally, growing lesions causing symptoms by compression can be considered for resection. FNH is a benign tumor present in <2% of the population and occurring mostly in females aged 40–50 years. FNH is a polyclonal hepatocellular proliferation due to an arterial malformation. MRI has the highest diagnostic accuracy with a specificity of 100%, when typical imaging features are present (homogeneous enhancement in the arterial phase with a central scar). Atypical FNH requires biopsy for diagnosis. Treatment is not recommended since these tumors do not degenerate or cause complications. In exceptional cases of expanding symptomatic lesions, surgery is the treatment of choice.

Hepatic adenomas are clonal benign proliferations resulting from single gene driver mutations. HCA have a low prevalence of 0.001% of the population and are frequently diagnosed in women aged 35–40 years. The female:male ratio is 10:1, and the main risk factors are oral contraceptives in females and use of anabolic androgenic steroids in male body builders. HCA have the potential for hemorrhage and HCC development, particularly when sized >5 cm. Nowadays, there is a clear understanding of the molecular classification of HCA: (a) HCA with *CTNNB1* mutations (10–20%) are at-risk of HCC development and are present in men treated with androgens; (b) inflammatory adenomas (50–60%) are associated with single mutations (*Gp130*: 65%) and are more prevalent in females with obesity or diabetes; and (c) adenomas with inactivated *HNF-1A*. Diagnosis is based on MRI imaging, which is able to correlate with molecular subtypes in 80% of cases (Inflammatory and *HNF-1A* type). For defining HCA with *CTNNB1* mutations, biopsy is required. Upon diagnosis, discontinuation of oral contraceptives and weight loss is recommended. Resection is indicated in all cases of size >5 cm or men or *CTNNB1* mutation. For HCA <5 cm, 1-year follow-up is recommended. In case of active HCA bleeding, embolization followed by resection is the treatment of choice. The presence of multiple HCA is common, and guidelines endorse treating them based on the size of the main nodule.

FURTHER READING

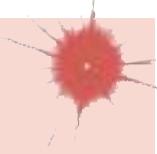
- BRIDGEWATER J et al: Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 60:1268, 2014.
- BRUIX J, SHERMAN M: Management of hepatocellular carcinoma. *Hepatology* 42:1208, 2005.
- BRUIX J et al: Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): A randomized, double-blind, placebo-controlled, phase 3 trial. *Lancet* 389:56, 2017.

- CLAVIEN PA et al: Recommendations for liver transplantation for hepatocellular carcinoma: An international consensus conference report. Lancet Oncol 13:e11, 2012.
- EASL-EORTC CLINICAL PRACTICE GUIDELINES: Management of hepatocellular carcinoma. J Hepatol 2018.
- FORNER A et al: Hepatocellular carcinoma. Lancet 379:1245, 2012.
- LLOVENT JM et al: Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 359:378, 2008.
- LLOVENT JM et al: Hepatocellular carcinoma. Nat Rev Disease Primers 2:16018, 2016.
- MAZZAFERRO V et al: Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 334:693, 1996.
- RAZUMILAVA N, GORES G: Cholangiocarcinoma. Lancet 383:2168, 2014.
- SCHULZE K et al: Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 47:505, 2015.
- VALLE J et al: Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N Engl J Med 362:1273, 2010.
- ZUCMAN-RONZI J et al: The genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology 149:1226, 2015.

79

Pancreatic Cancer

Daniel D. Von Hoff



Pancreatic cancer is the third leading cause of death from cancer in the United States with >53,000 Americans diagnosed and >43,000 dying from the disease each year. Unfortunately, pancreatic cancer is projected to be the second leading cause of death from cancer in the United States by 2030. Worldwide pancreatic cancer is the eleventh most common cancer with 338,000 new patients diagnosed and >334,000 deaths (seventh cause of cancer deaths). Pancreatic cancer currently has the worst survival rate of any cancer with an overall 5-year survival (regardless of stage) of ~8.2%. However, that situation is changing because some advances have been made against the disease with some improvements in survival (see below) that may affect the 5-year survival statistics. In particular, knowledge about specific molecular subsets of the disease has become crucial so that one can provide the best possible care for their patients with pancreatic cancer.

EPIDEMIOLOGY

Pancreatic cancer accounts for 3.2% of all new cancer cases in the United States and for 7.2% of all deaths from cancer in the United States. The lifetime risk of developing pancreatic cancer is ~1.6%. The incidence of pancreatic cancer has been increasing between 0.5 and 1% per year. Pancreatic cancer is more common with increasing age and more common in men than in women. The 5-year survival rate for all stages has increased from 3% in 1975 to 8.2% in 2013. The latest information from the U.S. Surveillance, Epidemiology, and End Results (SEER) database predicts that the 5-year survival for patients with localized pancreatic cancer is about 31.5%, 11.5% for those with regional disease, and 2–5% for patients with advanced metastatic disease. Pancreatic cancer is more common in developed countries (although generally it tracks with the prevalence of smoking). The incidence is highest in North America and Western Europe followed by other areas in Europe, Australia, New Zealand, and South-Central Asia. Of note is that the population at greatest risk are women living in Scandinavian countries, while the lowest risk is seen for women living in middle Africa.

RISK FACTORS

Age is one of the greatest risk factors for pancreatic cancer with median age at diagnosis of 70 years (the disease is most frequently diagnosed in the 65–74 age group). The number of new cases per 100,000 persons and the number of deaths per 100,000 persons are higher for males

and blacks of both sexes. Both the number of cases and the number of deaths per 100,000 people are lower for American Indian/Alaskan natives and Asian Pacific Islanders. Both the number of cases and deaths are intermediate for the Hispanic population.

Environment The greatest risk factor for pancreatic cancer is cigarette smoking. The risk correlates with the increased number of cigarettes smoked. It has been estimated that 30% of pancreatic cancer is caused by smoking. Exposure to cadmium as part of cigarette smoking or via exposure to welding, soldering, or dietary exposure has been weakly associated with an increased risk of pancreatic cancer.

Although dietary factors are often difficult to interpret, evidence suggests that high intakes of fat or meat (particularly well-done barbecued meat) are risk factors. High intakes of fruits and vegetables are associated with a decreased risk. Coffee and low-to-moderate alcohol consumption have been determined not to be associated with an increased risk for pancreatic cancers, while consumption of sugary fizzy drinks has been associated with an increased risk.

Microbiome To date, there is no solid evidence of an association between *Helicobacter pylori* infection and pancreatic cancer. Some data link the oral microbiome associated with poor dentition to pancreatic cancer but the evidence is very thin.

Hereditary/Genetics Hereditary factors may account for 10–16% of all pancreatic cancers. It is very important to recognize these factors for determining risk for family members of the patient affected with pancreatic cancer. (These family members should seek participation in an early detection program with genetic counseling, definition of risk and perhaps, if appropriate, periodic MRI screening of the abdomen, though this recommendation is not based on research data.) In addition, the identification of any of these germ-line mutations can lead to specific and effective new therapeutics for patients with these abnormalities in their tumors. Table 79-1 identifies the various germ-line mutations along with their familial cancer syndromes where an increased risk for pancreatic cancer is known.

Knowing the patient has a BRCA2 or PALB2 germ-line mutation or any of the above mutations should lead one to not only refer the patient's relatives to an early detection or high-risk individual clinic but also realize that for the BRCA2/PALB2 germ-line mutation patients consideration for treatment with a poly (ADP-ribose) polymerase (PARP) inhibitor should be entertained. Other germ-line mutations are under study to determine their increased risk of pancreatic cancer including CDK4, FANCC, PALLD, APC, ATM, BMPR1A, BRCA1, EPCAM, MEN1, MLH1, MSH2, MSH6, NF1, PMS2, SMAD4, TP53, TSC1, TSC2, and VHL. Some of these mutations are associated with pancreatic neuroendocrine tumors (Chap. 80).

TABLE 79-1 Germ-line Mutations, Their Familial Cancer Syndrome, and Fold Risk of Pancreatic Cancer

GERM-LINE MUTATION	FAMILIAL CANCER SYNDROME	ESTIMATED INCREASED RISK (FOLD) OF PANCREATIC CANCER
BRCA2 ^a	Familial breast/ovarian cancer	3.5–10
PALB2 (partner and localizer of BRCA2)	Familial breast cancer and others	~sixfold
p16/CDKN2A	Familial atypical multiple mole melanoma (FAMMM)	13–38
STK11 (LKB1)	Peutz-Jeghers syndrome	132
PRSS1 or SPIN11 ^b	Hereditary (familial) pancreatitis	53
ATM	Ataxia-telangiectasia	Not yet established
MLH1, MSH2, MSH6, PMS2	Hereditary nonpolyposis colorectal syndrome or Lynch syndrome ^c	9–30

^aParticularly common in individuals with Ashkenazi Jewish heritage. ^bForty percent chance of pancreatic cancer by the age of 70. ^cVery important because this is associated with microsatellite instability, which is a marker for response to an anti-PD-1/PD-L1 agent.

In addition to the recognized genetic syndromes, other possible familial pancreatic cancer genes have not yet been discovered. For example, a family history of pancreatic cancer is associated with a 13-fold increase in the disease. If you have one first-degree relative, the risk is increased 4.6-fold, 2 first-degree relatives 6.4-fold, and ≥ 3 first degree relatives a 32-fold increase. The risk is also increased if a relative developed pancreatic cancer at < 55 years old.

Other Considerations Most patients with pancreatic cancer relate that they have had developing symptoms over the last few years. However, Yachida and colleagues suggest that pancreatic cancer could be growing over a period of 21 years. Thus, there is a possibility for early detection of the disease.

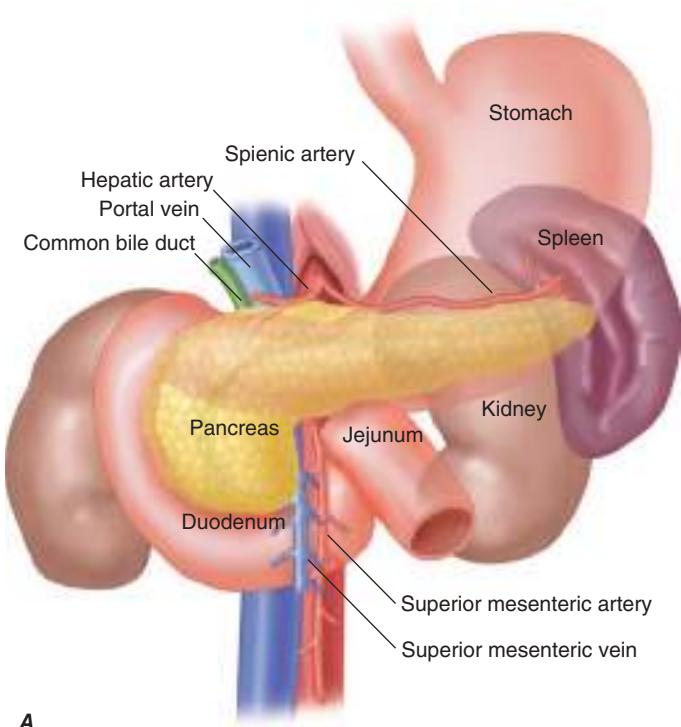
Medical Conditions Chronic pancreatitis that is nonfamilial is also associated with an increased risk of pancreatic cancer (2.3–16.5-fold increase). It is also increased in people with chronic pancreatitis associated with cystic fibrosis or tropical pancreatitis.

A clear association exists between diabetes mellitus and pancreatic cancer. Whether this is a causal association or whether the diabetes is the result of the cancer is not exactly clear. What is clear is that when a person presents with new onset diabetes, they should be considered at risk for having pancreatic cancer. The excessive insulin or insulin-like growth factors associated with adult onset diabetes and metabolic syndrome may promote pancreatic carcinogenesis.

Obesity is considered a possible risk factor for pancreatic cancer. A high body mass index (BMI) ≥ 30 is associated with a doubling of the risk of pancreatic cancer. Since obesity is a risk factor for diabetes, the contribution of obesity alone is unclear. Interestingly, patients with severe obesity who undergo a gastric bypass experience a reduction in the incidence of gastrointestinal (GI) cancer including pancreatic cancer by $> 30\%$ in the first 3 years (along with a dramatic decrease in their Hgb A1c and blood glucose). Physical inactivity also has been associated with an increased risk in pancreatic cancer.

PATHOLOGY AND MOLECULAR CONSIDERATION

Location The posterior location of the pancreas in the abdomen is likely one of the issues that leads to a late diagnosis (Fig. 79-1A).



A

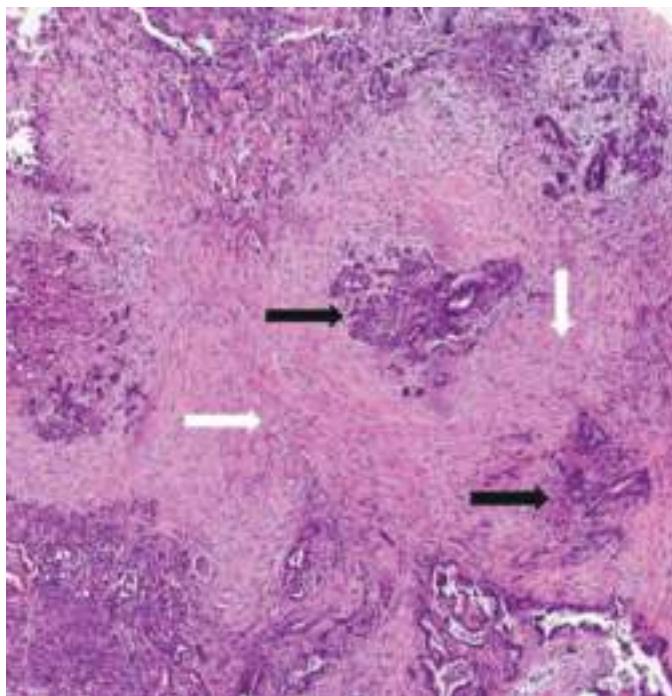
FIGURE 79-1 A. Note the relationship of the pancreas to the major vessels of the retroperitoneum. B. Ductal adenocarcinoma of the pancreas (black arrows), with intense stromal component (white arrows). (Part A is courtesy of Mary Kay Washington, MD, PhD Vanderbilt University. Part B is courtesy of Haiyong Han, PhD Translational Genomics Research Institute [TGen].)

Pathology Cancers of the pancreas can be divided into neoplasms of the endocrine pancreas (Chap. 80) and tumors of the exocrine pancreas. The most common neoplasm of the exocrine pancreas and most deadly is pancreatic infiltrating ductal adenocarcinoma. These tumors arise in the head, body, or tail of the pancreas and are characterized by infiltrating desmoplastic stromal reactions (Fig. 79-1B).

Other subtypes of nonneuroendocrine pancreatic cancers include acinar cell carcinoma (tumors of the exocrine enzyme producing cell); medullary carcinoma, adenosquamous, and other rare subtypes. Each of these are different in their behaviors and in their molecular characteristics and often require specific other types of treatment.

Molecular Characteristics The molecular characteristics of pancreatic ductal adenocarcinoma reveal four genes that are commonly mutated or inactivated (sometimes referred to as the “4 horsemen”). The most common is KRAS (usually in codon12), which is seen in virtually 100% of pancreatic adenocarcinomas. In fact, with the deep sequencing now available, if a KRAS mutation is not detected in the patient’s tumor, one should consider the tumor being of a different origin (such as small bowel, gallbladder, or cholangiocarcinoma)—all of which could require different treatments. *p16/CDKN2A* is also noted in $> 90\%$ of invasive pancreatic adenocarcinomas. *TP53* and *DPC4/MADH4* are mutated in about half of these tumors. As a reference point, the *BRCA2* gene noted in Table 79-1 is mutated in 7–10% of pancreatic adenocarcinomas.

Precursor Lesions Many pancreatic adenocarcinomas seem to arise from noninvasive epithelial precursor lesions. Detection of these could allow for early diagnosis of pancreatic cancer. These pancreatic intraepithelial neoplasias (PanINs) have varying degrees of dysplasia designated as PanINs 1–3 (and constitute a progression model for pancreatic cancer). Genetic alterations become more frequent as the PanIN grade increases (e.g., grade 3). Not all PanIN lesions progress to invasive malignancy. PanINs that are ≥ 1 cm are called *intraductal papillary neoplasms* and are usually noninvasive. If the intraductal tumor is in a branch duct, it is usually noninvasive; however, if the intraductal tumor is in a main duct and is large and nodular, it is more likely to have malignant behavior.



B

One other pancreatic tumor is the mucinous cystic neoplasm; they may be seen as incidental findings on scans. These lesions are less likely invasive (20%) unless they are large and have nodules in them.

■ CLINICAL FEATURES

History and Physical The classic presentation for a patient with pancreatic cancer has been abdominal pain and weight loss with or without jaundice. The pain is midepigastic (sometimes described as a “boring-like” pain). Often the pain is in the back (due to retroperitoneal invasion of the splanchnic nerve plexus). The pain may be exacerbated by eating or lying flat. Other items of note in a history is lightening in stool color (steatorrhea also causes malodorous stools), or the onset of diabetes in the prior year. Jaundice, first detectable with a bilirubin of 2.5–3.0 mg/dL, is usually associated with tumor in the head of the pancreas. In some instances, depression is noted (with a higher subsequent number of suicides). Pruritis may be seen when the bilirubin reaches 6–8 mg/dL.

Physical signs include jaundice, signs of weight loss, a palpable gallbladder (Courvoisier’s sign), hepatomegaly, an abdominal mass, and even an enlarged spleen (usually indicating a portal vein thrombosis). Migratory superficial thrombophlebitis can also be seen (Trousseau’s syndrome). Signs of late disease include a lymph node palpable in the supraclavicular fossa (usually on the left where the thoracic duct enters the subclavian vein). This is clinically referred to as Virchow’s node. Occasionally one can palpate subcutaneous metastases in the perumbilical area referred to as a Sister Mary Joseph’s node—named after one of the scrub nurses on the Mayo Clinic Operative Team who noted that when she prepped that area and felt those nodules, the patient often had peritoneal metastases.

The history and symptoms noted above may lead a person to see a physician; often CT and MRI scanning detects the disease before advanced disease symptoms appear.

■ DIAGNOSTIC WORKUP

Imaging Diagnostic imaging plays a major role in diagnosing pancreatic cancer and other intraabdominal diseases. The best technique is the use of a dual-phase contrast-enhanced spiral CT using the pancreatic cancer protocol which allows arterial phase enhancement and portal venous phase enhancement. This special protocol can provide helpful prospective staging and assessment of resectability. **Figure 79-2** demonstrates such a CT scan (with vascular involvement). **Figure 79-3** demonstrates the use of an 18F glucose positron emission tomography (PET) scan.

Histologic Diagnosis A histologic (tissue) diagnosis is essential and should be obtained with a cutting biopsy needle (not a skinny needle with cytology). Misdiagnosis is more common based on only fine-needle aspirates. Obtaining a tissue diagnosis allows not only for accuracy but also for molecular testing for KRAS mutations, microsatellite instability, and other important molecular abnormalities. Those molecular abnormalities and others will be increasingly important as more targeted therapies are developed for patients with pancreatic cancer.

The core needle (16–18 gauge) biopsy can be obtained via endoscopic ultrasound-guided techniques for a tumor localized to the pancreas or, if there are liver lesions or Virchow’s node, via percutaneous biopsy by interventional radiologists.

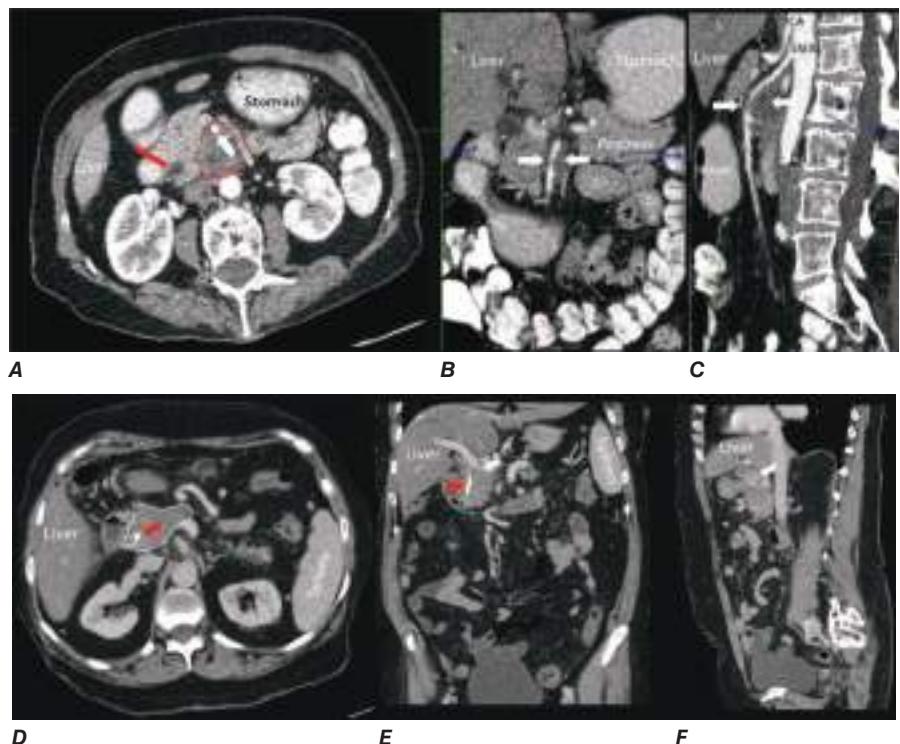


FIGURE 79-2 Selected images from contrast-enhanced CT in patients with locally advanced adenocarcinoma of the pancreas. A high-quality contrast-enhanced CT scan (arterial phase in Panels A–C and portal venous phase in Panels D–F) is required for optimal staging of pancreas cancer. Panel A demonstrates the typical features of adenocarcinoma of the pancreas on arterial phase axial CT scans (dotted outline) with tumor encasement of the superior mesenteric artery (white arrow). Note the dilatation of the common bile duct (red arrow). Panels B (magnified coronal) and C (sagittal) show reconstruction of CT images into additional orthogonal planes with exquisite details to confirm the unresectable nature of the tumor due to vascular encasement. Panel D demonstrates the typical features of adenocarcinoma of the pancreas on portal venous phase axial CT scans in a different subject. The dotted line outlines a pancreas cancer lesion in the pancreatic head, which is encasing the portal splenic confluence (dotted outline). Panels E (white arrow) and F show the pinched appearance of the portal splenic confluence by tumor abutment and invasion of the superior mesenteric vein (white arrow) on coronal and sagittal views. Note the presence of a stent in the common bile duct (red arrow) to help relieve biliary obstruction caused by the tumor. CA, celiac axis; SMA, superior mesenteric artery.

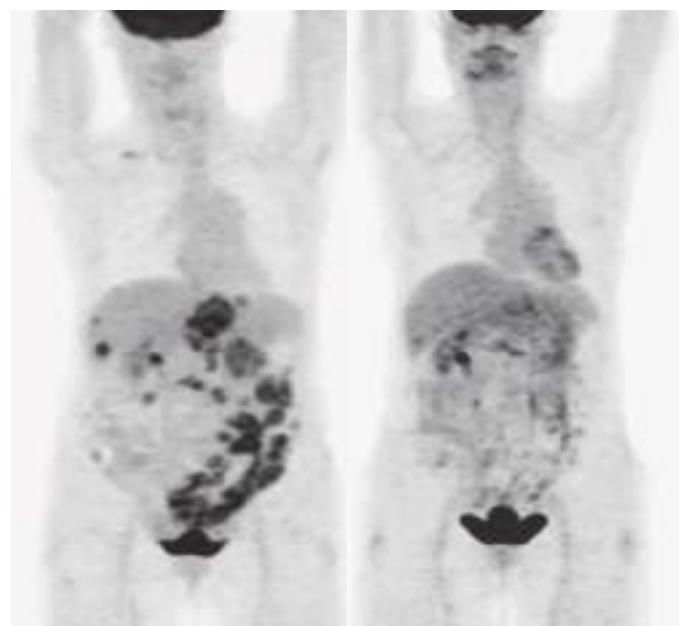


FIGURE 79-3. PET scan demonstrating metastatic disease—baseline and after 6 weeks of chemotherapy with some resolution of liver metastases.

Serum Markers Before treatment, a serum sample should be obtained for levels of CA19-9, carcinoembryonic antigen (CEA), or if both are negative, for CA125 (can be positive when the CA19-9 is negative due to the patient not being a Lewis antigen secretor). These markers are not useful for staging but can be useful in following the course of pancreatic cancer.

■ IMPORTANT IMMEDIATE CONSIDERATIONS IN PATIENT CARE

While the patient is being evaluated and staged, one must be alert for biliary tract obstruction (and the attendant risk for sepsis from the biliary tree). A stent can be placed (plastic if temporary or metal if needed longer) to relieve the jaundice and pruritus. If surgery is being contemplated, an early surgical consultation is in order as there are surgeons who will want to proceed to surgery without placement of a stent.

Patients with pancreatic cancer are often hypercoagulable and frequently have migratory thrombophlebitis (Trousseau's sign) as well as deep vein thrombosis with pulmonary emboli (a frequent cause of death). Appropriate examinations plus being alert to thromboses on the routine workup are mandatory so appropriate management can be put in place.

Control of pain or of any of the symptoms should be obtained if at all possible to help patients be as comfortable as possible for their decision-making. Sometimes simple approaches like the use of a replacement pancreatic enzyme (at good therapeutic doses) can relieve the bloating,

cramping, and diarrhea these patients suffer from. Early involvement of a palliative care team can improve a patient's quality of life and sometimes even its length.

■ CLINICAL STAGING

The clinical staging of pancreatic cancer according to the American Joint Commission on cancer staging is presented in **Table 79-2**.

Table 79-3 presents another clinical way to express extent of disease as well as therapeutic approaches (to be discussed later).

For proper staging, some physicians believe that a laparoscopy either before or at the time of contemplated surgery is important. If metastatic disease is found at laparoscopy, one can avoid surgery that would not be helpful because disease is already advanced.

TREATMENT

Resectable Disease

For patients with resectable disease (as defined in Table 79-3), the best option is surgery. Only a small percentage of patients are in this category (10–20%). The surgery for patients with tumors in the head or uncinate body of the pancreas is usually a pylorus-sparing pancreaticoduodenectomy (a modified Whipple procedure). For tumors in the body or tail, a distal pancreatectomy is usually performed. Clinical and pathologic findings of the resection are defined as either, an Ro

TABLE 79-2 Definition of Primary Tumor (T)

T CATEGORY	T CRITERIA
TX	Primary tumor cannot be assessed
TO	No evidence of primary tumor
Tis	Carcinoma in situ This includes high-grade pancreatic intraepithelial neoplasia (PanIn-3), intraductal papillary mucinous neoplasm with high-grade dysplasia, intraductal tubulopapillary neoplasm with high-grade dysplasia, and mucinous cystic neoplasm with high-grade dysplasia
T1	Tumor ≤2 cm in greatest dimension
T1a	Tumor is ≤0.5 cm in greatest dimension
T1b	Tumor >0.5 cm and <1 cm in greatest dimension
T1c	Tumor 1–2 cm in greatest dimension
T2	Tumor >2 cm and ≤4 cm in greatest dimension
T3	Tumor >4 cm in greatest dimension
T4	Tumor involves celiac axis, superior mesenteric artery, and/or common hepatic artery, regardless of size
M CATEGORY	M CRITERIA
M0	No distant metastasis
M1	Distant metastasis
N CATEGORY	N CRITERIA
NX	Regional lymph nodes cannot be assessed
NO	No regional lymph node metastases
N1	Metastasis in one to three regional lymph nodes
N2	Metastasis in four or more regional lymph nodes

AJCC Prognostic Stage Groups

WHEN T IS...	AND N IS...	AND M IS...	THEN THE STAGE GROUP IS....
Tis	NO	MO	0
T1	NO	MO	IA
T1	N1	MO	IIB
T1	N2	MO	III
T2	NO	MO	IB
T2	N1	MO	IIB
T2	N2	MO	III
T3	NO	MO	IIA
T3	N1	MO	IIB
T3	N2	MO	III
T4	Any N	MO	III
Any T	Any N	M1	IV

TABLE 79-3 Extent of Disease and Therapeutic Approach

DESIGNATION (MEDIAN SURVIVAL)	THERAPEUTIC APPROACHES
1. Resectable (localized): (18–23 mo) <ul style="list-style-type: none"> No encasement of celiac axis or superior mesenteric artery (SMA) Patent superior mesenteric—portal veins No extrapancreatic disease 	Surgical option (or preoperative-neoadjuvant therapy first). Surgery is followed by postsurgery adjuvant therapy <ul style="list-style-type: none"> Currently gemcitabine + capecitabine
2. Locally advanced: (6–10 mo) <ul style="list-style-type: none"> Encasement of arteries Venous occlusion (superior mesenteric vein [SMV] or portal) No extrapancreatic disease 	Either chemotherapy or chemotherapy + radiation therapy
3. Metastatic: (8.3–12.8 mo)	Systemic chemotherapy

resection (no macroscopic or microscopic disease left after surgery); an R1 resection refers to residual disease likely left behind. Patients with smaller tumors and lymph node negative disease have a better survival (median of about 18–23 months with 5-year survival of about 20%).

Two approaches are being explored to try to improve on this.

- (1) Postoperative adjuvant therapy. The standard of care is to use 6 months of adjuvant treatment with gemcitabine + capecitabine (referred to as the ESPAC4 trial). The median survival was 28 months (95% CI 23.5–31.5) for the combination of gemcitabine + capecitabine versus 25.5 months CI with (22.7–27.1) for the gemcitabine alone—hazard ratio 0.82 95% (0.6–0.98) ($p = 0.032$). Toxicities were manageable.
- (2) A more experimental approach is the use of neoadjuvant chemotherapy (chemotherapy given before surgery) to try to shrink the tumor and normalize the patient's serum CA19-9 level. Studies of neoadjuvant chemotherapy are ongoing.

Locally Advanced Disease (30% of Patients) For patients with locally advanced disease, the median survival is also quite poor (6–10 months) because many of the patients die with local problems (portal vein thrombosis with bleeding varices, obstruction, sepsis, etc.). The approach has been to try to reduce the bulk of the disease with use of radiation therapy plus chemotherapy or chemotherapy alone, hoping the disease could become resectable. No standard therapy has been agreed upon, but experimental approaches are applying some of the treatments that show promise in advanced metastatic disease.

Advanced Metastatic Disease (60% of Patients) Only a few of the many phase III randomized trials in patients with advanced pancreatic cancer have led to meaningful increases in survival. We have learned that a regimen needs to have at least a 50% improvement in overall survival or 90% improvement in 1-year survival in a pilot trial to predict for any degree of success in large randomized phase III trials.

Patients with the best chance of receiving a benefit from treatment have a good performance status (functioning up and around at least 70% of the day), have a reasonable albumin level (≥ 3.0 g/dL), and a neutrophil/lymphocyte ratio of ≤ 5.0 .

Single-agent gemcitabine achieves a median survival of 6 months and a 1-year survival rate of 18%. **Table 79-4** details three combination regimens that have further improved survival modestly. Median overall survival still ranges from 6 to 11 months. However, 1-year survival is now approaching 35% for these combination regimens with some long-term 4+ year survivors.

Liposomal irinotecan has been approved by the U.S. Food and Drug Administration (FDA) in combination with 5 fluorouracil + leucovorin for patients whose tumors have progressed on gemcitabine based on improved overall survival. PARP inhibitors have clinical activity against pancreatic cancers having mutations in BRCA2 or PALB2 (i.e., defective DNA repair proteins). In addition, tumors

TABLE 79-4 Combination Chemotherapy Regimens that Have an Impact on Survival

STUDY DESIGN (AUTHOR/REF)	NO. OF PATIENTS	MEDIAN SURVIVAL (MONTHS)
Gemcitabine + erlotinib vs Gemcitabine, (Moore et al: J Clin Oncol. 26:1960, 2007.)	569	6.24 vs 5.91 (HR 0.82; 95% CI 0.69–0.99, $p = 0.038$)
Folfirinox (folinic acid + 5FU + irinotecan + oxaliplatin) vs Gemcitabine, (Conroy et al: N Engl J Med 364:1817, 2011.)	342	11.1 vs 6.8 (HR 0.57; 95% CI 0.45–0.70, $p < 0.001$)
Nab-paclitaxel + gemcitabine vs gemcitabine, (Von Hoff et al: N Engl J Med 369:1691, 2013.)	861	8.5 vs 6.7 (HR 0.72; 95% CI 0.62–0.83, $p < 0.001^a$)

^aThe 2-year survival rate with this regimen is 9% and the 3+ year rate is 4%. Other studies have not reported on these parameters.

with microsatellite instability often have more mutations and such tumors appear to have a higher response rate to immunotherapy with checkpoint inhibitors, anti-PD-1 (pembrolizumab, nivolumab), and anti-PD-L1 antibodies.

Other Potential Factors Influencing Survival Preclinical studies have suggested that vitamin D can inhibit the development and growth of cancer. In models of pancreatic cancer, synthetic analogs of vitamin D had an effect on both tumor cells and on the tumor microenvironment. Clinical studies are conflicting as to whether circulating levels of plasma 25-hydroxyvitaminD (25(OH)D) affect the incidence of pancreatic cancer. However, patients with prediagnostic levels of 25(OH)D that are in the normal range have a longer survival than those who have reduced levels (35% lower hazard for death).

FUTURE DIRECTIONS

Death from pancreatic cancer is often due to progressive inanition. The metabolic consequences of this cancer are being examined. The tumor can be fatal at a modest level of tumor burden based on the profound metabolic effects. Other promising areas of investigation include addressing the florid stromal reaction around the tumor cells (believed to act as a physical barrier to drug delivery and as an immune sanctuary for the tumor cells). This attack on the stroma is being done with enzymatic (hyaluronidase) and other (antisuper enhancer genes) approaches. Also, utilization of hypomethylating and histone deacetylase inhibitors to correct epigenetic defects in the tumor microenvironment are under active study.

ACKNOWLEDGMENT

Thank you to Jennifer Byrne, BA, for assistance in the preparation of this chapter, and Drs. Elizabeth Washington, Ron Korn, and Haiyong Han and the American Joint Committee on Cancer for providing the figures.

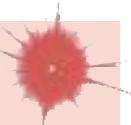
FURTHER READING

- CONROY T et al: FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med 364:1817, 2011.
- HRUBAN RJ et al: Genetic progression in the pancreatic ducts. Am J Pathol 156:1821, 2000.
- ILIC M, ILIC J: Epidemiology of pancreatic cancer. World J Gastroenterol 22:9694, 2016.
- NEOPOLEMUS JP et al: Comparison of adjuvant gemcitabine and capecitabine with gemcitabine nanotherapy in patients with resected pancreatic cancer (ESPAC4) a multicenter, open labeled, randomized phase 3 trial. Lancet 389:1011, 2017.
- RAHIB L et al: Evaluation of pancreatic cancer clinical trials and benchmarks for clinically meaningful future trials: A systemic review. JAMA Oncol 2:1209, 2016.
- SOLOMON S et al: Inherited pancreatic cancer syndromes. Cancer J 18:485, 2012.
- VON HOFF D et al: Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med 369:1691, 2013.

80

Neuroendocrine Tumors of the Gastrointestinal Tract and Pancreas

Robert T. Jensen



GENERAL FEATURES OF GASTROINTESTINAL NEUROENDOCRINE TUMORS

Gastrointestinal (GI) neuroendocrine tumors (NETs) are tumors derived from the diffuse neuroendocrine system of the GI tract, which is composed of amine- and acid-producing cells with different hormonal profiles, depending on the site of origin. NETs of the GI tract share many features with other NETs throughout the body and were historically divided into GI-NETs (in the GI tract) (also frequently called *carcinoid tumors*) and pancreatic neuroendocrine tumors (pNETs), although in newer pathologic classifications they are all classified as NETs (Table 80-1). These tumors originally were classified as APUDomas (for amine precursor uptake and decarboxylation), as were pheochromocytomas, NETs in other locations, melanomas, and medullary thyroid carcinomas, because they share certain cytochemical features as well as various pathologic, biologic, and molecular features. It was originally proposed that

APUDomas had a similar embryonic origin from neural crest cells, but it is now known that the peptide-secreting cells are not of neuroectodermal origin.

CLASSIFICATION/PATHOLOGY/TUMOR BIOLOGY OF NETS

NETs generally are composed of monotonous sheets of small round cells with uniform nuclei, and mitoses are uncommon. They can be frequently recognized on routine histology; however, these tumors are now recognized principally by their histologic staining patterns due to shared cellular proteins. Historically, silver staining was used, and tumors were classified as showing an argentaffin reaction if they took up and reduced silver or as being argyrophilic if they did not reduce it. Currently, immunocytochemical localization of chromogranins (A, B, C) and synaptophysin are routinely used. Chromogranins are acidic monomeric soluble proteins found in the large secretory granules. Synaptophysin is an integral membrane glycoprotein of 38,000 molecular weight found in small vesicles of neurons and NET. Chromogranin A is the most widely used.

Ultrastructurally, these tumors possess electron-dense neurosecretory granules and frequently contain small clear vesicles that correspond to synaptic vesicles of neurons. NETs synthesize numerous peptides, growth factors, and bioactive amines that may be ectopically secreted, giving rise to a specific clinical syndrome (Table 80-2). The diagnosis of the specific syndrome requires the clinical features of the disease (Table 80-2) and cannot be made from the immunocytochemistry results alone. The presence or absence of a specific clinical syndrome also cannot be predicted from the immunocytochemistry alone.

NETs of the GI tract (GI-NETs) have been classified according to their anatomic area of origin (i.e., foregut, midgut, hindgut) because tumors with similar areas of origin share functional manifestations, histochemistry, and secretory products (Table 80-3). Foregut tumors generally have a low serotonin (5-HT) content, are argentaffin-negative but argyrophilic, occasionally secrete adrenocorticotrophic hormone (ACTH) or 5-hydroxytryptophan (5-HTP), causing an atypical carcinoid syndrome (Fig. 80-1); these are often multihormonal and may metastasize to bone. They may produce a clinical syndrome due to the secreted products. Midgut carcinoids are argentaffin-positive, have a high serotonin content, most frequently cause the typical carcinoid syndrome when they metastasize (Table 80-3, Fig. 80-1), release serotonin and tachykinins (substance P, neuropeptide K, substance K), rarely secrete 5-HTP or ACTH, and less commonly metastasize to bone. Hindgut carcinoids (rectum, transverse and descending colon) are argentaffin-negative, are often argyrophilic, rarely contain serotonin or cause the carcinoid syndrome (Fig. 80-1, Table 80-3), rarely secrete 5-HTP or ACTH, contain numerous peptides, and may metastasize to bone.

However, the classification of GI NETs into foregut, midgut, or hindgut even though widely used, has not proved useful for prognostic or therapeutic purposes. More general classifications have been developed that allow NETs with similar features in different locations to be compared, have proven prognostic and tumor management value, and are now recommended in all recent guidelines and have become an essential requirement for management of these patients. The World Health Organization (WHO), European Neuroendocrine Tumor Society (ENETS), and the American Joint Committee on Cancer/International Union Against Cancer (AJCC/UICC) have developed classification systems (Table 80-1). Although there are some differences between these different classification systems, each uses similar information, and it is now recommended that the basic data underlying the classification be included in all standard pathology reports. These classification systems divide NETs from all sites into those that are well differentiated (low grade [G1] or intermediate grade [G2]) and those that are poorly differentiated (high grade [G3] divided into either small-cell carcinoma or large-cell neuroendocrine carcinoma [NEC]) (Table 80-1). In these classification systems, both pNETs and GI-NETs (carcinoids) are classified as NETs, and the old term of carcinoid is equivalent to well-differentiated NETs of the GI tract. These classification systems are based on not only the differentiation of the NET, but also a grading system assessing proliferative indices (Ki_{67} and the mitotic count)

TABLE 80-1 Comparison of the Criteria for the Tumor Category in the ENETS AJCC TNM Classifications of Pancreatic and Appendiceal NETs (Top Panel) and the WHO/ENETS Grading and Classification (Bottom Panel)

A. TNM Classification

	ENETS TNM	AJCC/UICC TNM
pNETs		
T1	Confined to pancreas, <2 cm	Confined to pancreas, <2 cm
T2	Confined to pancreas, 2–4 cm	Confined to pancreas, >2 cm
T3	Confined to pancreas, >4 cm, or invasion of duodenum or bile duct	Peripancreatic spread, but without major vascular invasion (truncus coeliacus, superior mesenteric artery)
T4	Invasion of adjacent organs or major vessels	Major vascular invasion
Appendiceal NETs		
T1	≤1 cm; invasion of muscularis propria	T1a, ≤1 cm; T1b, >1–2 cm
T2	≤2 cm and <3 mm invasion of subserosa/mesoappendix	>2–4 cm or invasion of cecum
T3	>2 cm or >3 mm invasion of subserosa/mesoappendix	>4 cm or invasion of ileum
T4	Invasion of peritoneum/other organs	Invasion of peritoneum/other organs

Abbreviations: AJCC, American Joint Committee on Cancer; ENETS, European Neuroendocrine Tumor Society; NET, neuroendocrine tumor; pNET, pancreatic neuroendocrine tumor; TNM, tumor, node, metastasis; UICC, International Union Against Cancer.

Source: Modified from DS Klimstra: Semin Oncol 40:23, 2013 and G Kloppel et al: Virchow Arch 456:595, 2010.

B. Grading

CLASSIFICATION	GRADE	MITOTIC COUNT (per 10 HPF)	KI ₆₇ INDEX (%)
NET	G1	<2	≤2
NET	G2	2–20	3–20
NEC (small cell and large cell)	G3	>20	>20

Abbreviations: HPF, high-power field; NEC, neuroendocrine carcinoma; NET, neuroendocrine tumor.

TABLE 80-2 Gastrointestinal Neuroendocrine Tumor Syndromes

Name	Biologically Active Peptide(s) Secreted	Incidence (New Cases/10 ⁶ Population/Year)	Tumor Location	Malignant, %	Associated with MEN 1, %	Main Symptoms/Signs
I. Established Specific Functional Syndromes						
A. Carcinoid syndrome due to GI-NET						
Carcinoid syndrome	Serotonin, possibly tachykinins, motilin, prostaglandins	0.5–2	Midgut (75–87%) Foregut (2–33%) Hindgut (1–8%) Unknown (2–15%)	95–100	Rare	Diarrhea (32–84%) Flushing (63–75%) Pain (10–34%) Asthma (4–18%) Heart disease (11–41%)
B. Well-established functional pNET syndromes						
Zollinger-Ellison syndrome	Gastrin	0.5–1.5	Duodenum (70%) Pancreas (25%) Other sites (5%)	60–90	20–25	Pain (79–100%) Diarrhea (30–75%) Esophageal symptoms (31–56%)
Insulinoma	Insulin	1–2	Pancreas (>99%)	<10	4–5	Hypoglycemic symptoms (100%)
VIPoma (Verner-Morrison syndrome, pancreatic cholera, WDHA)	Vasoactive intestinal peptide	0.05–0.2	Pancreas (90%, adult) Other (10%, neural, adrenal, periganglionic)	40–70	6	Diarrhea (90–100%) Hypokalemia (80–100%) Dehydration (83%)
Glucagonoma	Glucagon	0.01–0.1	Pancreas (100%)	50–80	1–20	Rash (67–90%) Glucose intolerance (38–87%) Weight loss (66–96%)
Somatostatinoma	Somatostatin	Rare	Pancreas (55%) Duodenum/jejunum (44%)	>70	45	Diabetes mellitus (63–90%) Cholelithiasis (65–90%) Diarrhea (35–90%)
GRFoma	Growth hormone-releasing hormone	Unknown	Pancreas (30%) Lung (54%) Jejunum (7%) Other (13%)	>60	16	Acromegaly (100%)
ACTHoma	ACTH	Rare	Pancreas (4–16% all ectopic Cushing's)	>95	Rare	Cushing's syndrome (100%)
pNET causing carcinoid syndrome	Serotonin, ?tachykinins	Rare (<100 cases)	Pancreas (<1% all carcinoids)	60–88	Rare	Same as carcinoid syndrome above
pNET causing hypercalcemia	PTHRP Others unknown	Rare	Pancreas (rare cause of hypercalcemia)	84	Rare	Abdominal pain due to hepatic metastases
II. Rare Specific Functional Syndromes						
pNET secreting renin	Renin	Rare	Pancreas	Unknown	No	Hypertension
pNET secreting luteinizing hormone	Luteinizing hormone	Rare	Pancreas	Unknown	No	Anovulation, virilization (female); reduced libido (male)
pNET secreting erythropoietin	Erythropoietin	Rare	Pancreas	100	No	Polycythemia
pNET secreting IGF-II	Insulin-like growth factor II	Rare	Pancreas	Unknown	No	Hypoglycemia
pNET secreting GLP-1	Glucagon-like peptide-1	Rare	Pancreas	Unknown	No	Hypoglycemia, diabetes
pNET secreting enteroglucagon	Enteroglucagon	Rare	Pancreas, small intestine	Unknown	Rare	Small intestinal hypertrophy, intestinal stasis, malabsorption
pNET secreting Cholecystokinin	Cholecystokinin	Rare	Pancreas	Unknown	No	Diarrhea, gallstones, peptic ulcer, weight loss
III. Possible Specific Functional pNET Syndromes						
pNET secreting calcitonin	Calcitonin	Rare	Pancreas (rare cause of hypercalcitonemia)	>80	16	Diarrhea (50%)
pNET secreting neuropeptid	Neurotensin	Rare	Pancreas (100%)	Unknown	No	Motility disturbances, vascular symptoms
pNET secreting pancreatic polypeptide (PPoma)	Pancreatic polypeptide	1–2	Pancreas	>60	18–44	Watery diarrhea
pNET secreting ghrelin	Ghrelin	Rare	Pancreas	Unknown	No	Effects on appetite, body weight
pNET secreting secretin	Secretin	Rare	Pancreas	Unknown	unknown	Watery diarrhea

(Continued)

TABLE 80-2 Gastrointestinal Neuroendocrine Tumor Syndromes (Continued)

Name	Biologically Active Peptide(s) Secreted	Incidence (New Cases/10 ⁶ Population/Year)	Tumor Location	Malignant, %	Associated with Men 1, %	Main Symptoms/Signs
IV. Nonfunctional Syndrome pNET						
PPoma/nonfunctional ^a	None	1–2	Pancreas (100%)	>60	18–44	Weight loss (30–90%) Abdominal mass (10–30%) Pain (30–95%)

^aPancreatic polypeptide–secreting tumors (PPomas) are listed in two places because most authorities classify these as not associated with a specific hormonal syndrome (nonfunctional); however, rare cases of watery diarrhea proposed to be due to PPomas have been reported.

Abbreviations: ACTH, adrenocorticotrophic hormone; GRFoma, growth hormone–releasing factor secreting pancreatic endocrine tumor; IGF-II, insulin-like growth factor II; MEN, multiple endocrine neoplasia; pNET, pancreatic neuroendocrine tumor; PPoma, tumor secreting pancreatic polypeptide; PTHrP, parathyroid hormone–related peptide; VIPoma, tumor secreting vasoactive intestinal peptide; WDHA, watery diarrhea, hypokalemia, and achlorhydria syndrome.

(Table 80-1). Based on these proliferative indices, NETs are classified as low grade (G1), intermediate grade (G2), or high grade (G3) (Table 80-1). In addition to the grading system, a TNM (TNM=tumor staging, T=primary size, N=regional lymph node involvement, M=distant metastases) classification has been proposed that is based on the level of tumor invasion, tumor size, and tumor extent (Table 80-1). Because of the proven prognostic value of these classification and grading systems, as well as the fact that NETs with different classifications/grades respond differently to treatments, these classification systems are now essential for the management of all NETs.

GI-NETs may or may not be associated with a specific functional syndrome (Table 80-2). In the case of pNETs the type of functional syndrome present is used to classify them into nine well-established specific functional syndromes (Table 80-2), seven additional very rare specific functional syndromes (less than five cases described), five possible specific functional syndromes (pNETs secreting calcitonin, neurotensin, pancreatic polypeptide [PP], ghrelin) (Table 80-2), and nonfunctional pNETs. Other functional hormonal syndromes due to nonpancreatic tumors (usually intra-abdominal in location) have been described only rarely and are not included in (Table 80-2). These include secretion by intestinal and ovarian tumors of peptide tyrosine tyrosine (PYY), which results in altered motility and constipation, and ovarian tumors secreting renin or aldosterone causing alterations in blood pressure or somatostatin causing diabetes or reactive hypoglycemia. Each of the functional syndromes listed in Table 80-2 is associated with symptoms due to the specific hormone released. In contrast, nonfunctional

pNETs release no products that cause a specific clinical syndrome. “Nonfunctional” is a misnomer in the strict sense because those tumors frequently ectopically secrete a number of peptides (PP, chromogranin A, ghrelin, neurotensin, α subunits of human chorionic gonadotropin, and neuron-specific enolase); however, they cause no specific clinical syndrome. The symptoms caused by nonfunctional pNETs are entirely due to the tumor per se. pNETs frequently ectopically secrete PP (60–85%), neurotensin (30–67%), calcitonin (30–42%), and to a lesser degree, ghrelin (5–65%). Whereas a few studies have proposed that their secretion can cause a specific functional syndrome, most studies support the conclusion that their ectopic secretion is not associated with a specific clinical syndrome, and thus they are listed in Table 80-2 as possible clinical syndromes. Because a large proportion of nonfunctional pNETs (60–90%) secrete PP, these tumors are often referred to as PPomas (Table 80-2). pNETs can secrete secretin (secretinoma) producing watery diarrhea; however, only two possible cases are described.

GI-NETs (carcinoids) can occur in almost any GI tissue (Table 80-3); however, at present, most (70%) have their origin in one of three sites: bronchus, jejunum, or colon/rectum. In the past, GI-NET (carcinoids) most frequently were reported in the appendix (i.e., 40%); however, at present they account for <5% (Table 80-3). Overall, the GI tract is the most common site for NETs, accounting for 64%, with the respiratory tract a distant second at 28%. Both race and sex can affect the frequency as well as the distribution of GI-NETs (carcinoids). African Americans have a higher incidence of carcinoids. Race is particularly important for rectal carcinoids, which are found in 41%

TABLE 80-3 GI-NET (Carcinoid) Location, Frequency of Metastases, and Association with the Carcinoid Syndrome

	Location (% of Total)	Incidence of Metastases	Incidence of Carcinoid Syndrome
Foregut			
Esophagus	<0.1	—	—
Stomach	4.6	10	9.5
Duodenum	2.0	—	3.4
Pancreas	0.7	71.9	20
Gallbladder	0.3	17.8	5
Bronchus, lung, trachea	27.9	5.7	13
Midgut			
Jejunum	1.8	{ 58.4	9
Ileum	14.9		9
Meckel's diverticulum	0.5	—	13
Appendix	4.8	38.8	<1
Colon	8.6	51	5
Liver	0.4	32.2	—
Ovary	1.0	28	50
Testis	<0.1	—	50
Hindgut			
Rectum	13.6	3.9	—

Abbreviation: GI-NET, gastrointestinal neuroendocrine tumor.

Source: Location is from the PAN-SEER data (1973–1999), and incidence of metastases is from the SEER data (1992–1999), reported by IM Modlin et al: Cancer 97:934, 2003. Incidence of carcinoid syndrome is from 4349 cases studied from 1950 to 1971, reported by JD Godwin: Cancer 36:560, 1975.

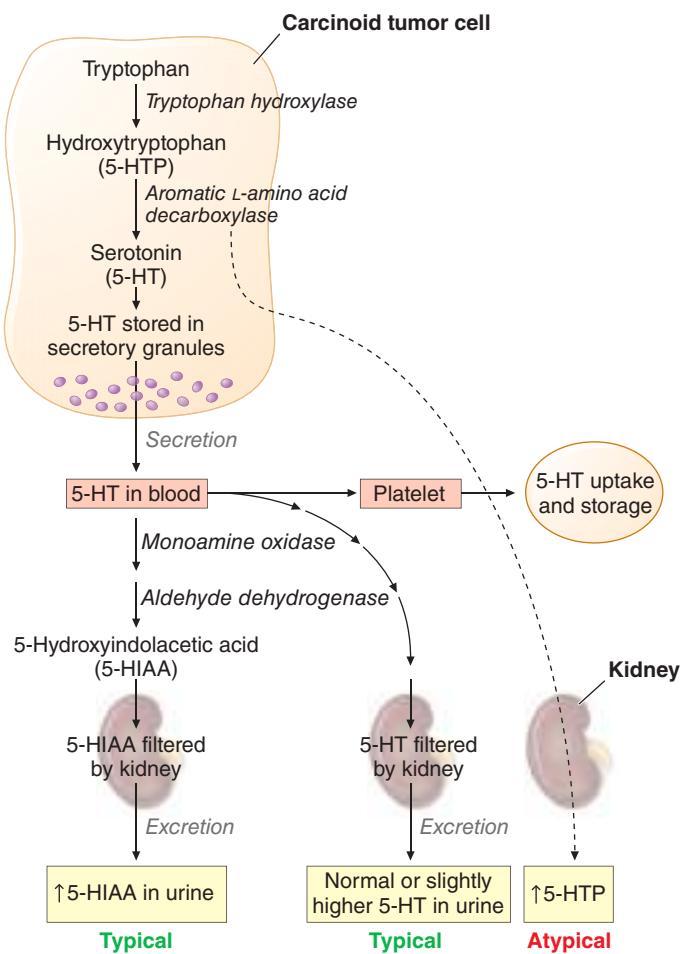


FIGURE 80-1 Synthesis, secretion, and metabolism of serotonin (5-HT) in patients with typical and atypical carcinoid syndromes. 5-HIAA, 5-hydroxyindolacetic acid.

of Asians/Pacific Islanders with NETs compared to 32% of American Indians/Alaskan natives, 26% of African Americans, and 12% of white Americans. Females have a lower incidence of small intestinal and pancreatic carcinoids.

The term *pancreatic neuroendocrine* or *endocrine tumor*, although widely used and therefore retained here, is also a misnomer, strictly speaking, because these tumors can occur either almost entirely in the pancreas (insulinomas, glucagonomas, nonfunctional pNETs, pNETs causing hypercalcemia) or at both pancreatic and extrapancreatic sites (gastrinomas, VIPomas [vasoactive intestinal peptide], somatostatinomas, GRFomas [growth hormone-releasing factor]). pNETs are also called islet cell tumors; however, the use of this term is discouraged because it is not established that they originate from the islets, and many can occur at extrapancreatic sites.

In addition to these classification/grading systems, a number of other factors have been identified that provide important prognostic information that can guide treatment (Table 80-4).

The exact incidence of GI-NETs (carcinoids) or pNETs varies according to whether only symptomatic tumors or all tumors are considered. The incidence of clinically significant carcinoids is 7–13 cases/million population per year, whereas any malignant carcinoids at autopsy are reported in 21–84 cases/million population per year. The incidence of GI-NETs (carcinoids) is ~25–50 cases per million in the United States, which makes them less common than adenocarcinomas of the GI tract. However, their incidence has increased sixfold in the last 30 years. In an analysis of 35,825 GI-NETs (carcinoids) (2004) from the U.S. Surveillance, Epidemiology, and End Results (SEER) database which includes predominantly malignant NETs, their incidence was 5.25/100,000 per year, and the 29-year prevalence was 35/100,000. Clinically significant pNETs have a prevalence of 10 cases/million population, with insulinomas, gastrinomas, and nonfunctional pNETs having an incidence of 0.5–2 cases/million population per year (Table 80-2). NF-pNETs are

predominating, often making up 50–80% of the series, and increasingly found when asymptomatic. pNETs account for 1–10% of all tumors arising in the pancreas and 1.3% of tumors in the SEER database. VIPomas are 2–8 times less common, glucagonomas are 17–30 times less common, and somatostatinomas are the least common. In autopsy studies, 0.5–1.5% of all cases have a pNET; however, in <1 in 1000 cases was a functional tumor thought to occur.

Both GI-NETs (carcinoids) and pNETs commonly show malignant behavior (Tables 80-2 and 80-3). With pNETs, except for insulinomas in which <10% are malignant, 50–100% in different series are malignant. With GI-NETs (carcinoids), the percentage showing malignant behavior varies in different locations (Table 80-3). For the three most common sites of NET's occurrence, the incidence of metastases varies greatly from the jejunum (58%), lung/bronchus (6%), and rectum (4%) (Table 80-3). With both GI-NETs (carcinoids) and pNETs, a number of factors (Table 80-4) are important prognostic factors in determining survival and the aggressiveness of the tumor. Patients with pNETs (excluding insulinomas) generally have a poorer prognosis than do patients with GI-NETs (carcinoids). The presence of liver metastases is the single most important prognostic factor in single and multivariate analyses for both GI-NETs (carcinoids) and pNETs. Particularly important in the development of liver metastases is the size of the primary tumor. For example, with small intestinal carcinoids, which are the most common cause of the carcinoid syndrome due to metastatic disease in the liver (Table 80-2), metastases occur in 15–25% if the tumor is <1 cm in diameter, 58–80% if it is 1–2 cm in diameter, and >75% if it is >2 cm in diameter. Similar data exist for gastrinomas and other pNETs; the size of the primary tumor is an independent predictor of the development of liver metastases. The presence of lymph node metastases, their ratio or presence of extra-hepatic metastases; the depth of invasion; the rapid rate of growth; various histologic features (differentiation, mitotic rates, growth indices, vessel density, vascular endothelial growth factor [VEGF], CD10 metalloproteinase expression, abnormal expression of p53, retinoblastoma or SMAD, and low expression of p27 nuclear staining, low progesterone receptor expression); necrosis; presence of cytokeratin; elevated serum alkaline phosphatase levels; older age; presence of circulating tumor cells; increased uptake on ¹⁸F-FDG-PET/CT scanning or low uptake (SUV_{max}) on ⁶⁸Ga-DOTANOC PET/CT scanning, and flow cytometric results, such as the presence of aneuploidy, are all important prognostic factors for the development of metastatic disease (Table 80-4). For patients with GI-NETs (carcinoids), additional associations with a worse prognosis include the development of the carcinoid syndrome (especially the development of carcinoid heart disease); male sex; the presence of a symptomatic tumor, a secondary malignancy, or greater increases in a number of tumor markers (5-hydroxyindolacetic acid [5-HIAA], neuropeptide K, chromogranin A), and the presence of various molecular features. With pNETs or gastrinomas, a worse prognosis is associated with female sex, overexpression of the Ha-ras oncogene or p53, the absence of Multiple Endocrine Neoplasia type 1 (MEN 1), presence of a NF-pNET, higher levels of various tumor markers (i.e., chromogranin A, gastrin, C-reactive protein), and presence of various histologic features (immunohistochemistry for c-KIT, low cyclin B1 or ATM, loss of PTEN/TSC-2, expression of fibroblast growth factor-13) and various molecular features (Table 80-4). The WHO, ENETS, and AJCC/UICC TNM classification systems and the grading systems (G1–G3) have important prognostic value and use in determining therapeutic management, that they are now generally routinely required.

A number of diseases due to various genetic disorders are associated with an increased incidence of NETs (Table 80-5). Each one is caused by a loss of a possible tumor-suppressor gene. The most important is MEN 1, which is an autosomal dominant disorder due to a defect in a 10-exon gene on 11q13, which encodes for a 610-amino-acid nuclear protein, menin (Chap. 381). Patients with MEN 1 develop hyperparathyroidism due to parathyroid hyperplasia in 95–100% of cases, pNETs in 80–100%, pituitary adenomas in 54–80%, adrenal adenomas in 27–36%, bronchial carcinoids in 8%, thymic carcinoids in 8% (predominately males), gastric carcinoids in 13–30% of patients with Zollinger-Ellison syndrome, skin tumors (angiofibromas [88%], collagenomas [72%]),

TABLE 80-4 Prognostic Factors in Neuroendocrine Tumors**I. Both GI-NETs (carcinoids) and pNETs**

Symptomatic presentation ($p < 0.05$)
 Performance status ($p < 0.04$)
 Presence/extent of liver metastases ($p < 0.01$)
 Presence of lymph node metastases or lymph node positive ratio ($p < 0.001$)
 Development of bone or extrahepatic metastases ($p < 0.01$)
 Depth of invasion ($p < 0.001$)
 Rapid rate of tumor growth
 Elevated serum alkaline phosphatase levels ($p = 0.003$)
 Primary tumor site/site ($p < 0.005$)
 High serum chromogranin A level ($p < 0.01$)
 Presence of one or more circulating tumor cells ($p < 0.001$)
 Increased uptake on (18)F-FDG PET scanning
 Low uptake (SUVmax) on (68)Ga-DOTANOC PET/CT scanning
 Various histologic/molecular features
 Tumor differentiation ($p < 0.001$)
 High growth indices (high Ki₆₇ index, PCNA expression)
 High mitotic counts ($p < 0.001$)
 Low progesterone receptor expression ($p < 0.001$)
 Necrosis present
 Presence of cytokeratin 19 ($p < 0.02$)
 Vascular or perineural invasion ($p < 0.02$)
 Vessel density (low microvessel density, increased lymphatic density)
 High CD10 metalloproteinase expression (in series with all grades of NETs)
 Flow cytometric features (i.e., aneuploidy)
 High VEGF expression (in low-grade or well-differentiated NETs only)
 Abnormal expression of p53, Rb, SMADs
 Loss of p27 expression (nuclear) ($p < 0.001$)
 WHO, ENETS, AJCC/UICC stage, and grade
 Presence of a pNET rather than GI-NET associated with poorer prognosis ($p = 0.0001$)
 Older age ($p < 0.01$)

II. GI-NETs (carcinoids)

Location of primary: appendix < lung, rectum < small intestine < pancreas
 Presence of carcinoid syndrome
 Laboratory results (urinary 5-HIAA levels [$p < 0.01$], plasma neuropeptide K [$p < 0.05$], serum chromogranin A [$p < 0.01$])
 Presence of a second malignancy
 Male sex ($p < 0.001$)
 Molecular findings (TGF- α expression [$p < .05$], chr 16q LOH or gain chr 4p [$p < .05$], gain in chr 14, loss of 3p13, loss of succinate dehydrogenase expression [ileal carcinoid], upregulation of Hoxc6), molecular profiling category (mutations, epigenetic changes, copy number-small intestinal NETs)

III. pNETs

Location of primary: duodenal (gastrinoma) better than pancreatic
 Ha-ras oncogene or p53 overexpression
 Female sex
 MEN 1 syndrome absent
 Presence of nonfunctional tumor (some studies, not all)
 Various histologic features: IHC positivity for c-KIT, low cyclin B1 or ATM expression ($p < 0.01$), loss of PTEN or of tuberous sclerosis-2 IHC, expression of fibroblast growth factor-13; high SSTR2 expression ($p = .001$);
 Laboratory findings (increased chromogranin A in some studies; gastrinomas—increased gastrin level, increased CRP ($p < 0.001$))
 Molecular findings (increased HER2/neu expression [$p = 0.032$], chr 1q, 3p, 3q, or 6q LOH [$p = 0.0004$], EGF receptor overexpression [$p = 0.034$], gains in chr 7q, 17q, 17p, 20q; alterations in the VHL gene [deletion, methylation]; presence of FGFR4-G388R single-nucleotide polymorphism); loss of ATRX/DAXX or positive for alternative lengthening of telomeres ($p < 0.001$); high nuclear surviving expression ($p < 0.01$)
 PHLDA3 LOH; altered miRNA expression (Inc miRNA-21, miRNA-196)

Abbreviations: 5-HIAA, 5-hydroxyindoleacetic acid; AJCC, American Joint Committee on Cancer; ATM, ataxia telangiectasia mutated kinase; ATRX, a-thalassemia/mental retardation X-linked; chr, chromosome; CRP, C-reactive protein; DAXX, death domain-associated protein; EGF, epidermal growth factor; FGFR, fibroblast growth factor receptor; GI-NET, gastrointestinal neuroendocrine tumor; IHC, immunohistochemistry; Ki₆₇, proliferation-associated nuclear antigen recognized by Ki₆₇ monoclonal antibody; LOH, loss of heterozygosity; MEN, multiple endocrine neoplasia; NET, neuroendocrine tumors; PCNA, proliferating cell nuclear antigen; PHLDA3, Pleckstrin homology-like domain family A, member 3; pNET, pancreatic neuroendocrine tumor; PTEN, phosphatase and tensin homologue deleted from chromosome 10; Rb, retinoblastoma; SSTR2, somatostatin receptor subtype 2; TGF- α , transforming growth factor α ; TNM, tumor, node, metastasis; UICC, International Union Against Cancer; VEGF, vascular endothelial growth factor; WHO, World Health Organization.

central nervous system (CNS) tumors (meningiomas, ependymomas, schwannomas [$<8\%$]), and smooth-muscle tumors (leiomyomas, leiomyosarcomas [$1\%-7\%$]). Among patients with MEN 1, 80–100% develop nonfunctional pNETs (most are microscopic with 0–13% large/symptomatic), and functional pNETs occur in 20–80% in different series,

with a mean of 54% developing Zollinger-Ellison syndrome, 18% insulinomas, 3% glucagonomas, 3% VIPomas, and $<1\%$ GFRomas or somatostatinomas. MEN 1 is present in 20–25% of all patients with Zollinger-Ellison syndrome, 4% of patients with insulinomas, and a low percentage ($<5\%$) of patients with other pNETs.

TABLE 80-5 Genetic Syndromes Associated with an Increased Incidence of Neuroendocrine Tumors (NETs) (GI-NETs [Carcinoids] or pNETs)

SYNDROME	LOCATION OF GENE MUTATION AND GENE PRODUCT	NETS SEEN/ FREQUENCY
Multiple endocrine neoplasia type 1 (MEN 1)	11q13 (encodes 610-amino-acid protein, menin)	80–100% develop pNETs (microscopic), 20–80% (clinical): (nonfunctional > gastrinoma > insulinoma) GI-NETs (Carcinoids): gastric (13–30%), bronchial/thymic (8%)
von Hippel–Lindau disease	3q25 (encodes 213-amino-acid protein)	12–17% develop pNETs (almost always nonfunctional)
von Recklinghausen's disease (neurofibromatosis 1 [NF-1])	17q11.2 (encodes 2485-amino-acid protein, neurofibromin)	0–10% develop pNETs, primarily duodenal somatostatinomas (usually nonfunctional) Rarely insulinoma, gastrinoma
Tuberous sclerosis	9q34 (TSC1) (encodes 1164-amino-acid protein, hamartin), 16p13 (TSC2) (encodes 1807-amino-acid protein, tuberin)	0–9% develop pNETs (nonfunctional and functional [insulinoma, gastrinoma])

Abbreviations: GI, gastrointestinal; PNETs, pancreatic neuroendocrine tumors.

Three phacomatoses associated with NETs are von Hippel–Lindau disease (VHL), von Recklinghausen's disease (neurofibromatosis type 1 [NF-1]), and tuberous sclerosis (Bourneville's disease) (Table 80-5). VHL is an autosomal dominant disorder due to defects on chromosome 3p25, which encodes for a 213-amino-acid protein that interacts with the elongin family of proteins as a transcriptional regulator (Chaps. 86, 309, 380, 381). In addition to cerebellar hemangioblastomas, renal cancer, and pheochromocytomas, 10–17% develop a pNET. Most are non-functional, although insulinomas and VIPomas have been reported. Patients with NF-1 (von Recklinghausen's disease) have defects in a gene on chromosome 17q11.2 that encodes for a 2845-amino-acid protein, neurofibromin, which functions in normal cells as a suppressor of the *ras* signaling cascade (Chap. 86). Up to 10% of these patients develop an upper GI-NET (carcinoid), characteristically in the periampullary region (54%). Many are classified as somatostatinomas because they contain somatostatin immunocytochemically; however, they uncommonly secrete somatostatin and rarely produce a clinical somatostatinoma syndrome. NF-1 has rarely been associated with insulinomas and Zollinger-Ellison syndrome. NF-1 accounts for 48% of all duodenal somatostatinomas and 23% of all ampullary GI-NETs (carcinoids). Tuberous sclerosis is caused by mutations that alter either the 1164-amino-acid protein hamartin (TSC1) or the 1807-amino-acid protein tuberin (TSC2) (Chap. 86). Both hamartin and tuberin interact in a pathway related to phosphatidylinositol 3-kinases and mammalian target of rapamycin (mTOR) signaling cascades. A few cases including nonfunctional and functional pNETs (insulinomas and gastrinomas) have been reported in these patients (Table 80-6).

A few other syndromes involving GI-NETs have been described due to various mutations, but no inherited cases have been reported. Mahvash disease is associated with the development of α -cell hyperplasia, hyperglucagonemia, without features of the glucagonoma syndrome, the development of NF pNETs, and is due to inactivating mutations of the human glucagon receptor. The polycythemia-paraganglioma-somatostatinoma (SSoma) syndrome involves childhood polycythemia, later the development of norepinephrine-producing paragangliomas (mean age 17 years) then SSomas (mean age 29 years) with gallbladder disease. In 28% of the patients, a somatic mutation in hypoxia-inducible factor 2 alphas is found.

Most GI NETs (carcinoids) are sporadic, although as discussed above gastric carcinoids occur (type 2) in MEN1 syndrome, duodenal

TABLE 80-6 Clinical Characteristics in Patients with Carcinoid Syndrome

	PERCENTAGE (RANGE)	
	AT PRESENTATION	DURING COURSE OF DISEASE
Symptoms/signs		
Diarrhea	32–93%	68–100%
Flushing	23–100%	45–96%
Pain	10%	34%
Asthma/wheezing	4–14%	3–18%
Pellagra	0–7%	0–5%
None	12%	22%
Carcinoid heart disease present	11–40%	14–41%
Demographics		
Male	46–59%	46–61%
Age		
Mean	57 years	59.2 years
Range	25–79 years	18–91 years
Tumor location		
Foregut	5–14%	0–33%
Midgut	57–87%	60–100%
Hindgut	1–7%	0–8%
Unknown	2–21%	0–26%

carcinoids (SSomas) occur in NF-1 and a small percentage of patients (<3%) with small intestinal carcinoids have a familial form of the disease, which in one family was due to mutations in the inositol polyphosphate multikinase gene (IMPK).

Mutations in common oncogenes (*ras*, *myc*, *fos*, *src*, *jun*) or common tumor-suppressor genes (*p53*, retinoblastoma susceptibility gene) are not commonly found in either pNETs or GI-NETs (carcinoids). However, frequent (70%) gene amplifications in *MDM2*, *MDM4*, and *WIP1* inactivating the *p53* pathway are noted in well-differentiated pNETs, and the retinoblastoma pathway is altered in the majority of pNETs. In addition to these genes, additional alterations that may be important in their pathogenesis include changes in the *MEN1* gene, *p16/MTS1* tumor-suppressor gene, and *DPC4/Smad4* gene; amplification of the HER-2/neu protooncogene; alterations in transcription factors (Hoxc6 [GI carcinoids]), growth factors, and their receptors; methylation of a number of genes that probably results in their inactivation; and deletions of unknown tumor-suppressor genes as well as gains in other unknown genes. The clinical antitumor activity of everolimus, an mTOR inhibitor, and sunitinib, a tyrosine kinase inhibitor (PDGFR, VEGFR1, VEGFR2, c-KIT, FLT-3), support the importance of the mTOR-AKT pathway and tyrosine kinase receptors in mediating growth of malignant NETs (especially pNETs). The importance of the mTOR pathway in pNET growth is further supported by the finding that a single-nucleotide polymorphism (FGFR4-G388R, in fibroblast growth factor receptor 4) affects selectivity to the mTOR inhibitor and can result in significantly higher risk of advanced pNET stage and liver metastases (Table 80-4). Comparative genomic hybridization, genome-wide allelotyping studies, and genome-wide single-nucleotide polymorphism analyses have shown that chromosomal losses and gains are common in pNETs and GI-NETs (carcinoids), but they differ between these two NETs, and some have prognostic significance (Table 80-4). Mutations in the *MEN1* gene are probably particularly important. Loss of heterozygosity at the MEN 1 locus on chromosome 11q13 is noted in 93% of sporadic pNETs (i.e., in patients without MEN 1) and in 26–75% of sporadic GI-NETs (carcinoids). Mutations in the *MEN1* gene are reported in 31–34% of sporadic gastrinomas. Exomic sequencing of sporadic pNETs found that the most frequently altered gene was *MEN1*, occurring in 44% of patients, followed by mutations in 43% of patients in genes encoding for two subunits of a transcription/chromatin remodeling complex consisting of DAXX (death-domain-associated protein) and ATRX (α -thalassemia/mental retardation

most GI NETs (carcinoids) are sporadic, although as discussed above gastric carcinoids occur (type 2) in MEN1 syndrome, duodenal

syndrome X-linked) and in 15% of patients in the mTOR pathway. The presence of a number of these molecular alterations in pNETs or GI-NETs (carcinoids) correlates with tumor growth, tumor size, and disease extent or invasiveness and may have prognostic significance (Table 80-4).

GI NETs (carcinoids) have frequently loss of chromosome 18 (>60%) as well as losses on Chr 9p, 16q and chromosomal gains of 17q, 19p (57%) and lesser gains on 4q, 14q, and Chr 5, but the exact genes mediating possible effects on the tumor in these areas are still unclear. In contrast to pNETs, mutations in GI-NETs (carcinoids) are uncommon, and in one study of 180 small intestinal carcinoids, with exome and genome-sequencing analysis recurrent mutations were only observed in the CDKN1B gene (cyclin-dependent kinase inhibitor 1B [p27^{Kip1}]) in 8%. Integrative genomic analysis incorporating DNA methylation, show that small intestinal GI carcinoids commonly have epigenetic changes and three molecular subgroups with differing clinical course and outcomes have been identified (Table 80-4).

■ CHARACTERISTICS OF THE MOST COMMON GI-NETs (CARCINOIDS)

Appendiceal NETs (Carcinoids) Appendiceal NETs (carcinoids) occur in 1 in every 200–300 appendectomies, usually in the appendiceal tip, have an incidence of 0.15/100,000 per year, comprise 2–5% of all GI-NETs (carcinoids), and comprise 32–80% of all appendiceal tumors. The mean age at diagnosis is 38–51 years. Most (i.e., >90%) are <1 cm in diameter without metastases in older studies, but more recently, 2–35% have had metastases (Table 80-3). In the SEER data of 1570 appendiceal carcinoids, 62% were localized, 27% had regional metastases, and 8% had distant metastases. The risk of metastases increases with size, with those <1 cm having a 0 to <10% risk of metastases and those >2 cm having a 25–44% risk. Besides tumor size, other important prognostic factors for metastases include basal NET location, invasion of mesoappendix, poor differentiation, advanced stage or WHO/ENETS classification, older age, and positive resection margins. The 5-year survival is 88–100% for patients with localized disease, 78–100% for patients with regional involvement, and 12–28% for patients with distal metastases. In patients with tumors <1 cm in diameter, the 5-year survival is 95–100%, whereas it is 29% if tumors are >2 cm in diameter. Most tumors are well-differentiated G1 tumors (87%) (Table 80-1), with the remainder primarily well-differentiated G2 tumors (13%); with poorly differentiated G3 tumors uncommon (<1%). Their percentage of the total number of carcinoids decreased from 43.9% (1950–1969) to 2.4% (1992–1999). Appendiceal goblet cell (GC) NETs (carcinoids)/carcinomas are a rare subtype (<5%) that are mixed adeno-neuroendocrine carcinomas. They are malignant and are thought to comprise a distinct entity; they frequently present with advanced disease and are recommended to be treated as adenocarcinomas, not carcinoid tumors.

■ SMALL INTESTINAL NETs (CARCINOIDS)

Small intestinal (SI) NETs (carcinoids) have a reported incidence of 0.67/100,000 in the United States, 0.32/100,000 in England, and 1.12/100,000 in Sweden and comprise >50% of all SI tumors. There is a male predominance (1.5:1), and race affects frequency, with a lower frequency in Asians and greater frequency in African Americans. The mean age of presentation is 52–63 years, with a wide range (1–93 years). Familial SI carcinoid families exist but are very uncommon. SI NETs (carcinoids) are frequently multiple; 9–18% occur in the jejunum, 70–80% are present in the ileum, and 70% occur within 6 cm (2.4 in.) of the ileocecal valve. Forty percent are <1 cm in diameter, 32% are 1–2 cm, and 29% are >2 cm. They are characteristically well differentiated; however, they are generally invasive, with 1.2% being intramucosal in location, 27% penetrating the submucosa, and 20% invading the muscularis propria. Metastases occur in a mean of 47–58% (range 20–100%). Liver metastases occur in 38%, to lymph nodes in 37% and more distant in 20–25%. They characteristically cause a marked fibrotic reaction, which can lead to intestinal obstruction. Tumor size is an important variable in the frequency of metastases. However, even small NETs (carcinoids) of the small intestine (<1 cm) have metastases in 15–25% of cases, whereas the proportion increases to 58–100% for tumors 1–2 cm in diameter.

Carcinoids also occur in the duodenum, with 31% having metastases. Duodenal tumors <1 cm rarely metastasize, whereas 33% of those >2 cm had metastases. SI NETs (carcinoids) are the most common cause (60–87%) of the carcinoid syndrome (Table 80-6). Important prognostic factors are listed in Table 80-4, and particularly important are the tumor extent, proliferative index by grading, and stage (Table 80-1). SI NETs (carcinoid) are well differentiated in 99% of cases with 61–72% being G1 and 11–37% being G2. The overall survival at 5 years is 55–75%; however, it varies markedly with disease extent, being 65–90% with localized disease, 66–72% with regional involvement, and 36–43% with distant disease.

Rectal NETs (Carcinoids) Rectal NETs (carcinoids) comprise 27% of all GI-NETs (carcinoids) and 16% of all NETs and are increasing in frequency. In Europe they comprise 5–14% of all NETs and in some Asian series (Japan, China, Korea) they comprise 60–89% of all NETs. In the U.S. SEER data, they currently have an incidence of 0.86/100,000 per year (up from 0.2/100,000 per year in 1973) and represent 1–2% of all rectal tumors. They are found in ~1 in every 1500/2500 proctoscopies/colonoscopies or 0.05–0.07% of individuals undergoing these procedures. Nearly all occur between 4 and 13 cm above the dentate line. Most are small, with 66–80% being <1 cm in diameter, and rarely metastasize (5%). Tumors between 1 and 2 cm can metastasize in 5–30%, and those >2 cm, which are uncommon, in >70%. Most invade only to the submucosa (75%), with 2.1% confined to the mucosa, 10% to the muscular layer, and 5% to adjacent structures. Histologically, most are well differentiated (98%) with 72% ENETS/WHO grade G1 and 28% grade G2 (Table 80-1). Overall survival is 88%; however, it is very much dependent of the stage, with 5-year survival of 91% for localized disease, 36–49% for regional disease, and 20–32% for distant disease. Risk factors are listed in Table 80-4 and particularly include tumor size, depth of invasion, presence of metastases, differentiation, and recent TNM classification and grade.

Bronchial NETs (Carcinoids) Bronchial NETs (carcinoids) comprise 25–33% of all well-differentiated NETs and 90% of all the poorly differentiated NETs found, likely due to a strong association with smoking. Their incidence ranges from 0.2 to 2/100,000 per year in the United States and European countries and is increasing at a rate of 6% per year. They are slightly more frequent in females and in whites compared with those of Hispanic/Asian/African descent, and are most commonly seen in the sixth decade of life, with a younger age of presentation for typical carcinoids (45 years) compared to atypical carcinoids (55 years).

A number of different classifications of bronchial GI-NETs (carcinoids) have been proposed. The principal factors used in classifying lung NETs include: morphology, presence or absence of necrosis, mitotic rate and size. In some studies, they are classified into four categories: typical carcinoid (also called bronchial carcinoid tumor, Kulchitsky cell carcinoma I [KCC-I]), atypical carcinoid (also called well-differentiated NEC [KC-II]), intermediate small-cell NEC, and small-cell neuroendocarcinoma (KC-III). Another proposed classification includes three categories of lung NETs: benign or low-grade malignant (typical carcinoid), low-grade malignant (atypical carcinoid), and high-grade malignant (poorly differentiated carcinoma of the large-cell or small-cell type). The WHO classification includes four general categories: typical carcinoid, atypical carcinoid, large-cell NEC, and small-cell carcinoma. The ratio of typical to atypical carcinoids is 8–10:1, with the typical carcinoids comprising 1–2% of lung tumors, atypical 0.1–0.2%, large-cell NETs 0.3%, and small-cell lung cancer 9.8% of all lung tumors. These different categories of lung NETs have different prognoses, varying from excellent for typical carcinoid to poor for small-cell NECs. The occurrence of large-cell and small-cell lung carcinoids, but not typical or atypical lung carcinoids, is related to tobacco use. The 5-year survival is very much influenced by the classification of the tumor, with survival of 92–100% for patients with a typical carcinoid, 61–88% with an atypical carcinoid, 13–57% with a large-cell neuroendocrine tumor, and 5% with a small-cell lung cancer. Typical/atypical lung carcinoids are generally well-differentiated with typical lung carcinoids sharing some homologies with G1 NETs, atypical

sharing some homologies with G2 NETs of the GI tract, whereas small cell and large cell lung NECs are poorly differentiated and correspond to the G3 NEC category of the GI Tract (Table 80-1).

Gastric NET (Carcinoids) Gastric NETs (carcinoids) account for 3 of every 1000 gastric neoplasms and 1.3–2% of all carcinoids, and their relative frequency has increased three- to fourfold over the last five decades (2.2% in 1950 to 9.6% in 2000–2007, SEER data). At present, it is unclear whether this increase is due to better detection with the increased use of upper GI endoscopy or to a true increase in incidence. Gastric NETs (carcinoids) are generally classified into three different categories, and this has important implications for pathogenesis, prognosis, and treatment. Each originates from gastric enterochromaffin-like (ECL) cells, one of the six types of gastric neuroendocrine cells, in the gastric mucosa. Two subtypes are associated with hypergastrinemic states, either chronic atrophic gastritis (type I) (70–80% of all gastric NETs [carcinoids]) or Zollinger-Ellison syndrome, which is almost always a part of the MEN 1 syndrome (type II) (5–6% of all cases). These tumors generally pursue a benign course, with type I uncommonly (<10%) associated with metastases, whereas type II tumors are slightly more aggressive, with 10–30% associated with metastases. Gastric carcinoids type 1 and type 2 are usually multiple, small, and infiltrate only to the submucosa. The third subtype of gastric NETs (carcinoids) (type III) (sporadic) occurs without hypergastrinemia (14–25% of all gastric carcinoids) and has an aggressive course, with 54–66% developing metastases. Sporadic carcinoids are usually single, large tumors; 50% have atypical histology, and they can be a cause of the carcinoid syndrome. Five-year survival is 99–100% in patients with type I, 60–90% in patients with type II, and 50% in patients with type III gastric NETs (carcinoids). Type 1 gastric carcinoids are usually grade G1 and well differentiated; type 2 are well-differentiated grade G1 or G2; and type 3 are characteristically NEC G3 with poor differentiation.

■ CLINICAL PRESENTATION OF NETs (CARCINOIDS)

GI/Lung NET (Carcinoid) Without the Carcinoid Syndrome

The age of patients at diagnosis ranges from 10 to 93 years, with a mean age of 63 years for the small intestine, 43–60 years for bronchial, and 66 years for the rectum. The presentation is diverse and is related to the site of origin and the extent of malignant spread. In the appendix, NETs (carcinoids) usually are found incidentally during surgery for suspected appendicitis. SI NETs (carcinoids) in the jejunileum present with periodic abdominal pain (51%), intestinal obstruction with ileus/invagination (31%), an abdominal tumor (17%), or GI bleeding (11%). Because of the vagueness of the symptoms, the diagnosis usually is delayed ~2 years from onset of the symptoms, with a range up to 20 years. Duodenal, gastric, and rectal NETs (carcinoids) are most frequently found by chance at endoscopy. The most common symptoms of rectal carcinoids are melena/bleeding (39%), constipation (17%), and diarrhea (12%). Bronchial NETs (carcinoids) frequently are discovered as a lesion on a chest radiograph, and 31–43% of the patients are asymptomatic. Thymic NETs (carcinoids) present as anterior mediastinal masses, usually on chest radiograph or computed tomography (CT) scan. Ovarian and testicular NETs (carcinoids) usually present as masses discovered on physical examination or ultrasound. Metastatic NETs (carcinoids) in the liver frequently present as hepatomegaly in a patient who may have minimal symptoms and nearly normal liver function test results.

■ GI-NETs (CARCINOIDS) WITH SYSTEMIC SYMPTOMS DUE TO SECRETED PRODUCTS

GI/lung NETs (carcinoids) immunocytochemically can contain numerous GI peptides: gastrin, insulin, somatostatin, motilin, neurotensin, tachykinins (substance K, substance P, neuropeptide K), glucagon, gastrin-releasing peptide, vasoactive intestinal peptide (VIP), PP, ghrelin, other biologically active peptides (ACTH, calcitonin, growth hormone, GRF), prostaglandins, and bioactive amines (serotonin). These substances may or may not be released in sufficient amounts to cause symptoms. In various studies of patients with GI-NETs (carcinoids), elevated serum levels of PP were found in 43%, motilin

in 14%, gastrin in 15%, and VIP in 6%. Foregut NETs (carcinoids) are more likely to produce various GI peptides than are midgut NETs (carcinoids). Ectopic ACTH production causing Cushing's syndrome is seen increasingly with foregut carcinoids (respiratory tract primarily) and, in some series, has been the most common cause of the ectopic ACTH syndrome, accounting for 64% of all cases. Acromegaly due to growth hormone-releasing factor release occurs with foregut NETs (carcinoids), as does the somatostatinoma syndrome, but rarely occurs with duodenal NETs (carcinoids). The most common systemic syndrome with GI-NETs (carcinoids) is the carcinoid syndrome, which is discussed in detail in the next section.

■ CARCINOID SYNDROME

Clinical Features The cardinal features from a number of series at presentation as well as during the disease course are shown in Table 80-6.

Flushing and diarrhea are the two most common symptoms, occurring in a mean of 69–70% of patients initially and in up to 78% of patients during the course of the disease. The characteristic flush is of sudden onset; it is a deep red or violaceous erythema of the upper body, especially the neck and face, often associated with a feeling of warmth and occasionally associated with pruritus, lacrimation, diarrhea, or facial edema. Flushes may be precipitated by stress; alcohol; exercise; certain foods, such as cheese; or certain agents, such as catecholamines, pentagastrin, and serotonin reuptake inhibitors. Flushing episodes may be brief, lasting 2–5 min, especially initially, or may last hours, especially later in the disease course. Flushing usually is associated with metastatic midgut NETs (carcinoids) but can also occur with foregut NETs (carcinoids). With bronchial NETs (carcinoids), the flushes frequently are prolonged for hours to days, reddish in color, and associated with salivation, lacrimation, diaphoresis, diarrhea, and hypotension. The flush associated with gastric NETs (carcinoids) can also be reddish in color, but with a patchy distribution over the face and neck, although the classic flush seen with midgut NETs (carcinoids) can also be seen with gastric NETs (carcinoids). It may be provoked by food and have accompanying pruritus.

Diarrhea usually occurs with flushing (85% of cases). The diarrhea usually is described as watery, with 60% of patients having <1 L/d of diarrhea. Steatorrhea is present in 67%, and in 46%, it is >15 g/d (normal <7 g). Abdominal pain may be present with the diarrhea or independently in 10–34% of cases.

Cardiac manifestations occur initially in 11–40% (mean 26%) of patients with carcinoid syndrome and in 14–41% (mean 30%) at some time in the disease course. The cardiac disease is due to the formation of fibrotic plaques (composed of smooth-muscle cells, myofibroblasts, and elastic tissue) involving the endocardium, primarily on the right side, although lesions on the left side also occur occasionally (mean 11%, range 0–25), especially if a patent foramen ovale exists. The dense fibrous deposits are most commonly on the ventricular aspect of the tricuspid valve and less commonly on the pulmonary valve cusps. They can result in constriction of the valves, and pulmonic stenosis is usually predominant, whereas the tricuspid valve is often fixed open, resulting in regurgitation predominating. Overall, in patients with carcinoid heart disease, 90–100% have tricuspid insufficiency, 43–59% have tricuspid stenosis, 50–81% have pulmonary insufficiency, 25–59% have pulmonary stenosis, and 11% (0–25%) left-side lesions. Up to 80% of patients with cardiac lesions develop heart failure. Lesions on the left side are much less extensive, occur in 30% at autopsy, and most frequently affect the mitral valve. Up to 80% of patients with cardiac lesions have evidence of heart failure. At diagnosis in various series, 27–43% of patients are in New York Heart Association class I, 30–40% are in class II, 13–31% are in class III, and 3–12% are in class IV. At present, carcinoid heart disease is reported to be decreasing in frequency and severity, with mean occurrence in 20% of patients and occurrence in as few as 3–4% in some reports. Whether this decrease is due to the widespread use of somatostatin analogues, which control the release of bioactive agents thought involved in mediating the heart disease, is unclear.

Other clinical manifestations include wheezing or asthma-like symptoms (8–18%), pellagra-like skin lesions (2–25%), and impaired

cognitive function. A variety of noncardiac problems due to increased fibrous tissue have been reported, including retroperitoneal fibrosis causing urethral obstruction, Peyronie's disease of the penis, intraabdominal fibrosis, and occlusion of the mesenteric arteries or veins.

Pathobiology Carcinoid syndrome occurred in 8% of 8876 patients with GI-NETs (carcinoids), with a rate of 1.7–18.4% in different studies. It occurs only when sufficient concentrations of products secreted by the tumor reach the systemic circulation. In 91–100% of cases, this occurs after distant metastases to the liver. Rarely, primary GI-NETs (carcinoids) with nodal metastases with extensive retroperitoneal invasion, pNETs (carcinoids) with retroperitoneal lymph nodes, or NETs (carcinoids) of the lung, testis or ovary with direct access to the systemic circulation can cause the carcinoid syndrome without hepatic metastases. All GI-NETs (carcinoids) do not have the same propensity to metastasize and cause the carcinoid syndrome (Table 80-3). Midgut NETs (carcinoids) account for 57–67% of cases of carcinoid syndrome, foregut NETs (carcinoids) for 0–33%, hindgut for 0–8%, and an unknown primary location for 2–26% (Tables 80-3 and 80-6).

One of the main secretory products of GI-NETs (carcinoids) involved in the carcinoid syndrome is serotonin (5-HT) (Fig. 80-1), which is synthesized from tryptophan. Up to 50% of dietary tryptophan can be used in this synthetic pathway by tumor cells, and this can result in inadequate supplies for conversion to niacin; hence, some patients (2.5%) develop pellagra-like lesions. Serotonin has numerous biologic effects, including stimulating intestinal secretion with inhibition of absorption, stimulating increases in intestinal motility, and stimulating fibrogenesis. In various studies, 56–88% of all GI-NETs (carcinoids) were associated with serotonin overproduction; however, 12–26% of the patients did not have the carcinoid syndrome. In one study, platelet serotonin was elevated in 96% of patients with midgut NETs (carcinoids), 43% with foregut tumors, and 0% with hindgut tumors. In 90–100% of patients with the carcinoid syndrome, there is evidence of serotonin overproduction. Serotonin is thought to be predominantly responsible for the diarrhea. Patients with the carcinoid syndrome have increased colonic motility with a shortened transit time and possibly a secretory/absorptive alteration that is compatible with the known actions of serotonin in the gut mediated primarily through 5-HT₃ and, to a lesser degree, 5-HT₄ receptors. Serotonin receptor antagonists (especially 5-HT₃ antagonists) relieve the diarrhea in many, but not all, patients. A tryptophan 5-hydroxylase inhibitor, telotristat (LX-10310), which inhibits serotonin synthesis in peripheral tissues, caused a decrease in bowel movement frequency in 40–50% of patients with the carcinoid syndrome. Additional studies suggest that tachykinins may be important mediators of diarrhea in some patients. In one study, plasma tachykinin levels correlated with symptoms of diarrhea. Serotonin does not appear to be involved in the flushing in most patients because serotonin receptor antagonists do not relieve flushing. In patients with gastric carcinoids, the characteristic red, patchy pruritic flush is thought due to histamine release because H₁ and H₂ receptor antagonists can prevent it. Numerous studies have shown that tachykinins (substance P, neuropeptide K) are stored in GI-NETs (carcinoids) and released during flushing. However, some studies have demonstrated that octreotide can relieve the flushing induced by pentagastrin in these patients without altering the stimulated increase in plasma substance P, suggesting that other mediators must be involved in the flushing. A correlation between plasma tachykinin levels (but not substance P levels) and flushing has been reported. Prostaglandin release could be involved in mediating either the diarrhea or flush, but conflicting data exist. Both histamine and serotonin may be responsible for the wheezing as well as the fibrotic reactions involving the heart, causing Peyronie's disease and intraabdominal fibrosis.

The exact mechanism of the heart disease remains unclear, although increasing evidence supports a central role for serotonin. Patients with heart disease have higher plasma levels of neurokinin A, substance P, plasma atrial natriuretic peptide (ANP), pro-brain natriuretic peptide, chromogranin A, and activin A as well as higher urinary 5-HIAA excretion.

The valvular heart disease caused by the appetite-suppressant drugs dexfenfluramine and benfluorex is histologically indistinguishable

from that observed in carcinoid disease. Furthermore, ergot-containing dopamine receptor agonists used for Parkinson's disease (pergolide, cabergoline) cause valvular heart disease that closely resembles that seen in the carcinoid syndrome. Furthermore, in animal studies, the formation of valvular plaques/fibrosis occurs after prolonged treatment with serotonin as well as in animals with a deficiency of the 5-HIAA transporter gene, which results in an inability to inactivate serotonin. Metabolites of fenfluramine, as well as the dopamine receptor agonists, have high affinity for serotonin receptor subtype 5-HT_{2B} receptors, whose activation is known to cause fibroblast mitogenesis. Serotonin receptor subtypes 5-HT_{1B,1D,2A,2B} normally are expressed in human heart valve interstitial cells. High levels of 5-HT_{2B} receptors are known to occur in heart valves and occur in cardiac fibroblasts and cardiomyocytes. Studies of cultured interstitial cells from human cardiac valves have demonstrated that these valvulopathic drugs induce mitogenesis by activating 5-HT_{2B} receptors and stimulating upregulation of transforming growth factor β and collagen biosynthesis. These observations support the conclusion that serotonin overproduction by GI-NETs (carcinoids) is important in mediating the valvular changes, possibly by activating 5-HT_{2B} receptors in the endocardium. Both the magnitude of serotonin overproduction and prior chemotherapy are important predictors of progression of the heart disease, whereas patients with high plasma levels of ANP have a worse prognosis. Plasma connective tissue growth factor levels are elevated in many fibrotic conditions; elevated levels occur in patients with carcinoid heart disease and correlate with the presence of right ventricular dysfunction and the extent of valvular regurgitation in patients with GI-NETs (carcinoids).

Patients may develop either a typical or, rarely, an atypical carcinoid syndrome (Fig. 80-1). In patients with the typical form, which characteristically is caused by midgut NETs (carcinoids), the conversion of tryptophan to 5-HTP by tryptophan hydroxylase is the rate-limiting step (Fig. 80-1). Once 5-HTP is formed, it is rapidly converted to 5-HT and stored in secretory granules of the tumor or in platelets. A small amount remains in plasma and is converted to 5-HIAA, which appears in large amounts in the urine. These patients have an expanded serotonin pool size, increased blood and platelet serotonin, and increased urinary 5-HIAA. Some GI-NETs (carcinoids) cause an atypical carcinoid syndrome that is thought to be due to a deficiency in the enzyme dopa decarboxylase; thus, 5-HTP cannot be converted to 5-HT (serotonin), and 5-HTP is secreted into the bloodstream (Fig. 80-1). In these patients, plasma serotonin levels are normal but urinary levels may be increased because some 5-HTP is converted to 5-HT in the kidney. Characteristically, urinary 5-HTP and 5-HT are increased, but urinary 5-HIAA levels are only slightly elevated. Foregut carcinoids are the most likely to cause an atypical carcinoid syndrome; however, they also can cause a typical carcinoid syndrome.

One of the most immediate life-threatening complications of the carcinoid syndrome is the development of a carcinoid crisis. This is more common in patients who have intense symptoms or have greatly increased urinary 5-HIAA levels (i.e., >200 mg/d). The crisis may occur spontaneously; however, it is usually provoked by procedures such as anesthesia, chemotherapy, surgery, biopsy, endoscopy, or radiologic examinations such as during biopsies, hepatic artery embolization, and vessel catheterization. It can be provoked by stress or procedures as mild as repeated palpation of the tumor during physical examination. Patients develop intense flushing, diarrhea, abdominal pain, cardiac abnormalities including tachycardia, hypertension, or hypotension, and confusion or stupor. If not adequately treated, this can be a terminal event.

■ DIAGNOSIS OF THE CARCINOID SYNDROME AND GI-NETs (CARCINOIDS)

The diagnosis of carcinoid syndrome relies on measurement of urinary or plasma serotonin or its metabolites in the plasma or urine. The measurement of urinary 5-HIAA is used most frequently. False-positive elevations may occur if the patient is eating serotonin-rich foods such as bananas, pineapples, walnuts, pecans, avocados, or hickory nuts or is taking certain medications (cough syrup containing guaifenesin, acetaminophen, salicylates, serotonin reuptake inhibitors, or L-dopa).

The normal range for daily urinary 5-HIAA excretion is 2–8 mg/d. Serotonin overproduction was noted in 92% of patients with carcinoid syndrome in one study, and in another study, 5-HIAA had 73% sensitivity and 100% specificity for carcinoid syndrome. Serotonin overproduction is *not* synonymous with the presence of clinical carcinoid syndrome because 12–26% of patients with serotonin overproduction do not have clinical evidence of the carcinoid syndrome.

Most physicians use only the urinary 24-h 5-HIAA excretion rate; even though a recent study shows an overnight urinary collection is just as accurate. Assessment of plasma and platelet serotonin levels and plasma 5-HIAA, if available, may provide additional information and/or substitute for the 24 h urinary 5-HIAA study. Platelet serotonin levels are more sensitive than urinary 5-HIAA but are not generally available. A single plasma 5-HIAA determination was found to have similar sensitivity/specification to that with the 24-h urinary 5-HIAA assessment, suggesting this could replace the standard urinary collection because of its greater convenience and avoidance of incomplete or improper collections. It, however, could be affected by renal disease. Because patients with foregut NETs (carcinoids) may produce an atypical carcinoid syndrome, if this syndrome is suspected and the urinary 5-HIAA is minimally elevated or normal, other urinary metabolites of tryptophan, such as 5-HTP and 5-HT, should be measured (Fig. 80-1).

Flushing occurs in a number of other diseases, including systemic mastocytosis, chronic myeloid leukemia with increased histamine release, menopause, reactions to alcohol or glutamate, and side effects of chlorpropamide, calcium channel blockers, and nicotinic acid. None of these conditions causes increased urinary 5-HIAA.

The diagnosis of carcinoid tumor can be suggested by the carcinoid syndrome, recurrent abdominal symptoms in a healthy-appearing individual, or the discovery of hepatomegaly or hepatic metastases associated with minimal symptoms. Ileal NETs (carcinoids), which make up 25% of all clinically detected carcinoids, should be suspected in patients with bowel obstruction, abdominal pain, flushing, or diarrhea.

Serum chromogranin A levels are elevated in 56–100% of patients with GI-NETs (carcinoids), and the level correlates with tumor bulk. Serum chromogranin A levels are not specific for GI-NETs (carcinoids) because they are also elevated in patients with pNETs and other NETs. Furthermore, a major problem is caused by potent acid antisecretory drugs such as proton pump inhibitors (omeprazole and related drugs) because they almost invariably cause elevation of plasma chromogranin A levels; the elevation occurs rapidly (3–5 days) with continued use, and the elevated levels overlap with the levels seen in many patients with NETs. Plasma neuron-specific enolase levels are also used as a marker of GI-NETs (carcinoids) but are less sensitive than chromogranin A, being increased in only 17–47% of patients. Newer markers have been proposed including pancreatic polypeptide (a chromogranin A breakdown product), and activin A. The former is not affected by proton pump inhibitors; however, its sensitivity and specificity are not established. Plasma activin elevations are reported to correlate with the presence of cardiac disease with a sensitivity of 87% and specificity of 57%. Plasma levels of N-terminal pro brain natriuretic peptide moderately correlate with carcinoid heart disease severity.

TREATMENT

Carcinoid Syndrome and Nonmetastatic Gastrointestinal Neuroendocrine Tumors (Carcinoids)

CARCINOID SYNDROME

Treatment includes avoiding conditions that precipitate flushing, dietary supplementation with nicotinamide, treatment of heart failure with diuretics, treatment of wheezing with oral bronchodilators, and control of the diarrhea with antidiarrheal agents such as loperamide and diphenoxylate. If patients still have symptoms, somatostatin analogues or less frequently, serotonin receptor antagonists, are the drugs of choice (Fig. 80-2). An additional point dealt with in later sections, is the fact that most patients who develop the carcinoid syndrome have metastatic disease to the liver. Numerous

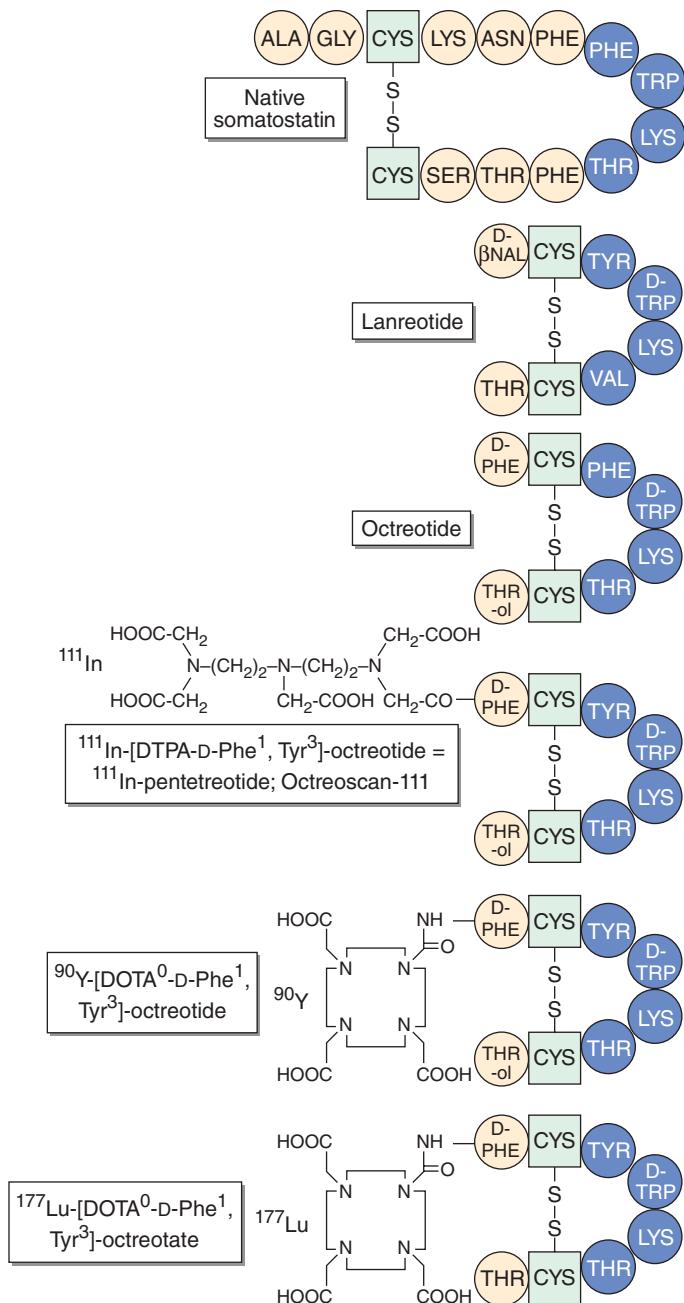


FIGURE 80-2 Structure of somatostatin and synthetic analogues used for diagnostic or therapeutic indications.

antitumor therapies (liver-directed therapies, PRRT, surgery, chemotherapy/targeted drug therapies) also can ameliorate the severity of the carcinoid syndrome.

There are 14 subclasses of serotonin receptors, and antagonists for many are not available. The 5-HT₁ and 5-HT₂ receptor antagonists methysergide, cyproheptadine, and ketanserin have all been used to control the diarrhea but usually do not decrease flushing. The use of methysergide is limited because it can cause or enhance retroperitoneal fibrosis. Ketanserin diminishes diarrhea in 30–100% of patients. 5-HT₃ receptor antagonists (ondansetron, tropisetron, alosetron) can control diarrhea and nausea in up to 100% of patients and occasionally ameliorate the flushing. A combination of histamine H₁ and H₂ receptor antagonists (i.e., diphenhydramine and cimetidine or ranitidine) may control flushing in patients with foregut carcinoids. A phase 3 prospective, double-blind study provides evidence the peripheral tryptophan hydroxylase inhibitor, telotristat, will be useful to control the diarrhea in many of these patients.

Synthetic analogues of somatostatin (octreotide, lanreotide) are now the most widely used agents to control the symptoms of patients with

carcinoid syndrome (Figs. 80-1 and 80-2). These drugs are effective at relieving symptoms and decreasing urinary 5-HIAA levels in patients with this syndrome. Octreotide-LAR (10–30 mg i.m., monthly) and lanreotide-SR/autogel (Somatuline) (60–120 mg sc-deep, monthly) (sustained-release formulations allowing monthly injections) (Fig. 80-2), control symptoms in 74 and 68% of patients, respectively, with carcinoid syndrome and show a biochemical response in 51 and 64%, respectively. Patients with mild to moderate symptoms usually are treated initially with octreotide 50–100 µg SC every 8 h or lower doses or low doses of the long-acting formulations, and then receive higher doses as needed of the long-acting monthly depot forms (octreotide-LAR or lanreotide-autogel). Forty percent of patients escape control after a median time of 4 months, and the depot dosage may have to be increased as well as supplemented with the shorter-acting formulation, SC octreotide. Pasireotide (SOM230) is a somatostatin analogue with broader selectivity (high-affinity somatostatin receptors [sst_1 , sst_2 , sst_3 , sst_5]) than octreotide/lanreotide (sst_2 , sst_5). In a phase II study of patients with refractory carcinoid syndrome, pasireotide controlled symptoms in 27%.

Carcinoid heart disease is associated with a decreased mean survival (3.8 years), and therefore, it should be sought for and carefully assessed in all patients with carcinoid syndrome. Transthoracic echocardiography remains a key element in establishing the diagnosis of carcinoid heart disease and determining the extent and type of cardiac abnormalities. Treatment with diuretics and somatostatin analogues can reduce the negative hemodynamic effects and secondary heart failure. It remains unclear whether long-term treatment with these drugs or with the tryptophan hydroxylase inhibitor, telotristat, when it becomes available, will decrease the progression of carcinoid heart disease. Balloon valvuloplasty for stenotic valves or cardiac valve surgery may be required.

To prevent as well as treat patients with carcinoid crises, somatostatin analogues are recommended, although there is controversy of how effective they are and what dosage should be used. To prevent carcinoid crises development, treatment with somatostatin analogues is recommended prior to the possible precipitating event such as surgery, anesthesia, chemotherapy, and stress. It is generally recommended that octreotide 150–250 µg SC every 6–8 h be used 24–48 h before anesthesia and then continued throughout the procedure. Another commonly used protocol is to use 100 µg/h by continuous infusion with or without a preoperative bolus.

Currently, sustained-release preparations of both octreotide (octreotide-LAR [long-acting release], 10, 20, 30 mg) and lanreotide (lanreotide-PR [prolonged release, lanreotide-autogel], 60, 90, 120 mg) are available and widely used because their use greatly facilitates long-term treatment. Octreotide-LAR (30 mg/month) gives a plasma level ≥ 1 ng/mL for 25 days, whereas this requires 3–6 injections a day of the non-sustained-release form. Lanreotide-autogel (Somatuline) is given every 4–6 weeks.

Short-term side effects occur in up to one-half of patients. Pain at the injection site and side effects related to the GI tract (59% discomfort, 15% nausea, diarrhea) are the most common. They are usually short-lived and do not interrupt treatment. Important long-term side effects include gallstone formation, steatorrhea, and deterioration in glucose tolerance. The overall incidence of gallstones/biliary sludge varies from 5 to 66%, with 7% having symptomatic disease that required surgical treatment in one study.

Interferon α is reported to be effective in controlling symptoms of the carcinoid syndrome either alone or combined with hepatic artery embolization. With interferon α alone, the clinical response rate is 30–70%, and with interferon α with hepatic artery embolization, diarrhea was controlled for 1 year in 43% and flushing was controlled in 86%. Side effects develop in almost all patients, with the most frequent being a flu-like syndrome (80–100%), followed by anorexia and fatigue, even though these frequently improve with continued treatment. Other more severe side effects include bone marrow toxicity, hepatotoxicity, autoimmune disorders, and rarely CNS side effects (depression, mental disorders, visual problems).

Hepatic artery embolization alone or with chemotherapy (chemoembolization) has been used to control the symptoms of carcinoid syndrome. Embolization alone is reported to control symptoms in up to 76% of patients, and chemoembolization (5-fluorouracil, doxorubicin, cisplatin, mitomycin) controls symptoms in 60–75% of patients. Hepatic artery embolization can have major side effects, including nausea, vomiting, pain, and fever. In two studies, 5–7% of patients died from complications of hepatic artery occlusion.

Other drugs have been used successfully in small numbers of patients to control the symptoms of carcinoid syndrome. Parachlorophenylalanine can inhibit tryptophan hydroxylase and therefore the conversion of tryptophan to 5-HTP. However, its severe side effects, including psychiatric disturbances, make it intolerable for long-term use. α -Methyldopa inhibits the conversion of 5-HTP to 5-HT, but its effects are only partial.

Peptide radioreceptor therapy (PRRT; using radiotherapy with radiolabeled somatostatin analogues), cytoreductive surgery, the use of radiolabeled microspheres, and other methods for treatment of advanced metastatic disease can facilitate control of the carcinoid syndrome (see below).

GI-NETS (CARCINOIDS) (NONMETASTATIC)

Surgery is the only potentially curative therapy. Because with most GI-NETs (carcinoids), the probability of metastatic disease increases with increasing size, in most guidelines, the therapeutic approach is determined accordingly. Furthermore, the grade of the tumor is having an increasingly important role in determining the therapeutic approach. With well-differentiated (G1/G2) GI-NETs the size of the primary NET plays an important role. With appendiceal NETs (carcinoids) < 1 cm, simple appendectomy was curative in 103 patients followed for up to 35 years. With rectal NETs (carcinoids) < 1 cm, local resection is curative. With SI NETs (carcinoids) < 1 cm, there is not complete agreement. Because 15–69% of SI NETs (carcinoids) this size have metastases in different studies, most recommend a wide resection with en bloc resection of the adjacent lymph-bearing mesentery. If the tumor is > 2 cm for rectal, appendiceal, or SI NETs (carcinoids), a full cancer operation should be done. This includes a right hemicolectomy for appendiceal NETs (carcinoids), an abdominoperineal resection or low anterior resection for rectal NETs (carcinoids), and an en bloc resection of adjacent lymph nodes for SI NETs (carcinoids). For appendiceal NETs (carcinoids) 1–2 cm in diameter, a simple appendectomy is proposed by some, whereas others favor a formal right hemicolectomy. For 1–2 cm rectal NETs (carcinoids), it is recommended that a wide, local, full-thickness excision be performed.

With well-differentiated (G1/G2) type I or II gastric NETs (carcinoids), which are usually < 1 cm, endoscopic removal is recommended. In type I or II gastric carcinoids, if the tumor is > 2 cm or if there is local invasion, some recommend total gastrectomy, whereas others recommend antrectomy in type I to reduce the hypergastrinemia, which has led to regression of the carcinoids in a number of studies. For types I and II gastric NETs (carcinoids) of 1–2 cm, there is no agreement, with some recommending endoscopic treatment followed by chronic somatostatin treatment and careful follow-up and others recommending surgical treatment. With type III gastric NETs (carcinoids) > 2 cm, excision and regional lymph node clearance are recommended. Most tumors < 1 cm are treated endoscopically. Type 1 and 2 gastric carcinoids tend to recur after endoscopic treatments so patients need to continue to be followed. Treatment of type 1 or 2 gastric carcinoids using a CCK_B (gastrin) receptor antagonist, netazepide (not yet FDA approved) decreased the size and number of gastric carcinoids. However, netazepide needed to be continued or they would return. Poorly differentiated G3 carcinoids of the GI tract are treated like G3 tumors in other locations, which involve primarily chemotherapy and will be discussed in a later section on treatment of advanced/aggressive disease.

Resection of isolated or limited hepatic metastases may be beneficial and will be discussed in a later section on treatment of advanced disease.

PANCREATIC NEUROENDOCRINE TUMORS (pNETs)

Functional pNETs (F-pNETs) usually present clinically with symptoms due to the hormone-excess state (Table 80-2). Only late in the course of the disease does the tumor per se cause prominent symptoms such as abdominal pain. In contrast, all the symptoms due to nonfunctional pNETs (NF-pNET) are due to the tumor per se. The overall result of this is that some F-pNETs may present with severe symptoms with a small or undetectable primary tumor, whereas NF-pNETs usually present late in the disease course with large tumors, which are frequently metastatic. The mean delay between onset of continuous symptoms and diagnosis of a F-pNET syndrome is 4–7 years. Therefore, the diagnoses frequently are missed for extended periods.

TREATMENT

Pancreatic Neuroendocrine Tumor (General Points)

Treatment of pNETs requires two different strategies. First, treatment must be directed at the hormone-excess state such as the gastric acid hypersecretion in gastrinomas or the hypoglycemia in insulinomas. Ectopic hormone secretion usually causes the presenting symptoms and can cause life-threatening complications. Second, with all the tumors except insulinomas, >50% are malignant (Table 80-2); therefore, treatment must also be directed against the tumor per se. Because in many patients these tumors are not surgically curable due to the presence of advanced disease at diagnosis, surgical resection for cure, which addresses both treatment aspects, is often not possible.

GASTRINOMA (ZOLLINGER-ELLISON SYNDROME)

A gastrinoma is an NET that secretes gastrin; the resultant hypergastrinemia causes gastric acid hypersecretion (Zollinger-Ellison syndrome [ZES]). The chronic hypergastrinemia results in marked gastric acid hypersecretion and growth of the gastric mucosa with increased numbers of parietal cells and proliferation of gastric ECL cells. The gastric acid hypersecretion characteristically causes peptic ulcer disease (PUD), often refractory and severe, as well as diarrhea. The most common presenting symptoms are abdominal pain (70–100%), diarrhea (37–73%), gastroesophageal reflux disease (GERD) (30–35%), and 10–20% of patients have diarrhea only. Although peptic ulcers may occur in unusual locations, most patients have a typical duodenal ulcer. Important observations that should suggest this diagnosis include PUD with diarrhea; PUD in an unusual location or with multiple ulcers; PUD refractory to treatment or persistent; PUD associated with prominent gastric folds; PUD associated with findings suggestive of MEN 1 (endocrinopathy, family history of ulcer or endocrinopathy, nephrolithiasis); and PUD without *Helicobacter pylori* present. *H. pylori* is present in >90% of idiopathic peptic ulcers but is present in <50% of patients with gastrinomas. Chronic unexplained diarrhea also should suggest ZES.

Approximately 20–25% of patients with ZES have MEN 1 (MEN1/ZES), and in most cases, hyperparathyroidism is present before the ZES develops. In older studies it was generally reported that almost all MEN/ZES presented with the hyperparathyroidism, but in a number of recent series up to one-third of these patients present with the ZES, and while the hyperparathyroidism is present it may be mild and difficult to diagnose without appropriate testing. These patients are treated differently from those without MEN 1 (sporadic ZES); therefore, MEN 1 should be sought in all patients with ZES by family history and by measuring plasma ionized calcium and prolactin levels and plasma hormone levels (parathormone, growth hormone).

Most gastrinomas (50–90%) in sporadic ZES are present in the duodenum, followed by the pancreas (10–40%) and other intraabdominal sites (mesentery, lymph nodes, biliary tract, liver, stomach, ovary). Rarely, the tumor may involve extraabdominal sites (heart, lung cancer). In MEN 1/ZES the gastrinomas are also usually in the duodenum (80–100%), followed by the pancreas (0–20%), and are almost always multiple. About 60–90% of gastrinomas are malignant (Table 80-2) with

metastatic spread to lymph nodes and liver. Distant metastases to bone occur in 12–30% of patients with liver metastases.

Diagnosis

The diagnosis of ZES requires the demonstration of inappropriate fasting hypergastrinemia, usually by demonstrating hypergastrinemia occurring with an increased basal gastric acid output (BAO) (hyperchlorhydria). More than 98% of patients with ZES have fasting hypergastrinemia, although in 40–60% the level may be elevated less than tenfold. Therefore, when the diagnosis is suspected, a fasting gastrin is usually the initial test performed. It is important to remember that potent gastric acid suppressant drugs such as proton pump inhibitors (PPIs) (omeprazole, esomeprazole, pantoprazole, lansoprazole, rabeprazole) can suppress acid secretion sufficiently to cause hypergastrinemia; because of their prolonged duration of action, these drugs have to be tapered or frequently discontinued for a week before the gastrin determination. Withdrawal of PPIs should be performed carefully because PUD complications can rapidly develop in some patients and is best done in consultation with GI units with experience in this area. The widespread use of PPIs can confound the diagnosis of ZES. First, by raising a false-positive diagnosis by causing hypergastrinemia in a patient being treated with idiopathic PUD (without ZES). Second, by leading to a false-negative diagnosis because at routine doses used to treat patients with idiopathic PUD, PPIs control symptoms in most ZES patients and thus mask the diagnosis. If ZES is suspected and the gastrin level is elevated, it is important to show that it is increased when gastric pH is ≤2.0 because physiologically hypergastrinemia secondary to achlorhydria (atrophic gastritis, pernicious anemia) is one of the most common causes of hypergastrinemia. Nearly all ZES patients have a fasting pH ≤2 when off antisecretory drugs. If the fasting gastrin is >1000 pg/mL (increased tenfold) and the pH is ≤2.0, which occurs in 40–60% of patients with ZES, the diagnosis of ZES is established after the possibility of retained antrum syndrome has been ruled out by history. In patients with hypergastrinemia with fasting gastrins <1000 pg/mL (<tenfold increased) and gastric pH ≤2.0, other conditions, such as *H. pylori* infections, antral G-cell hyperplasia/hyperfunction, gastric outlet obstruction, and, rarely, renal failure, can masquerade as ZES. To establish the diagnosis in this group, a determination of BAO and a secretin provocative test should be done. In patients with ZES without previous gastric acid-reducing surgery, the BAO is usually (>90%) elevated (i.e., >15 mEq/h). The secretin provocative test is usually positive, with the criterion of a >120-pg/mL increase over the basal level having the highest sensitivity (94%) and specificity (100%). Unfortunately the diagnosis of ZES is becoming increasing more difficult. This is due not only to the widespread use of PPIs (leading to false-positive results as well as masking ZES presentation), but also recent studies demonstrate that many of the commercial gastrin kits that are used by most laboratories to measure fasting serum gastrin levels are not reliable. In one study, 7 of the 12 tested commercial gastrin kits inaccurately assessed the true serum concentration of gastrin primarily because the antibodies used had inappropriate specificity for the different circulating forms of gastrin and were not adequately validated. Both underestimation and overestimation of fasting serum gastrin levels occurred using these commercial kits. To circumvent this problem, it is either necessary to use one of the five reliable kits identified or, alternatively, to refer the patient to a center with expertise in making the diagnosis in your area, or if this is not possible, to contact such a center and use the gastrin assay they recommend. An accurate gastrin assay is essential for accurate measurement of fasting serum gastrin level as well as for assessing gastrin levels during the secretin provocative test, and thus, the diagnosis of ZES cannot reliably be made without one.

TREATMENT

Zollinger-Ellison Syndrome

Gastric acid hypersecretion in patients with ZES can be controlled in almost every case by oral gastric antisecretory drugs. Because of their long duration of action and potency, which allows dosing once or twice a day, the PPIs (H^+ , K^+ -ATPase inhibitors) are the drugs of

choice. Histamine H₂-receptor antagonists are also effective, although more frequent dosing (q 4–8 h) and high doses are required. In patients with MEN1/ZES with hyperparathyroidism, correction of the hyperparathyroidism increases the sensitivity to gastric antisecretory drugs and decreases the basal acid output. Long-term treatment with PPIs (>15 years) has proved to be safe and effective, without development of tachyphylaxis. Although patients with ZES, especially those with MEN 1/ZES, more frequently develop gastric NETs (carcinoids), no data suggest that the long-term use of PPIs increases this risk in these patients. With long-term PPI use in ZES patients, vitamin B₁₂ deficiency can develop; thus, vitamin B₁₂ levels should be assessed during follow-up. Long-term PPI use may be associated with a number of side-effects including; an increased incidence of bone fractures; *Clostridium difficile* infections; dementia; hypomagnesemia; renal disease; and numerous drug interactions; however, at present, there is no report these are increased in ZES patients.

With the increased ability to control acid hypersecretion, >50% of patients who are not cured (>60% of patients) will die from tumor-related causes. At presentation, careful imaging studies are essential to localize the extent of the tumor to determine the appropriate treatment. A third of patients present with hepatic metastases, and in <15% of those patients, the disease is limited, so that surgical resection may be possible. Surgical short-term cure is possible in 60% of all patients without MEN 1/ZES or liver metastases (40% of all patients) and in 30% of patients long term. In patients with MEN 1/ZES, long-term surgical cure is rare without aggressive resection (i.e., Whipple resections), because the tumors are small, multiple, and frequently with lymph node metastases. At present the role of routine surgery for removal of the gastrinoma in MEN1/ZES patients is controversial for the above reason, with most guidelines recommending attempted gastrinoma resection only in MEN1/ZES patients with pNETs ≥1.5–2 cm in diameter. Surgical studies demonstrate that successful resection of the gastrinoma not only decreases the chances of developing liver metastases but also increases the disease-related survival rate. Therefore, all patients with gastrinomas without MEN 1/ZES or a medical condition that limits life expectancy should undergo surgery by a surgeon experienced in the treatment of these disorders.

INSULINOMAS

An insulinoma is an NET of the pancreas that is thought to be derived from beta cells that ectopically secrete insulin, which results in hypoglycemia. The average age of occurrence is 40–50 years old. The most common clinical symptoms are due to the effect of the hypoglycemia on the CNS (neuroglycemic symptoms) and include confusion, headache, disorientation, visual difficulties, irrational behavior, and even coma. Also, most patients have symptoms due to excess catecholamine release secondary to the hypoglycemia, including sweating, tremor, and palpitations. Characteristically, these attacks are associated with fasting.

Insulinomas are generally small (>90% are <2 cm) and usually not multiple (90%); only 5–15% are malignant, and they almost invariably occur only in the pancreas, distributed equally in the pancreatic head, body, and tail. They are associated with the MEN1 syndrome in 4%.

Insulinomas should be suspected in all patients with hypoglycemia, especially when there is a history suggesting that attacks are provoked by fasting, or with a family history of MEN 1. Insulin is synthesized as proinsulin, which consists of a 21-amino-acid α chain and a 30-amino-acid β chain connected by a 33-amino-acid connecting peptide (C peptide). In insulinomas, in addition to elevated plasma insulin levels, elevated plasma proinsulin levels are found, and C-peptide levels are elevated.

Diagnosis The diagnosis of insulinoma requires the demonstration of an elevated plasma insulin level at the time of hypoglycemia. A number of other conditions may cause fasting hypoglycemia, such as the inadvertent or surreptitious use of insulin or oral hypoglycemic agents, severe liver disease, alcoholism, poor nutrition, and other extrapancreatic tumors. Furthermore, postprandial hypoglycemia can be caused by a number of conditions that confuse the diagnosis of insulinoma. Particularly important here is the increased occurrence of hypoglycemia

after gastric bypass surgery for obesity, which is now widely performed. A new entity, insulinomatosis, was described that can cause hypoglycemia and mimic insulinomas. It occurs in 10% of patients with persistent hyperinsulinemic hypoglycemia and is characterized by the occurrence of multiple macro-/microadenomas expressing insulin, and it is not clear how to distinguish this entity from insulinoma preoperatively. The most reliable test to diagnose insulinoma is a fast up to 72 h with serum glucose, C-peptide, proinsulin, and insulin measurements every 4–8 h. If at any point the patient becomes symptomatic or glucose levels are persistently <2.2 mmol/L (40 mg/dL), the test should be terminated, and repeat samples for the above studies should be obtained before glucose is given. Some 70–80% of patients will develop hypoglycemia during the first 24 h, and 98% by 48 h. In nonobese normal subjects, serum insulin levels should decrease to <43 pmol/L (<6 μU/mL) when blood glucose decreases to <2.2 mmol/L (<40 mg/dL) and the ratio of insulin to glucose is <0.3 (in mg/dL). In addition to having an insulin level >6 μU/mL when blood glucose is <40 mg/dL, some investigators also require an elevated C-peptide and serum proinsulin level, an insulin/glucose ratio >0.3, and a decreased plasma β-hydroxybutyrate level for the diagnosis of insulinomas. A commonly used set of criteria to make the diagnosis include: low blood glucose levels (<2.2 mmol/L (<40 mg/dL); concomitant insulin levels ≥6 U/L (≥36 pmol/L; ≥3 U/L by ICMA); C-peptide levels ≥200 pmol/L; proinsulin levels ≥5 pmol/L; β-hydroxybutyrate levels ≤2.7 mmol/L; and absence of sulfonylurea metabolites) in the plasma and/or urine.

Surreptitious use of insulin or hypoglycemic agents may be difficult to distinguish from insulinomas. The combination of proinsulin levels (normal in exogenous insulin/hypoglycemic agent users), C-peptide levels (low in exogenous insulin users), antibodies to insulin (positive in exogenous insulin users), and measurement of sulfonylurea levels in serum or plasma will allow the correct diagnosis to be made. The diagnosis of insulinoma has been complicated by the introduction of specific insulin assays that do not also interact with proinsulin, as do many of the older radioimmunoassays (RIAs), and therefore give lower plasma insulin levels. The increased use of these specific insulin assays has resulted in increased numbers of patients with insulinomas having lower plasma insulin values (<6 μU/mL) than levels proposed to be characteristic of insulinomas by RIA. In these patients, the assessment of proinsulin and C-peptide levels at the time of hypoglycemia is particularly helpful for establishing the correct diagnosis. An elevated proinsulin level when the fasting glucose level is <45 mg/dL is sensitive and specific.

TREATMENT

Insulinomas

Only 5–15% of insulinomas are malignant; therefore, after appropriate imaging (see below), surgery should be performed. In different studies, 75–100% of patients are cured by surgery. Before surgery, the hypoglycemia can be controlled by frequent small meals and the use of diazoxide (150–800 mg/d). Diazoxide is a benzothiadiazide whose hypoglycemic effect is attributed to inhibition of insulin release. Its side effects are sodium retention and GI symptoms such as nausea. Approximately 50–60% of patients respond to diazoxide. Other agents effective in some patients to control the hypoglycemia include verapamil and diphenylhydantoin. Long-acting somatostatin analogues such as octreotide and lanreotide are acutely effective in 40% of patients. However, octreotide must be used with care because it inhibits growth hormone secretion and can alter plasma glucagon levels; therefore, in some patients, it can worsen the hypoglycemia.

For the 5–15% of patients with malignant insulinomas, these drugs or somatostatin analogues are used initially. In a small number of patients with malignant tumors, mammalian target of rapamycin (mTOR) inhibitors (everolimus, rapamycin) are reported to control the hypoglycemia. If they are not effective, various anti-tumor treatments such as hepatic arterial embolization, chemoembolization, chemotherapy, and peptide receptor radiotherapy with radiolabeled somatostatin analogues (PRRT) have been used and can be effective, particularly PRRT.

Insulinomas, which are usually benign (>90%) and intrapancreatic in location, are increasingly resected using a laparoscopic approach, which has lower morbidity rates. This approach requires that the insulinoma be localized on preoperative imaging studies.

GLUCAGONOMAS

A glucagonoma is NET of the pancreas that secretes excessive amounts of glucagon, which causes a distinct syndrome characterized by dermatitis, glucose intolerance or diabetes, and weight loss. Glucagonomas principally occur between 45 and 70 years of age. The tumor is clinically heralded by a characteristic dermatitis (migratory necrolytic erythema) (67–90%), accompanied by glucose intolerance (40–90%), weight loss (66–96%), anemia (33–85%), diarrhea (15–29%), and thromboembolism (11–24%). The characteristic rash usually starts as an annular erythema at intertriginous and periorificial sites, especially in the groin or buttock. It subsequently becomes raised, and bullae form; when the bullae rupture, eroded areas form. The lesions can wax and wane. The development of a similar rash in patients receiving glucagon therapy suggests that the rash is a direct effect of the hyperglucagonemia. A characteristic laboratory finding is hypoaminoacidemia, which occurs in 26–100% of patients.

Glucagonomas are generally large tumors at diagnosis (5–10 cm). Some 50–80% occur in the pancreatic tail. From 50 to 82% have evidence of metastatic spread at presentation, usually to the liver. Glucagonomas are rarely extrapancreatic, usually occur singly, and <3% are associated with the MEN1 syndrome.

Two new entities have been described that can also cause hyperglucagonemia and may mimic glucagonomas. Mahvash disease is due to an inactivating mutation (homozygous P86S mutation) of the human glucagon receptor. It is associated with the development of α -cell hyperplasia, hyperglucagonemia, and the development of nonfunctioning pNETs. Subsequently other patients with other inactivating mutations of the human glucagon receptor have been described with similar findings, leading to the suggestion that a hepato-pancreatic feedback regulation of the cells, possibly involving amino acids, may exist in humans. A second disease called *glucagon cell adenomatosis* can mimic glucagonoma syndrome clinically and is characterized by the presence of hyperplastic islets staining positive for glucagon instead of a single glucagonoma.

Diagnosis The diagnosis is confirmed by demonstrating an increased plasma glucagon level. Characteristically, plasma glucagon levels exceed 1000 pg/mL (normal is <150 pg/mL) in 90%; 7% are between 500 and 1000 pg/mL, and 3% are <500 pg/mL. A trend toward lower levels at diagnosis has been noted in the last decade. A plasma glucagon level >1000 pg/mL is considered diagnostic of glucagonoma. Other diseases causing increased plasma glucagon levels include cirrhosis, diabetic ketoacidosis, celiac disease, renal insufficiency, acute pancreatitis, hypercorticism, hepatic insufficiency, severe stress, and prolonged fasting or familial hyperglucagonemia, as well as danazol treatment. With the exception of cirrhosis, these disorders do not increase plasma glucagon >500 pg/mL.

Necrotic migratory erythema is not pathognomonic for glucagonoma and occurs in myeloproliferative disorders, hepatitis B infection, malnutrition, short-bowel syndrome, inflammatory bowel disease, zinc deficiency, and malabsorption disorders.

TREATMENT

Glucagonomas

In 50–80% of patients, hepatic metastases are present, and so curative surgical resection is not possible. Surgical debulking in patients with advanced disease or other antitumor treatments may be beneficial as well as PRRT with radiolabeled somatostatin analogues (see below). Long-acting somatostatin analogues such as octreotide and lanreotide improve the skin rash in 75% of patients and may improve the weight loss, pain, and diarrhea, but usually do not improve the glucose intolerance.

SOMATOSTATINOMA SYNDROME

The somatostatinoma syndrome is due to an NET that secretes excessive amounts of somatostatin, which causes a distinct syndrome characterized by diabetes mellitus, gallbladder disease, diarrhea, and steatorrhea. There is no general distinction in the literature between a tumor that contains somatostatin-like immunoreactivity (somatostatinoma) and does (11–45%) or does not (55–90%) produce a clinical syndrome (somatostatinoma syndrome) by secreting somatostatin. In a review of 173 cases of somatostatinomas, only 11% were associated with the somatostatinoma syndrome. The mean age is 51 years. Somatostatinomas occur primarily in the pancreas and small intestine, and the frequency of the symptoms and occurrence of the somatostatinoma syndrome differ in each. Each of the usual symptoms is more common in pancreatic than in intestinal somatostatinomas: diabetes mellitus (95% vs 21%), gallbladder disease (94% vs 43%), diarrhea (92% vs 38%), steatorrhea (83% vs 12%), hypochlorhydria (86% vs 12%), and weight loss (90% vs 69%). The somatostatinoma syndrome occurs in 30–90% of pancreatic and 0–5% of SI somatostatinomas. In various series, 43% of all duodenal NETs contain somatostatin; however, the somatostatinoma syndrome is rarely present (<2%). Somatostatinomas occur in the pancreas in 56–74% of cases, with the primary location being the pancreatic head. The tumors are usually solitary (90%) and large (mean size 4.5 cm). Liver metastases are common, being present in 69–84% of patients. Somatostatinomas are rare in patients with MEN 1, occurring in only 0.65%.

The existence of a somatostatinoma syndrome (SSoma syndrome) has been called into question. This occurred because in one review of 821 patients with duodenal or pancreatic NETs, a proportion of which showed predominant somatostatin expression, none had the SSoma syndrome leading to the conclusion it is either very rare or nonexistent. However, in other studies, a proportion of the patients with somatostatin positive pancreatic or duodenal NETs had number of the proposed features of the SSoma syndrome.

Somatostatin is a tetradecapeptide that is widely distributed in the CNS and GI tract, where it functions as a neurotransmitter or has paracrine and autocrine actions. It is a potent inhibitor of many processes, including release of almost all hormones, acid secretion, intestinal and pancreatic secretion, and intestinal absorption. Most of the clinical manifestations are directly related to these inhibitory actions.

Diagnosis In most cases, somatostatinomas have been found by accident either at the time of cholecystectomy or during endoscopy. The presence of psammoma bodies in a duodenal tumor should particularly raise suspicion. Duodenal somatostatin-containing tumors are increasingly associated with von Recklinghausen's disease (NF-1) (Table 80-5). Most of these tumors (>98%) do not cause the SSoma syndrome. The diagnosis of the SSoma syndrome requires the demonstration of elevated plasma somatostatin levels.

TREATMENT

Somatostatinomas

Pancreatic tumors are frequently (70–92%) metastatic at presentation, whereas 30–69% of SI somatostatinomas have metastases. Surgery is the treatment of choice for those without widespread hepatic metastases. Symptoms in patients with the SSoma syndrome are also improved by octreotide treatment.

VIPomas

VIPomas are NETs that secrete excessive amounts of vasoactive intestinal peptide (VIP), which causes a distinct syndrome characterized by large-volume diarrhea, hypokalemia, and dehydration. This syndrome also is called Verner-Morrison syndrome, pancreatic cholera, and WDHA syndrome for watery diarrhea, hypokalemia, and achlorhydria, which some patients develop. The mean age of patients with this syndrome is 49 years; however, it can occur in children, and when it does, it is usually caused by a ganglioneuroma or ganglioneuroblastoma.

The principal symptoms are large-volume diarrhea (100%) severe enough to cause hypokalemia (80–100%), dehydration (83%),

hypochlorhydria (54–76%), and flushing (20%). The diarrhea is secretory in nature, persisting during fasting, and is almost always >1 L/d and in 70% is >3 L/d. In a number of studies, the diarrhea was intermittent initially in up to half the patients. Most patients do not have accompanying steatorrhea (16%), and the increased stool volume is due to increased excretion of sodium and potassium, which, with the anions, accounts for the osmolality of the stool. Patients frequently have hyperglycemia (25–50%) and hypercalcemia (25–50%).

VIP is a 28-amino-acid peptide that is an important neurotransmitter, ubiquitously present in the CNS and GI tract. Its known actions include stimulation of SI chloride secretion as well as effects on smooth-muscle contractility, inhibition of acid secretion, and vasodilatory effects, which explain most features of the clinical syndrome.

In adults, 80–90% of VIPomas are pancreatic in location, with the rest due to VIP-secreting pheochromocytomas, intestinal carcinoids, and rarely ganglioneuromas. These tumors are usually solitary, 50–75% are in the pancreatic tail, and 37–68% have hepatic metastases at diagnosis. In children <10 years old, the syndrome is usually due to ganglioneuromas or ganglioblastomas and is less often malignant (10%).

Diagnosis The diagnosis requires the demonstration of an elevated plasma VIP level and the presence of large-volume diarrhea. A stool volume <700 mL/d is proposed to exclude the diagnosis of VIPoma. When the patient fasts, a number of diseases can be excluded that can cause marked diarrhea because the high volume of diarrhea is not sustained during the fast. Other diseases that can produce a secretory large-volume diarrhea include gastrinomas, chronic laxative abuse, carcinoid syndrome, systemic mastocytosis, rarely medullary thyroid cancer, diabetic diarrhea, sprue, and AIDS. Among these conditions, only VIPomas caused a marked increase in plasma VIP. Chronic surreptitious use of laxatives/diuretics can be particularly difficult to detect clinically. Hence, in a patient with unexplained chronic diarrhea, screens for laxatives should be performed; they will detect many, but not all, laxative abusers. Elevated plasma levels of VIP should not be the only basis of the diagnosis of VIPomas because they can occur with some diarrheal states including inflammatory bowel disease, post small bowel resection, and radiation enteritis. Furthermore, nesidioblastosis can mimic VIPomas by causing elevated plasma VIP levels, diarrhea, and even false-positive location in the pancreatic region on somatostatin receptor scintigraphy (SRS).

TREATMENT

VIPomas

The most important initial treatment in these patients is to correct their dehydration, hypokalemia, and electrolyte losses with fluid and electrolyte replacement. These patients may require >5 L/d of fluid and >350 mEq/d of potassium. Because 37–68% of adults with VIPomas have metastatic disease in the liver at presentation, a significant number of patients cannot be cured surgically. In these patients, long-acting somatostatin analogues such as octreotide and lanreotide (Fig. 80-2) are the drugs of choice.

Octreotide/lanreotide will control the diarrhea short- and long-term in 75–100% of patients. In nonresponsive patients, the combination of glucocorticoids and octreotide/lanreotide has proved helpful in a small number of patients. Other drugs reported to be helpful in small numbers of patients include prednisone (60–100 mg/d), clonidine, indomethacin, phenothiazines, loperamide, lidamidine, lithium, propranolol, and metoclopramide. Treatment of advanced disease with cytoreductive surgery, embolization, chemoembolization, chemotherapy, radiotherapy, radiofrequency ablation (RFA), and peptide receptor radiotherapy may be helpful (see below). Control of the diarrhea in VIPoma patients using the tyrosine kinase inhibitor, sunitinib, has been described in case reports.

NONFUNCTIONAL PANCREATIC NEUROENDOCRINE TUMORS (NF-pNETs)

NF-pNETs are NETs that originate in the pancreas and either secretes no products or their products do not cause a specific clinical syndrome.

Their symptoms are due entirely to the tumor per se. NF-pNETs secrete chromogranin A (90–100%), chromogranin B (90–100%), α -HCG (human chorionic gonadotropin) (40%), neuron-specific enolase (31%), and β -HCG (20%), and because 40–90% secrete PP, they are also often called PPomas. A proportion also secrete ghrelin, neurotensin, calcitonin and other GI hormones/neurotransmitters, which are generally accepted as not causing a distinct clinical syndrome. Because the symptoms are due to the tumor mass, patients with NF-pNETs usually present late in the disease course with invasive tumors and hepatic metastases (64–92%), and the tumors are usually large (72% >5 cm). An increasing proportion of NF-pNETs are asymptomatic (up to 30–50%) and are found at screening for various nonspecific symptoms. NF-pNETs are usually solitary except in patients with MEN 1, in which case they are multiple. They occur primarily in the pancreatic head. Even though these tumors do not cause a functional syndrome, immunocytochemical studies show that they synthesize numerous peptides and cannot be distinguished from functional pNETs by immunocytochemistry. In MEN 1, 80–100% of patients have microscopic NF-pNETs, but they become large or symptomatic in a minority (0–13%) of cases. In VHL, 12–17% develop NF-pNETs, and in 4%, they are ≥ 3 cm in diameter.

The most common symptoms are abdominal pain (30–80%), jaundice (20–35%), and weight loss, fatigue, or bleeding. The average time from the beginning of symptoms to diagnosis is 5 years.

Diagnosis The diagnosis is established by histologic confirmation in a patient without either the clinical symptoms or the elevated plasma hormone levels of one of the established syndromes. The principal difficulty in diagnosis is to distinguish an NF-pNET from a nonendocrine pancreatic tumor, which is more common, as well as from a F-pNET. Even though chromogranin A levels are elevated in almost every patient, this is not specific for this disease as it can be found in F-pNETs, GI-NETs (carcinoids), and other neuroendocrine disorders, as well in patients without any of these, but being treated with PPIs. Plasma PP elevations should strongly suggest the diagnosis in a patient with a pancreatic mass because it is usually normal in patients with pancreatic adenocarcinomas. Elevated plasma PP is not diagnostic of this tumor because it is elevated in a number of other conditions, such as chronic renal failure, old age, inflammatory conditions, alcohol abuse, pancreatitis, hypoglycemia, postprandially, and diabetes. A positive somatostatin receptor scan in a patient with a pancreatic mass should suggest the presence of pNET/NF-pNET rather than a nonendocrine tumor.

TREATMENT

Nonfunctional Pancreatic Neuroendocrine Tumors (NF-pNETs)

Overall survival in patients with sporadic NF-pNET is 30–63% at 5 years, with a median survival of 6 years. Unfortunately, surgical curative resection can be considered only in a minority of these patients because 30–92% present with diffuse metastatic disease. Treatment needs to be directed against the tumor per se using the various modalities discussed below for advanced disease. Whereas the treatment of NF-pNETs in either MEN 1 patients or patients with VHL has remained controversial for a number of years, the treatment in sporadic cases has also become controversial. In these inherited disorders, most recommend surgical resection for any tumor >2 –3 cm in diameter; however, there is controversy in patients with smaller NF-pNETs (≤ 1.5 –2 cm), with most guidelines recommending careful surveillance of these patients. This approach is taken because patients with these inherited diseases are not curable without aggressive surgery with its associated mortality/morbidity, because of the multiplicity of the small NF-pNETs; studies show these patients with NF-pNETs ≤ 2 cm have no increased mortality; and most are slow growing. Most of these are low- or intermediate-grade lesions, and <7% are malignant. Similarly in patients with sporadic NF-pNETs in the past almost all were operated on; however, because

of the generally benign course of those that are asymptomatic and ≤ 2 cm in diameter, increasingly they are not operated, but followed closely. No consensus exists on this point with the result that some advocate a non-operative approach with careful, regular follow-up, whereas others recommend an operative or laparoscopic.

GRFomas

GRFomas are NETs that secrete excessive amounts of growth hormone-releasing factor (GRF) that cause acromegaly. GRF is a 44-amino-acid peptide, and 25–44% of pNETs have GRF immunoreactivity, although it is uncommonly secreted. GRFomas are lung tumors in 47–54% of cases, pNETs in 29–30%, SI carcinoids in 8–10%; and up to 12% occur at other sites. Patients have a mean age of 38 years, and the symptoms usually are due to either acromegaly or the tumor per se. The acromegaly caused by GRFomas is indistinguishable from classic acromegaly. The pancreatic tumors are usually large (>6 cm), and liver metastases are present in 39%. They should be suspected in any patient with acromegaly and an abdominal tumor, a patient with MEN 1 with acromegaly, or a patient without a pituitary adenoma with acromegaly or associated with hyperprolactinemia, which occurs in 70% of GRFomas. GRFomas are an uncommon cause of acromegaly. GRFomas occur in <1% of MEN 1 patients. The diagnosis is established by performing plasma assays for GRF and growth hormone. Most GRFomas have a plasma GRF level >300 pg/mL (normal <5 pg/mL men, <10 pg/mL women). Patients with GRFomas also have increased plasma levels of insulin-like growth factor type I (IGF-I) similar to those in classic acromegaly. Surgery is the treatment of choice if diffuse metastases are not present. Long-acting somatostatin analogues such as octreotide and lanreotide are the agents of choice, with 75–100% of patients responding.

OTHER RARE PANCREATIC NEUROENDOCRINE TUMOR SYNDROMES

Cushing's syndrome (ACTHoma) due to a pNET occurs in 4–16% of all ectopic Cushing's syndrome cases. It occurs in 5% of cases of sporadic gastrinomas, almost invariably in patients with hepatic metastases, and is an independent poor prognostic factor. Paraneoplastic hypercalcemia due to pNETs releasing parathyroid hormone-related peptide (PTHRP), a PTH-like material, or unknown factor, is rarely reported. The tumors are usually large, and liver metastases are usually present. Most (88%) appear to be due to release of PTHRPs. pNETs occasionally can cause the carcinoid syndrome and this may occur without the presence of liver metastases. A number of very rare pNET syndromes involving a few cases (less than five) have been described; these include a renin-producing pNET in a patient presenting with hypertension; pNETs secreting luteinizing hormone, resulting in masculinization or decreased libido; a pNET secreting erythropoietin, resulting in polycythemia; pNETs secreting IGF-II, causing hypoglycemia; pNETs secreting enteroglucagon, causing small intestinal hypertrophy, colonic/SI stasis, and malabsorption and a pNET secreting cholecystokinin (CCKoma) which can mimic ZES clinically with patients presenting with severe peptic ulcer disease, diarrhea, weight loss, and gallstones, but with a normal fasting gastrin level (Table 80-2). A number of other possible functional pNETs have been proposed, but most authorities classify these as unclear or as a nonfunctional pNET because in each case numerous patients have been described with similar plasma hormone elevations that do not cause any symptoms. These include pNETs secreting calcitonin, neurotensin (neurotensinoma), PP (PPoma), and ghrelin (Table 80-2).

TUMOR LOCALIZATION

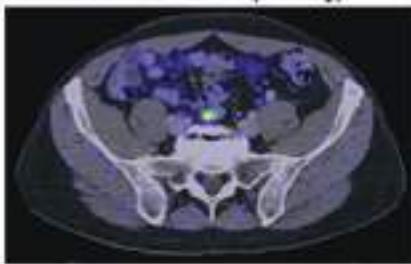
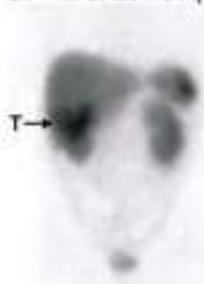
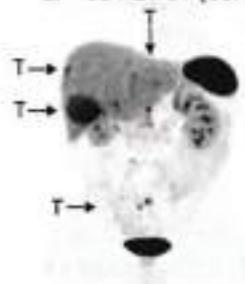
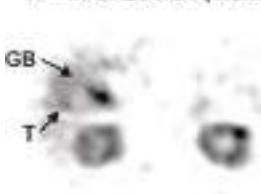
Localization of the primary tumor and knowledge of the extent of the disease are essential to the proper management of all GI-NETs (carcinoids) and pNETs. Without proper localization studies, it is not possible to determine whether the patient is a candidate for surgical resection (curative or cytoreductive) or requires antitumor treatment, to determine whether the patient is responding to antitumor therapies, whether postresection recurrent disease is present or to appropriately classify/stage the patient's disease to assess prognosis.

Numerous tumor localization methods are used in both types of NETs, including cross-sectional imaging studies (CT, magnetic

resonance imaging [MRI], trans-abdominal ultrasound), selective angiography, SRS, and positron emission tomography. In pNETs, endoscopic ultrasound (EUS) and functional localization by measuring venous hormonal gradients are also reported to be useful. Bronchial carcinoids are usually detected by standard chest radiography and assessed by CT. Rectal, duodenal, colonic, and gastric carcinoids are usually detected by GI endoscopy. Because of their wide availability, CT and MRI are generally initially used to determine the location of the primary NETs and the extent of disease. NETs are hypervascular tumors, and with both MRI and CT, contrast enhancement is essential for maximal sensitivity, and it is recommended that generally triple-phase scanning be used. The ability of cross-sectional imaging and, to a lesser extent, SRS to detect NETs is a function of NET size. With CT and MRI, <10% of tumors <1 cm in diameter are detected, 30–40% of tumors 1–3 cm are detected, and >50% of tumors >3 cm are detected. Many primary GI-NETs (carcinoids) are small, as are insulinomas and duodenal gastrinomas, and are frequently not detected by cross-sectional imaging, whereas most other pNETs present late in the course of their disease and are large (>4 cm). Selective angiography is more sensitive, localizing 60–90% of all NETs; however, it is now used infrequently. For detecting liver metastases, CT and MRI are more sensitive than ultrasound, and with recent improvements, 5–25% of patients with liver metastases will be missed by CT and/or MRI.

pNETs, as well as GI-NETs (carcinoids), frequently (>80%) overexpress high-affinity sst in both the primary tumors and the metastases. Of the five types of somatostatin receptors (sst_{1–5}), radiolabeled octreotide binds with high affinity to sst₂ and sst₅, has a lower affinity for sst₃, and has a very low affinity for sst₁ and sst₄. Between 80 and 100% of well-differentiated (G1, G2 grades) GI-NETs (carcinoids) and pNETs possess sst₂, and many also have some of the other four sst subtypes. Interaction with these receptors can be used to treat these tumors as well as to localize NETs by using radiolabeled somatostatin analogues (i.e. somatostatin receptor imaging [SRI]). In contrast, only 50–70% of poorly differentiated (G3 grade) NETs have sst₂ receptors. In the United States, (¹¹³In-DTPA-d-Phe¹) octreotide (octreoscan) (Fig. 80-2) is still generally used with gamma camera detection using single-photon emission computed tomography (SPECT) imaging. Using gallium-68-labeled somatostatin analogues and positron emission tomography (⁶⁸Ga-PET/CT) detection has greater sensitivity than using ¹¹³In-labeled somatostatin analogues (¹¹³In-SPECT/CT) (Fig. 80-3). It (NEWSPOT) is now approved for use in the United States. Because of its sensitivity and ability to localize tumor throughout the body, SRI is the initial imaging modality of choice for localizing both the primary tumor and metastatic NETs. SRI localizes tumor in 73–95% of patients with GI-NETs (carcinoids) and in 56–100% of patients with pNETs, except insulinomas. Insulinomas are usually small and have low densities of sst receptors, resulting in SRI being positive in only 12–50% of patients with insulinomas. SRS identifies >90–95% of patients with liver metastases due to NETs. Figure 80-3 shows an example with SRI of the increased sensitivity of ⁶⁸Ga-PET/CT over ¹¹³In-SPECT-CT and CT scanning to localize both the primary NET and liver/bone metastases in a patient with a metastatic small intestinal carcinoid (GI-NET). Occasional false-positive responses with SRI can occur (12% in one study) because numerous other normal tissues as well as diseases can have high densities of sst receptors, including granulomas (sarcoid, tuberculosis, etc.), thyroid diseases (goiter, thyroiditis), activated lymphocytes (lymphomas, wound infections), splenunculi, increased osteoblastic activity, meningiomas, and increased physiological uptake in the pancreatic uncinate process (⁶⁸Ga-DOTATAE PET/CT). If liver metastases are identified by SRI (performed without hybrid CT), to plan the proper treatment, either a CT or an MRI (with contrast enhancement) is recommended to assess the size and exact location of the metastases, because SRI does not provide information on tumor size. For pNETs in the pancreas, EUS is highly sensitive, localizing 77–100% of insulinomas, which occur almost exclusively within the pancreas. EUS is less sensitive for extrapancreatic tumors. It is increasingly used in patients with MEN 1, and to a lesser extent VHL, to detect small pNETs not seen with other modalities or for serial pNET assessments to determine size changes or rapid growth in patients in

A. CT (Primary)

B. ^{68}Ga -PET (Primary)C. Fusion ^{68}Ga -PET/CT (Primary)D. ^{111}In -SPECT/CT (Coronal)E. ^{68}Ga -PET/CT (Coronal)F. ^{111}In -SPECT/CT (Transv.)G. ^{68}Ga -PET (Transv.)

H. CT (Transv.)

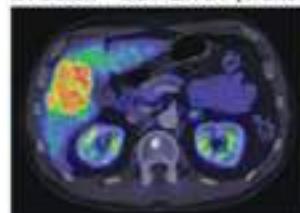
I. Fusion ^{68}Ga -PET/CT (Transv.)

FIGURE 80-3 Enhanced sensitivity of ^{68}Ga -PET/CT to localize lesions in patient with a metastatic small intestinal carcinoid (GI-NET). Panels A–C show the ability of the ^{68}Ga -PET (transverse images) to localize the primary (T) when the CT is negative. Panels D and E (coronal views—maximum intensity projections) show the greater resolution and ability to localize more metastatic lesions (T) in the liver and bone of ^{68}Ga -PET than the ^{111}In -SPECT/CT scanning, which has been generally used in the United States until recently. Panels F–I (transverse images) show the increased sensitivity of ^{68}Ga -PET over the ^{111}In -SPECT/CT scanning in identifying the extent of the liver metastases as well as identifying bone metastasis. GB, gallbladder; T, tumor; Transv, transverse images. (Results kindly provided by Prof. Anders Sundin, Department of Radiology, Uppsala University Hospital, Uppsala, Sweden.)

whom surgery is deferred. EUS with cytologic evaluation also is used frequently to distinguish an NF-pNET from a pancreatic adenocarcinoma or another nonendocrine pancreatic tumor. Not infrequently patients present with liver metastases due to an NET and the primary site is unclear. Occult small intestinal NETs (carcinoids) are increasingly detected by double-balloon enteroscopy or capsule endoscopy.

Insulinomas frequently overexpress receptors for glucagon-like peptide-1 (GLP-1), and radiolabeled GLP-1 analogues have been developed that can detect occult insulinomas not localized by other imaging modalities. This study is only performed in a few specialty centers. Functional localization by measuring hormonal gradients is now uncommonly used with gastrinomas (after intra-arterial secretin injections) but is still frequently used in insulinoma patients in whom other imaging studies are negative (assessing hepatic vein insulin concentrations post-intra-arterial calcium injections). Functional localization measuring hormone gradients in insulinomas or gastrin gradients in gastrinoma is a sensitive method, being positive in 80–100% of patients. The intra-arterial calcium test may also allow differentiation of the cause of the hypoglycemia and indicate whether it is due to an insulinoma or a nesidioblastosis. The latter entity is becoming increasingly important because hypoglycemia after gastric bypass surgery for obesity is increasing in frequency, and it is primarily due to nesidioblastosis, although it can occasionally be due to an insulinoma.

PET and use of hybrid scanners such as CT and SRI has sensitivity because of the greater resolution of PET scanning. PET scanning with ^{18}F -fluoro-DOPA in patients with carcinoids or with ^{11}C -5-HTP in patients with pNETs or GI-NETs (carcinoids) has greater sensitivity than cross-sectional imaging studies and may be used increasingly in

the future. PET/CT scanning using ^{18}F -FDG is receiving increasing attention in patients with NETs. It was initially thought that this would not be useful in NETs with the majority being well differentiated (G1, G2 grades, >85–98%) and having a low proliferative rate. However, ^{18}F -FDG PET/CT can identify higher grade NETs, and is particularly helpful for imaging G3 NETs, which are more frequently negative with SRS. There ^{18}F -FDG positivity can not only provide imaging information on location and tumor size, but also prognostic information because the relative survival of patients with the different NET grades is G1>G2>G3.

TREATMENT

Advanced and /or Aggressive Disease (Diffuse Metastatic Disease)

The single most important prognostic factor for survival is the presence of liver metastases (Fig. 80-4A, B, D, and E). For patients with foregut carcinoids without hepatic metastases, the 5-year survival in one study was 95%, and with distant metastases, it was 20% (Fig. 80-4A, B). With gastrinomas, the 5-year survival without liver metastases is 98%; with limited metastases in one hepatic lobe, it is 78%; and with diffuse metastases, 16%. In a large study of 156 patients (67 pNETs, rest carcinoids), the overall 5-year survival rate was 77%; it was 96% without liver metastases, 73% with liver metastases, and 50% with distant disease.

The recent introduction and validation of the prognostic value of the different classification and grading systems (WHO, ENETS, the American Joint Committee on Cancer/International Union Against

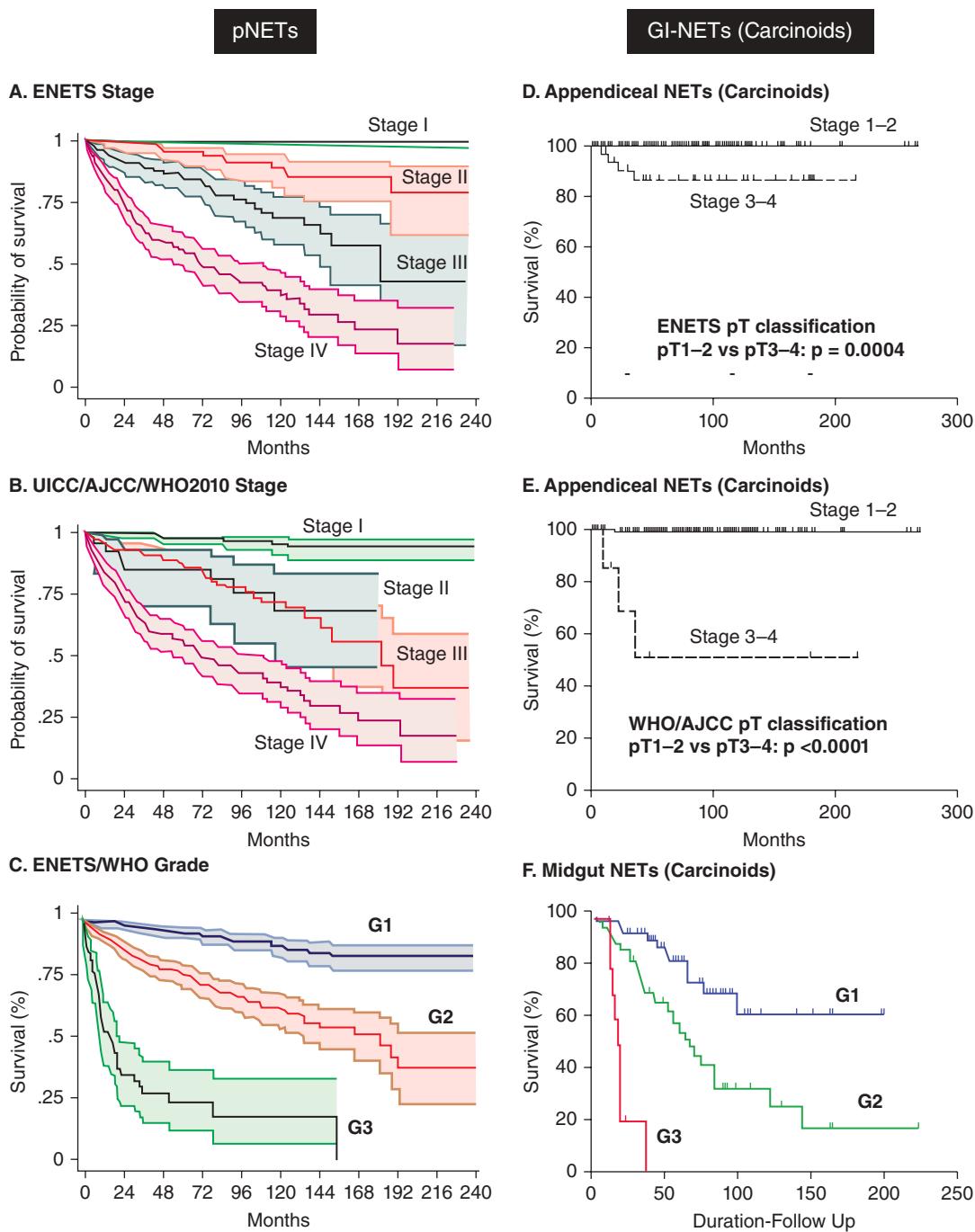


FIGURE 80-4 Survival (Kaplan-Meier plots) of patients with pancreatic neuroendocrine tumors (pNETs; n = 1072) (**A-C**) or gastrointestinal neuroendocrine tumors (GI-NETs; carcinoids) (appendix, n = 138; midgut, n = 238) (**D-F**) stratified according to recent proposed classification and grading systems. (Panels A-C are drawn from data in G Rindi et al: *J Natl Cancer Inst* 104:764, 2012; panels D and E are drawn from data in M Volante et al: *Am J Surg Pathol* 37:606, 2013; and panel F is drawn from data in MS Khan: *Br J Cancer* 108:1838, 2013.)

Cancer [AJCC/UICC]) are proving essential to stratify patients into different risk groups. A particular important prognostic factor is whether the NET is well differentiated (G1/G2) or poorly differentiated (<1% of all NETs) (G3) (Fig. 80-4C, F). In various series overall, well-differentiated NETs, which are aggressive tumors, have a 5-year survival of 50–80%, whereas poorly differentiated NETs survival of only 0–15% at 5 years (Fig. 80-4C, F).

Therefore, treatment for advanced metastatic disease is an important challenge. A number of different modalities are reported to be effective, including cytoreductive surgery (surgically or by RFA), treatment with chemotherapy, somatostatin analogues, interferon α , hepatic embolization alone or with chemotherapy (chemoembolization), molecular targeted therapy, radiotherapy with radiolabeled beads/microspheres, PRRT, and liver transplantation.

SPECIFIC ANTITUMOR TREATMENTS

Cytoreductive surgery is considered if either all of the visible metastatic disease or at least 90% is thought resectable; however, unfortunately, this is possible in only the 9–22% of patients who present with limited hepatic metastases. Although no randomized studies have proven that it extends life, results from a number of studies suggest that it may increase survival; therefore, it is recommended, if possible. RFA can be applied to NET liver metastases if they are limited in number (usually <5) and size (usually <3.5 cm in diameter). It can be used at the time of surgery (either general or laparoscopic) or using radiologic guidance. Response rates are >80%, the responses can last up to 3 years, the morbidity rate is low, and this procedure may be particularly helpful in patients with F-pNETs that are difficult to control medically. Although RFA has not been established in

a controlled trial, both the European and North American Neuroendocrine Tumor Society guidelines (ENETS, NANETS) state it can be an effective antitumor treatment for both refractory functional syndromes and for palliative treatment.

Although there are no controlled, long-term trials, palliative surgical resection of the small intestinal primary and surrounding tumor is generally recommended in most guidelines and expert opinion reviews for patients with midgut carcinoids with carcinoid syndrome, who almost invariably have unresectable liver metastases. Systematic analysis of existing data supports the conclusion surgical resection of the primary prevents complications (obstruction, etc.) and also prolongs survival in some studies. At the time of this resection, a cholecystectomy is recommended to possible biliary complications from long-term somatostatin therapy.

Chemotherapy plays a different role in the treatment of patients with pNETs and GI-NETs (carcinoids). Chemotherapy continues to be widely used in the treatment of patients with advanced pNETs with moderate success (response rates 20–70%). However, in general, its results in patients with metastatic GI-NETs (carcinoids) have been disappointing, with response rates of 0–30% with various two- and three-drug combinations, and thus, it is infrequently used in these patients. An important distinction in patients with pNETs is whether the tumor is well differentiated (G1/G2) or poorly differentiated (G3). The chemotherapeutic approach is different for these two groups. The current regimen of choice for patients with well-differentiated pNETs is the combination of streptozotocin and doxorubicin with or without 5-fluorouracil. Streptozotocin is a glucosamine nitrourea compound originally found to have cytotoxic effects on pancreatic islets, and later in studies with doxorubicin with or without 5-fluorouracil, it produced response rates of 20–45% in advanced pNETs. Streptozotocin causes considerable morbidity, with 70–100% of patients developing side effects (most prominent being nausea/vomiting in 60–100% or leukopenia/thrombocytopenia) and 15–70% of patients developing some degree of renal dysfunction (primarily proteinuria and/or decreased creatinine clearance). The combination of temozolamide (TMZ) with capecitabine is receiving increased attention as a possible alternative to streptozotocin-based therapies. Experience is still limited with this protocol and it is being evaluated in a number of current studies; however, analysis of larger retrospective studies shows responded rates from 48 to 70%. The use of TMZ or another alkylating agent in advanced pNETs is supported by some, but not all, studies that show low levels of the DNA repair enzyme O⁶-methylguanine DNA methyltransferase in pNETs correlate with the sensitivity to TMZ. Grade G3 NETs are primarily treated by chemotherapy (see below).

In addition to the effectiveness in controlling the functional hormonal state, long-acting somatostatin analogues such as octreotide and lanreotide are increasingly used for their antiproliferative effects. Whereas somatostatin analogues rarely decrease tumor size (i.e., 0–17%), these drugs have tumorstatic effects, stopping additional growth in 26–95% of patients with NETs. In a randomized, double-blind study in patients with metastatic midgut carcinoids (PROMID study), octreotide-LAR demonstrated a marked lengthening of time to progression (14.3 vs 6 months, $p = 0.000072$). This improvement was seen in patients with limited liver involvement. This study did not assess whether such treatment will extend survival. A double-blind, randomized, placebo-controlled, phase III study in patients with well-differentiated, metastatic, inoperable pNETs (45%) or GI-NETs (carcinoids) (55%) (CLARINET study) showed that monthly treatment with lanreotide-autogel reduced tumor progression or death by 53%. Somatostatin analogues can induce apoptosis in GI-NETs (carcinoids), which probably contributes to their tumorstatic effects. Treatment with somatostatin analogues is generally well-tolerated, with most side effects being mild and uncommonly leading to stopping the drug. Potential long-term side effects include diabetes/glucose intolerance, steatorrhea, and the development of gallbladder sludge/gallstones (10–80%), although only 1% of patients develop symptomatic gallbladder

disease. Because of these phase III studies, somatostatin analogues are generally recommended as first-line treatment for patients with well-differentiated metastatic NETs.

Interferon α , similar to somatostatin analogues, is effective at controlling the hormonal excess symptoms of NETs and has anti-proliferative effects in NETs, which primarily result in disease stabilization (30–80%), with a decrease in tumor size in <15% of patients. Interferon can inhibit DNA synthesis, block cell cycle progression in the G₁ phase, inhibit protein synthesis, inhibit angiogenesis, and induce apoptosis. Interferon α treatment results in side effects in the majority of patients, with the most frequent being a flu-like syndrome (80–100%), anorexia with weight loss, and fatigue. These side effects frequently decrease in severity with continued treatment. In addition, patients become accommodated to the symptoms. More serious side effects include hepatotoxicity (31%), hyperlipidemia (31%), bone marrow toxicity, thyroid disease (19%), and rarely CNS side effects (depression, mental/visual disorders). ENETS 2016 guidelines conclude that in patients with well-differentiated NETs that are slowly progressive, interferon α treatment should be considered if the tumor is somatostatin receptor negative or if somatostatin or targeted therapy (everolimus, sunitinib) treatment fails.

Molecular targeted medical treatment with either an mTOR inhibitor (everolimus) or a tyrosine kinase inhibitor (sunitinib) is now approved treatment in the United States and Europe for patients with metastatic unresectable pNET, each supported by a phase III, double-blind, prospective, placebo-controlled trial. Furthermore, a Phase 3 double-blind study (RADIANT-4) also demonstrated the effectiveness of everolimus in advanced, non-functional NETs of the lung or GI-tract. In this study involving patients with advanced, progressive well differentiated, NF-lung/GI-NETs, everolimus significantly ($p < 0.000001$) improved progression-free survival and led to FDA approval for its use.

mTOR is a serine-threonine kinase that plays an important role in proliferation, cell growth, and apoptosis in both normal and neoplastic cells. Activation of the mTOR cascade is important in mediating NET cell growth. A number of mTOR inhibitors have shown promising antitumor activity in NETs including everolimus and temsirolimus, with the former undergoing two phase III trials (RADIANT-3/ RADIANT-4) studies in patients with advance progressive NETs (RADIANT-3=pNETs, RADIANT-4=lung, GI NF-NETs). In the RADIANT-III study which involved 410 patients with advanced, progressive pNETs, everolimus caused significant improvement in progression-free survival (11 vs 4.6 months, $p < 0.001$) and increased by a factor of 3.7 the proportion of patients progression-free at 18 months (37% vs 9%). Everolimus treatment was associated with frequent side effects, causing a twofold increase in adverse events, with the most frequent being grade 1 or 2. Grade 3 or 4 side effects included hematologic, GI (diarrhea, stomatitis, or hypoglycemia occurring in 3–7% of patients. Most grade 3 or 4 side effects were controlled by dose reduction or drug interruption. Similar side-effects were found in the RADIANT-4 study. The ENETS 2016 guidelines conclude that everolimus, similar to sunitinib (below), can be considered as a first-line treatment in well-differentiated pNETs that are unresectable especially if somatostatin analogues are not an option. However, these guidelines recommended that somatostatin analogues be the initial treatment because of their low incidence of side-effects. In patients with GI-NETs, the ENET 2016 guidelines recommended that everolimus could be recommended as second-line therapy after somatostatin analogues.

Like other normal and neoplastic cells, NETs frequently possess multiple types of the 20 different tyrosine kinase (TK) receptors that are known and mediate the action of different growth factors. Numerous studies demonstrate that TK receptors in normal and neoplastic tissues as well as NETs are especially important in mediating cell growth, angiogenesis, differentiation, and apoptosis. Whereas a number of TK inhibitors show antiproliferative activity in NETs, only sunitinib has undergone a phase III controlled trial. Sunitinib is

an orally active small-molecule inhibitor of TK receptors (PDGFRs, VEGFR-1, VEGFR-2, c-KIT, FLT-3). In a phase III study in which 171 patients with progressive, metastatic, nonresectable pNETs were treated with sunitinib (37.5 mg/d) or placebo, sunitinib treatment caused a doubling of progression-free survival (11.4 vs 4.5 months, $p < 0.001$), an increase in objective tumor response rate (9% vs 0%, $p = 0.007$), and an increase in overall survival. Sunitinib treatment was associated with an overall threefold increase in side effects, although most were grade 1 or 2. The most frequent grade 3 or 4 side effects were neutropenia (12%) and hypertension (9.6%), which were controlled by dose reduction or temporary interruption. There is no consensus regarding the order of sunitinib or everolimus use in patients with advanced, well-differentiated, progressive pNETs.

In patients with liver-predominant metastatic disease, a number of locoregional strategies have been used including: transarterial arterial embolization (TAE) alone or with chemotherapeutic agents (TACE); and selective internal radiation therapy (SIRT) or radioembolization. TACE/TAE can be effective because the blood supply to normal liver tissue is primarily from the portal vein whereas tumors receive 70–80% of their supply from the hepatic artery. Occlusion of selective branches of the hepatic artery is now generally performed radiologically. Contraindications include >50–75% liver involvement by tumor, portal vein thrombosis, post-biliary reconstructive surgery, liver failure, and a poor performance rating. Results include a symptomatic response rate of 50–100%, and an objective response rate of 25–86% with a mean duration of response of 6–45 months. Complications include a postembolization syndrome with pain, nausea/vomiting and fever in 10–80% with <6% mortality. SIRT using yttrium-90 (⁹⁰Y) glass or resin microspheres is a relatively newer approach being evaluated in patients with unresectable NET liver metastases. The treatment requires careful evaluation for vascular shunting before treatment and a pretreatment angiogram to evaluate placement of the catheter and is generally reserved for patients without extrahepatic metastatic disease and with adequate hepatic reserve. One of two types of ⁹⁰Y microspheres is used: either microspheres with a 20- to 60- μ m diameter and 50 Bq/sphere (SIR-Spheres) or glass microspheres (TheraSpheres) with a 20- to 30- μ m diameter and 2500 Bq/sphere. The ⁹⁰Y-microspheres are delivered to the liver by intraarterial injection from percutaneously placed catheters. The response rate varied from 50 to 61% (partial or complete), tumor stabilization occurred in 22–41%, 60–100% had symptomatic improvement, and overall survival varied from 25 to 70 months. Side effects include postembolization syndrome (pain, fever, nausea/vomiting [frequent]), which is usually mild, although grade 2 (43%) or grade 3 (1%) symptoms can occur; radiation-induced liver disease (<1%); and radiation pneumonitis (<1%). Contraindications to use include excess shunting to the GI tract or lung, inability to isolate the liver arterial supply, and inadequate liver reserve. Because of the limited data available in the ENETS 2012 guidelines, treatment with SIRTs is considered experimental.

PRRT for NETs with radiolabeled somatostatin analogues is now being increasing considered for patients with advanced NETs. The success of this approach is based on the finding that somatostatin SST are overexpressed or ectopically expressed by 60–100% of all NETs, which allows the targeting of cytotoxic, radiolabeled somatostatin receptor ligands. Three different radionuclides have been used including: high doses of [¹¹¹In-DTPA-d-Phe¹] octreotide, which emits γ -rays, internal conversion, and Auger electrons; ⁹⁰Yttrium, which emits high-energy β -particles coupled by a DOTA chelating group to octreotide or octreotate; and ¹⁷⁷Lutetium-coupled analogues, which emit both (Fig. 80-2). At present, the ¹⁷⁷lutetium-coupled analogues are the most widely used and although not approved for general use in any country, they are frequently available in speciality centers on a special or compassionate basis. A double-blind, prospective, randomized trial (NETTER-1 Study) (using ¹⁷⁷Lu-Dotatacept [Lutathera]) has supported the efficacy and safety of this approach in patients with advanced inoperable, progressive midgut GI-NETs (carcinoids). In this trial, which included 229 patients with grade G1,2 metastatic

midgut carcinoids, a marked increased in progressive-free survival ($p < 0.0001$) was seen with PRRT treatment with a ¹⁷⁷Lu-labeled-somatostatin-analog, with an acceptable safety profile and with a suggestion of an improved survival, although final survival analysis is not yet complete. In a number of retrospective, non-blinded trials, ¹¹¹Indium-, ⁹⁰Yttrium-, and ¹⁷⁷lutetium-labeled compounds caused tumor stabilization in patients with advanced, progressive NETs in 41–81%, 44–88%, and 23–51%, respectively, and a decrease in tumor size in 8–30%, 6–37%, and 38%, respectively, of patients. In one large study involving 504 patients with malignant NETs, ¹⁷⁷lutetium-labeled analogues produced a reduction of tumor size of >50% in 30% of patients (2% complete) and tumor stabilization in 51% of patients. The ENETS 2016, NANETS, Nordic 2010, and European Society for Medical Oncology (ESMO) guidelines list PRRT as an experimental or investigational treatment at present.

The use of liver transplantation has been abandoned for treatment of most metastatic tumors to the liver. However, for metastatic NETs, it is still a consideration by many centers although its use is controversial. An analysis of data from a number of centers showed that the overall 5-year survival is 47–58%, but varies widely in different studies from 36 to 97%; the 5-year disease-free survival was usually 20–30%, but varied from 9 to 77%, with a postoperative mortality <15%. With pNETs the 5-year survival rate varies from 30 to 50% and for GI-NETs from 60 to 90%. In various studies, important prognostic factors for a poor outcome include a major resection performed in addition at the time of the liver transplant; poor tumor differentiation; hepatomegaly; age >45 years; a primary NET in the duodenum or pancreas; the presence of extrahepatic metastatic disease or extensive liver involvement (>50%); Ki₆₇ proliferative index >10%; and abnormal E-cadherin staining. The ENETS 2016 guidelines conclude that liver transplantation should be viewed as an option in highly selected patients, preferably in young patients with functional syndromes demonstrating early resistance to medical therapies.

The management and treatment of patients with G3 NETs (Ki₆₇>20) (WHO classification as NECs) has undergone a number of changes because of some important new insights. It is now realized that G3 NETs are heterogeneous and this has resulted in a proposal that they be divided into at least two categories; this division has important management ramifications because it is proposed they be treated differently. In reviews of G3 patients from a number of centers, a group of patients have G3 grading but with well-differentiated morphology (usually with a Ki₆₇ 20–55) and it is proposed these be called G3 NET. These G3 NET patients have a better prognosis than poorly differentiated G3 tumors (usually with Ki₆₇>55), which are proposed to be called G3 NEC tumors. Pathology studies show that G3 NETs frequently have loss of ATRX/DAXX, whereas the G3 NEC poorly differentiated tumors have abnormal expression of p53, retinoblastoma and/or SMAD4. Most patients with G3 NETs have regional or distant metastases at the time of diagnosis and surgery is rarely curative, with the result that chemotherapy is usually recommended. This new subclassification has therapeutic implications because it is proposed to treat the G3 NET tumors similar to treatment for well-differentiated G2 tumors, whereas for G3 NEC tumors, treatment with cisplatin-based regimens with etoposide or other agents (vincristine, paclitaxel) is recommended. The response rates with this protocol are 40–70%; however, responses are generally short-lived (<12 months). This chemotherapy regimen can be associated with significant toxicity including GI toxicities (nausea, vomiting), myelosuppression.

FURTHER READING

- BASUROY R et al: Neuroendocrine tumors. *Gastroenterol Clin North Am* 45:487, 2016.
- CIVES M, STROSBERG J: Treatment strategies for metastatic neuroendocrine tumors of the gastrointestinal tract. *Curr Treat Options Oncol* 18:1, 2017.

- Ito T, JENSEN RT: Molecular imaging in neuroendocrine tumors: Recent advances, controversies, unresolved issues, and roles in management. *Curr Opin Endocrinol Diabetes Obes* 24:15, 2017.
- KLIMSTRA DS: Pathologic classification of neuroendocrine neoplasms. *Hematol Oncol Clin North Am* 30:1, 2016.
- OBERG K, SUNDIN A: Imaging of Neuroendocrine Tumors. *Front Horm Res* 45:142, 2016.
- O'TOOLE D et al: ENETS 2016 Consensus Guidelines for the Management of Patients with Digestive Neuroendocrine Tumors: An Update. *Neuroendocrinology* 103:113, 2016.
- PAVEL M et al: ENETS Consensus Guidelines Update for the Management of Distant Metastatic Disease of Intestinal, Pancreatic, Bronchial Neuroendocrine Neoplasms (NEN) and NEN of Unknown Primary Site. *Neuroendocrinology* 103:172, 2016.
- STROSBERG J et al: Phase 3 Trial of ¹⁷⁷Lu-Dotatate for Midgut Neuroendocrine Tumors. *N Engl J Med* 376:125, 2017.
- TAMBURRINO D et al: Surgical management of neuroendocrine tumors. *Best Pract Res Clin Endocrinol Metab* 30:93, 2016.
- YAO JC et al: Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): A randomised, placebo-controlled, phase 3 study. *Lancet* 387:968, 2016.

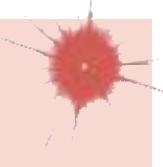
tenth most common in females; the male-to-female ratio is 2:1. Though this malignancy may be diagnosed at any age, it is uncommon in those aged <45 years, and incidence peaks between the ages of 50 and 70 years. Many factors have been investigated as possible contributing causes; associations include cigarette smoking, obesity, and hypertension. Risk is also increased for patients who have acquired cystic disease of the kidney associated with end-stage renal disease and for those with tuberous sclerosis.

Most cases of renal cell carcinoma are sporadic, although familial forms have been reported (**Table 81-1**). One such form is associated with von Hippel-Lindau (VHL) syndrome, an autosomal dominant disorder. Genetic studies identified the *VHL* gene on the short arm of chromosome 3. Approximately 35% of individuals with VHL disease develop clear cell renal cell carcinoma. 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 kidney tumors. The renal tumors are usually of the chromophobe type, but they can exist as hybrids with other cell types. This disorder is associated with mutations in the *FLCN* gene, which codes for folliculin.

81

Renal Cell Carcinoma

Robert J. Motzer



Renal cell carcinomas account for 90–95% of malignant neoplasms arising from the kidney. Notable features include diagnosis without symptoms, resistance to cytotoxic agents, infrequent responses to biologic response modifiers such as interleukin (IL)-2, robust activity of antiangiogenesis-targeted agents, and a variable clinical course for patients with metastatic disease, including anecdotal reports of spontaneous regression. 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. 82** for transitional cell carcinomas.

EPIDEMIOLOGY

The incidence of renal cell carcinoma rose for three decades but has now reached a plateau of ~63,000 cases annually in the United States, resulting in >14,000 deaths per year. It is the ninth most common cancer overall in the United States, seventh most common in males, and

PATHOLOGY AND GENETICS

Renal cell neoplasia represents a heterogeneous group of tumors with distinct histopathologic, genetic, and clinical features ranging from benign to high-grade malignant (**Table 81-2**). They are classified on the basis of morphology and histology. Categories include clear cell carcinoma (70% of cases), papillary tumors (10%), chromophobe tumors (<5%), oncocytomas (5–10%), collecting duct or Bellini duct tumors (<1%), and translocation carcinoma (<1%). Papillary tumors tend to be bilateral and multifocal. Chromophobe tumors have a more indolent clinical course, and oncocytomas are considered benign neoplasms. In contrast, Bellini duct carcinomas, which are thought to arise from the collecting ducts within the renal medulla, are rare but often very aggressive. Medullary carcinoma has histopathologic and clinical features similar to those of Bellini duct carcinoma, but it is associated with sickle cell trait.

Clear cell tumors, the predominant histology, are found in >80% of patients who develop metastases. Clear cell tumors arise from the epithelial cells of the proximal tubules and usually show chromosome 3p deletions. Deletions of 3p21–26 (where the *VHL* gene maps) are identified in patients with familial as well as sporadic tumors. *VHL* encodes a tumor suppressor protein that is involved in regulating the transcription of vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF), and a number of other hypoxia-inducible proteins. Inactivation of *VHL* leads to overexpression of

TABLE 81-1 Hereditary Renal Cell Tumors

SYNDROME	CHROMOSOME(S)	GENE	PROTEIN	KIDNEY TUMOR TYPE	ADDITIONAL FINDINGS
von Hippel-Lindau syndrome	3p25	<i>VHL</i>	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	<i>MET</i>	MET	Papillary (type I)	
Hereditary leiomyomatosis and RCC	1q42	<i>FH</i>	Fumarate hydratase	Papillary (non-type I)	Leiomyoma; uterine leiomyoma/leiomyosarcoma
Birt-Hogg-Dubé syndrome	17p11	<i>FLCN</i>	Folliculin	Chromophobe, oncocytoma	Facial fibrofolliculoma; pulmonary cysts
Tuberous sclerosis	9q34 16p13	<i>TSC1</i> <i>TSC2</i>	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
Constitutional chromosome 3 translocations	3p13-14	Unknown	Unknown	Clear cell	

Abbreviation: RCC, renal cell carcinoma.

TABLE 81-2 Classification of Epithelial Neoplasms Arising from the Kidney

CARCINOMA TYPE	CHARACTERISTICS GROWTH PATTERN	CELL OF ORIGIN	CYTOGENETICS
Clear cell	Acinar or sarcomatoid	Proximal tubule	3p-, 5q+, 14q-
Papillary	Papillary or sarcomatoid	Proximal tubule	+7, +17, -Y
Chromophobe	Solid, tubular, or sarcomatoid	Distal tubules/cortical collecting duct	Whole arm losses (1, 2, 6, 10, 13, 17, and 21)
Oncocytic	Tumor nests	Cortical collecting duct	-1, -14, -Y; rearrangement involving 11q13; or normal karyotype
Collecting duct	Papillary or sarcomatoid	Medullary collecting duct	Variable or undetermined
MitF Translocation ^a	Clear and papillary	Undetermined	Gene fusions involving Xp11 (<i>TFE3</i>) or t(6;11) (<i>MALAT1-TFE3</i>)

^aMicrophthalmia transcription factor gene family.

these agonists of the VEGF and PDGF receptors, which promote tumor angiogenesis and tumor growth. Agents that inhibit proangiogenic growth factor activity show antitumor effects. Enormous genetic variability has been documented in tumors within individual patients. Although the tumors have a clear clonal origin and often contain *VHL* mutations in common, different portions of the primary tumor and different metastatic sites may have wide variation in genetic lesions they contain. This tumor heterogeneity may underlie the emergence of treatment resistance.

While *VHL* is the gene most frequently mutated in clear cell renal cell carcinoma (52% of cases), other genes are implicated as well: *PBRM1* in 40% of cases, *SETD2* in 15% of cases, and *BAP1* in 15% of cases. These three genes, all part of the chromatin remodeling/histone methylation pathway, are also located within a 50-Mb region on the short arm of chromosome 3p. Mutations in *BAP1* have been linked to shorter survival in renal cancer. In a subset of clear cell renal cell carcinomas, alterations have been found in components of the mammalian target of rapamycin (mTOR) pathway, spurring the study of mTOR inhibitors in renal cancer.

Approximately 10% of renal cell carcinomas are of the papillary subtype, where the most common copy-number events are gain of chromosome 7 (where MET is located) and chromosome 17. Alterations in MET are associated with type I papillary renal cell carcinoma, whereas type II papillary tumors are characterized by NFR2-antioxidant response element alterations. In the chromophobe subtype, which comprises ≤5% of cases of renal cell carcinoma, two mutations have been noted: TP53 in 32% of cases and PTEN in 9%.

CLINICAL PRESENTATION

The presenting signs and symptoms include hematuria, flank or abdominal pain, and a flank or abdominal mass. Other symptoms are fever, weight loss, anemia, and a varicocele. The tumor is most commonly detected as an incidental finding on a radiograph. Widespread use of radiologic cross-sectional imaging procedures (computed tomography [CT], ultrasound, magnetic resonance imaging [MRI]) contributes to earlier detection, including incidental renal masses detected 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 renal cell carcinoma 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 ~3% of patients. Anemia, a sign of metastatic 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.

The standard evaluation of patients with suspected renal cell tumors includes a CT scan of the abdomen and pelvis, chest radiograph, urine analysis, and urine cytology. 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. The differential diagnosis of a renal mass includes cysts, benign neoplasms (adenoma, angiomyolipoma, oncocytoma), inflammatory lesions (pyelonephritis or abscesses), and other primary or metastatic cancers. Other malignancies that may involve the kidney include transitional cell carcinoma of the renal pelvis, sarcoma, lymphoma, and Wilms' tumor. All of these are less common causes of renal masses than is renal cell cancer.

STAGING AND PROGNOSIS

Staging is based on the American Joint Committee on Cancer (AJCC) staging system (Fig. 81-1). 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 (IIIa) or involve a single hilar lymph node (N1), and stage IV disease includes tumors that have invaded adjacent organs or involve multiple 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 74% across all renal cell carcinomas, but it varies by stage: 81% for stage I, 74% for stage II, 53% for stage III, and 8% for stage IV.

Prognostic risk models are helpful for counseling patients, and for anticipating survival rates when designing a clinical trial. The most widely used prognostic model, developed by investigators at Memorial Sloan Kettering Cancer Center, incorporated five factors shown to correlate with worse survival in advanced renal cell carcinoma: poor performance status, high serum lactate dehydrogenase, high serum calcium, low hemoglobin concentration, and <1-year interval from diagnosis to treatment. Patients with zero risk factors had significantly longer median survival (30 months) than those with one or two risk factors (14 months) and those with three to five risk factors (5 months).

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, the ipsilateral adrenal gland in some cases, and adjacent hilar lymph nodes. Open, laparoscopic, or robotic surgical techniques may be used to perform radical nephrectomy. The role of a regional lymphadenectomy is controversial. Extension into the renal vein or inferior vena cava (stage III disease) does not preclude resection even if cardiopulmonary bypass is required. If the tumor is resected, half of these patients have prolonged survival.

Nephron-sparing approaches via open or laparoscopic surgery may be appropriate for patients who have impaired renal function or only one kidney, depending on the size and location of the lesion. A nephron-sparing approach can also be used for patients

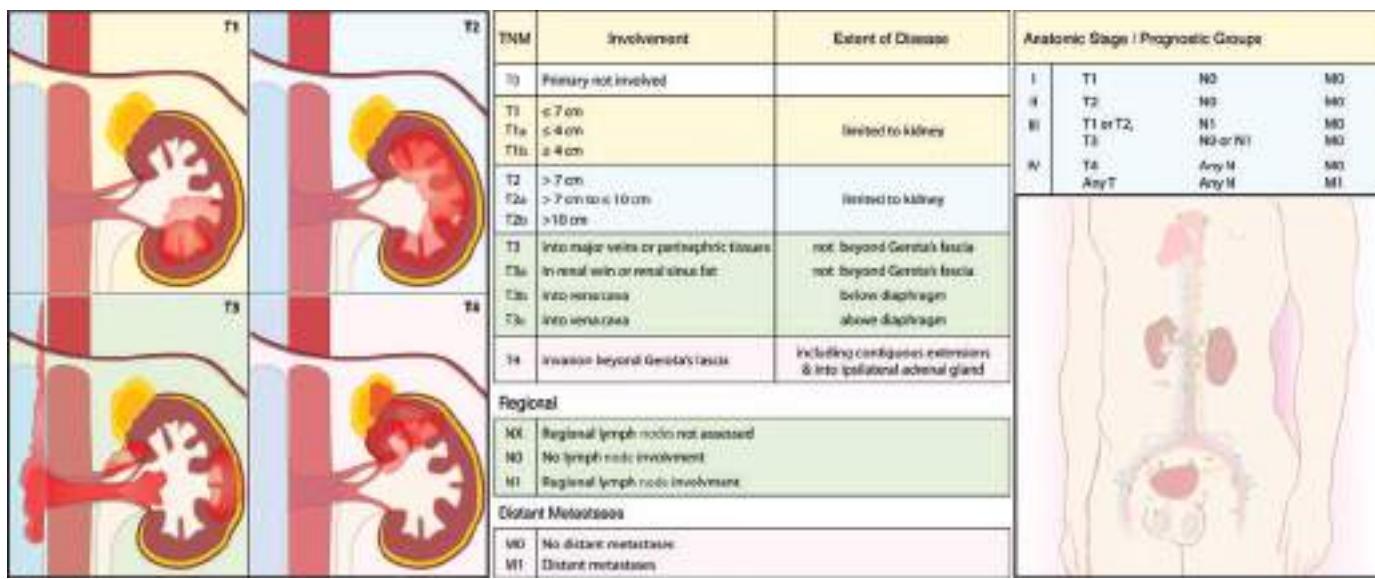


FIGURE 81-1 Renal cell carcinoma staging. TNM, tumor, node, metastasis.

with bilateral tumors. Partial nephrectomy techniques are applied electively to resect small masses for patients with a normal contralateral kidney. Radical nephrectomy can lead to an increased risk for chronic kidney disease and is associated with increased risks of cardiovascular morbidity and mortality. When compared with radical nephrectomy, partial nephrectomy can achieve preserved renal function, and reduced frequency of late cardiovascular events.

Adjuvant therapy with interferon- α or radiation therapy following this surgery does not improve outcome, even in cases with a poor prognosis. Adjuvant trials with sunitinib, an orally administered antiangiogenesis inhibitor, do not consistently show a benefit in prolonging time to relapse following nephrectomy.

METASTATIC DISEASE

Surgery has a limited role for patients with metastatic disease. Long-term survival may occur in patients who relapse after nephrectomy in a solitary site that is removed. One indication for nephrectomy with metastases at initial presentation is to alleviate pain or hemorrhage of a primary tumor. Also, a cytoreductive nephrectomy before systemic treatment improves survival for carefully selected patients with stage IV tumors. 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. It may be best to document progression before considering systemic treatment.

Radiation therapy is generally used for palliation of bone or brain metastases. The types of radiotherapy most commonly used are external beam therapy and stereotactic radiotherapy. In select cases, stereotactic ablative radiotherapy to a metastatic site may result in local control with relatively minimal toxicity.

Metastatic renal cell carcinoma is refractory to cytotoxic chemotherapy. Cytokine therapy with IL-2 or interferon- α produces regression in 10–15% of patients. IL-2 produces durable complete remission in a small proportion of cases. In general, cytokine therapy is considered unsatisfactory for most patients due to high levels of toxicity and the unpredictability of response.

The situation changed dramatically when two large-scale randomized trials established a role for antiangiogenic therapy, as predicted by the genetic studies. These trials separately evaluated two orally administered antiangiogenic agents, sorafenib and sunitinib, that inhibited receptor tyrosine kinase signaling through the VEGF and PDGF receptors. Both showed efficacy as second-line treatment following progression during cytokine treatment, resulting in approval by regulatory authorities for the treatment of metastatic renal cell carcinoma. A

randomized phase III trial comparing sunitinib to interferon- α showed superior efficacy for sunitinib with an acceptable safety profile. This trial resulted in a change in the standard first-line treatment from interferon to sunitinib.

These were followed by eight new systemic agents for metastatic renal cell carcinoma (Table 81-3): pazopanib, axitinib, cabozantinib, and lenvatinib, also tyrosine kinase inhibitors; the antiangiogenic bevacizumab that inhibits the VEGF ligand; the mTOR inhibitors temsirolimus and everolimus; and nivolumab that inhibits PD-1. While the improvements in 5-year renal cancer survival rates over the past decades (50% in the mid-1970s, 57% in the late 1980s, and 74% for 2005–2012) can be attributed to widespread imaging leading to earlier discovery of tumors, the new agents are likely playing a part as well.

Pazopanib was compared to sunitinib in a randomized first-line phase III trial. Efficacy was similar, and there was less fatigue and skin toxicity, resulting in better quality-of-life scores for pazopanib compared with sunitinib. Temsirolimus and everolimus show activity in patients with untreated poor-prognosis tumors and in sunitinib/sorafenib-refractory tumors. Patients benefit from the sequential use of axitinib and everolimus following progression with sunitinib or pazopanib first-line therapy. Nivolumab, cabozantinib, and lenvatinib plus everolimus were compared to everolimus in randomized trials and showed that patients lived longer with each of these agents compared to patients treated with everolimus.

Biomarkers are needed to select appropriate treatment for individual patients and to get quicker confirmation of whether treatment is working. However, though a number of predictive biomarker candidates have been tested in metastatic renal cell carcinoma patients receiving various systemic therapies, none have been validated for clinical use.

GLOBAL CONSIDERATIONS

Worldwide, ~340,000 patients are diagnosed every year with malignant tumors arising from the kidney, resulting in >140,000 deaths annually. Kidney cancer is the ninth most common cancer in men and the fourteenth most common cancer in women. Higher incidence is observed in developed countries, including the United States, Northern Europe, Eastern Europe, and Australia. 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 access to and availability of 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.

TABLE 81-3 Approved Systemic Therapies for Metastatic Renal Cell Carcinoma

CLASS	DRUG	FIRST FDA APPROVAL FOR RCC	ORIGINALLY APPROVED FOR	CURRENTLY USED FOR
Cytokines	High-dose interleukin-2 ^a	1992	Advanced RCC	Advanced RCC first-line
	Interferon- α	2009	Advanced RCC, in combination with bevacizumab	Advanced RCC, in combination with bevacizumab first-line
Antiangiogenic: tyrosine kinase inhibitors	Sorafenib	2005	Advanced RCC second-line	Advanced RCC third-line or later
	Sunitinib	2006	Advanced RCC second-line	Advanced RCC first-line
	Pazopanib	2009	Advanced RCC first-line or after cytokine therapy	Advanced RCC first-line
	Axitinib	2012	Advanced RCC second-line	Advanced RCC second-line
	Cabozantinib	2016	Advanced RCC second-line	Advanced RCC second- or third-line
	Lenvatinib	2016	Advanced RCC second-line in combination with everolimus	Advanced RCC second- or third-line in combination with everolimus
Antiangiogenic: VEGF ligand antibody	Bevacizumab	2009	Advanced RCC first-line in combination with interferon- α	Advanced RCC first-line in combination with interferon- α
mTOR inhibitors	Temsirolimus ^b	2007	Advanced RCC	Advanced RCC with poor prognosis features first-line
	Everolimus	2009	Metastatic RCC second-line	Advanced RCC third-line or later
PD-1 inhibitor	Nivolumab	2015	Advanced RCC second-line	Advanced RCC second- or third-line

^aOption only for patients with good performance status, no significant comorbidity, and access to medical centers experienced with this agent. ^bOption for poor-risk patients.

Abbreviations: FDA, Food and Drug Administration; mTOR, mammalian target of rapamycin; PD-1, programmed cell death-1; RCC, renal cell carcinoma; VEGF, vascular endothelial growth factor.

FURTHER READING

CANCER GENOME ATLAS RESEARCH NETWORK: Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499:43, 2013.

CHOUEIRI TK, MOTZER RJ: Systemic therapy for metastatic renal cell carcinoma. *N Engl J Med* 376:354, 2017.

GERLINGER M et al: Intratumor heterogeneity and branched evolution revealed by multi-region sequencing. *N Engl J Med* 366:883, 2012.

GILL IS et al: Clinical practice. Small renal mass. *N Engl J Med* 362:634, 2010.

MOTZER RJ et al: Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med* 369:722, 2013.

MOTZER RJ et al: Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 373:1803, 2015.

SRIGLEY JR et al: The International Society of Urological Pathology (ISUP) Vancouver Classification of renal neoplasia. *Am J Surg Pathol* 37:1469, 2013.

ZNAOR A et al: International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol* 67:531, 2015.

the fifth most common cancer diagnosis annually in the United States with >76,000 new cases and 16,000 deaths every year. Because cancers of the renal pelvis are often lumped in with all kidney cancers, the true incidence and mortality from nonbladder urinary tract cancers are less precise. While less frequent than bladder cancer, an additional 20,000 new cases and 5000 deaths are estimated every year. While significant advances in therapy options and improvements in patient outcomes have rapidly occurred in many cancers in the past decade, progress in urinary tract cancers has lagged. Fortunately, an accelerated understanding of the molecular underpinnings of bladder and urinary tract cancer biology has led to a significant increase in clinical trials with the first U.S. Food and Drug Administration (FDA) approval of a new drug for advanced bladder and urinary tract cancers in over 25 years with many more expected to follow. This chapter reviews the established, current, and emerging evidence that serves as the basis for the rapidly evolving standards of care for patients with bladder and urinary tract cancers.

CLINICAL EPIDEMIOLOGY AND RISK FACTORS

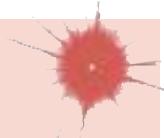
Bladder cancer typically affects older patients with a median age at diagnosis of 73 years. Males are four times more frequently affected than females. Similarly, bladder cancer is more common in Caucasians than in Asian patients. Singular inheritable genetic risk factors are rare in patients with bladder or urinary tract cancers. Patients with defects in mismatch repair genes leading to microsatellite instability (*MLH1*, *MSH2*, *MSH6*, etc.) as part of the familial cancer Lynch syndrome are at particular risk of upper urinary tract cancers of the renal pelvis and ureter. Additionally, patients with Cowden disease (*PTEN* mutations) or retinoblastoma (*RB1* mutations) are at increased risk for developing bladder cancer.

Historically, associations have existed between environmental toxic exposures and higher rates of developing bladder cancer. Carcinogenic agents associated with increased risk of bladder cancer have included the aromatic amines benzidine and beta-naphthylamine that can be present in industrial dyes as well as arsenic that can be found in some drinking water supplies in underdeveloped countries. Other chemicals in the leather, paint, rubber, textiles, and printing industries have been associated with bladder cancer. More recently, associations with exposures to hair dyes and hair sprays in workers in the hairstyling field have been suggested. Additionally, much concern has been raised regarding use of the antidiabetic medication, pioglitazone, and bladder cancer risk. Extensive review of population data by leading

82

Cancer of the Bladder and Urinary Tract

Noah M. Hahn



GLOBAL CONSIDERATIONS

Within the United States, urothelial carcinoma of the bladder and urinary tract are most closely related to tobacco smoking history. However, within developing countries water supplies contaminated with arsenic or schistosomiasis parasites also are major carcinogenic contributors.

INTRODUCTION

Cancers of the urinary tract including the bladder, renal pelvis, ureter, and urethra occur frequently, and they represent the second most common class of genitourinary cancers. Bladder cancer alone represents



bladder cancer experts has produced mixed associations. An association between chronic inflammatory states and the development of squamous bladder cancer clearly exists in underdeveloped countries in patients chronically infected with the parasitic disease schistosomiasis and in paraplegic patients with chronic indwelling catheters. Above and beyond each of these associations, however, smoking of tobacco products (cigarettes, cigars, pipes, etc.) has been and continues to remain the overwhelming leading risk factor for development of bladder cancer. Among new bladder cancer diagnoses, 90% of cases occur in current or former smokers. Toxicologists have estimated that over 70 confirmed carcinogenic toxins are present within tobacco smoke. It is estimated that one-third of bladder cancer cases could be prevented through simple modification of lifestyle choices, in particular cessation of smoking.

■ CLINICAL PRESENTATION AND DIAGNOSTIC WORKUP

Occasionally, patients will present with flank pain in association with an upper tract renal pelvis or ureter cancer or due to hydronephrosis in association with a bladder tumor obstructing the orifice of the ureter within the bladder. Only in rare cases do patients present with significant cachexia and widespread metastatic disease. For most patients, painless hematuria (either gross or microscopic) represents the initial manifestation of an underlying urinary tract cancer. In females, hematuria due to malignancy can often be mistaken for a urinary tract infection or menstrual bleeding. While treatment with antibiotics is warranted if a concurrent urinary tract infection is noted on initial urinalysis, persistent hematuria requires further workup. Painless hematuria in males is almost always abnormal and should be worked up. Initial investigations in patients of either sex should include urine cytology and visual examination of the bladder by cystoscopy. Cytology is successful in identifying cancer in only 50% of individuals with high-grade bladder cancers. In addition to urine cytology, radiographic evaluation of the kidneys and upper urinary tract by CT urogram should be performed. Because of the increased sensitivity and reduced IV contrast loads, CT urograms have largely replaced IV pyelograms as the preferred upper urinary tract imaging modality. A magnetic resonance (MR) urogram may be substituted in patients with poor renal function. Additional diagnostic testing of the urine to assess for cancer-associated chromosomal changes by fluorescent in situ hybridization, increased levels of nuclear mitotic proteins, increased bladder tumor-associated antigens, or higher levels of staining on cells shed by the bladder may identify some cancers missed by traditional cytology testing. However, they may also produce abnormal results in patients who do not have cancer. For now, these adjunct molecular tests are primarily utilized in detecting recurrent cancer in patients with a prior diagnosis of urinary tract cancer. Small tumors, particularly flat noninvasive tumors of the bladder, may be detected at higher rates with the use of blue light cystoscopy or narrow-band imaging cystoscopy. Both blue light and narrow-band imaging cystoscopies are now used routinely in the initial workup and subsequent monitoring of patients with bladder cancer. For patients with no bladder abnormalities in whom upper tract tumors are suspected, visualization

of the upper urinary tracts and renal pelvises should be performed by ureteroscopy or retrograde pyelography.

In all patients with abnormalities noted in the bladder or upper urinary tracts, complete endoscopic resection for histologic diagnosis and staging should be performed when possible via either transurethral resection of bladder tumor (TURBT) or endoscopic resection of upper tract tumors.

■ HISTOLOGY

Urothelial carcinoma, formerly referred to as *transitional cell carcinoma*, is the most common urinary tract cancer histology that is observed in ~90% of cases. Squamous, glandular, micropapillary, plasmacytoid, sarcomatoid, and other variant features can often be found in portions of urothelial carcinoma tumors; however, pure variant histologies are rare. The presence of some variant histologies including micropapillary and plasmacytoid has been associated with worse surgical outcomes compared to urothelial carcinoma. Nonurothelial variant histologies including squamous cell carcinoma, adenocarcinoma, small-cell carcinoma, and carcinosarcoma collectively account for ≤10% of urinary tract tumors. Examples of traditional urothelial carcinoma and some of the variant histologies are shown in Fig. 82-1.

■ MOLECULAR BIOLOGY

Clinically, urothelial carcinoma of the bladder displays a biphasic phenotype characterized by (1) low-grade papillary tumors that frequently recur but rarely invade or metastasize and (2) high-grade sometimes flat tumors that invade early leading to lethal metastatic disease. In both of these phenotypes, loss of portions of chromosomes 9q and 9p by loss of heterozygosity analyses is an early molecular event, whose exact significance is not clear. Potential candidate regulatory genes in these

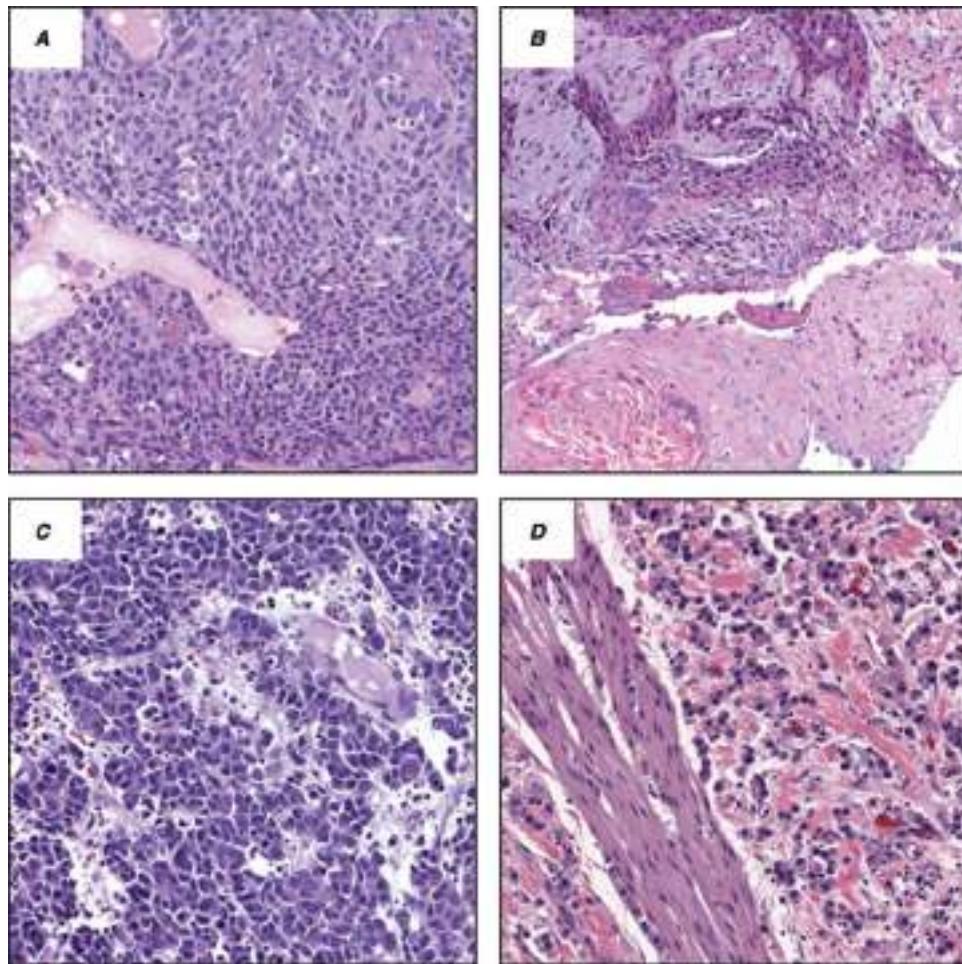


FIGURE 82-1 Bladder and urinary tract cancer histologies. **A.** Urothelial carcinoma; **B.** squamous cell carcinoma; **C.** small-cell carcinoma; **D.** plasmacytoid variant. (Courtesy of Alex Baras, MD, PhD, Johns Hopkins University Department of Pathology.)

genomic regions include *CDNK2A* and *TSC1*. Early investigations have demonstrated that low-grade tumors are characterized by alterations in the *RAS/RAF* signaling pathway with activating *FGFR3* mutations or gene fusions present in 60–80% of patients. In contrast, the high-grade invasive phenotype is notable for early deleterious mutations in *TP53* and *RB1*, alterations in *CDH1*, and increased expression of *VEGFR2*. In urothelial carcinoma of the renal pelvis and ureter, 10–20% of cases may be associated with Lynch syndrome hereditary defects in the *MLH1*, *MSH2*, or *MSH6* mismatch repair genes leading to microsatellite instability and frequent DNA mutations. Testing for germline mutations in these genes is recommended in patients with upper urinary tract urothelial carcinoma under the age of 60 at diagnosis, with a first-degree relative with a Lynch syndrome-associated cancer diagnosed under the age of 50, or with two first-degree relatives with a Lynch syndrome-associated cancer regardless of the age at diagnosis.

As genomic analysis technologies have improved, so has our understanding of the molecular biology unique to urothelial carcinoma. In 2014, the initial bladder cancer results of The Cancer Genome Atlas (TCGA) project were published. This effort comprehensively analyzed gene mutations, fusions, expression, copy number variations, methylation, and microRNA across the genome of patients with bladder urothelial carcinoma treated with surgery. While this data set will continue to be analyzed for years to come, the initial findings include (1) genomic alterations in genes (e.g., *FGFR3*, *EGFR*, *ERBB2*, *ERBB3*, *PIK3CA*, *TSC1*, etc.) targetable by currently approved drugs or drugs in development in 69% of patients; (2) genomic alterations in chromatin modifying genes (*KDM6A*, *MLL2*, *CREBBP*, *EP300*, etc.) in 89% of patients; (3) hypermethylation of tumor suppressor genes in 34% of patients; and (4) the identification by RNA sequencing of four distinct intrinsic molecular subtypes (luminal 1, luminal 2, basal 3, and basal 4) closely resembling luminal and basal sub-classifications of breast cancers. These initial bladder TCGA findings have led to clinical trial designs enriching for patients with specific gene mutation profiles as well as interrogation of candidate biomarkers according to intrinsic molecular subtypes.

■ STAGING AND OUTCOMES BY STAGE

The staging of bladder cancer is dependent on the depth of invasion within the bladder wall, involvement of lymph nodes, and spread to surrounding and distant organs as depicted in Fig. 82-2. Approximately 75% of bladder cancer presents with non-muscle-invasive bladder cancer (NMIBC), 18% with disease invading into or through the muscular wall of the bladder, and only 3% presenting with metastatic spread to distant organs. NMIBC is defined by tumors that involve only the immediate epithelial layer of cells (carcinoma in situ [CIS] and Ta) or that only penetrate into the connective tissue below the urothelium (T1) but not into the muscular layer known as the *muscularis propria*. Muscle-invasive bladder cancer (MIBC) is defined by tumors that invade into the *muscularis propria* (T2), through the *muscularis propria* to involve the surrounding serosa (T3), or into immediately adjacent pelvic organs such as the rectum, prostate, vagina, or cervix (T4). Lymph node staging is classified according to involvement of a solitary node within the true pelvis (N1), multiple nodes involved in the true pelvis (N2), or involvement of the common iliac nodes (N3). Any disease that has spread beyond the true pelvis is considered metastatic (M1). The staging of bladder cancer is driven primarily by the T-stage of the tumor with stages 0a–III defined entirely by the T-stage in the absence of nodal or metastatic disease. Conversely, involvement of either nodal or distant metastases qualifies as stage IV disease. Clinical outcomes of patients with bladder cancer correlate

closely with staging at diagnosis with 5-year overall survival rates of 80% for disease confined to the bladder (stage I-II), 35–50% for disease that penetrates through the bladder (stage III), and only 10–20% for disease extending to surrounding organs, lymph nodes, or metastatic sites (stage IV).

■ TREATMENT APPROACHES

Early-Stage Disease For NMIBC, removal of all visible tumors by TURBT in the operating room is considered the mainstay of surgical treatment. Risk of recurrence can be classified as low, intermediate, or high depending on the presence of features summarized in Table 82-1. For patients with low-risk disease meta-analyses have demonstrated a 12% reduction in early relapses when a single chemotherapy treatment of mitomycin C, epirubicin, or gemcitabine was instilled directly into the bladder (intravesical therapy) within 24 hours of the TURBT. For patients with intermediate- or high-risk tumors, weekly intravesical instillations for 6 consecutive weeks of the attenuated mycobacterium strain known as *Bacille-Calmette Guerin* (BCG) reduce the risk of recurrence at 12 months from 56 to 29%. In addition, BCG treatment has been shown to decrease the rate of progression to MIBC by 27%. Intravesical BCG is generally well tolerated. Side effects can include dysuria, urinary frequency, bladder spasms, hematuria, and, in rare cases (<5%), a systemic inflammatory response that can mimic disseminated BCG infection. Following a 6-week induction BCG schedule, additional maintenance BCG treatments given according to the Southwest Oncology Group schedule further reduce the risk of recurrent NMIBC compared to induction BCG alone. In patients with

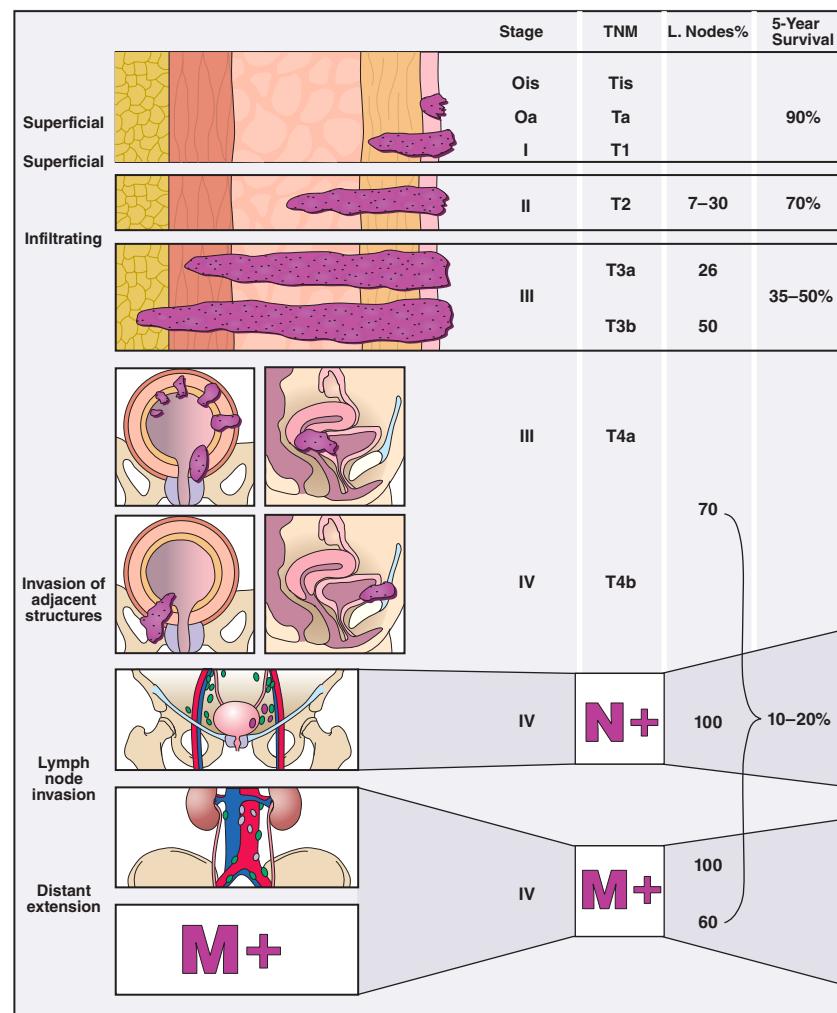


FIGURE 82-2 Bladder cancer staging. TNM, tumor, node metastasis. (Reprinted with permission from HI Scher, JE Rosenberg, RJ Motzer: Bladder and renal cell carcinomas, in DL Kasper et al [eds]: Harrison's Principles of Internal Medicine, 19th ed. New York, McGraw-Hill, Chap. 114, 2015.)

TABLE 82-1 Non-Muscle-Invasive Bladder Cancer Recurrence Risk Groups

RISK GROUP	CHARACTERISTICS
Low risk	Initial tumor, solitary tumor, low grade, <3 cm, no CIS
Intermediate risk	All tumors not defined in the two adjacent categories (between the category of low and high risk)
High risk	Any of the following: <ul style="list-style-type: none"> • T1 tumor • High-grade • CIS • Multiple and recurrent and large (>3 cm) Ta low-grade tumors (all conditions must be met for this point on Ta low-grade tumors)

Abbreviation: CIS, carcinoma in situ.

NMIBC that recurs long after initial BCG treatment, a repeat course of BCG can be considered. For patients with recurrence after a second induction course of BCG or with relapsed NMIBC within 6 months of initial BCG exposure, surgical removal of the entire bladder by cystectomy is recommended due to the high risk of progression to MIBC and potentially metastatic disease. For patients who are not fit enough for or who refuse cystectomy, non-BCG alternative intravesical agents (mitomycin C, gemcitabine, docetaxel, valrubicin) can achieve temporary tumor responses.

In patients with urothelial carcinoma of the renal pelvis or ureter, endoscopic tissue acquisition and staging are more challenging than primary tumors located in the bladder. Tumors possessing all of the following are considered low risk: solitary tumor, low grade, size <1 cm, no invasive component on imaging. Low-risk tumors can successfully be treated by laser ureteroscopic ablation or surgical resection and reanastomosis of the remaining ureter ends in tumors that cannot be successfully eradicated endoscopically.

Muscle-Invasive Disease In patients with urothelial carcinoma of the bladder that invades into or through the muscularis propria but with no evidence of metastatic spread, more aggressive therapy options summarized in **Table 82-2** are required to achieve cure. In carefully selected patients with no evidence of CIS or hydronephrosis, bladder-sparing combined modality therapy with concurrent chemotherapy and radiation can achieve cure in ~65% of patients. Various chemotherapy regimens have been utilized in combination with radiation including cisplatin, carboplatin, 5-fluorouracil, mitomycin C, paclitaxel, and gemcitabine. It is important to note that a maximal debulking of all visible tumor by TURBT is required prior to initiation of combined modality therapy. In patients who achieve a complete response to combined modality therapy, regular cystoscopic

monitoring of the bladder is required with salvage cystectomy offered to patients who develop MIBC in follow-up.

In a similar fashion, bladder-sparing partial cystectomy can be performed in a very small subset of MIBC patients. The ideal patient for partial cystectomy is the patient with a solitary, clinical T2 urothelial carcinoma in the dome of the bladder. In such patients, the tumor and immediate surrounding urothelium can be resected with reconstruction of the remaining bladder to maintain near physiologic urinary function.

In the majority of patients, however, resection of the entire bladder is required. In males, a cystoprostatectomy with removal of the bladder, prostate, and pelvic lymph nodes is performed while in females an anterior exenteration with removal of the bladder, uterus, ovaries, cervix, and pelvic lymph nodes is performed. With the bladder removed, three options exist to re-route the urine outflow. In an ileostomy, the bilateral ureters are connected to a portion of ileum that is brought through an incision in the abdominal wall to create a stoma that drains urine into an affixed bag outside of the body. In a continent urinary reservoir or “Indiana pouch,” the ureters are connected to a portion of ileum that has been separated on both ends from the rest of the small bowel transit to form a urinary reservoir. The remaining small bowel is reanastomosed, and the urinary reservoir is brought up just beneath the abdominal wall muscles with patients catheterizing the urinary reservoir several times per day via a small stoma tract. Last, in a neobladder, the same urinary reservoir described previously is brought down into the pelvis and is anastomosed to the remaining urethra to provide the opportunity to the patient to void urine through the urethra. The choice of which urinary reconstruction to perform is affected not only by patient choice but also by anatomic tumor considerations and urologist experience with each procedure. Regardless of the type of surgery performed, all patients undergo a significant catabolic change in their metabolism following removal of the bladder. While many MIBC patients are affected by weight loss preoperatively, it is not uncommon for postcystectomy patients to lose an additional 10–15 lb in the first month postoperatively. In addition, patients can experience long-term nutritional changes such as low B₁₂ levels due to alterations in small bowel physiology caused by all of the urinary diversion options.

Despite aggressive surgery, only half of patients undergoing cystectomy are cured by surgery alone. Therefore, many clinical trials have investigated the role of systemic chemotherapy before (neoadjuvant) or after (adjuvant) surgery. Meta-analyses have shown a 5–10% absolute overall survival advantage when combination chemotherapy regimens utilizing cisplatin have been used before surgery. A similar benefit exists with cisplatin-based combination chemotherapy given after surgery. However, the data in the adjuvant setting are based on smaller, older trials. Furthermore, in the postoperative setting, some patients may not recover sufficiently from their surgery within a time frame optimal for chemotherapy administration. Importantly, non-cisplatin-containing chemotherapy regimens have proven inferior to cisplatin-containing regimens. Therefore, if patients are not suitable candidates for cisplatin administration due to poor functional status or comorbidities (e.g., poor renal function), patients should proceed directly to surgery and forego neoadjuvant therapy.

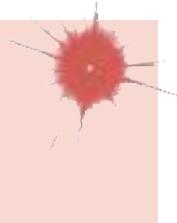
For patients with high-risk urothelial carcinoma of the upper urinary tract, resection of the kidney and ureter (including the ureter bladder cuff) by nephroureterectomy is preferred. Segmental ureterectomy may be appropriate in patients with decreased renal function in which nephron-sparing outcomes are critical to prevent the need for dialysis. Similarly, in CIS patients, administration of BCG therapy via a nephrostomy tube can be considered to preserve intact renal function. In retrospective series, the use of cisplatin-based neoadjuvant chemotherapy has been associated with a pathologic complete response at surgery of 14% in upper tract urothelial carcinoma patients. While adjuvant chemotherapy can be considered in patients with locally advanced stages (T3, T4, or node positive), due to the removal of a kidney at surgery and subsequent associated drop in renal function, many postoperative upper tract patients cannot receive cisplatin-based regimens.

TABLE 82-2 Treatment Approaches to MIBC Patients

TREATMENT	PATIENT SELECTION	CLINICAL OUTCOMES
Bladder-sparing chemoradiation	No CIS, no hydronephrosis, maximal TURBT required	65% cure, 55% bladder intact, highly dependent on patient selection
Bladder-sparing partial cystectomy	Solitary tumors in dome of bladder are ideal	Variable, highly dependent on patient selection
Cystectomy	Any MIBC patient	50% cure with surgery alone, highly dependent on pathologic stage
Neoadjuvant cisplatin-based chemotherapy	Cisplatin-eligible MIBC patients	5–10% improvement in overall survival compared to cystectomy alone
Adjuvant cisplatin-based chemotherapy	Cisplatin-eligible high-risk postcystectomy MIBC patients (pT3-4, N+)	Similar improvement as neoadjuvant treatment, data less robust, many patients not suitable for adjuvant treatment

Abbreviations: CIS, carcinoma in situ; MIBC, muscle-invasive bladder cancer; TURBT, transurethral resection of bladder tumor.

Metastatic Disease For patients with metastatic urothelial carcinoma regardless of primary tumor origin, systemic chemotherapy is



the most established standard of care. In a randomized phase 3 clinical trial, the combination of methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC) demonstrated an improvement in median overall survival from 8.2 to 12.5 months compared to single-agent cisplatin. In a head-to-head randomized phase 3 clinical trial, the combination of cisplatin and gemcitabine (CG) demonstrated similar overall survival compared to MVAC with a more favorable side-effect profile. Since 2000, treatment with either MVAC or CG has remained the standard for first-line treatment of patients with metastatic urothelial carcinoma with adequate renal function and functional status suitable for cisplatin therapy. For patients with lymph node only metastases and good functional status, cure is achieved in 15–20% of such patients. Unfortunately, only ~5% of metastatic patients fulfill both these criteria. For most patients, chemotherapy may prolong survival, but disease resistance proving lethal eventually develops. Furthermore, approximately half of patients with urothelial carcinoma have renal insufficiency, comorbidities, or frail functional status, and are not candidates for cisplatin treatment. In cisplatin-ineligible patients, carboplatin-based chemotherapy regimens are most often used with median overall survival rates decreased to 9.3 months.

Following front-line chemotherapy treatment, second-line chemotherapy regimens have shown modest 10–20% response rates, but no overall survival benefit. In recent years, exponential development of novel immunotherapy approaches has occurred for patients with urothelial carcinoma. The immune checkpoint targets programmed cell death protein 1 (PD-1) and programmed death ligand 1 (PD-L1) have demonstrated the most encouraging clinical benefits. In normal physiology, PD-1/PD-L1 are upregulated in response to inflammation to dampen and prevent an overactive inflammatory response. In cancers including urothelial carcinoma, however, PD-1/PD-L1 are often upregulated on the tumor surface or immune cells in the tumor microenvironment. Upregulated PD-1/PD-L1 in this situation serves as a mechanism of immune escape that facilitates tumor growth. Atezolizumab (an anti-PD-L1 antibody) was the first drug approved in the United States for metastatic urothelial carcinoma in over two decades based on a response rate of 15% in postplatinum patients. Subsequently, pembrolizumab (an anti-PD-1 antibody) demonstrated an improvement in overall survival from 7.4 to 10.3 months compared to standard second-line chemotherapy options. Multiple other PD-1/PD-L1 agents have demonstrated clinical responses in urothelial carcinoma. In addition, clinical trials investigating immunotherapy, chemotherapy, and radiation combinations are ongoing. Last, leveraging the molecular knowledge gained from the TCGA project, clinical trials are also investigating the role of molecularly targeted therapies in patients with metastatic urothelial carcinoma harboring specific genetic alterations predictive of clinical benefit (e.g., activating FGFR3 mutations or gene fusions). Collectively, these new emerging options for metastatic urothelial carcinoma patients offer hope for improved outcomes for patients with urothelial carcinoma of all stages in the future.

FURTHER READING

- BABJUK M et al: EAU Guidelines on Non-Muscle-Invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur Urol* 2016; Available from <http://dx.doi.org/10.1016/j.eururo.2016.05.041>. Accessed December 3, 2016.
- CANCER GENOME RESEARCH ATLAS COLLABORATORS: Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507:315, 2014.
- HOWLADER N et al: SEER Cancer Statistics Review, 1975–2013. Available from <https://seer.cancer.gov/statfacts/html/urinb.html>. Accessed December 3, 2016.
- KAMAT AM et al: Bladder cancer. *Lancet* 388:2796, 2016.
- KNOWLES MA, HURST CD: Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat Rev Cancer* 15:25, 2015.
- ROUPRÉT M et al: European Association of Urology Guidelines on Upper Urinary Tract Urothelial Cell Carcinoma: 2015 update. *Eur Urol* 68:868, 2015.
- SIEGEL RL et al: Cancer statistics, 2016. *CA Cancer J Clin* 66:7, 2016.

Benign and malignant changes in the prostate increase with age. Autopsies of men in the eighth decade of life show hyperplastic changes in >90% and malignant changes in >70% of individuals. The high prevalence of these diseases among the elderly, who often have competing causes of morbidity and mortality, mandates a risk-adapted approach to diagnosis and treatment. This can be achieved by considering these diseases as a series of states. Each state represents a distinct clinical milestone for which therapy(ies) may be recommended based on disease extent, current symptoms, the risk of developing symptoms, or the risk of death from disease in relation to death from other causes within a given time frame. For benign proliferative disorders, symptoms of urinary frequency, infection, and potential for obstruction are weighed against the side effects and complications of medical or surgical intervention. For prostate malignancies, the likelihood that a clinically significant cancer is present in the gland and the concomitant risk of symptoms or death from cancer are balanced against the morbidities of the recommended treatments and preexisting comorbidities.

ANATOMY AND PATHOLOGY

The prostate is located in the pelvis and is surrounded by the rectum, the bladder, the periprostatic and dorsal vein complexes and neurovascular bundles that are responsible for erectile function, and the urinary sphincter that is responsible for passive urinary control. The prostate is composed of branching tubuloalveolar glands arranged in lobules surrounded by fibromuscular stroma. The acinar unit includes an epithelial compartment made up of epithelial, basal, and neuroendocrine cells and separated by a basement membrane, and a stromal compartment that includes fibroblasts and smooth-muscle cells. Prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP) are produced in the epithelial cells. Both prostate epithelial cells and stromal cells express androgen receptors (ARs) and depend on androgens for growth. Testosterone, the major circulating androgen, is converted by the enzyme 5 α -reductase to dihydrotestosterone in the gland.

The periurethral portion of the gland increases in size during puberty and after the age of 55 years due to the growth of nonmalignant cells in the transition zone of the prostate that surrounds the urethra. Most cancers develop in the peripheral zone, and cancers in this location may be palpated during a digital rectal examination (DRE).

PROSTATE CANCER

In 2017, ~161,360 prostate cancer cases were diagnosed and 26,730 men died from prostate cancer in the United States. The absolute number of prostate cancer deaths has decreased in the past 10 years, attributed by some to the widespread use of PSA-based detection strategies. However, the paradox of management is that although 1 in 6 men will eventually be diagnosed with prostate cancer, and the disease remains the second leading cause of cancer deaths in men, only 1 man in 30 with prostate cancer will die of his disease.

EPIDEMIOLOGY

Epidemiologic studies show that the risk of being diagnosed with prostate cancer increases 2.5-fold if one first-degree relative is affected and fivefold if two or more are affected. Current estimates are that 40% of early-onset and 5–10% of all prostate cancers are hereditary. Prostate cancer affects ethnic groups differently. Matched for age, African-American males have a higher incidence and present at a more advanced stage with higher-grade, more aggressive tumors. A high risk in families has been linked to the HPC1 (hereditary prostate cancer 1) susceptibility locus in *RNASEL*. Genome-wide association studies (GWAS) have identified >40 prostate cancer susceptibility loci that are estimated to explain up to 25% of prostate cancer risk. Among

the genes implicated in variations in incidence and outcome are single-nucleotide polymorphisms (SNPs) in the vitamin D receptor in African-Americans and variants in the AR, CYP3A4, both involved in the deactivation of testosterone, as well as CYP17, which is involved in steroid biosynthesis.

The prevalence of autopsy-detected cancers is similar around the world, while the incidence of clinical disease varies. Thus, environmental and dietary factors may play a role in prostate cancer growth and progression. High consumption of dietary fats, such as α -linoleic acid or polycyclic aromatic hydrocarbons that form when red meats are cooked, is believed to increase risk. Similar to breast cancer in Asian women, the risk of prostate cancer in Asian men increases when they move to Western environments. Protective factors include consumption of the isoflavanoid genistein (which inhibits 5 α -reductase), cruciferous vegetables with isothiocyanate sulforaphane, lycopene found in tomatoes, and inhibitors of cholesterol biosynthesis (e.g., statin drugs). The development of prostate cancer is a multistep process. One early change is hypermethylation of the GSTP1 gene promoter, which leads to loss of function of a gene that detoxifies carcinogens. The finding that many prostate cancers develop adjacent to a lesion termed PIA (proliferative inflammatory atrophy) suggests a role for inflammation. Not smoking, regular exercise, and maintaining a healthy body weight may reduce the risk of progression.

■ DIAGNOSIS AND TREATMENT BY CLINICAL STATE

The prostate cancer continuum—from the appearance of a preneoplastic and invasive lesion that is localized to the gland, to a metastatic lesion causing symptoms and, ultimately, mortality—can span decades. To limit overdiagnosis of clinically insignificant cancers, and for disease management in general, competing risks are considered in the context of a series of clinical states (Fig. 83-1). The states are defined operationally on the basis of whether or not a cancer diagnosis has been established and, for those with a diagnosis, whether or not metastases are detectable on imaging studies and the measured level of testosterone in the blood. With this approach, an individual resides in only one state and remains in that state until he has progressed. At each assessment, the decision to offer treatment and the specific form of treatment are based on the risk posed by the cancer relative to competing causes of morbidity and mortality that may be present in that individual. It follows that the more advanced the disease, the greater the need for treatment.

For those without a cancer diagnosis, the decision to undergo testing to detect a cancer is based on the individual's estimated life expectancy and, separately, the probability that a clinically significant cancer may

be present. For those with a prostate cancer diagnosis, the clinical states model considers the probability of developing symptoms or dying from the disease. Thus, a patient with localized tumor that has been surgically removed remains in the state of localized disease as long as the PSA remains undetectable. The time within a state then becomes a measure of the efficacy of an intervention, though the effect may not be assessable for years. Because many men with active cancer are not at risk for developing metastases, symptoms, or death, the clinical states model allows a distinction between *cure*—the elimination of all cancer cells, the primary therapeutic objective of treatment for most cancers—and *cancer control*, by which the tempo of the illness is determined to be so slow or has been altered to the point where it is unlikely to cause symptoms, to metastasize, or to shorten a patient's life expectancy. Importantly, from a patient standpoint, both outcomes can be considered equivalent therapeutically, assuming the patient has not experienced symptoms of the disease or the treatment needed to control it. Even when a recurrence is documented, immediate therapy is not always necessary. Rather, as at the time of diagnosis, the need for intervention is based on the tempo of the illness as it unfolds in the individual, relative to the risk-to-benefit ratio of the intervention being considered.

■ NO CANCER DIAGNOSIS

Prevention No agent is currently approved for the prevention of prostate cancer. The results from several large double-blind, randomized chemoprevention trials have established 5 α -reductase inhibitors (5ARI) as the predominant therapy to reduce the future risk of a prostate cancer diagnosis. The Prostate Cancer Prevention Trial (PCPT), in which men aged >55 years received placebo or the 5ARI finasteride, which inhibits the type 1 isoform, showed a 25% (95% confidence interval 19–31%) reduction in prostate cancer incidence from 24% with placebo to 18% with finasteride. In REDUCE (Reduction by Dutasteride of Prostate Cancer Events trial), a reduction in incidence from 25% with placebo to 20% with dutasteride was found ($p = 0.001$). Dutasteride inhibits both the type 1 and type 2 5ARI isoforms. While both studies met their endpoint, there was concern that most of the cancers that were prevented were low-risk and that there was a slightly higher rate of clinically significant cancers (those with higher Gleason score) in the treatment arm. Neither drug is approved for prostate cancer prevention. In comparison, the Selenium and Vitamin E Cancer Prevention Trial (SELECT), which enrolled African-American men aged ≥ 50 years and others aged ≥ 55 years, showed no difference in cancer incidence in patients receiving vitamin E (4.6%) or selenium (4.9%) alone or in combination (4.6%) relative to placebo (4.4%). A similar lack of benefit for vitamin E, vitamin C, and selenium was seen in the Physicians Health Study II.

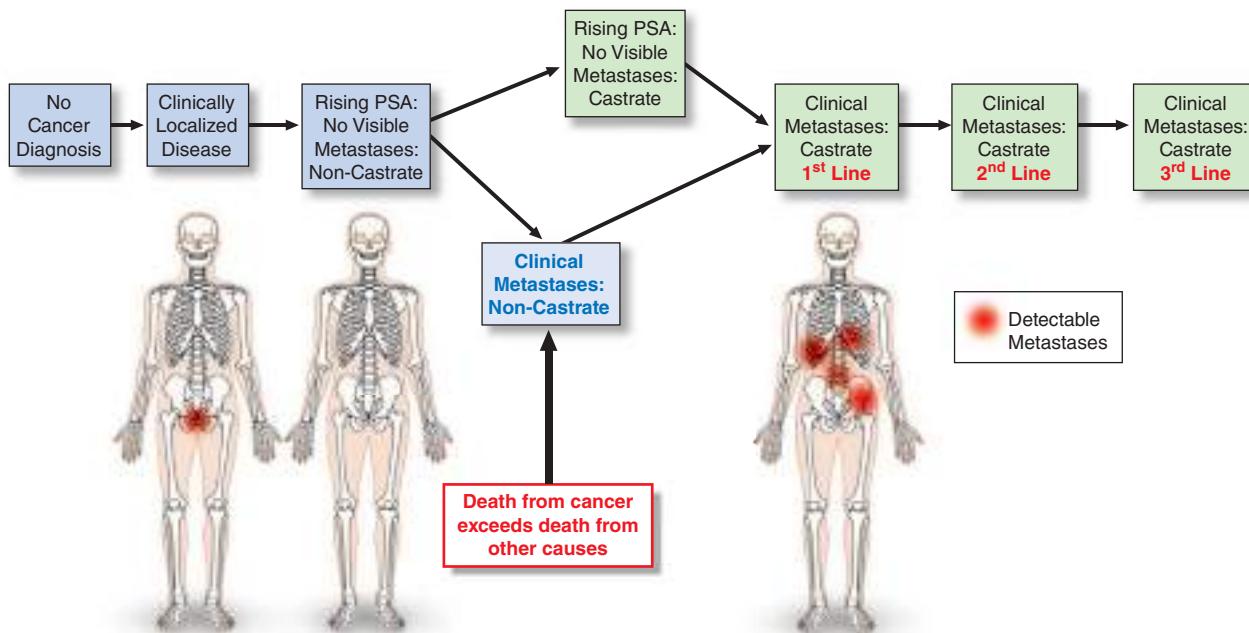


FIGURE 83-1 Clinical states of prostate cancer. PSA, prostate-specific antigen.

Screening/Early Detection and Diagnosis The need to pursue a diagnosis of prostate cancer must balance the benefit from detecting and treating clinically significant cancers that, left untreated, would adversely affect patients' quality and duration of life against the morbidity associated with overdiagnosis and overtreatment of clinically insignificant cancers that are highly prevalent in the general population. The balance is best approached through shared decision making between the patient and physician. Considerations for whether to pursue a diagnosis include symptoms, an abnormal DRE, or more typically, a change in or an elevated serum PSA. Genetic risk is also considered. The urologic history should focus on symptoms of outlet obstruction, continence, potency, or change in ejaculatory pattern.

PHYSICAL EXAMINATION The DRE focuses on prostate size and consistency and abnormalities within or beyond the gland. Many cancers occur in the peripheral zone and may be palpated on DRE. Carcinomas are characteristically hard, nodular, and irregular, while induration may also be due to benign prostatic hyperplasia (BPH) or calculi. Overall, 20–25% of men with an abnormal DRE have prostate cancer.

PROSTATE-SPECIFIC ANTIGEN PSA (kallikrein-related peptidase 3; *KLK3*) is a kallikrein-related serine protease that causes liquefaction of seminal coagulum. It is produced by both nonmalignant and malignant epithelial cells and, as such, is prostate-specific, not prostate cancer-specific. Serum levels may also increase from prostatitis and BPH. Serum levels are not significantly affected by DRE, but the performance of a cystoscopy or prostate biopsy can increase PSA levels up to tenfold for 8–10 weeks. PSA circulating in the blood is inactive and mainly occurs as a complex with the protease inhibitor α_1 -antichymotrypsin and as free (unbound) PSA forms. The formation of complexes between PSA, α_1 -macroglobulin, or other protease inhibitors is less significant. Free PSA is rapidly eliminated from the blood by glomerular filtration with an estimated half-life of 12–18 h. Elimination of PSA bound to α_1 -antichymotrypsin is slow (estimated half-life of 1–2 weeks) as it too is largely cleared by the kidneys. Levels should be undetectable after about 6 weeks if the prostate has been completely removed (radical prostatectomy). Immunohistochemical staining for PSA can be used to establish a prostate cancer diagnosis.

PSA testing was approved by the U.S. Food and Drug Administration (FDA) in 1994 for early detection of prostate cancer, and the widespread use of the test has played a significant role in the proportion of men diagnosed with early-stage cancers: more than 70–80% of newly diagnosed cancers are clinically organ confined. The level of PSA in blood is strongly associated with the risk and outcome of prostate cancer. A single PSA measured at age 60 is associated (area under the curve [AUC] of 0.90) with lifetime risk of death from prostate cancer. Most (90%) prostate cancer deaths occur among men with PSA levels in the top quartile (>2 ng/mL), although only a minority of men with PSA >2 ng/mL will develop lethal prostate cancer. Despite this and mortality rate reductions reported from large randomized prostate cancer screening trials, routine use of the test remains controversial.

In 2012, the United States Preventive Services Task Force (USPSTF) published a review of the evidence for PSA-based screening for prostate cancer and made a clear recommendation against screening. By giving a grade of "D" in the recommendation statement that was based on this review, the USPSTF concluded that "there is moderate or high certainty that this service has no net benefit or that the harms outweigh the benefits." In 2013, the American Urological Association (AUA) updated their consensus statement regarding prostate cancer screening. They concluded that the quality of evidence for the benefits of screening was moderate for men aged 55–69 years. For men outside this age range, evidence was lacking for benefit, but the harms of screening, including overdiagnosis and overtreatment, remained. The AUA recommends shared decision-making for men aged 55–69 years considering PSA-based screening, a target age group for whom benefits may outweigh harms. Outside this age range, PSA-based screening as a routine was not recommended. The entire guideline is available at [http://www.auanet.org/guidelines/early-detection-of-prostate-cancer-\(2013-reviewed-and-validity-confirmed-2015\)](http://www.auanet.org/guidelines/early-detection-of-prostate-cancer-(2013-reviewed-and-validity-confirmed-2015)). As of 2017, the USPSTF has issued a draft of a revised recommendation with a grade of "C" for PSA-based

prostate cancer screening for men aged 55–69. They recommend shared decision-making for men aged 55–69 and do not recommend screening for men aged ≥ 70 ; this is now roughly in agreement with the 2013 AUA guideline. The USPSTF notes that the increased use of active surveillance (observation with selective delayed treatment) for low-risk prostate cancer has reduced the risks of screening.

We believe that implementation of the following three guidelines will further improve PSA screening outcomes in the United States and will have a greater practical impact on men's health than the USPSTF and AUA recommendations that are based almost solely on age. First, avoid PSA tests in men with little to gain. There is no rationale for recommending PSA screening in asymptomatic men with a short life expectancy. Hence, men aged >75 years should only be tested in special circumstances, such as higher than median PSAs measured before age 70 or excellent overall health. In addition, because a baseline PSA is a strong predictor of the future risk of lethal prostate cancer, men with low PSAs, for example <1 ng/mL, can undergo testing less frequently, perhaps every 5 years, with screening possibly ending at age 60 if the PSA remains at ≤ 1 ng/mL. Men with PSAs that are above age median but below biopsy thresholds can be counseled about their elevated risk and actively encouraged to return for regular screening and more comprehensive risk assessment. Second, do not treat those who do not need treatment. High proportions of men with screen-detected prostate cancer do not need immediate treatment and can be managed by active surveillance. Third, refer men who do need treatment to high-volume centers. Although it is clearly not feasible to restrict treatment exclusively to high-volume centers, shifting treatment trends so that more patients are treated at such centers by high-volume providers will improve cancer control and decrease complications. The goal of prostate cancer screening should be to maximize the benefits of PSA testing and minimize its harms. Following the three rules outlined here should continue to improve the ratio of harms to benefits from PSA screening.

The PSA criteria used to recommend a diagnostic prostate biopsy have evolved over time. However, based on the commonly used cut-point for prostate biopsy (a total PSA ≥ 4 ng/mL), most men with a PSA elevation do not have histologic evidence of prostate cancer at biopsy. In addition, many men with PSA levels below this cut-point harbor cancer cells in their prostate. Information from the Prostate Cancer Prevention Trial demonstrates that there is no PSA below which the risk of prostate cancer is zero. Thus, the PSA level establishes the likelihood that a man will harbor cancer if he undergoes a prostate biopsy. The goal is to increase the sensitivity of the test for younger men more likely to die of the disease and to reduce the frequency of detecting cancers of low malignant potential in elderly men more likely to die of other causes. Patients with symptomatic prostatitis should have a course of antibiotics before biopsy. However, the routine use of antibiotics in an asymptomatic man with an elevated PSA level is strongly discouraged.

SECOND-LINE SCREENING TESTS The 4Kscore® Test (OPKO Lab, Nashville, TN) measures four prostate-specific kallikreins (total PSA, free PSA, intact PSA, and human kallikrein 2). The results are combined with clinical information in an algorithm that estimates an individual's percent risk of being found to harbor an aggressive prostate cancer should that individual opt for a prostate biopsy. The 4Kscore test has also been shown to identify the likelihood that an individual will develop aggressive prostate cancer, defined as high grade prostate cancer pathology and/or poor prostate cancer clinical outcomes, within 20 years.

Prostate Health Index (PHI™, Innovative Diagnostic Laboratory, Richmond, VA) is a blood test that estimates the risk of having prostate cancer. The PHI test is a combination of the free PSA, total PSA, and the [-2]proPSA isoform of free PSA. These three tests are combined in a formula that calculates the PHI score. The PHI score is a better predictor of prostate cancer than the total PSA test alone or the free PSA test alone.

PROSTATE BIOPSY A diagnosis of cancer is established by an image-guided needle biopsy. Direct visualization by transrectal ultrasound (TRUS), magnetic resonance imaging (MRI), or fusion of the ultrasound and MRI images ensures that all areas of the gland including suspicious areas are sampled. Contemporary schemas advise an extended-pattern 12-core biopsy that includes sampling from the peripheral zone as

well as a lesion-directed palpable nodule or suspicious image-guided sampling. Because a prostate biopsy is subject to sampling error, men with an abnormal PSA and negative biopsy are frequently advised to undergo additional testing which may include a 4Kscore Test, PHI, prostate MRI, and/or repeat biopsy.

PATHOLOGY Each core of the biopsy is examined for the presence of cancer, and the amount of cancer is quantified based on the length of the cancer within the core and the percentage of the core involved. Of the cancers identified, >95% are adenocarcinomas; the rest are squamous or transitional cell tumors or, rarely, carcinosarcomas. Metastases to the prostate are rare, but in some cases colon cancers or transitional cell tumors of the bladder invade the gland by direct extension.

When prostate cancer is diagnosed, a measure of histologic aggressiveness is assigned using the *Gleason grading system*, in which the dominant and secondary glandular histologic patterns are scored from 1 (well-differentiated) to 5 (undifferentiated) and summed to give a total score of 2–10 for each tumor. The most poorly differentiated area of tumor (i.e., the area with the highest histologic grade) often determines biologic behavior. The presence or absence of perineural invasion and extracapsular spread are also recorded.

Over the years, the Gleason grading system has undergone several changes. Currently, Gleason total scores 2–5 are no longer assigned and in practice the lowest total score is now assigned a 6, although the scale continues to range from 2 to 10. This leads to a logical yet incorrect assumption on the part of patients that their Gleason 6 cancer is in the middle of the scale, triggering the fear that their cancer is serious and the assumption that treatment is necessary despite Gleason score 6 actually being favorable risk. To address these issues, a new 5-grade group system has been developed:

- Grade Group 1 (Gleason score ≤6)
- Grade Group 2 (Gleason score 3+4 = 7)
- Grade Group 3 (Gleason score 4+3 = 7)
- Grade Group 4 (Gleason score 4+4 = 8)
- Grade Group 5 (Gleason scores 9 and 10)

The new system simplifies the grading of prostate cancer, appropriately classifies the lowest risk as Grade Group 1 (rather than Gleason score 6), and accurately predicts prognosis.

PROSTATE CANCER STAGING The TNM (tumor, nodes, metastasis) staging system includes categories for cancers that are identified solely on the basis of an abnormal PSA (T1c), those that are palpable but clinically confined to the gland (T2), and those that have extended outside the gland (T3 and T4) (Table 83-1, Fig. 83-2). DRE alone is inaccurate in determining the extent of disease within the gland, the presence or absence of capsular invasion, involvement of seminal vesicles, and extension of disease to lymph nodes. Because of the inadequacy of DRE for staging, the TNM staging system was modified to include the results of imaging. Unfortunately, no single test has proven to accurately indicate the stage or the presence of organ-confined disease, seminal vesicle involvement, or lymph node spread.

TRUS is the imaging technique most frequently used to assess the primary tumor, but its chief use is directing prostate biopsies, not staging. No TRUS finding consistently indicates cancer with certainty. Computed tomography (CT) lacks sensitivity and specificity to detect extraprostatic extension and is inferior to MRI in visualization of lymph nodes. In general, MRI is superior to CT to detect cancer in the prostate and to assess local disease extent. T1-weighted MRI produces a high signal in the periprostatic fat, periprostatic venous plexus, perivesicular tissues, lymph nodes, and bone marrow. T2-weighted MRI demonstrates the internal architecture of the prostate and seminal vesicles. Most cancers have a low signal, while the normal peripheral zone has a high signal, although the technique lacks sensitivity and specificity. MRI is also useful for the planning of surgery and radiation therapy.

Radionuclide bone scans (bone scintigraphy) are used to evaluate spread to osseous sites. This test is sensitive but relatively nonspecific because areas of increased uptake are not always related to metastatic disease. Healing fractures, arthritis, Paget's disease, and other conditions will also cause abnormal uptake. True-positive bone scans are uncommon when the PSA is <10 ng/mL unless the tumor is high-grade.

TABLE 83-1 TNM Classification

TNM (tumor, nodes, metastasis) Staging System for Prostate Cancer ^a	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Localized Disease	
T1	Clinically inapparent tumor, neither palpable nor visible by imaging
T1a	Tumor incidental histologic finding in ≤5% of resected tissue; not palpable
T1b	Tumor incidental histologic finding in >5% of resected tissue
T1c	Tumor identified by needle biopsy (e.g., because of elevated PSA)
T2	Tumor confined within prostate ^b
T2a	Tumor involves half of one lobe or less
T2b	Tumor involves more than one half of one lobe, not both lobes
T2c	Tumor involves both lobes
Local Extension	
T3	Tumor extends through the prostate capsule ^c
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumor invades seminal vesicles
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall
Metastatic Disease	
N1	Positive regional lymph nodes
M1	Distant metastases

^aRevised from SB Edge et al (eds): AJCC Cancer Staging Manual, 7th ed. New York, Springer, 2010. ^bTumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c. ^cInvasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.

Abbreviation: PSA, prostate-specific antigen.

TREATMENT

LOCALIZED DISEASE OR CLINICALLY LOCALIZED DISEASE

Cancers are those that appear to be nonmetastatic after staging studies are performed. Patients with clinically localized disease are managed by radical prostatectomy, radiation therapy, or active surveillance. Choice of therapy requires the consideration of several factors: the presence of symptoms, the probability that the untreated tumor will adversely affect the quality or duration of survival and thus require treatment, and the probability that the tumor can be cured by single-modality therapy directed at the prostate versus requiring both local and systemic therapy to achieve cure.

Data from the literature (such as the ProtecT trial) do not provide clear evidence for the superiority of any one form of local therapy relative to another. This is due to the lack of prospective randomized trials, referral bias and physician bias, variation in the experience of the treating teams, and differences in trial endpoints and the definitions of cancer control. Often, PSA relapse-free survival is used because an effect on metastatic progression or survival may not be apparent for years. For many patients, however, a PSA recurrence does not necessarily mean that the disease will cause symptoms or shorten survival. After radical surgery to remove all prostate tissue, PSA should become undetectable in the blood within 6 weeks. If PSA remains or becomes detectable after radical prostatectomy, the patient is considered to have persistent or recurrent disease. After radiation therapy, in contrast, PSA does not become undetectable because the remaining nonmalignant elements of the gland continue to produce PSA even if all cancer cells have been eliminated. Similarly, cancer control is not well defined for a patient managed by active surveillance because PSA levels may continue to rise in the absence of therapy. Other outcomes are time to objective progression (local or systemic), cancer-specific survival, and overall survival; however, these outcomes may take years to assess.

The more extensive the local disease, the higher the probability of regional lymph node involvement even when imaging studies are

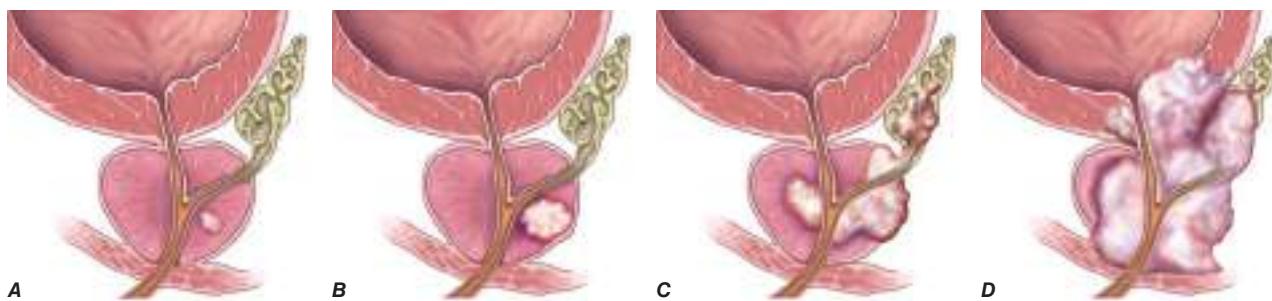


FIGURE 83-2 T stages of prostate cancer. **A.** T1—Clinically inapparent tumor, neither palpable nor visible by imaging; **B.** T2—Tumor confined within prostate; **C.** T3—Tumor extends through prostate capsule and may invade the seminal vesicles; **D.** T4—Tumor is fixed or invades adjacent structures. Eighty percent of patients present with local disease (T1 and T2), which is associated with a 5-year survival rate of 100%. An additional 12% of patients present with regional disease (T3 and T4 without metastases), which is also associated with a 100% survival rate after 5 years. Four percent of patients present with distant disease (T4 with metastases), which is associated with a 30% 5-year survival rate. (Three percent of patients are ungraded.) (Data from AJCC, <http://seer.cancer.gov/statfacts/html/prost.html>. Figure © Memorial Sloan-Kettering Cancer Center Medical Graphics; used with permission.)

normal, the lower the probability of local control, and the higher the probability of systemic relapse. More important is that within the categories of T1, T2, and T3 disease are cancers with a range of prognoses. Some T3 tumors are curable with therapy directed solely at the prostate, and some T1 lesions have a high probability of systemic relapse that requires the integration of local and systemic therapy to achieve cure. For T1c cancers in particular, stage alone is inadequate to predict outcome and select treatment; other factors must be considered.

To better assess risk and guide treatment selection, many groups have developed prognostic models or nomograms that use a combination of the initial clinical T stage, biopsy Gleason score, the number of biopsy cores in which cancer is detected, and baseline PSA. Some use discrete cut-points (PSA <10 or ≥10 ng/mL; Gleason score of ≤6, 7, or ≥8); others employ nomograms that use PSA and Gleason score as continuous variables. More than 100 nomograms have been reported to predict (a) the probability that a clinically significant cancer is present, (b) disease extent (organ-confined vs non-organ-confined, node-negative or -positive), or (c) the probability of treatment success for specific local therapies using pretreatment variables. Considerable controversy exists over what constitutes “high risk” based on a predicted probability of success or failure. In these situations, nomograms and predictive models can only go so far. Exactly what probability of success or failure would lead a physician to recommend and a patient to seek alternative approaches is controversial. As an example, it may be appropriate to recommend radical surgery for a younger patient with a low probability of cure. Nomograms are being refined continually to incorporate additional clinical parameters, biologic determinants, and year of treatment, which can also affect outcomes, making treatment decisions a dynamic process.

The frequency of adverse events varies by treatment modality and the experience of the treating team. For example, following radical prostatectomy, incontinence rates range from 2 to 47% and impotence rates range from 25 to 89%. Part of the variability relates to how the complication is defined and whether the patient or physician is reporting the event. The time of the assessment is also important. After surgery, impotence is immediate but may reverse over time, while with radiation therapy impotence is not immediate but may develop over time. Of greatest concern to patients are the effects on continence, sexual potency, and bowel function.

Radical Prostatectomy The goal of radical prostatectomy is to excise the cancer completely with a clear margin, to maintain continence by preserving the external sphincter, and to preserve potency by sparing the autonomic nerves in the neurovascular bundle. The procedure is advised for patients with a life expectancy of 10 years or more and is performed via a retropubic or perineal approach, or via a minimally invasive robotic-assisted or hand-held laparoscopic approach. Outcomes can be predicted using postoperative nomograms that consider pretreatment factors and the pathologic findings at surgery. PSA failure is usually defined as a value >0.1 or 0.2 ng/mL. Specific criteria to guide the choice of one approach over another are lacking. Minimally invasive approaches offer the advantage of a shorter

hospital stay and reduced blood loss. Rates of cancer control, recovery of continence and recovery of erectile function are comparable. The individual surgeon rather than the surgical approach used is most important in determining outcomes after surgery.

Neoadjuvant hormonal treatment with gonadotropin-releasing hormone (GnRH) agonists/antagonists alone has also been explored in an attempt to improve the outcomes of surgery for high-risk patients using a variety of definitions. The results of several large trials testing 3 or 8 months of androgen depletion before surgery showed that serum PSA levels decreased by 96%, prostate volumes decreased by 34%, and margin positivity rates decreased from 41 to 17%. Unfortunately, these findings have not been shown to improve PSA relapse-free survival.

Factors associated with incontinence following radical prostatectomy include older age and urethral length, which impacts the ability to preserve the urethra beyond the apex and the distal sphincter. The skill and experience of the surgeon are also factors.

The likelihood of recovery of erectile function is associated with younger age, quality of erections before surgery, and the absence of damage to the neurovascular bundles. In general, erectile function begins to return about 6 months after surgery if neurovascular tissue has been preserved. Potency is reduced by half if at least one neurovascular bundle is sacrificed. Overall, with the availability of drugs such as sildenafil, intraurethral inserts of alprostadil, and intracavernosal injections of vasodilators, many patients recover satisfactory sexual function.

Radiation Therapy Radiation therapy is given by external beam, by radioactive sources implanted into the gland, or by a combination of the two techniques.

External beam radiation therapy Contemporary external beam intensity-modulated radiation therapy (IMRT) permits shaping of the dose, and allows the delivery of higher doses to the prostate and a dramatic reduction in normal tissue exposure compared to three-dimensional conformal treatment alone. These advances have enabled the safe administration of doses >80 Gy and resulted in higher local control rates and fewer side effects.

Cancer control after radiation therapy has been defined by various criteria, including a decline in PSA to <0.5 or 1 ng/mL, “nonrising” PSA values, and a negative biopsy of the prostate 2 years after completion of treatment. The current standard definition of biochemical failure (the Phoenix definition) is a rise in PSA by ≥2 ng/mL higher than the lowest PSA achieved. The date of failure is “at call” and not backdated.

Radiation dose is critical to the eradication of prostate cancer. In a representative study, a PSA nadir of <1.0 ng/mL was achieved in 90% of patients receiving 75.6 or 81.0 Gy vs 76 and 56% of those receiving 70.2 and 64.8 Gy, respectively. Positive biopsy rates at 2.5 years were 4% for those treated with 81 Gy vs 27 and 36% for those receiving 75.6 and 70.2 Gy, respectively.

More recently, hypofractionation schedules, utilizing fewer treatments of higher radiation doses, have been evaluated and shown to provide good cancer control rates based on post-treatment

biopsies showing no evidence of cancer, with no apparent increase in treatment-related morbidity. Hypofractionated treatments can range from as few as 5 treatments to upwards of 26 treatments, both regimens representing substantial reductions in treatment length.

Multiple clinical trials have evaluated the use of androgen deprivation therapy (ADT) in combination with radiation. In patients with intermediate-risk prostate cancer, short-course ADT (6 months), when combined with external beam radiotherapy, has demonstrated significant improvements in overall survival. In patients with high-risk disease, longer courses of ADT (18–36 months) have proven superior to shorter courses and represent the current standard of care when combined with radiotherapy.

Neoadjuvant hormone therapy before radiation therapy is used to decrease the size of the prostate and, consequently, to reduce the exposure of normal tissues to full-dose radiation, to increase local control rates, and to decrease the rate of systemic failure. Short-term hormone therapy can reduce toxicities and improve local control rates, but long-term treatment (2–3 years) is needed to prolong the time to PSA failure and lower the risk of metastatic disease in men with high-risk cancers. The impact on survival has been less clear.

The Prostate Testing for Cancer and Treatment (ProtecT) trial investigated the effects of active monitoring, radical prostatectomy, and radical radiotherapy with hormones on patient-reported outcomes in men diagnosed with low- and intermediate-risk prostate cancer (about 75% with Gleason score 6 or Gleason Grade Group 1 cancer). Patient-reported outcomes among 1643 men who completed questionnaires before diagnosis, at 6 and 12 months, and annually thereafter were compared. Of the three treatments, prostatectomy had the greatest negative effect on sexual function and urinary continence, and although there was some recovery, these outcomes remained worse in the prostatectomy group than in the other groups throughout the trial. The negative effect of radiotherapy on sexual function was greatest at 6 months, but sexual function then recovered somewhat and was stable thereafter; radiotherapy had little effect on urinary continence. Sexual and urinary function declined gradually in the active-monitoring group. Bowel function was worse in the radiotherapy group at 6 months than in the other groups but then recovered somewhat, except for the increasing frequency of bloody stools; bowel function was unchanged in the other groups. Urinary voiding and nocturia were worse in the radiotherapy group at 6 months but then mostly recovered and were similar to the other groups after 12 months. Effects on quality of life mirrored the reported changes in function. No significant differences were observed among the groups in measures of anxiety, depression, or general health-related or cancer-related quality of life.

Brachytherapy Brachytherapy is the direct implantation of radioactive sources (seeds) into the prostate. It is based on the principle that the deposition of radiation energy in tissues decreases as a function of the square of the distance from the source (Chap. 69). The goal is to deliver intensive irradiation to the prostate, minimizing the exposure of the surrounding tissues. The current standard technique achieves a more homogeneous dose distribution by placing seeds according to a customized template based on imaging assessment of the cancer and computer-optimized dosimetry. The implantation is performed transperineally as an outpatient procedure with real-time imaging.

Improvements in brachytherapy techniques have resulted in fewer complications and a marked reduction in local failure rates. In a series of 197 patients followed for a median of 3 years, 5-year actuarial PSA relapse-free survival for patients with pretherapy PSA levels of 0–4, 4–10, and >10 ng/mL were 98, 90, and 89%, respectively. In a separate report of 201 patients who underwent posttreatment biopsies, 80% were negative, 17% were indeterminate, and 3% were positive. The results did not change with longer follow-up. Nevertheless, many physicians feel that implantation is best reserved for patients with good or intermediate prognostic features.

Brachytherapy is well tolerated, although most patients experience urinary frequency and urgency that can persist for several months. Incontinence has been seen in 2–4% of cases. Higher complication rates are observed in patients who have undergone a prior

transurethral resection of the prostate (TURP), while those with obstructive symptoms at baseline are at a higher risk for retention and persistent voiding symptoms. Proctitis has been reported in <2% of patients.

Active surveillance Although prostate cancer is the most common form of cancer affecting men in the United States, patients are being diagnosed earlier and more frequently present with early-stage disease. Active surveillance, described previously as *watchful waiting* or *deferred therapy*, evolved from (1) studies that evaluated predominantly elderly men with well-differentiated tumors who demonstrated no clinically significant progression for protracted periods, (2) recognition of the contrast between incidence and disease-specific mortality, (3) the high prevalence of autopsy cancers, and (4) an effort to reduce overtreatment and treatment-related side effects. In practice, active surveillance is the treatment recommended to patients with cancers of low aggressiveness that can be safely monitored at fixed intervals with DREs, PSA measurements, imaging (usually prostate MRI), and repeat prostate biopsies as indicated until histopathologic or serologic changes correlative of progression warrant treatment with curative intent.

Case selection is critical, and determining clinical parameters predictive of cancer aggressiveness that can be used to reliably select men most likely to benefit from treatment by active surveillance is an area of intense study. In one prostatectomy series, it was estimated that 10–15% of those treated had “insignificant” disease. One set of criteria includes men with clinical T1c tumors that are biopsy Gleason grade 6 (Grade Group 1) involving 3 or fewer cores, each core having <50% involvement by tumor, and a PSA density of <0.15.

Concerns about active surveillance include the limited ability to predict pathologic findings by needle biopsy even when multiple cores are obtained, the recognized multifocality of the disease, and the possibility of a missed opportunity to cure the disease. Nomograms to help predict which patients can safely be managed by active surveillance continue to be refined, and as their predictive accuracy improves, it can be anticipated that more patients will be candidates.

RISING PSA AFTER DEFINITIVE LOCAL THERAPY

It includes patients in whom the sole manifestation of disease is a rising PSA after surgery and/or radiation therapy. By definition, there is no evidence of disease on imaging studies. For these patients, the central issue is whether the rise in PSA results from persistent disease in the primary site, systemic disease, or both. In theory, disease in the primary site may still be curable by additional local treatment.

The decision to recommend radiation therapy after prostatectomy is guided by the pathologic findings at surgery and an MRI of the prostate or prostate bed, as CT and radionuclide bone scan are typically uninformative. Others recommend that a biopsy of the urethrovesical anastomosis be obtained before considering radiation. New PET tracers such as C-11 choline, F-18 fluciclovine, (both FDA approved) and F-18 or Ga-68 PSMA (prostate-specific membrane antigen) are more sensitive and can detect low-volume disease in the prostate bed or other sites to better inform the decision to recommend additional local therapies. Detection rates, both in and outside the prostate bed, correlate with the absolute level of PSA. Factors that predict for response to salvage radiation therapy are a positive surgical margin, lower Gleason score in the radical prostatectomy specimen, long interval from surgery to PSA failure, slow PSA doubling time, and low (<0.5–1 ng/mL) PSA value at the time of radiation treatment. Radiation therapy is generally not recommended if the PSA was persistently elevated after surgery, which usually indicates that the disease had spread outside of the area of the prostate bed and is unlikely to be controlled with radiation therapy. As is the case for other disease states, nomograms to predict the likelihood of success are available.

For patients with a rising PSA after radiation therapy, salvage local therapy can be considered if the disease was “curable” at the outset, if persistent disease has been documented by a biopsy of the prostate or by PET or other imaging, and if no disease is detectable outside of the prostate bed or regional lymph nodes. Unfortunately, case selection

is poorly defined in most series, and morbidities are significant. Options include salvage radical prostatectomy, salvage cryotherapy, salvage radiation therapy and salvage irreversible electroporation.

The rise in PSA after surgery or radiation therapy may indicate subclinical or micrometastatic disease with or without local recurrence. In these cases, the need for treatment depends, in part, on the estimated probability that the patient will show evidence of metastatic disease on a scan and in what time frame. That immediate therapy is not always required was shown in a series where patients received no systemic therapy until metastatic disease was documented. Overall, the median time to metastatic progression was 8 years, and 63% of the patients with rising PSA values remained free of metastases at 5 years. Factors associated with progression included the Gleason score of the radical prostatectomy specimen, time to recurrence, and PSA doubling time. For those with Gleason grade ≥ 8 , the probability of metastatic progression was 37, 51, and 71% at 3, 5, and 7 years, respectively. If the time to recurrence was <2 years and PSA doubling time was long (>10 months), the proportion with metastatic disease at the same time intervals was 23, 32, and 53%, vs 47, 69, and 79% if the doubling time was short (<10 months). PSA doubling times are also prognostic for survival. In one series, all patients who succumbed to disease had PSA doubling times of ≤ 3 months. Most physicians advise treatment when PSA doubling times are ≤ 12 months. A difficulty with predicting the risk of metastatic spread, symptoms, or death from disease in the rising PSA state is that most patients receive some form of therapy before the development of metastases. Nevertheless, predictive models continue to be refined.

METASTATIC DISEASE: NONCASTRATE

The state of *noncastrate metastatic disease* includes men with metastases visible on an imaging study at the time of diagnosis or after local therapy(ies), and noncastrate levels of testosterone (>150 ng/dL). Symptoms of metastatic disease include pain from osseous spread, although many patients are asymptomatic despite extensive spread. Less common are symptoms related to marrow infiltration by tumor (myelophthisis), coagulopathy, or spinal cord compression.

Standard treatment is to deplete/lower androgens by medical or surgical means, the latter being the least acceptable to patients. A less frequently used treatment is to block androgen binding to the AR with antiandrogens. More than 90% of male hormones originate in the testes; $<10\%$ are synthesized in the adrenal gland (Fig. 83-3). Survival benefits were shown for the combination of ADT plus docetaxel, and separately for ADT plus abiraterone plus prednisone in large scale randomized phase 3 trials.

Testosterone-Lowering Agents Medical therapies that lower testosterone levels include the GnRH agonists/antagonists, 17,20-lyase inhibitors, CYP-17 inhibitors, and estrogens such as diethylstilbestrol (DES). The last have fallen out of favor due to the risk of vascular complications such as fluid retention, phlebitis, emboli, and stroke. GnRH agonists/antagonists (leuprolide acetate and goserelin acetate) initially produce a rise in luteinizing hormone and follicle-stimulating hormone followed by a downregulation of receptors in the pituitary gland, which effects a chemical castration. They were approved on the basis of randomized comparisons showing an improved safety profile (specifically, reduced cardiovascular toxicities) relative to DES, with equivalent potency. The initial rise in testosterone may result in a clinical flare of the disease and as such are relatively contraindicated in men with significant obstructive symptoms, cancer-related pain, or spinal cord compromise. Pure androgen antagonists such as bicalutamide can be used to prevent flare. GnRH antagonists such as degarelix achieve castrate levels of testosterone within 48 h without the initial rise in serum testosterone and may be associated with a lower risk of cardiovascular complications.

Agents that lower testosterone are associated with an androgen-deprivation syndrome that includes hot flushes, weakness, fatigue, loss of muscle mass, anemia, change in personality, and depression. Changes in lipids, obesity, and insulin resistance, along with an

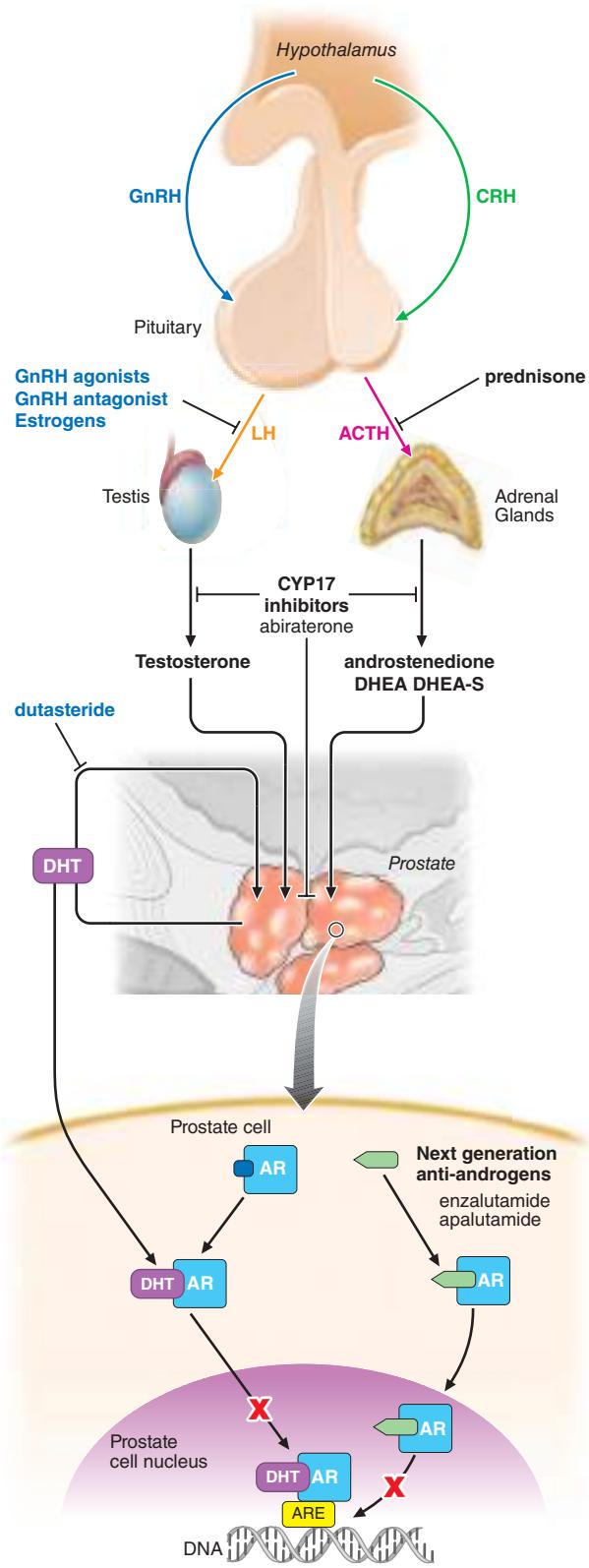


FIGURE 83-3 Sites of action of different hormone therapies.

increased risk of diabetes and cardiovascular disease are also seen, along with a decrease in bone density that worsens over time and results in an increased risk of clinical fractures. This is a particular concern in men with preexisting osteopenia that results from hypogonadism or steroid or alcohol use, and which is significantly underappreciated. Baseline fracture risk can be assessed using the FRAX scale, and to minimize fracture risk patients are advised calcium and vitamin D supplementation, along with a bisphosphonate, RANK-ligand inhibitor (denosumab), or torimafene.

Antiandrogens Nonsteroidal first-generation antiandrogens such as bicalutamide and nilutamide block the ligand binding to the AR and were initially approved to block the flare associated with the rise in serum testosterone associated with GnRH agonist/antagonist therapy. When antiandrogens are given alone, testosterone levels increase, but relative to testosterone-lowering therapies, they cause fewer hot flushes, less of an effect on libido, less muscle wasting, fewer personality changes, and less bone loss. Gynecomastia remains a significant problem but can be prevented in part by tamoxifen or prophylactic breast irradiation.

Most reported randomized trials suggest that the cancer-specific outcomes are inferior when antiandrogens are used alone. Bicalutamide, even at a dose of 150 mg (three times the approved dose for use in combination in GnRH agonists), was associated with a shorter time to progression and inferior survival compared to surgical castration for patients with established metastatic disease.

Improving on the outcomes with ADT alone has been a focus of the field for decades. One approach was to combine a first-generation antiandrogen (flutamide, bicalutamide, or nilutamide) with a GnRH analogue or surgical orchiectomy, which has not been shown to be superior to ADT alone. As a result, use of these first-generation compounds is largely limited to the first 2–4 weeks of treatment, to protect against the flare.

More recently, significant improvements in time to progression and overall survival were reported in large-scale trials for the combination of ADT with docetaxel or with abiraterone acetate plus prednisone, relative to ADT alone. Docetaxel was the first systemic therapy shown to prolong life in metastatic castration-resistant prostate cancer (mCRPC) and was approved in 2004. Abiraterone acetate (a CYP-17 inhibitor shown to reduce androgen levels to the 1–2 ng/dL range) plus prednisone was approved for mCRPC in 2011. With docetaxel, the greatest benefit was seen for patients with “high-volume” disease defined as the presence of ≥4 lesions on radionuclide bone scan or visceral disease. For abiraterone acetate and prednisone, benefit was seen across disease states ranging from high-risk localized to metastatic disease.

Intermittent Androgen Deprivation Therapy (IADT) One way to reduce the side effects of androgen depletion is to administer antiandrogens on an intermittent basis. This was proposed as a way to prevent the selection of cells that are resistant to androgen depletion. The hypothesis is that by allowing endogenous testosterone levels to rise, the cells that survive androgen depletion will induce a normal differentiation pathway. In this way, the surviving cells that are allowed to proliferate in the presence of androgen will retain sensitivity to subsequent androgen depletion. Applied in the clinic, androgen depletion is continued for 2–6 months beyond the point of maximal response. Once treatment is stopped, endogenous testosterone levels increase, and the symptoms associated with hormone treatment abate. PSA levels also begin to rise, and at some level treatment is restarted. With this approach, multiple cycles of regression and proliferation have been documented in individual patients. Unknown is whether the intermittent approach increases, decreases, or does not change the overall duration of sensitivity to androgen depletion. The approach is safe, but long-term data are needed to assess the course in men with low PSA levels. A trial to address this question is ongoing.

Outcomes of Androgen Deprivation The anti-prostate cancer effects of the various androgen depletion strategies are similar, and the clinical course is predictable: an initial response, then a period of stability in which tumor cells are dormant and nonproliferative, followed after a variable period of time by a rise in PSA and regrowth that is visible on a scan as a castration-resistant lesion. Androgen depletion is not curative because cells that survive castration are present when the disease is first diagnosed. Considered by disease manifestation, PSA levels return to normal in 60–70% of patients, and measurable disease regression occurs in 50%; improvements in bone scan occur in 25% of cases, but the majority remain stable. Duration of survival is inversely proportional to disease extent at the

time androgen depletion is first started and the nadir level of PSA at 6 months. Patients with nadir values above a certain threshold have markedly inferior survival times and should be considered for alternative approaches.

An unresolved question remains on how early systemic therapies should be offered to patients: in the adjuvant setting after surgery or radiation treatment of the primary tumor; at the time that a PSA recurrence is documented; or wait until metastatic disease or symptoms of disease are manifest? Trials in support of early therapy have been largely underpowered relative to the reported benefit or have been criticized on methodologic grounds. One trial which showed a survival benefit for patients treated with radiation therapy and 3 years of ADT, relative to radiation alone, was criticized for the poor outcomes of the control group. Another trial showing a survival benefit for patients with positive lymph nodes who were randomized to immediate medical or surgical castration compared to observation ($p = .02$) was criticized because the confidence intervals around the 5- and 8-year survival distributions for the two groups overlapped.

METASTATIC DISEASE: CASTRATE

Castration-resistant prostate cancer (CRPC), disease that progresses while the measured levels of testosterone in the blood are 50 ng/mL or lower, can produce some of the most feared complications of the disease and is lethal for most men. The most common manifestation is a rising PSA, frequently occurring with progression in bone. Nodal and/or visceral spread is less frequent and symptoms may or may not be present. The bone- and PSA-dominant pattern limits the ability to assess treatment effects reliably because traditional bone imaging is inaccurate and no PSA-based outcome has been shown to be a true surrogate for survival. It is essential to define therapeutic objectives based on the manifestations of the disease in the individual. As such, for the patient with symptomatic bone disease, relief of pain can be more clinically relevant than lowering the PSA. Naturally, for all patients the central focus is delaying or preventing disease progression, symptom development, and death from cancer.

Through 2010, docetaxel was the only FDA-approved life-prolonging therapy for CRPC. Since then, our understanding of the biology of the disease has increased significantly, which in turn has led to improved therapies. In particular, it is now recognized that the majority of mCRPCs continue to express the AR, which in upwards of 50% of cases harbors a series of oncogenic changes including overexpression of the receptor itself and the enzymes in the androgen biosynthesis pathways. These oncogenic changes have been successfully targeted with the next-generation antiandrogen enzalutamide and the CYP-17 inhibitor abiraterone acetate (given in combination with prednisone), both of which have been proven to prolong life and are FDA-approved for use in CRPC in both the pre- and post-chemotherapy setting. More recently, the results of large-scale molecular profiling efforts have led to biologically based pathway-focused classification that showed a markedly higher than expected frequency of germline and somatic BRCA2 alterations, along with other genes in the DNA damage repair pathway, that has been successfully treated with poly ADP ribose polymerase (PARP) inhibitors. Other classes of therapy that have been approved based on a demonstrated survival benefit include the biologic agent sipuleucel-T, the second-generation taxane cabazitaxel, and the alpha-emitting bone targeting radiopharmaceutical radium-223. An intense focus of current CRPC research is to understand the optimal sequence in which to utilize these agents to maximize benefit for the individual patient.

Pain Management Pain secondary to osseous metastases is one of the most feared complications of the disease and a major cause of morbidity, worsened by the narcotics needed to control symptoms. Management requires accurate diagnoses because non-cancer etiologies including degenerative disease, spinal stenosis, and vertebral collapse secondary to bone loss are common. Neurologic symptoms, including those suggestive of base of skull disease or spinal cord compromise, require emergency evaluation because loss of function may be permanent if not addressed quickly. Neurologic symptoms

or loss of function are best treated with external beam radiation, as are single sites of pain. Diffuse symptoms in the absence of neurologic deficits can be treated with bone-seeking radioisotopes such as radium-223 or the beta emitter ^{153}Sm -EDTMP, or mitoxantrone, or other systemic therapies such as abiraterone acetate, enzalutamide, and docetaxel. Radium-223 is indicated for patients with symptoms while ^{153}Sm -EDTMP and mitoxantrone are approved for the palliation of pain but not shown to prolong life. Abiraterone, enzalutamide, and docetaxel do not have a formal indication for pain, but were shown to palliate pain in the registration trials that led to their approval by showing a survival benefit.

Other bone-targeting agents, including bisphosphonates such as zoledronic acid and the RANK-ligand inhibitor denosumab, have been shown to reduce the frequency and development of skeletal complications including pain requiring analgesia, neurologic compromise from epidural extension of tumor, and/or the need for surgery or radiation therapy to treat symptomatic osseous disease. It is important to note that for all of these agents, the direct effect on the tumor is modest and benefits are seen without declines in PSA or improvements on imaging.

BENIGN DISEASE

BENIGN PROSTATIC HYPERPLASIA

BPH is a pathologic process that contributes to the development of lower urinary tract symptoms (LUTS) in men. LUTS, arising from lower urinary tract dysfunction, are further subdivided into obstructive symptoms (urinary hesitancy, straining, weak stream, terminal dribbling, prolonged voiding, incomplete emptying) and irritative symptoms (urinary frequency, urgency, nocturia, urge incontinence, small voided volumes). LUTS and other sequelae of BPH are not just due to a mass effect, but also likely due to a combination of the prostatic enlargement and age-related detrusor dysfunction.

Diagnostic Procedures and Treatment LUTS symptoms are generally measured using a validated, reproducible index that is designed to determine disease severity and response to therapy—the American Urological Association's Symptom Index (AUASI), also adopted as the International Prostate Symptom Score (IPSS) (**Table 83-2**). Serial AUASI is particularly useful in following patients as they are treated with various forms of therapy. Asymptomatic patients do not require treatment

regardless of the size of the gland, while those with an inability to urinate, gross hematuria, recurrent infection, or bladder stones may require surgery. In patients with symptoms, uroflowmetry can identify those with normal flow rates who are unlikely to benefit from treatment, and bladder ultrasound can identify those with high postvoid residuals who may need intervention. Pressure-flow (urodynamic) studies detect primary bladder dysfunction. Cystoscopy is recommended if hematuria is documented and to assess the urinary outflow tract before surgery. Imaging of the upper tracts is advised for patients with hematuria, a history of calculi, or prior urinary tract problems.

Symptomatic relief is the most common reason men seek treatment for BPH, and therefore symptomatic relief is usually the goal of therapy for BPH. Alpha-adrenergic receptor antagonists are thought to treat the dynamic aspect of BPH by reducing sympathetic tone of the bladder outlet, thereby decreasing resistance and improving urinary flow. 5ARIs are thought to treat the static aspect of BPH by reducing prostate volume and having a similar, albeit delayed effect. They have also proven to be beneficial in the prevention of BPH progression, as measured by prostate volume, the risk of developing acute urinary retention, and the risk of having BPH-related surgery. The use of an alpha-adrenergic receptor antagonist and a 5ARI as combination therapy seeks to provide symptomatic relief while preventing progression of BPH.

Another class of medications that has shown improvement in LUTS secondary to BPH is phosphodiesterase-5 (PDE5) inhibitors, used currently in the treatment of erectile dysfunction. All four of the PDE5 inhibitors available in the United States, sildenafil, vardenafil, tadalafil, and avanafil, appear to be effective in the treatment of LUTS secondary to BPH. The use of PDE5 inhibitors is not without controversy, however, given the fact that short-active phosphodiesterase inhibitors such as sildenafil need to be dosed separately from alpha blockers such as tamsulosin because of potential hypotensive effects.

Symptoms due to BPH often coexist with symptoms due to overactive bladder, and the most common pharmacologic agents for the treatment of overactive bladder symptoms are anticholinergics. This has led to multiple studies evaluating the efficacy of anticholinergics for the treatment of LUTS secondary to BPH.

Surgical therapy is now considered second-line therapy and is usually reserved for patients after a trial of medical therapy. The goal of surgical therapy is to reduce the size of the prostate, effectively reducing resistance to urine flow. Surgical approaches include TURP, transurethral incision, or removal of the gland via a retropubic, suprapubic, or

TABLE 83-2 AUA Symptom Index

QUESTIONS TO BE ANSWERED	AUA SYMPTOM SCORE (CIRCLE 1 NUMBER ON EACH LINE)					
	NOT AT ALL	LESS THAN 1 TIME IN 5	LESS THAN HALF THE TIME	ABOUT HALF THE TIME	MORE THAN HALF THE TIME	ALMOST ALWAYS
Over the past month, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0+	1	2	3	4	5
Over the past month, how often have you had to urinate again less than 2 h after you finished urinating?	0	1	2	3	4	5
Over the past month, how often have you found you stopped and started again several times when you urinated?	0	1	2	3	4	5
Over the past month, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
Over the past month, how often have you had a weak urinary stream?	0	1	2	3	4	5
Over the past month, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?	(None)	(1 time)	(2 times)	(3 times)	(4 times)	(5 times)
Sum of 7 circled numbers (AUA Symptom Score): _____						

Abbreviation: AUA, American Urological Association.

Source: MJ Barry et al: J Urol 148:1549, 1992. Used with permission.

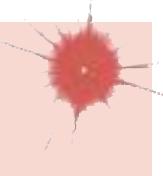
perineal approach. Also used are transurethral ultrasound-guided laser-induced prostatectomy (TULIP), stents, and hyperthermia.

FURTHER READING

- BARRY MJ, SIMMONS LH: Prevention of prostate cancer morbidity and mortality: Primary prevention and early detection. *Med Clin North Am* 101:787, 2017.
- BUYYOUNOUSKI MK et al: Prostate cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 67:245, 2017.
- FIZAZI K et al: Abiraterone plus prednisone in metastatic, castration-sensitive prostate cancer. *N Engl J Med* 377:352, 2017.
- GARTRELL BA et al: Metastatic prostate cancer and the bone: Significance and therapeutic options. *Eur Urol* 68:850, 2015.
- HIGANO CS: Intermittent versus continuous androgen deprivation therapy. *J Natl Compr Canc Netw* 12:727, 2014.
- JAMES ND et al: Abiraterone for prostate cancer not previously treated with hormone therapy. *N Engl J Med* 377:338, 2017.
- MCGINLEY KF et al: Prostate cancer in men of African origin. *Nat Rev Urol* 13:99, 2016.
- PRITCHARD CC et al: Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N Engl J Med* 375:443, 2016.
- ROBINSON D et al: Integrated clinical genomics of advanced prostate cancer. *Cell* 161:1215, 2015.
- SWEENEY CJ et al: Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. *N Engl J Med* 373:737, 2015.

84 Testicular Cancer

David J. Vaughn



Testicular germ cell tumors (GCTs) represent 95% of all testicular neoplasms. Non-germ cell tumors of the testis are much less common. Approximately 5% of GCTs arise in extragonadal locations including the mediastinum, retroperitoneum, and pineal gland. Treatment for testicular GCTs is determined by pathology and stage. The development of effective chemotherapy for this disease represents a landmark achievement in oncology. About 95% of newly diagnosed patients with testicular GCTs will be cured. For this reason, testicular cancer has been called “a model for a curable neoplasm.”

INCIDENCE

In 2016, ~8700 cases of testicular GCTs will be diagnosed in the United States, with <400 deaths. These tumors are diagnosed most commonly in men between 20 and 40 years. It has recently been reported that the incidence of GCTs is increasing in men 50 years and older.

GLOBAL CONSIDERATIONS

 The incidence of testicular GCTs appears to be increasing worldwide. The disease has the highest incidence in Scandinavia, Western Europe, and Australia/New Zealand. Africa and Asia have the lowest incidence. The incidence in the United States and the United Kingdom is intermediate. While there does not appear to be a distinct biology related to geography, several countries have reported a migration to earlier stage disease in part related to public awareness and earlier diagnosis.

EPIDEMIOLOGY

GCTs are predominantly seen in young Caucasian men. The disease is much less commonly seen in African Americans. Although most patients with GCTs do not have a family history of this disease, there are rare familial cases. Interestingly, the risk of GCT is higher in male siblings and cousins than in offspring of the patient. Although epidemiological studies have been performed attempting to identify a

relationship with environmental exposures, no conclusive causal links have been established.

Risk Factors The strongest risk factors for testicular GCT include a prior history of the disease, cryptorchidism, and a history of testicular intratubular germ cell neoplasia (ITGCN). Patients with a prior history of testicular GCT have a 2% risk of developing a contralateral GCT. These are more commonly metachronous than synchronous. Men with cryptorchidism have approximately a four- to sixfold increased risk of developing testicular GCT. Orchidopexy before puberty decreases but does not eliminate this risk. Interestingly, the contralateral descended testis is also at risk for this disease. Men undergoing infertility evaluation in which a testicular biopsy demonstrates ITGCN have a 50% risk of developing GCT. Although scrotal ultrasound of patients with testicular GCT may demonstrate testicular microcalcifications that may be related to ITGCN, the significance of testicular microcalcifications in the general population is unclear.

BIOLOGY

The primordial germ cell is the cell of origin for GCTs. All malignant GCTs arise from ITGCN. The molecular events that result in the development of ITGCN and subsequent malignant GCT have not been fully determined. However, genetic analysis of GCTs have demonstrated an excess copy number of isochromosome 12p (i[12p]) in most cases. Several genome-wide association studies have identified independent loci associated with testicular GCT risk. The strongest of these is the *KITLG* (KIT ligand) locus on chromosome 12.

PATHOLOGY

GCTs are either seminomas or nonseminomas. For a tumor to be considered a seminoma, it must be 100% seminoma. Any mixed GCT is best approached as a nonseminomatous GCT (NSGCT). Seminomas represent ~50% of cases. Seminomas arise most commonly in patients in the fourth decade of life. Seminomas may contain syncytiotrophoblastic cells which may secrete β human chorionic gonadotropin (HCG). Seminomas do not secrete α fetoprotein (AFP). Seminomas are exquisitely sensitive to both chemotherapy and radiation therapy. Seminomas are believed to be a common precursor that subsequently differentiates into the NSGCT subtypes. NSGCTs are most commonly diagnosed in the third decade of life. The histologic subtypes include embryonal carcinoma, yolk sac tumor, choriocarcinoma, and teratoma. Embryonal carcinoma is the most undifferentiated NSGCT subtype with the potential to differentiate into the other subtypes. Embryonal carcinoma may secrete AFP, HCG, both, or neither. Yolk sac tumor (also referred to as endodermal sinus tumor) often secretes AFP. Choriocarcinoma is an aggressive subtype, often secreting HCG at very high levels. These NSGCT subtypes are all considered chemotherapy sensitive. Teratoma is composed of somatic cell types that are derived from two or more germinal layers (endoderm, mesoderm, and ectoderm). Teratomas are classified as mature, in which cell types resemble normal adult somatic tissue; immature, in which cell types resemble fetal somatic tissue; and malignant, in which the cell types have undergone malignant transformation into the malignant counterpart of the somatic tissue. Teratomas are chemotherapy resistant and must be approached surgically.

INITIAL PRESENTATION

Signs and Symptoms Although a painless testicular mass is pathognomonic of a GCT, most patients present with testicular swelling, firmness, discomfort, or a combination of these. The differential diagnosis may include epididymitis or orchitis and a trial of antibiotics may be considered. Patients with retroperitoneal metastases may complain of back or flank pain. Patients may have cough, shortness of breath, or hemoptysis as a result of lung metastases. In patients with elevation of serum HCG, gynecomastia may be present. Diagnostic delay is not uncommon, and may be associated with a more advanced stage at diagnosis.

Physical Examination Careful examination of the affected testis and the contralateral normal testis should be performed. Many tumors

will have a hard consistency to palpation. Some patients may show testicular atrophy. Evaluation for supraclavicular lymphadenopathy, gynecomastia, and abdominal mass should be performed. Inguinal lymphadenopathy is rare. Most patients with lung metastases will have normal auscultation of the lungs.

Diagnostic Testing If a firm testicular mass is identified, a scrotal ultrasound should be performed. Patients with suspected epididymitis or orchitis who do not respond to antibiotics should also undergo scrotal ultrasound. Scrotal ultrasound should include both testicles. On ultrasound, a testicular GCT is hypoechoic and may be multifocal. A solid mass identified on ultrasound should be considered malignant until otherwise proven. Transscrotal aspiration or biopsy of a testicular mass should never be performed. Such scrotal violation may result in tumor seeding of the scrotum or inguinal lymph nodes.

Serum Tumor Markers Serum AFP, HCG, and lactate dehydrogenase (LDH) should be measured in patients suspected of testicular GCT. AFP is elevated in ~60–70% of patients who present with NSGCTs. Seminomas never secrete AFP. A patient with a seminoma with elevation of AFP should be approached as having a NSGCT. The half-life of AFP is 5–7 days. A falsely elevated AFP may be seen in patients with hepatic disease or a condition called hereditary persistence of AFP in which patients may have baseline AFP levels that are mildly elevated. HCG may be elevated in both NSGCTs as well as seminomas. Patients with choriocarcinoma may have markedly elevated levels of HCG. The half-life for HCG is 24–36 h. False-positive elevation of HCG may be seen secondary to hypogonadism, marijuana use, or as a result of interfering substances measured by the assay. LDH is a nonspecific marker for GCT. Its principal use is to help in the assessment of the risk classification of a patient with metastatic disease. Although elevation of serum tumor markers support the diagnosis of a testicular GCT, it should be remembered that most patients with a seminoma and up to a third of patients with NSGCTs do not have elevated levels.

■ INITIAL MANAGEMENT

Inguinal Orchiectomy Prompt referral to urology should be performed if a testicular GCT is suspected. The initial treatment for most patients suspected of having a testicular GCT is radical inguinal orchiectomy with removal of the testicle and spermatic cord to the level of the internal inguinal ring. In patients who present with metastatic disease and the diagnosis of GCT is certain, orchiectomy may be deferred until completion of chemotherapy. Pathologic examination of the entire testicle is important, since testicular GCTs may be multifocal. Given the rarity of this cancer, review by an experienced pathologist is essential for accurate tumor classification. Serum tumor markers should be obtained before and after orchiectomy.

Staging The staging of testicular GCT is based upon an understanding of the pattern of spread. The initial spread is by the lymphatic route to the retroperitoneal lymph nodes. A left-sided testicular GCT spreads first to the primary landing zone of left paraaortic lymph nodes inferior to the left renal vessels. A right-sided testicular GCT spreads first to the primary landing zone of the aortocaval nodes inferior to the right renal vessels. Nodal metastases may extend into the iliac regions. If scrotal violation occurred, inguinal lymph node metastases may be seen. Subsequent lymphatic spread is to the retrocrural, mediastinal, and supraclavicular lymph nodes. Hematogenous spread to the lung is the next most common site of metastasis. Metastases to the liver, bone, and brain are less commonly seen. Patients with newly diagnosed testicular GCTs should undergo computed tomography (CT) scan of the abdomen and pelvis. Chest x-ray should be performed. CT scan of the chest is performed if retroperitoneal metastases are present or if lung nodules are identified on chest x-ray. Bone scan and magnetic resonance imaging (MRI) of the brain are not routinely performed unless clinically indicated. Positron emission tomography (PET) has little role in the initial staging of testicular GCTs.

The American Joint Committee on Cancer tumor/node/metastasis (TNM) staging classification is used. There are three main stages of testicular GCT. Stage I is limited to the testis; stage II involves the

retroperitoneal lymph nodes; stage III includes lymph node involvement beyond the retroperitoneum and/or distant metastatic disease.

■ STAGE-BASED MANAGEMENT

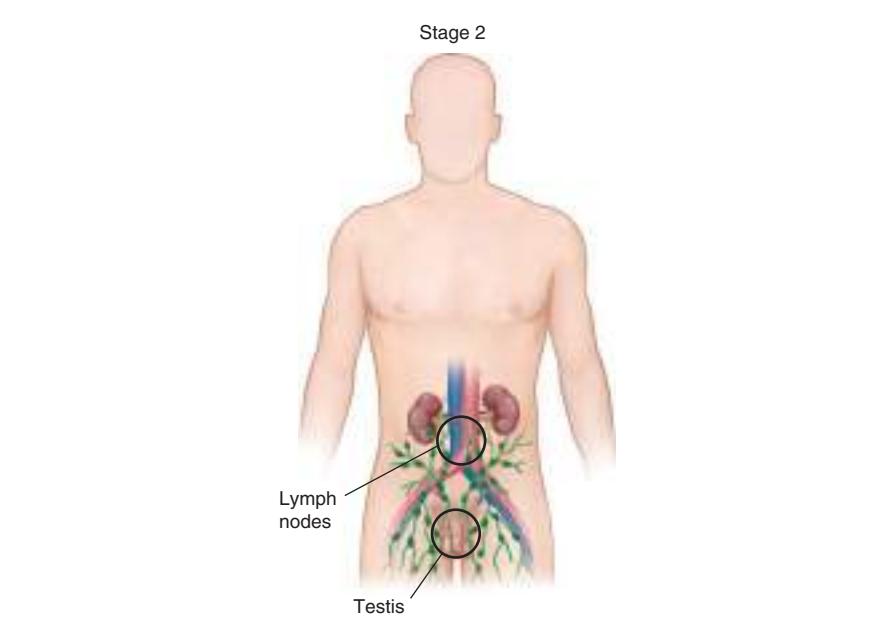
Treatment of testicular GCT is based upon two factors: (1) whether the tumor is seminoma or NSGCT, and (2) the stage of the patient. This is summarized in Fig. 84-1.

Stage I • SEMINOMA About 70% of newly diagnosed patients with seminoma present with stage I disease. This is defined as no evidence of metastatic disease on imaging of the chest, abdomen, and pelvis. If pre-orchiectomy serum HCG is elevated, this must normalize post-orchiectomy to be considered stage I. Approximately 15% of patients with stage I seminoma have metastatic disease at the microscopic level, usually in the retroperitoneum. Historically, patients with stage I seminoma were treated with a course of adjuvant radiation therapy to the paraaortic lymph nodes. While still an option, this is not usually performed because of concerns for late radiation-induced secondary malignancies. Active surveillance is the most common approach elected by these patients following orchiectomy. With active surveillance, interval physical examination and CT scan of the abdomen are performed. For the 15% of patients that develop metastatic disease during active surveillance, treatment with definitive radiation therapy or chemotherapy is curative in nearly all. A third option for clinical stage I seminoma is adjuvant chemotherapy with carboplatin monotherapy for one cycle. While effective in decreasing the risk of recurrence, it should be remembered that most patients are cured by orchiectomy alone, and therefore the additional treatment is unnecessary. In addition, long-term data on toxicity and efficacy are not available.

NSGCTs About 40% of newly diagnosed patients with NSGCTs present with stage I disease. Because NSGCTs have an increased potential for invasion and metastasis, spread to the retroperitoneum and beyond is more common than with seminoma. If pre-orchiectomy serum tumor markers are elevated, these must normalize post-orchiectomy to be considered stage I. Patients with persistently elevated or rising serum tumor markers after orchiectomy have stage IS disease and should be treated with cisplatin-based chemotherapy. If the tumor is pT1, defined as limited to testis and epididymis with no vascular or lymphatic invasion and no invasion into tunica vaginalis, the risk of recurrence is approximately 20%. However, if the tumor is pT2, defined as limited to testis and epididymis with vascular or lymphatic invasion, or tumor extension into tunica vaginalis, the risk of recurrence is ~50%. Historically, a prophylactic retroperitoneal lymph node dissection (RPLND) was performed. This surgery is not only diagnostic, but also therapeutic. In fact, most patients who undergo prophylactic RPLND will never require chemotherapy. While still an option, this approach subjects many patients to unnecessary major abdominal surgery. RPLND is also associated with a small risk of retrograde ejaculation due to nerve injury, and nerve sparing techniques have been developed. Active surveillance is frequently performed especially for patients with pT1 disease. Most patients who relapse will be treated with cisplatin-based chemotherapy and achieve cure rates approaching 100%. Active surveillance can also be employed for patients with pT2 disease, although the risk of progression is significantly higher. For this reason, some advocate adjuvant cisplatin-based chemotherapy such as BEP for one cycle for patients with pT2 disease. Other centers favor a prophylactic RPLND. Almost all patients who present with stage I NSGCTs will achieve cure.

Stage II • SEMINOMA Approximately 15–20% of newly diagnosed patients with seminoma present with stage II disease. Patients are subgrouped into IIA, IIB, or IIC based upon the size of the retroperitoneal nodes (2 cm or less, more than 2 to 5 cm, or >5 cm, respectively). Patients with stage IIA disease are usually treated with “dogleg” radiation therapy which includes the paraaortic and ipsilateral iliac nodes. Cisplatin-based chemotherapy may also be considered. Stage IIB disease is treated with cisplatin-based chemotherapy or, in select patients, radiation therapy. Most patients treated with radiation therapy that relapse will subsequently be cured with cisplatin-based chemotherapy.

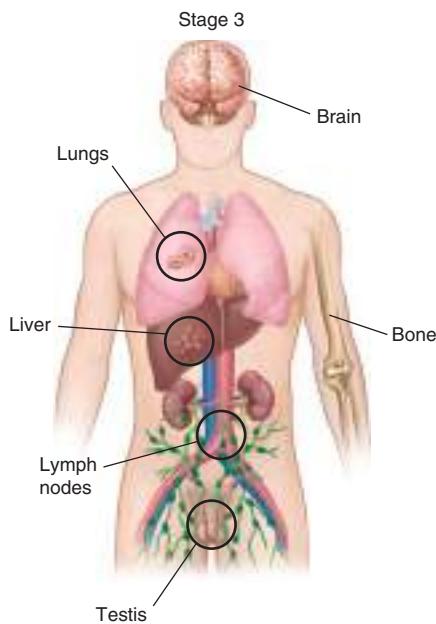
	Seminoma	NSGCT
Stage IA pT1: Testis only, no vascular/lymphatic invasion	Active surveillance; or, Adjuvant carboplatin x 1 cycle; or, Adjuvant para-aortic RT	Active surveillance; or, Nerve sparing RPLND
Stage IB pT2: Testis only, with vascular/lymphatic invasion or extension through tunica albuginea into tunica vaginalis	Active surveillance; or, Adjuvant carboplatin x 1 cycle; or, Adjuvant para-aortic RT	Adjuvant BEP x 1 cycle; or, Active surveillance; or, Nerve sparing RPLND
Stage IS Elevated serum tumor markers post-orchectomy	BEP x 3 cycles; or, EP x 4 cycles	BEP x 3 cycles; or, EP x 4 cycles

A

	Seminoma	NSGCT
Stage IIA N1: nodes ≤ 2 cm	Para-aortic and ipsilateral iliac RT; or, BEP x 3 cycles or EP x 4 cycles	Nerve-sparing RPLND; or, BEP x 3 cycles or EP x 4 cycles
Stage IIB N2: nodes > 2 to 5 cm	BEP x 3 cycles or EP x 4 cycles; or, Para-aortic and ipsilateral iliac RT	BEP x 3 cycles or EP x 4 cycles +/- post-chemotherapy RPLND
Stage IIC N3: nodes > 5 cm	BEP x 3 cycles or EP x 4 cycles	BEP x 3 cycles or EP x 4 cycles +/- post-chemotherapy RPLND

B

FIGURE 84-1 Stage-based management of testicular GCT.



	Seminoma	NSGCT
Stage IIIA (good-risk)	BEP x 3 cycles; or, EP x 4 cycles	BEP x 3 cycles; or, EP x 4 cycles; +/- Post-chemotherapy surgery
Stage IIB (intermediate-risk)	BEP x 4 cycles; or, VIP x 4 cycles	BEP x 4 cycles; or, VIP x 4 cycles +/- Post-chemotherapy surgery
Stage IIC (poor-risk)	N/A	BEP x 4 cycles; or, VIP x 4 cycles +/- Post-chemotherapy surgery

Abbreviations: BEP, bleomycin, etoposide, cisplatin; EP, etoposide, cisplatin; NSGCT, non-seminomatous germ cell tumor; RPLND, retroperitoneal lymph node dissection; RT, radiation therapy; VIP, etoposide, ifosfamide, cisplatin.

C

FIGURE 84-1 (Continued)

For patients with stage IIC disease, cisplatin-based chemotherapy should be used.

NSGCTs Approximately 15% of newly diagnosed patients with NSGCTs present with clinical stage II disease. Patients with stage IIA disease may be treated with primary RPLND. Alternatively, these patients may be treated with cisplatin-based chemotherapy. Patients with stage IIB and IIC disease are best initially managed with cisplatin-based chemotherapy.

Stage III Patients who present with stage III GCT (seminoma or NSGCT) are treated with cisplatin-based chemotherapy. These patients are classified into good-, intermediate-, or poor-risk categories using the International Germ Cell Consensus Classification system, which is based upon clinical factors including histology, site of primary, the presence of non-pulmonary visceral metastatic disease, and the level of post-orchiectomy serum tumor markers (Table 84-1). Most patients with stage III GCT present with good-risk disease; >90% will be cured. The remainder present with intermediate-risk or poor-risk disease, associated with 5-year survival rates of ~80% and 50%, respectively. Select patients with rapidly progressive metastatic disease and life-threatening symptoms such as hemoptysis in whom there is a high clinical suspicion of GCT should emergently initiate cisplatin-based chemotherapy, even without a tissue diagnosis.

Chemotherapy The development of cisplatin-based chemotherapy represents an important advance in cancer medicine. Through a series of carefully performed clinical trials with the aim of maximizing cure while minimizing the extent of treatment, the chemotherapy approach to the treatment of these patients has been standardized.

TABLE 84-1 International Germ Cell Consensus Classification System

RISK GROUP	SEMINOMA	NSGCT
Good	Any primary site; and normal AFP, any HCG, any LDH; and nonpulmonary visceral metastases absent	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and AFP <1000 ng/mL; and HCG <5000 mIU/mL; and LDH <1.5 × ULN
Intermediate	Any primary site; and normal AFP, any HCG, any LDH; and nonpulmonary visceral metastases present	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and one of the following: AFP 1000–10,000 ng/mL HCG 5000–50,000 mIU/mL LDH 1.5–10 × ULN
Poor	N/A	Mediastinal primary; or nonpulmonary visceral metastases present; or one of the following: AFP >10,000 ng/mL HCG >50,000 mIU/mL LDH >10 × ULN

Abbreviations: AFP, α fetoprotein; HCG, human chorionic gonadotropin; LDH, lactate dehydrogenase; N/A, not applicable; NSGCT, nonseminomatous germ cell tumor; ULN, upper limit normal. Nonpulmonary visceral metastases include liver, bone, and brain.

Source: International Germ Cell Cancer Collaborative Group: International Germ Cell Consensus Classification: A prognostic factor based staging system for metastatic germ cell tumors. *J Clin Oncol* 15:594, 1997.

Patients with good-risk metastatic GCT are treated with either three cycles of bleomycin, etoposide, cisplatin (BEP) or four cycles of etoposide, cisplatin (EP). Patients with intermediate- and poor-risk metastatic disease are treated with either four cycles of BEP or four cycles of etoposide, ifosfamide, cisplatin (VIP). Maintaining dose and schedule is important, as dose modifications and delays have been associated with inferior outcomes. Serum tumor markers should be monitored throughout treatment and should normalize during or after treatment. Cisplatin-based chemotherapy is associated with myelosuppression, nausea and vomiting, and alopecia. Cisplatin may result in nephrotoxicity, ototoxicity, and peripheral neuropathy. Bleomycin may result in pulmonary toxicity and risk factors for this include age greater than 40, renal failure, tobacco use, and the cumulative dose of bleomycin received. For patients at increased risk of bleomycin-induced pneumonitis, non-bleomycin-containing regimens as noted above may be given. Cisplatin-based chemotherapy is also associated with sterility. Approximately 30% of newly diagnosed testicular GCT patients have severe oligospermia or azospermia. For the remainder with normal baseline spermatogenesis who receive cisplatin-based chemotherapy, all will be azoospermic at the completion of therapy. Approximately 80% of these patients will recover spermatogenesis over a period of several years. For this reason, pre-chemotherapy sperm banking should be offered to all patients treated with chemotherapy.

Post-Chemotherapy Surgery Upon completion of cisplatin-based chemotherapy, many patients with normalized serum tumor markers will have radiographic evidence of residual masses. In approximately half of patients with NSGCT, the residual mass is composed of necrosis and/or fibrosis. About 40% will have residual teratoma and only 10% will have residual viable non-teratomatous GCT. Unfortunately, radiographic imaging cannot accurately differentiate between these entities. For this reason NSGCT patients with residual masses after chemotherapy undergo resection of all sites of disease. This most commonly includes a post-chemotherapy RPLND. However, thoracotomy and neck dissection are required in some patients. If the patients are found to have residual necrosis or teratoma, no additional therapy is required. However, for patients with residual viable non-teratomatous GCT, two additional cycles of chemotherapy are frequently administered. It should be noted that in most centers patients with minimal residual tumors defined as retroperitoneal lymph nodes of <10 mm in short axis will forego post-chemotherapy RPLND. Patients who experience normalization of serum tumor markers with first-line chemotherapy but have enlarging tumors, most often cystic masses in the retroperitoneum, may have "growing teratoma syndrome." These patients are best approached with surgery.

For patients with metastatic seminoma, ~20% of residual masses harbor viable tumor; the remainder have only necrosis. Patients with residual masses 3 cm or less may be observed without surgery. For patients with residual masses >3 cm, FDG-PET may be used to distinguish necrosis from viable seminoma and identify patients who should be considered for post-chemotherapy surgery.

■ RELAPSED DISEASE

Approximately 20–30% of patients with metastatic GCTs treated with cisplatin-based chemotherapy will not achieve durable disease control. Most of these patients will experience disease progression within 2 years following completion of chemotherapy. The International Prognostic Factors Study Group developed a risk stratification classification system for patients in first relapse. Contributors to a worsened prognosis include NSGCT histology, extragonadal primary, incomplete response to first-line chemotherapy, time to relapse of 3 months or less, level of serum tumor markers at relapse, and the presence of non-pulmonary visceral metastatic disease.

Patients in first relapse may be treated with either conventional-dose salvage chemotherapy or high-dose salvage chemotherapy with autologous stem cell rescue. There is controversy concerning which approach is optimal. Some institutions advocate for risk stratification, with more favorable prognosis patients receiving conventional-dose chemotherapy and worse prognosis patients receiving high-dose

chemotherapy. The most commonly utilized conventional-dose regimen includes paclitaxel, ifosfamide, and cisplatin (TIP). In one study of TIP in patients with more favorable risk disease, approximately two-thirds experienced 2-year progression-free survival. High-dose chemotherapy consists of initial salvage therapy followed by stem cell harvest and then two or three cycles of high-dose carboplatin and etoposide (CE) with stem cell rescue. The largest series of patients treated with high-dose chemotherapy was reported by researchers at Indiana University where this approach is considered standard for most patients in first relapse regardless of risk classification. In their study, ~70% of patients in first relapse achieved durable progression-free survival. A large retrospective analysis has compared conventional-dose salvage chemotherapy to high-dose salvage chemotherapy in patients in first relapse. This study reports a more favorable outcome with high-dose salvage chemotherapy across nearly all risk groups. However given the retrospective nature of this study and the controversy concerning optimal approaches, an international randomized trial comparing conventional dose chemotherapy (TIP) to high-dose chemotherapy with autologous stem cell rescue (TI-CE) has been initiated.

Some patients who experience disease progression after conventional-dose salvage chemotherapy may successfully be treated with high-dose salvage chemotherapy with autologous stem cell rescue. Patients with disease progression after high-dose salvage chemotherapy may be treated with subsequent chemotherapy regimens that include gemcitabine/oxaliplatin, gemcitabine/paclitaxel, epirubicin/cisplatin, and oral etoposide. While these patients may benefit from third-line chemotherapy, few will achieve durable disease control. Select patients with relapsed but resectable disease may be candidates for salvage or so-called "desperation" surgery.

Patients who experience disease progression >2 years after chemotherapy are considered to have "late relapse." Late relapse appears to have a different biology than early relapse. These patients tend to have more chemotherapy-resistant disease. Patients with late relapse usually have NSGCT with elevation of serum AFP. Many of these patients recur in the retroperitoneum many years after first-line chemotherapy, and this likely represents residual retroperitoneal disease that was not controlled after first-line therapy. These patients are best approached with salvage surgery.

■ EXTRAGONADAL GCTs

Approximately 5% of patients who present with GCTs have extragonadal primaries. These mainly originate in the mediastinum or retroperitoneum. Patients suspected of extragonadal GCT should undergo scrotal ultrasound to exclude a gonadal primary. Extranodal seminomas have a similar excellent prognosis as their gonadal counterparts and are approached the same. Mediastinal NSGCTs are classified as poor-risk and are treated with either four cycles of BEP or four cycles of VIP. These patients frequently require post-chemotherapy thoracic surgery for residual disease. For this reason, some advocate avoiding bleomycin in this patient population. Klinefelter's syndrome is associated with an increased risk of mediastinal NSGCTs. Rarely, mediastinal NSGCTs are associated with hematologic disorders including acute myelogenous leukemia. NSGCTs arising in the retroperitoneum do not have a worse prognosis than their gonadal counterparts. Many patients who present with extragonadal GCTs will undergo core needle biopsy for diagnosis. However, select patients with extragonadal tumors and definitive elevation of serum tumor markers may initiate chemotherapy without a tissue diagnosis.

Cancers of unknown primary are defined as histologically proven metastatic malignancy in which the primary site is not obvious. A subgroup of patients with cancer of unknown primary have occult GCTs. Male gender, age <65 years, midline tumors, and nonsmoking status increase the likelihood of this presentation. Pathology may demonstrate a poorly differentiated malignant neoplasm. Immunohistochemical staining is used to exclude lymphoma. Tumor may be analyzed by FISH for i(12p) which confirms the diagnosis. Even if the diagnosis is not certain, patients should be treated with cisplatin-based chemotherapy, which will cure up to 20% of this patient group.

■ TESTICULAR NON-GERM CELL TUMORS

Rarely, patients may develop testicular non-GCTs. These include lymphoma, most commonly occurring in men over the age of 50; sex cord stromal tumors including Leydig cell tumors and Sertoli cell tumors; mesothelioma of the tunica vaginalis; and, paratesticular sarcoma. Metastasis to the testis is rare, most commonly occurring in patients with advanced prostate cancer and melanoma.

■ SURVIVORSHIP AND LATE EFFECTS

Because most patients with testicular GCT will experience long-term survival, survivorship care is important. Since many of these patients will be followed by primary care physicians, an understanding of the physical, psychological, and social late effects is important. Late effects are defined as health problems that occur months or years after a disease is diagnosed or after treatment has ended. Late effects may be related to the underlying cancer or to the treatment the patient received. In long-term survivors of testicular GCT, increased cardiovascular risk and increased secondary malignancies have been reported. Patients treated with cisplatin-based chemotherapy have an increased risk of hypertension, hyperlipidemia, metabolic syndrome, and cardiovascular events. Patients treated with high cumulative doses of etoposide (such as patients who receive standard chemotherapy, relapse, and then receive salvage high dose chemotherapy) may experience up to a 1–2% risk of developing acute myelogenous leukemia, typically 2–3 years after completing therapy and associated with an 11q23 translocation. Patients treated with radiation therapy, cisplatin-based chemotherapy, or both have an increased risk of developing secondary solid malignancies.

■ FURTHER READING

- EINHORN LH et al: High-dose chemotherapy and stem-cell rescue for metastatic germ cell tumors. *N Engl J Med* 357:340, 2007.
- FELDMAN DR et al: Medical treatment of advanced testicular cancer. *JAMA* 299:272, 2008.
- HANNA NH, EINHORN LH: Testicular cancer—discoveries and updates. *N Engl J Med* 371:2005, 2014.
- INTERNATIONAL GERM CELL CANCER COLLABORATIVE GROUP: International Germ-Cell Consensus Classification: A prognostic factor based staging system for metastatic germ cell tumors. *J Clin Oncol* 15:594, 1997.
- INTERNATIONAL PROGNOSTIC FACTORS STUDY GROUP et al: Prognostic factors in patients with metastatic germ cell tumors who experienced treatment failure with cisplatin-based first-line chemotherapy. *J Clin Oncol* 28:4906, 2010.
- KANETSKY PA et al: Common variation in KITGL and at 5q31.3 predisposes to testicular germ cell cancer. *Nat Genet* 41:811, 2009.
- KOLLMANSBERGER C et al: Patterns of relapse in patients with clinical stage 1 testicular cancer managed with active surveillance. *J Clin Oncol* 33:51, 2015.
- LORCH A et al: Conventional-dose versus high-dose chemotherapy as first salvage treatment in male patients with metastatic germ cell tumors: Evidence from a large international database. *J Clin Oncol* 29:2178, 2011.
- TRAVIS LB et al: Testicular cancer survivorship: Research strategies and recommendations. *J Natl Cancer Inst* 102:1114, 2010.

menarche (11–13 years) and menopause (45–55 years), the ovary is responsible for follicle maturation associated with egg maturation, ovulation, and cyclical sex steroid hormone production. These complex biologic functions are linked to stromal and germ cells within the ovary. These cells can be broadly grouped into stromal cells and ovarian germ cells and the enveloping epithelial cells. Malignancies arising in each group include multiple histological variants with unique neoplastic behaviors. Epithelial tumors are the most common histological variant of ovarian neoplasms; they may be benign (50%), frankly malignant (33%), or of borderline malignancy of low malignant potential (16%). In adnexal masses detected by imaging or physical examination, age influences risk of malignancy; tumors in younger women are more likely benign. In the malignant group, the most common tumors are epithelial. In the group of the ovarian epithelial, malignancies are the serous tumors (60–70%); mucinous tumors (10%), endometrioid (10–15%), and clear cell (10–15%), tumors. The distribution of histologic types varies in different parts of the world. Less common stromal tumors arise from the ancillary, supportive cells such as steroid hormone-producing cells and likewise have different phenotypes and clinical presentations. Most stromal tumors do not produce estrogen, but ectopic hormone production can be seen in certain subtypes. Tumors arising in the ovarian germ cell lineage are generally similar in biology and behavior to testicular tumors in males, although their intraperitoneal location alters some metastatic behaviors (Chap. 84). Ovarian tissue may also host metastatic tumors arising from breast, colon, gastric, and pancreatic primaries. Bilateral ovarian masses from metastatic mucin-secreting gastrointestinal cancers are termed *Krukenberg tumors*. A survey of other potential primaries is commonly required during the diagnostic workup of ovarian masses.

■ OVARIAN CANCER OF EPITHELIAL ORIGIN

Epidemiology An American woman has ~1 in 72 lifetime risk (1.6%) of developing ovarian cancer, with the majority of affected women developing epithelial tumors. In 2017, 22,440 cases of ovarian cancer with 14,195 deaths are expected in the United States. Sporadic (not familial) epithelial tumors of the ovary have a peak incidence in women in their fifties and sixties, although age at presentation ranges from the third decade to the eighties and nineties. Ovarian cancer risk has been linked to an interactive mixture of epidemiologic, environmental, and genetic factors. Nulliparity, obesity, diet, infertility treatments, and possibly hormone replacement therapy have all been linked to an increase in risk. Protective factors include the use of oral contraceptives, multiparity, tubal ligation, aspirin use, and breast-feeding. Other epidemiologic factors such as the use of perineal talcum agents remain controversial. The mechanisms underlying the various protective factors are largely unknown, but theories include suppression of ovulation, modulation of gonadotropins and progestins, and perhaps reduction of ovarian inflammation and damage associated with the repair of the ovarian cortex associated with ovulation.

■ GENETICS AND PATHOGENESIS

Ovarian cancers are divided into type 1 cancers and a more aggressive type 2 variant. The type 1 cancers are characterized by low-grade histology and more indolent behavior. These tumors include the low malignant potential tumors, low-grade endometrial and mucinous histologies, and clear cell cancers. Genetic alterations commonly include mutations in *KRAS*, *BRAF*, *PTEN*, and *PIK3CA*. In contrast, studies have implicated serial genetic changes in the fallopian tube as the actual site of origin for most type 2 serous epithelial ovarian cancers. These aggressive tumors are more common and linked to losses in *TP53* and DNA repair capacity. Carcinoma *in situ* has been identified in the tubal epithelium with early losses in *TP53* and the *BRCA1/BRCA2* genes characterizing early tubal intraepithelial cancers. Following these two early genetic events, additional mutations in these transformed cells lead to tumor cell shedding, metastasis, and invasion. These type 2, poorly differentiated “ovarian” cancer cells can then spread to the ovaries, and the peritoneal cavity, aided by the ovarian cancer cell’s affinity for mesothelial lining cells.

85

Gynecologic Malignancies

David Spriggs

OVARIAN CANCER

■ INCIDENCE AND PATHOLOGY

Ovarian cancer remains a leading cause of cancer deaths in American women, ranking behind lung, breast, colon, and pancreatic cancers. The ovary is responsible for the hormone and egg production. Between

TABLE 85-1 Staging and Survival in Gynecologic Malignancies

STAGE	OVARIAN	5-YEAR SURVIVAL, %	ENDOMETRIAL	5-YEAR SURVIVAL, %	CERVIX	5-YEAR SURVIVAL, %
0	—	—	—	—	Carcinoma in situ	100
I	Confined to ovary	88–95	Confined to corpus	>90	Confined to uterus	85
II	Confined to pelvic organs	70–80	Involves corpus and cervix	~75	Invades beyond uterus but not to pelvic wall	65
III	Intraabdominal spread to omentum, diaphragm or lymph nodes	20–40	Extends outside the uterus but not outside the true pelvis	45–60	Extends to pelvic wall and/or lower third of vagina, or hydronephrosis	35
IV	Spread outside abdominal cavity, parenchymal spread + pleural effusion cytology or extra-abdominal lymph nodes (inguinal, thoracic or supraclavicular)	17	Extends outside the true pelvis or involves the bladder or rectum	~20	Invades mucosa of bladder or rectum or extends beyond the true pelvis	7

In work done as part of the Tumor Genome Atlas, type 2, serous ovarian cancer is principally a disease characterized by amplifications and deletions rather than point mutations. Damage to the tumor suppressor gene *TP53* occurs in >95% of serous ovarian cancers. Damage to homologous DNA repair genes including *BRCA1* and *BRCA2* was also common in these tumors. Low prevalence but statistically recurrent somatic mutations in seven other genes including *NF1*, *RB1*, and *CDK12* were also seen. The most common heritable abnormality linked to ovarian cancer is a germ-line mutation in either *BRCA1* (chromosome 17q12-21) or *BRCA2* (chromosome 13q12-13). These genes are important parts of the homologous DNA repair machinery for double-stranded DNA break repair. Individuals inheriting a single copy of a mutant allele (these act as autosomal dominant genes) have an increased lifetime risk of breast (46–87% for *BRCA1*; 38–84% for *BRCA2*) and ovarian cancer (39–63% for *BRCA1*; 16.5–27% for *BRCA2*). Many of these women have a family history that includes multiple cases of breast and/or ovarian cancer of at an early age. Male breast cancer, pancreatic cancer, and prostate cancer are also linked to familial *BRCA2* mutations. The most common malignancy in these women is breast carcinoma, although women harboring germ-line *BRCA1* mutations have a marked increased risk of developing ovarian malignancies in their forties and fifties. Women harboring a mutation in *BRCA2* have a lower penetrance of ovarian cancer with onset typically in their fifties or sixties. Other uncommon germ-line mutation of other genes encoding proteins linked to homologous DNA repair (e.g., *PALB2*) can also contribute to cancer risk although the frequency mutation and magnitude of risk increment is much lower and not well defined. Screening studies, even in the *mBRCA1/mBRCA2* families, suggest that any of the available screening techniques, including serial evaluation of the CA-125 tumor marker and transvaginal ultrasound, are insufficient to reliably detect early-stage ovarian cancer. Uniform germ-line *BRCA1/BRCA2* testing is recommended for all incident epithelial ovarian cancers to detect probands and identify relatives at risk. Women with these high-risk germ-line mutations are advised to undergo prophylactic removal of fallopian tubes and ovaries after completing childbearing and ideally before age 40. Early prophylactic salpingo-oophorectomy is highly protective. Salpingo-oophorectomy also appears to protect these women from subsequent breast cancer (risk reduction 50%). Prophylactic salpingectomy is almost certainly a key part of any surgical prophylaxis strategy for ovarian cancer, but the benefits of oophorectomy on either ovarian or breast cancer risk have not yet been clearly defined. Although less common, ovarian cancer is also another form of cancer (along with colorectal and endometrial cancer) that may develop in women with Lynch syndrome, type II, caused by mutations in one of the DNA mismatch repair genes (*MSH2*, *MLH1*, *MLH6*, *PMS1*, *PMS2*). Ovarian cancer may appear in women <50 years of age in this syndrome.

Presentation Neoplasms of the ovary tend to be painless unless they undergo torsion. Symptoms are therefore typically related to compression of local organs or due to symptoms from metastatic disease. Women with tumors localized to the ovary sometimes do have an increased incidence of symptoms including pelvic discomfort, bloating, and perhaps changes

in a woman's typical urinary or bowel pattern. Unfortunately, these symptoms are common in primary care and are frequently dismissed by either the woman or her health care team until later stages of disease. The pathogenic factors and timing of spread beyond the ovary are still not well understood. The most common symptoms at presentation include a period of progressive complaints that typically include some combination of nausea, early satiety, bloating, indigestion, constipation, and abdominal pain. Signs include the rapid increase in abdominal girth due to the accumulation of ascites that typically alerts the patient and her physician that the concurrent gastrointestinal symptoms are likely associated with malignant pathology. Radiologic evaluation typically demonstrates a complex adnexal mass with ascites, carcinomatosis, with pelvic, para-aortic and mesenteric adenopathy in advanced disease. Positron emission tomography (PET) scans are generally not required. Laboratory evaluation demonstrates a markedly elevated CA-125, a shed mucin (MUC16) associated with, but not specific for, ovarian cancer. Ovarian cancers are divided into four stages, with stage I tumors confined to the ovary, stage II malignancies confined to the pelvis, and stage III confined to the peritoneal cavity and retroperitoneal nodes (Table 85-1). These three stages are subdivided, with the most common presentation, stage IIIc, defined as tumors with bulky intraperitoneal disease or positive lymph node involvement. About 70% of women present with stage III disease. Stage IV disease includes women with parenchymal metastases (liver, lung, spleen) or, alternatively, abdominal wall or pleural disease. The 30% not presenting with stage III disease are roughly evenly distributed among the other stages.

Screening Ovarian cancer is a highly lethal condition, curable in early stages, and seldom curable in advanced stages; hence, screening is of considerable interest. Early-stage tumors often secrete excessive amounts of normal proteins that can be measured in the serum such as CA-125, mesothelin, and HE-4. Nevertheless, the incidence of ovarian cancer in the middle-aged female population is very low, with only ~1 in 2000 women between the ages of 50 and 60 carrying an asymptomatic and undetected tumor. Thus, effective screening techniques must be both sensitive and highly specific so to minimize the number of false positives. Panels of serum markers have not improved on CA-125 alone, although risk assessment by algorithms with multiple CA-125 is in advanced testing. No other screening strategies have been successful to date. Some large studies have suggested that low specificity screening might even worsen mortality in the screened population. Screening for ovarian cancer is currently not recommended outside of a clinical trial.

TREATMENT

Ovarian Cancer

TREATMENT

Epithelial ovarian cancer can be divided into distinct "disease states" for the purpose of treatment selection as shown in Fig. 85-1. Surgery by a skilled gynecologic oncologist remains the mainstay of initial therapy for ovarian cancer. The amount of residual visible

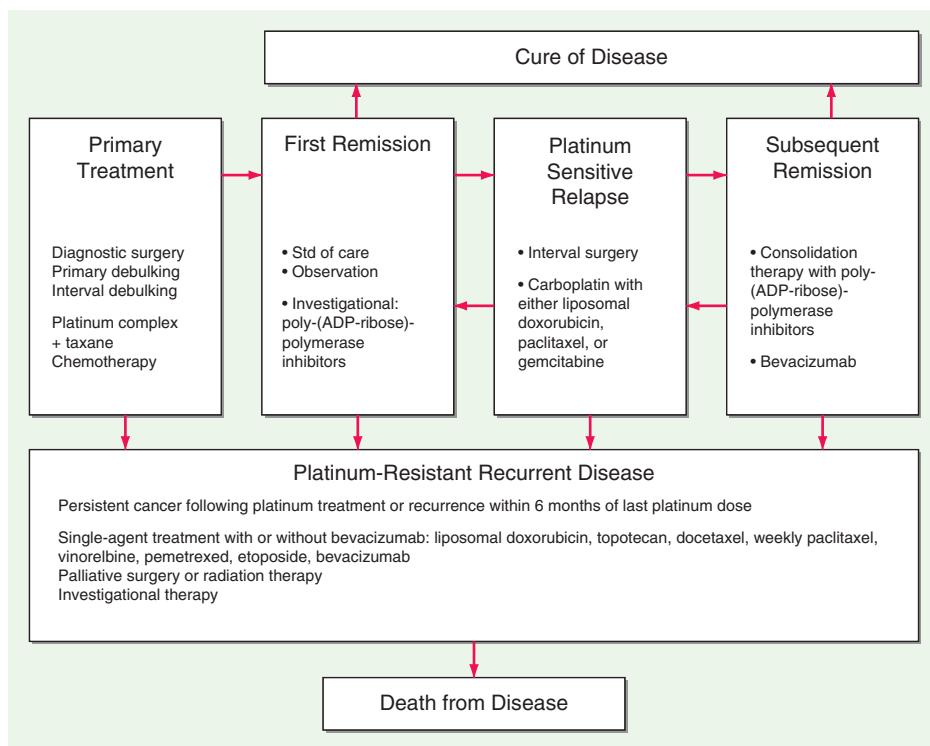


FIGURE 85-1 Disease states model of epithelial ovarian cancer and its treatment. Each box represents a relatively homogeneous group of patients that share a palette of potential treatment choices and have a similar prognosis. The arrows indicate that a single patient may move from one state to another during the course of her illness and the choice of treatments will become different in her new disease state.

cancer at the end of a primary operation is strongly predictive of outcome, and is paired with histology, grade, and stage to determine prognosis and treatment. In women presenting with a localized ovarian mass, the principal diagnostic and therapeutic maneuver is abdominal surgery to determine if the tumor is benign or malignant. In the event that the tumor is malignant, the surgical specimen will determine if the tumor arises in the ovary or is a site of metastatic disease. Metastatic disease to the ovary can be seen from primary tumors of the colon, appendix, stomach (Krukenberg tumors), and breast. Needle biopsy is contraindicated to avoid malignant contamination of the peritoneal cavity with malignant cells. Typically, women undergo laparoscopic evaluation and unilateral salpingo-oophorectomy for diagnostic purposes. If pathology reveals a primary ovarian malignancy or disseminated disease is present, then the procedure should be followed by a total hysterectomy, removal of the remaining tube and ovary, omentectomy, and pelvic node sampling along with biopsies of the peritoneal cavity and diaphragms. This extensive surgical procedure is performed because ~30% of tumors that by visual inspection appear to be confined to the ovary have already disseminated to the peritoneal cavity and/or surrounding lymph nodes. As with axillary dissections in breast cancer, node sampling is diagnostic but full lymphadenectomy appears to provide little or no additional therapeutic advantage over nodal sampling. The desired outcome of an ovarian cancer surgery is always an “R0” resection, with no visible residual cancer. The less favorable “optimal resection” (no disease greater than 1 cm in size) is still clinically useful and the prognosis of those patients is much better than the patients who are left with >1 cm disease at the end of surgery. These “suboptimally debulked” patients derive very little benefit from their surgery. Patients without gross residual disease after resection have a median survival in excess of 60 months, compared to 28–42 months for those left with macroscopic tumor.

After appropriate surgical treatment, primary chemotherapy will consist of combination treatment with paclitaxel and carboplatin. Primary chemotherapy can be delivered intravenously or alternatively, some therapy can be directly administered into the peritoneal cavity via an indwelling catheter. Several randomized studies have demonstrated improved survival with intraperitoneal therapy, but

this approach is technically difficult and is increasingly replaced by carboplatin and dose dense (weekly) paclitaxel, which appears to offer similar results in some studies.

With optimal debulking surgery and platinum-based chemotherapy (usually carboplatin dosed to an area under the curve [AUC] of 6.0 plus paclitaxel 175 mg/m² by 3-h infusion in monthly cycles), 70% of women who present with advanced-stage tumors respond, and 40–50% experience a complete remission with normalization of their CA-125, CT scans, and physical examination. These patients are sometimes enrolled in consolidation trials to extend remission and increase likelihood of cure. New immunotherapies and poly-ADP ribose polymerase inhibitors (PARPi) such as olaparib are in active testing in these patients because less than half of the complete responders are cured. Disease recurs within 1–4 years from the completion of their primary therapy. CA-125 levels often increase as a first sign of relapse and CT scan findings are confirmatory. Recurrent disease is managed, but rarely cured, with additional surgery and variety of chemotherapeutic agents. Eventually all of these women develop chemotherapy-refractory disease and refractory ascites, poor bowel motility, and obstruction or tumor-infiltrated aperistaltic bowel are all common premorbid events. Limited surgery to relieve intestinal obstruction, localized radiation therapy to relieve pressure or pain from masses, or palliative chemotherapy may be helpful. Agents with >15% response rates include gemcitabine, topotecan, liposomal doxorubicin, and bevacizumab.

Five-year survival correlates with the stage of disease: stage I, 90–95%; stage II, 70–80%; stage III, 25–40%; stage IV, 10–15% (Table 85-1). Prognosis is also influenced by histologic grade: 5-year survival is 88% for well-differentiated tumors, 58% for moderately differentiated tumors, and 27% for poorly differentiated tumors. Histologic type has less influence on outcome.

■ UNCOMMON OVARIAN TUMORS

Low Malignant Potential Tumors (Borderline Tumors)
These type 1 tumors are found in younger women (ages 40–50), indolent in behavior and few of these patients will succumb to their tumors (10 years survival may approach 98%) although recurrence is not uncommon.

Certain features like micropapillary histology and microinvasion are linked to a more aggressive behavior. Patients with tumors of low malignant potential are managed primarily by surgery; chemotherapy and radiation therapy do not substantially alter survival.

Stromal Tumors Approximately 7% of ovarian neoplasms are stromal tumors, with ~1800 cases expected each year in the United States. Ovarian stromal tumors or sex cord tumors are most common in women in their fifties or sixties, but tumors can present at any age. These tumors arise from the mesenchymal components of the ovary, including both steroid-producing cells and fibroblasts. Most of these tumors are indolent tumors with limited metastatic potential and present as unilateral solid masses. These tumors primarily are discovered by the detection of an abdominal mass sometimes with abdominal pain due to ovarian torsion, intratumoral hemorrhage, or rupture. Rarely, stromal tumors can produce estrogen and present with breast tenderness as well as precocious puberty in children, menstrual disturbances in reproductively active women, or postmenopausal bleeding. In some women, estrogen-associated secondary malignancies, such as endometrial or breast cancer, may present as synchronous malignancies. Sertoli-Leydig tumors often present with hirsutism, virilization due to increased production androgens. Hormonally inert tumors include fibroma that presents as a solitary mass often in association with ascites and occasionally hydrothorax also known as Meigs' syndrome. A subset of these tumors present in individuals with a variety of inherited disorders that predispose them to mesenchymal neoplasia including Ollier's disease (juvenile granulosa cell tumors) and Peutz-Jeghers syndrome (ovarian sex cord tumors). The treatment of these tumors is almost exclusively by surgical resection. Chemotherapy with carboplatin and paclitaxel is generally reserved for either unresectable or multiply recurrent tumors.

Germ Cell Tumors of the Ovary Germ cell tumors, like their counterparts in the testis, are cancers of germ cells. These totipotent cells contain the programming for differentiation to essentially all tissue types, and hence the germ cell tumors include a histologic menagerie of bizarre tumors, including benign teratomas (dermoid cysts) and a variety of malignant tumors, such as dysgerminoma, immature teratomas, yolk sac malignancies, and choriocarcinomas. Benign teratoma (or dermoid cyst) is the most common germ cell neoplasm of the ovary and often presents in young woman. These tumors include a complex mixture of differentiated tissue including tissues from all three germ layers. In older women these differentiated tumors can develop malignant transformation, most commonly squamous cell carcinomas. Malignant germ cell tumors include dysgerminomas, yolk sac tumors, immature teratomas, as well as embryonal and choriocarcinomas. Germ cell tumors can present at all ages, but the peak age of presentation tends to be in adolescents. Typically these tumors will become large ovarian masses, which eventually present as palpable low abdominal or pelvic masses. Like sex cord tumors, torsion or hemorrhage may present urgently or emergently as acute abdominal pain. Some of germ cell tumors produce elevated levels of human chorionic gonadotropin (hCG) or α fetoprotein (AFP). Unlike epithelial ovarian cancer, these tumors have a higher proclivity for nodal or hematogenous metastases. Germ cell tumors typically present in women who are of childbearing age, and because bilateral tumors are uncommon (except in dysgerminoma, 10–15%), the typical treatment is unilateral oophorectomy or salpingo-oophorectomy with lymph node sampling. Most commonly, women with advanced malignant germ cell tumors typically receive bleomycin, etoposide, and cisplatin (BEP) chemotherapy, in an analogous fashion to the treatment of testicular cancers. In the majority of these women, even those with advanced-stage disease, cure is expected. Dysgerminoma is the ovarian counterpart of testicular seminoma and is highly curable. Although the tumor is highly radiation-sensitive, radiation produces infertility in many patients. BEP chemotherapy is as effective or more so without causing infertility.

FALLOPIAN TUBE CANCER

Transport of the egg to the uterus occurs through the fallopian tube, with the distal ends of these tubes composed of fimbriae that drape about the ovarian surface and capture the egg as it erupts from the

ovarian cortex. As described previously, the majority of type 2 ovarian cancers are now thought to arise from the tubal epithelium. As might be expected, fallopian tube malignancies are typically serous histology and share the biology and recommended treatment as serous ovarian cancer. These tumors often present as clinically isolated adnexal masses, but like ovarian cancer, these tumors spread relatively early throughout the peritoneal cavity. Fallopian tubal cancers have a natural history and treatment that is essentially identical to ovarian cancer (Table 85-1).

CERVICAL CANCER

Etiology and Genetics

Cervical cancer is the second most common and the most lethal malignancy in women worldwide. Infection with high-risk strains of human papillomavirus (HPV) is the primary neoplastic-initiating event in the vast majority of women with invasive cervical cancer. This double-strand DNA virus infects epithelium near the transformation zone of the cervix where underlying columnar epithelium becomes squamous epithelium. More than 60 types of HPV are known, with ~20 types having the ability to generate high-grade dysplasia and malignancy. HPV16 and 18 are the types most frequently associated with high-grade dysplasia, but types 31, 33, 35, 52, and 58 are also considered to be high-risk variants. The large majority of sexually active adults are exposed to HPV, and most women clear the infection without specific intervention. The 8-kilobase HPV genome encodes seven early genes, most notably *E6* and *E7*, which can bind to *RB* and *p53*, respectively. High-risk types of HPV encode *E6* and *E7* molecules that are particularly effective at inhibiting the normal cell cycle checkpoint functions of these regulatory proteins, leading to immortalization but not full transformation of cervical epithelium. A minority of women will fail to clear the infection with subsequent HPV integration into the host genome. Over as little as a few months to several years, some of these persistently infected women develop worsening dysplasia, a premalignant condition that, untreated, can progress to cervical carcinoma. Complete transformation to cancer occurs over a period years and almost certainly requires the acquisition of other poorly defined genetic mutations within the infected and immortalized epithelium.

Approximately 528,000 new cases of cervical cancer were reported in 2012 worldwide with approximately an estimated 266,000 deaths. Cancer incidence is particularly high in women residing in central and South America, the Caribbean, and southern and eastern Africa. Mortality rate is disproportionately high in Africa. In the United States, an estimated 12,800 women will be diagnosed with cervical cancer this year, and 4210 women will die of the disease. Efforts in developed countries have looked at high-technology screening techniques for HPV involving polymerase chain reaction (PCR) and other molecular technologies.

In the integrated genomic characterization of cervical cancer by the Cancer Genome Atlas (TCGA), integration of HPV sequences was found in all of the HPV18 linked cancers and over 3 quarters of the HPV16 cancers. The cervical tumors also showed a characteristic APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; a family of cytidine deaminases that edit DNA and are endogenous mutagenic enzymes) pattern of mutagenesis with *ERBB3*, *CASP8*, and *TGFRB2* identified as significantly mutated genes presumably linked to progression from dysplasia to carcinoma. Amplification of immune targets PD-L1 and PD-L2 was also seen, which may suggest vulnerability to immunotherapy. In the much smaller number of HPV negative cancers, mutations in the common oncogenes *KRAS*, *ARID1A*, and *PTEN* were commonly seen. The clinical behavior of these cancers is likely to be different.

HPV Infection and Prevention

The Pap smear is the primary detection method for asymptomatic preinvasive cervical dysplasia of squamous epithelial lining during a gynecologic examination. Because of the delay between dysplasia and frank cervical cancer is years long; annual (or longer) screening and prevention strategies that detect precancerous dysplasia and carcinoma

in situ can be implemented successfully. Annual or biannual cervical scraping for cytology (Pap Smear) is highly effective in reducing the incidence of cervical cancer by early detection and subsequent surgical treatment. Although no randomized trial data demonstrate the utility of Pap smears, the dramatic drop in cervical cancer incidence and death in developed countries employing wide-scale screening provides strong evidence for its effectiveness. The incorporation of HPV testing by PCR or other molecular techniques increases the sensitivity of detecting cervical pathology but at the cost of lower sensitivity in that it identifies many women with transient infections who require no specific medical intervention. Unfortunately, both the collection of a Pap smear and its cytological evaluation require infrastructure beyond the means of many middle- and low-income countries. High-throughput, low-technology prevention strategies are needed to identify and treat women bearing high-risk but treatable cervical dysplasia.

A primary prevention strategy relies on HPV vaccines. Currently approved vaccines include the recombinant proteins to the late proteins, L1 and L2 of HPV-16 and -18 as well as other, less common cancer causing isotypes 11, 31, 33, 45, 52, and 58. Vaccination of girls aged 11–13 years with two injections (one year apart) before the initiation of sexual activity dramatically reduces the rate of high-risk HPV infection and subsequent dysplasia. There is also partial protection against other HPV types, although vaccinated women are still at risk for HPV infection and still benefit from standard Pap smear screening.

■ CLINICAL PRESENTATIONS

Risk Factors Clinical risk factors include many HPV infection-linked features: a high number of sexual partners, early age of first intercourse, and history of venereal disease. Smoking is a cofactor; heavy smokers have a higher risk of dysplasia with HPV infection. HIV infection, especially when associated with low CD4+ T-cell counts, is associated with a higher rate of high-grade dysplasia and likely a shorter latency period between infection and invasive disease. Histologically, the majority of cervical malignancies are squamous cell carcinomas associated with HPV, but adenocarcinomas are also HPV-related, and both arise in transitional zone of the endocervical canal; the lesions in the canal or cervical glands may not be seen by visual inspection of the cervix and can be missed by Pap smear screening. Uncommon malignancies including carcinoids, small cell carcinomas, sarcomas, and lymphomas are also found but are linked to HPV infection.

Diagnosis of Cervical Cancer Early cancer of the cervix is asymptomatic and this underlies the recommendations for routine gynecologic care. Larger, invasive carcinomas often have symptoms or signs including postcoital spotting or intermenstrual cycle bleeding or menometrorrhagia. Foul-smelling or persistent yellow discharge may also be seen. Presentations that include pelvic or sacral pain suggest lateral extension of the tumor into pelvic nerve plexus by either the primary tumor or a pelvic node and are signs of advanced-stage disease. Likewise, flank pain from hydronephrosis from ureteral compression or deep venous thrombosis from iliac vessel compression suggests either extensive nodal disease or direct extension of the primary tumor to the pelvic sidewall. The most common finding upon physical examination is a visible tumor on the cervix. Larger tumors may be identified by inspection and biopsied directly. Staging of cervical cancer is performed by clinical examination. Stage I cervical tumors are confined to the cervix, whereas stage II tumors extend into the upper vagina or paracervical soft tissue (Fig. 85-2).

Stage III tumors extend to the lower vagina or the pelvic sidewalls, whereas stage IV tumors invade the bladder or rectum or have spread to distant sites. While radiographic studies are not part of the formal clinical staging of cervical cancer, treatment planning requires them for appropriate therapy. CT can detect hydronephrosis indicative of pelvic sidewall disease but is not accurate at evaluating other pelvic structures. MRI is more accurate at estimating uterine extension and paracervical extension of disease into soft tissues typically bordered by broad and cardinal ligaments that support the uterus in the central pelvis. Very small stage I cervical tumors can be treated with a variety of surgical procedures. In young women desiring to maintain fertility, radical trachelectomy removes the cervix with subsequent anastomosis of the upper vagina to the uterine corpus; however, subsequent pregnancies may be more problematic. Large stage I cervical tumors (4 cm) confined to the cervix and all stage II–IV patients are treated with radiation therapy in combination with cisplatin-based chemotherapy. This multimodality treatment can offer the patient with advanced stage disease, a 40–80% of cure depending on the clinical circumstances. Platinum agents (cisplatin or carboplatin) combined with paclitaxel and bevacizumab are generally considered as the best palliative choice for metastatic cervical cancer patients. Secondary chemotherapy confers minimal improvement in most patients.

UTERINE CANCER

■ EPIDEMIOLOGY

Several different tumor types arise in uterine corpus. Most tumors arise in the glandular lining and are endometrial adenocarcinomas. Benign (leiomyomas) and malignant tumors (leiomyosarcomas) can also arise in the uterine smooth muscle and have very different clinical features. The endometrioid histologic subtype of endometrial cancer is the most common gynecologic malignancy in the United States. In 2017, over 60,000 new corpus cancers of uterus are projected for American women, but the surgical cure rate is high, and about 10,920 deaths from uterine cancers are predicted. Development of these tumors is a multistep process with estrogen playing an important early role in driving endometrial gland proliferation. Relative overexposure to this class of hormones is the principal risk factor for the subsequent development of endometrioid tumors. In contrast, progestins drive glandular maturation and are protective. Hence, women with high endogenous or pharmacologic exposure to estrogens, especially if unopposed by progesterone, are at higher risk for endometrial cancer. Obese women, women treated with postmenopausal estrogens or

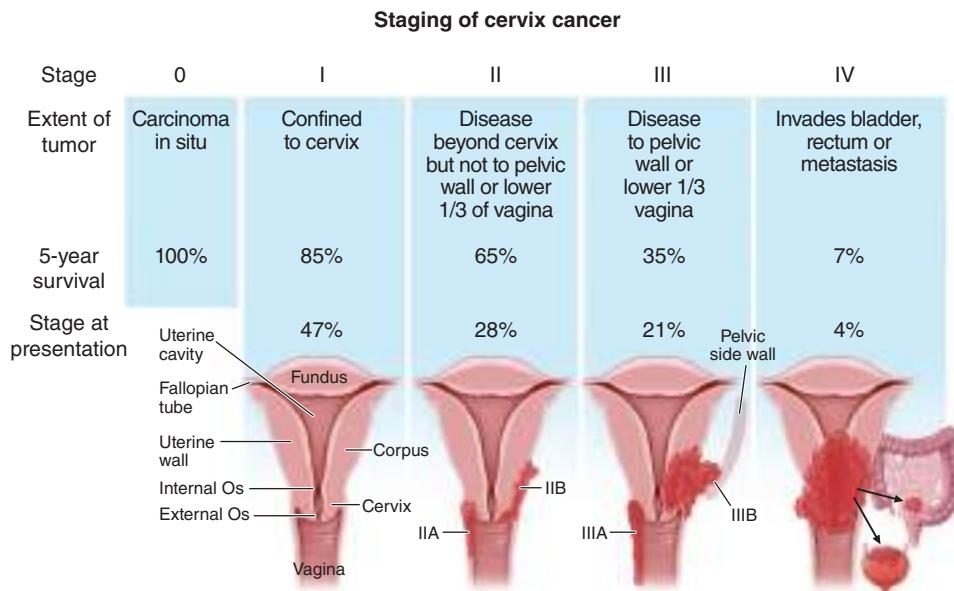


FIGURE 85-2 Anatomic display of the stages of cervix cancer defined by location, extent of tumor, frequency of presentation, and 5-year survival. (From MV Seiden: Gynecologic malignancies, in Harrison's Principles of Internal Medicine, 19th ed. New York, McGraw-Hill.)

women with estrogen-producing tumors are at higher risk for endometrial cancer. In addition, long-term treatment with tamoxifen, which has antiestrogenic effects in breast tissue but can show weak estrogenic effects in uterine epithelium, is associated with an increased risk of endometrial cancer.



Genetics Women with a germ-line mutation in one of the series of DNA mismatch repair genes associated with the Lynch syndrome, also known as hereditary nonpolyposis colon cancer (HNPCC) syndrome, are at increased risk for endometrioid endometrial carcinoma. These individuals have germ-line mutations in *MSH2*, *MLH1*, and in rare cases *PMS1* and *PMS2*. Individuals who carry these mutations typically have a family history of cancer and are at markedly increased risk for colon cancer and modestly increased risk for ovarian cancer and a variety of other tumors. Middle-aged women with HNPCC carry a 4% annual risk of endometrial cancer and a relative overall risk of approximately 200-fold as compared to age-matched women without HNPCC. In sporadic cancers, secondary events such as mutation of the *PI3K* gene or the loss of the *PTEN* tumor suppressor gene likely serve as secondary "hits" in the carcinogenesis of estrogenic excess. The molecular events that underlie less common endometrial cancers such as clear cell and papillary serous tumors of the uterine corpus are less well understood.

PATHOLOGY

Approximately 75–80% of endometrial cancers are adenocarcinomas and have been characterized as type 1 (estrogen-linked) endometrial cancers and type 2 cancers that have less clear associations with estrogens (clear cell cancers, serous cancers, and mucinous cancers). Endometrial serous cancers show TP53 functional loss and behave clinically like ovarian cancers. Serous endometrial cancers are marked by a much higher risk of distant recurrence and a lower risk for locoregional spread. Prognosis depends on stage, histologic grade, and depth of myometrial invasion.

CLINICAL PRESENTATION

The majority of women with tumors of the uterine corpus present with postmenopausal vaginal bleeding due to shedding of the malignant endometrial lining. Premenopausal women often will present with atypical bleeding between typical menstrual cycles. These signs typically bring a woman to the attention of a health care professional, and the majority of women present with early-stage disease in which the tumor is confined to the uterine corpus and the consequent high cure rate. Diagnosis is typically established by endometrial biopsy. Epithelial tumors may spread to pelvic or para-aortic lymph nodes. Serous tumors tend to have patterns of spread much more reminiscent of ovarian cancer, and patients may present with omental/peritoneal disease and sometimes ascites. Some women with endometrial cancer have a history of endometriosis. Some women presenting with uterine sarcomas will present with pelvic pain. Sarcomas commonly are found by detection of symptomatic large pelvic masses that may or may not be associated with dysfunctional bleeding.

TREATMENT

Uterine Cancer

Most women with endometrial cancer have disease that is localized to the uterus (75% are stage I, Table 85-1), and definitive treatment typically involves a hysterectomy with removal of the ovaries and fallopian tubes. The resection of lymph nodes does not improve outcome, but sentinel node resection does provide staging and prognostic information. Node involvement defines stage IIIC disease. Tumor grade and depth of invasion are the two key prognostic variables in early-stage tumors, and women with low-grade and/or minimally invasive tumors (<50% myometrial penetration) are typically observed after definitive surgical therapy. Patients with high-grade tumors or tumors that are deeply invasive (stage IB) are at higher risk for pelvic recurrence or recurrence at the vaginal cuff, which is typically prevented by intravaginal brachytherapy.

Women with regional metastases or metastatic disease (3% of patients) with low-grade tumors can be treated with progesterone or tamoxifen. Poorly differentiated tumors lack hormone receptors and are typically resistant to hormonal manipulation. The role of adjuvant chemotherapy in stage I-II disease is currently under investigation but is usually employed for advanced stage (III-IV) cancer and most tumors with serous histology. Carboplatin and paclitaxel combinations are the current standard of care. Chemotherapy for metastatic disease is delivered with palliative intent. Potentially active drugs include bevacizumab, mTOR inhibitors (e.g., temsirolimus). Patients with advanced cancer and known mismatch repair deficits may respond particularly well to immunotherapy with antagonists of the PD1/PDL1 axis.

Chemotherapy of leiomyosarcomas of the uterus with docetaxel/gemcitabine, ifosfamide/doxorubicin, and trabectedin can have substantial benefit. Carcinosarcomas of the uterus contain both mesenchymal and epithelial components but will often respond to paclitaxel and platinum complex therapy.

GESTATIONAL TROPHOBlastic TUMORS



Gestational trophoblastic diseases represent a spectrum of neoplasia from benign hydatidiform mole to choriocarcinoma due to persistent trophoblastic disease associated most commonly with molar pregnancy but occasionally seen after normal gestation. The most common presentations of trophoblastic tumors are partial and complete molar pregnancies. These represent ~1 in 1500 conceptions in developed Western countries. The incidence widely varies globally, with areas in Southeast Asia having a much higher incidence of molar pregnancy. Regions with high molar pregnancy rates are often associated with diets low in carotene and animal fats.

RISK FACTORS

Trophoblastic tumors result from the outgrowth or persistence of placental tissue. They arise most commonly in the uterus but can also arise in other sites such as the fallopian tubes due to ectopic pregnancy. Risk factors include poorly defined dietary and environmental factors as well as conceptions at the extremes of reproductive age, with the incidence particularly high in females conceiving younger than age 16 or older than age 50. In older women, the incidence of molar pregnancy might be as high as one in three, likely due to increased risk of abnormal fertilization of the aged ova. Most trophoblastic neoplasms are associated with complete moles, diploid tumors with all genetic material from the paternal donor (known as uniparental disomy). This is thought to occur when a single sperm fertilizes an enucleate egg that subsequently duplicates the paternal DNA. Trophoblastic proliferation occurs with exuberant villous stroma. If pseudopregnancy extends out past the 12th week, fluid progressively accumulates within the stroma leading to "hydropic changes." There is no fetal development in complete moles.

Partial moles arise from the fertilization of an egg with two sperm; hence two-thirds of genetic material is paternal in these triploid tumors. Hydropic changes are less dramatic, and fetal development can often occur through late first trimester or early second trimester at which point spontaneous abortion is common. Laboratory findings will include excessively high hCG and high AFP. The risk of persistent gestational trophoblastic disease after partial mole is ~5%. Complete and partial moles can be noninvasive or invasive. Myometrial invasion occurs in no more than one in six complete moles and a lower portion of partial moles.

PRESENTATION OF INVASIVE TROPHOBlastic DISEASE

The clinical presentation of molar pregnancy is changing in developed countries due to the early detection of pregnancy with home pregnancy kits and the very early use of Doppler and ultrasound to evaluate the early fetus and uterine cavity for evidence of a viable fetus. Thus, in these countries, the majority of women presenting with trophoblastic disease have their moles detected early and have typical symptoms of early pregnancy including nausea, amenorrhea, and breast tenderness.

With uterine evacuation of early complete and partial moles, most women experience spontaneous remission of their disease as monitored by serial hCG levels. These women require no chemotherapy. Patients with persistent elevation of hCG or rising hCG postevacuation have persistent or actively growing gestational trophoblastic disease and require therapy. Most series suggest that between 15 and 25% of women will have evidence of persistent gestational trophoblastic disease after molar evacuation.

In women who lack access to prenatal care, presenting symptoms can be life threatening including the development of preeclampsia or even eclampsia. Hyperthyroidism can also be seen. Evacuation of large moles can be associated with life-threatening complications including uterine perforation, volume loss, high-output cardiac failure, and adult respiratory distress syndrome (ARDS).

For women with evidence of rising hCG or radiologic confirmation of metastatic or persistent regional disease, prognosis can be estimated through a variety of scoring algorithms that identify those women at low, intermediate, and high risk for requiring multiagent chemotherapy. In general, women with widely metastatic nonpulmonary disease, very elevated hCG, and prior normal antecedent term pregnancy are considered at high risk and typically require multiagent chemotherapy at an expert center for cure. Even very advanced gestational trophoblastic disease is almost uniformly curable when managed by an expert in this rare malignancy.

TREATMENT

Invasive Trophoblastic Disease

Management of invasive trophoblastic disease should be 100% curative and complex patients should be managed by clinicians experienced in this disease. The management for a persistent and rising hCG postevacuation of a molar conception is typically chemotherapy, although surgery can play an important role for chemotherapy-resistant disease that is isolated in the uterus (especially if childbearing is complete) or to control hemorrhage. For women wishing to maintain fertility or with metastatic disease, the preferred treatment is chemotherapy. Trophoblastic disease is exquisitely sensitive to chemotherapy and guided by serial serum hCG testing, successful, curative treatment is the rule. Single-agent treatment with methotrexate or actinomycin D cures 90% of women with low-risk disease. Patients with high-risk disease (very high hCG levels, presentation 4 or more months after pregnancy, brain or liver metastases, failure of methotrexate therapy) are typically treated with multiagent chemotherapy (etoposide, methotrexate, and actinomycin D alternating with cyclophosphamide and vincristine [EMA-CO]), which is typically curative even in those women with extensive metastatic disease. Cisplatin and etoposide alternating with etoposide / methotrexate / actinomycin D is used for the highest risk patients. In the highest-risk patients with liver lung and brain metastases, hemorrhage from the rich tumor vasculature is a major risk during chemotherapy initiation. Cured women may get pregnant again without evidence of increased fetal or maternal complications.

ACKNOWLEDGMENT

Michael V. Seiden was the author of this chapter in the 19th edition. Material from his chapter has been included here.

FURTHER READING

- BROWN J et al: 15 years of progress in gestational trophoblastic disease: Scoring, standardization and salvage. *Gynecol Oncol* 49:241, 2017.
- CHUANG LT et al: Management and care of women with invasive cervical cancer: American Society of Clinical Oncology Resource Stratified Clinical Practice Guideline. *J Global Oncol* 2:311, 2016.
- HARTMANN LC, LINDOR NM: The role of risk reducing surgery in hereditary breast and ovarian cancer. *N Engl J Med* 374:454, 2016.
- JOURA EA et al: A 9 valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med* 372:711, 2015.
- MORICE P et al: Endometrial cancer. *Lancet* 2016, 387:1094, 2016.

86

Primary and Metastatic Tumors of the Nervous System

Lisa M. DeAngelis, Patrick Y. Wen

Primary brain tumors are diagnosed in ~78,000 people each year in the United States. At least 25,000 are malignant, and most of these are gliomas. Meningiomas account for 35%, vestibular schwannomas 10%, and central nervous system (CNS) lymphomas ~2%. Brain metastases are three times more common than all primary brain tumors combined and are diagnosed in ~150,000 people each year. Metastases to the leptomeninges and epidural space of the spinal cord each occur in ~3–5% of patients with systemic cancer and are also a major cause of neurologic disability.

APPROACH TO THE PATIENT

Primary and Metastatic Tumors of the Nervous System

CLINICAL FEATURES

Brain tumors of any type can present with a variety of symptoms and signs that fall into two categories: general and focal; patients often have a combination of the two (**Table 86-1**). General or nonspecific symptoms include headache, with or without nausea or vomiting, cognitive difficulties, personality change, and gait disorder. Generalized symptoms arise when the enlarging tumor and its surrounding edema cause an increase in intracranial pressure or compression of cerebrospinal fluid (CSF) circulation leading to hydrocephalus. The classic brain tumor headache predominates in the morning and improves during the day, but this pattern is seen only in a minority of patients. Headaches are often holocephalic but can be ipsilateral to the side of a tumor. Occasionally, headaches have features of a typical migraine with unilateral throbbing pain associated with visual scotoma. Personality changes may include apathy and withdrawal from social situations, mimicking depression. Focal or lateralizing findings include hemiparesis, aphasia, or visual field defect. Lateralizing symptoms are typically subacute and progressive; language difficulties may be mistaken for confusion. Seizures are common, occurring in ~25% of patients with brain metastases or malignant gliomas and are the presenting symptom in up to 90% of patients with a low-grade glioma. All seizures that arise from a brain tumor will have a focal onset whether or not it is apparent clinically.

NEUROIMAGING

Cranial magnetic resonance imaging (MRI) is the preferred diagnostic test for any patient suspected of having a brain tumor and should be performed with gadolinium contrast administration. Computed tomography (CT) scan should be reserved for those patients unable to undergo MRI. Malignant brain tumors—whether primary or metastatic—typically enhance with gadolinium, have central areas of necrosis, and are surrounded by edema of the neighboring white matter. Low-grade gliomas usually do not enhance with gadolinium and are best appreciated on fluid-attenuated inversion recovery (FLAIR) MRIs. Meningiomas have a typical appearance on MRI because they are dural-based enhancing tumors with a dural tail and compress but do not invade the brain. Dural metastases or a dural lymphoma can have a similar appearance. Imaging is characteristic for many primary and metastatic tumors and sometimes will suffice to establish a diagnosis when the location precludes surgical intervention (e.g., brainstem glioma). Functional MRI is useful in presurgical planning to define eloquent sensory, motor, or language cortex. Positron emission tomography (PET) is useful in determining

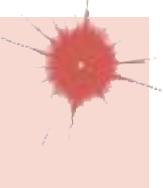


TABLE 86-1 Symptoms and Signs at Presentation of Brain Tumors

	HIGH-GRADE GLIOMA (%)	LOW-GRADE GLIOMA (%)	MENINGIOMA (%)	METASTASES (%)
Generalized				
Impaired cognitive function	50	10	30	60
Hemiparesis	40	10	36	60
Headache	50	40	37	50
Lateralizing				
Seizures	20	70+	17	18
Aphasia	20	<5	—	18
Visual field deficit	—	—	—	7

the metabolic activity of the lesions seen on MRI; MR perfusion and spectroscopy can provide information on blood flow or tissue composition. These techniques may help distinguish tumor progression from necrotic tissue as a consequence of treatment with radiation and chemotherapy. Neuroimaging is the only test necessary to diagnose a brain tumor. Laboratory tests are rarely useful, although patients with metastatic disease may have elevation of a serum tumor marker (e.g., β human chorionic gonadotropin [β -hCG] from testicular cancer). Additional testing such as cerebral angiogram, electroencephalogram (EEG), or lumbar puncture is rarely indicated or helpful.

TREATMENT

Brain Tumors

Therapy of any intracranial malignancy requires both symptomatic and definitive treatments. Definitive treatment is based on the specific tumor type and includes surgery, radiotherapy, and chemotherapy. However, symptomatic treatments apply to brain tumors of any type. Most high-grade malignancies are accompanied by substantial surrounding edema, which contributes to neurologic disability and raised intracranial pressure. Glucocorticoids are highly effective at reducing perilesional edema and improving neurologic function, often within hours of administration. Dexamethasone has been the glucocorticoid of choice because of its relatively low mineralocorticoid activity; initial doses are 8–16 mg/d. Glucocorticoids rapidly ameliorate symptoms and signs, but their long-term use causes substantial toxicity including insomnia, weight gain, diabetes mellitus, steroid myopathy, and personality changes. Consequently, a taper is indicated as definitive treatment is administered and the patient improves.

Patients with brain tumors who present with seizures require antiepileptic drug therapy. There is no role for prophylactic antiepileptic drugs in patients who have not had a seizure. The agents of choice are those drugs that do not induce the hepatic microsomal enzyme system. These include levetiracetam, topiramate, lamotrigine, valproic acid, and lacosamide (Chap. 418). Other drugs, such as phenytoin and carbamazepine, are used less frequently because they are potent enzyme inducers that can interfere with both glucocorticoid and chemotherapy metabolism. Venous thromboembolic disease occurs in 20–30% of patients with high-grade gliomas or brain metastases. Prophylactic anticoagulants should be used during hospitalization and in nonambulatory patients. Those who have had either a deep vein thrombosis or pulmonary embolus can receive therapeutic doses of anticoagulation safely and without increasing the risk for hemorrhage into the tumor. Inferior vena cava filters are reserved for patients with absolute contraindications to anticoagulation such as recent craniotomy.

PRIMARY BRAIN TUMORS

EPIDEMIOLOGY

No underlying cause has been identified for the majority of primary brain tumors. The only established risk factors are exposure to ionizing radiation (meningiomas, gliomas, and schwannomas) and

immunosuppression (primary CNS lymphoma). There is no proven evidence for any association with exposure to electromagnetic fields including cellular telephones, head injury, foods containing *N*-nitroso compounds, or occupational risk factors. A small minority of patients have a family history of brain tumors. Some of these familial cases are associated with genetic syndromes (Table 86-2).

MOLECULAR PATHOGENESIS

As with other neoplasms, brain tumors arise as a result of a multistep process driven by the sequential acquisition of genetic alterations. These include loss of tumor-suppressor genes (e.g., *p53*, cyclin-dependent kinase inhibitor 2A and 2B [*CDKN2A/B*], and phosphatase and tensin homolog on chromosome 10 [*PTEN*]) and amplification and overexpression of protooncogenes such as the epidermal growth factor receptor (*EGFR*) and the platelet-derived growth factor receptors (*PDGFR*). The accumulation of these genetic abnormalities results in uncontrolled cell growth and tumor formation.

Important progress has been made in understanding the molecular pathogenesis of several types of brain tumors, including glioblastoma and medulloblastoma, allowing them to be separated into different subtypes with different prognoses. This has led the World Health Organization (WHO) to issue an update on the classification of CNS tumors in 2016 that for the first time incorporates molecular parameters in addition to traditional histology into the diagnosis of brain tumors.

INTRINSIC “MALIGNANT” TUMORS

DIFFUSE GIOMAS

Gliomas are the most common type of malignant primary brain tumor and are derived, based on their presumed lineage, into astrocytomas and oligodendrogliomas. These tumors are classified based on two highly recurrent molecular alterations, isocitrate dehydrogenase (*IDH*) mutations and 1p/19q codeletion, in addition to more conventional histopathologic parameters. Most lower-grade astrocytomas have *IDH* mutations but intact 1p/19q, and often mutations in *ATRX* and *p53*. Oligodendrogliomas usually have *IDH* mutations and codeletion of 1p/19q.

ASTROCYTOMAS

These are infiltrative tumors with a presumptive glial cell of origin. WHO classifies astrocytomas into four prognostic grades based on histologic features: grade I (pilocytic astrocytoma, subependymal giant cell astrocytoma); grade II (astrocytoma); grade III (anaplastic astrocytoma); and grade IV (glioblastoma). Grades I and II are considered low-grade, and grades III and IV high-grade, astrocytomas.

Low-Grade Astrocytoma • GRADE I ASTROCYTOMAS Pilocytic astrocytomas (WHO grade I) are the most common tumor of childhood. They occur typically in the cerebellum but may also be found elsewhere in the neuraxis, including the optic nerves and brainstem. Frequently they appear as cystic lesions with an enhancing mural nodule. Often they have *BRAF* fusions or mutations. These are well-demarcated lesions that are potentially curable if they can be resected completely. Giant-cell subependymal astrocytomas are usually found in the ventricular wall of patients with tuberous sclerosis. They often do not require intervention but can be treated surgically or with inhibitors of the mammalian target of rapamycin (mTOR).

TABLE 86-2 Genetic Syndromes Associated with Primary Brain Tumors

SYNDROME	INHERITANCE	GENE/PROTEIN	ASSOCIATED TUMORS
Cowden's syndrome	AD	Mutations of <i>PTEN</i> (ch10p23)	Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease), meningioma, astrocytoma Breast, endometrial, thyroid cancer, trichilemmomas
Familial schwannomatosis	Sporadic Hereditary	Mutations in <i>INI1/SNF5</i> (ch22q11)	Schwannomas, gliomas
Gardner's syndrome	AD	Mutations in <i>APC</i> (ch5q21)	Medulloblastoma, glioblastoma, craniopharyngioma Familial polyposis, multiple osteomas, skin and soft tissue tumors
Gorlin syndrome (basal cell nevus syndrome)	AD	Mutations in <i>Patched 1</i> gene (ch9q22.3)	Medulloblastomas Basal cell carcinoma
Li-Fraumeni syndrome	AD	Mutations in <i>p53</i> (ch17p13.1)	Gliomas, medulloblastomas Sarcomas, breast cancer, leukemias, others
Multiple endocrine neoplasia 1 (Werner's syndrome)	AD	Mutations in <i>Menin</i> (ch11q13)	Pituitary adenoma, malignant schwannomas Parathyroid and pancreatic islet cell tumors
NF1	AD	Mutations in <i>NF1/neurofibromin</i> (ch17q12-22)	Schwannomas, astrocytomas, optic nerve gliomas, meningiomas Neurofibromas, neurofibrosarcomas, others
NF2	AD	Mutations in <i>NF2/merlin</i> (ch22q12)	Bilateral vestibular schwannomas, astrocytomas, multiple meningiomas, ependymomas
TSC (Bourneville disease)	AD	Mutations in <i>TSC1/TSC2</i> (ch9q34/16)	Subependymal giant-cell astrocytoma, ependymomas, glioma, ganglioneuroma, hamartoma
Turcot syndrome	AD AR	Mutations in <i>APC</i> ^a (ch5) <i>hMLH1</i> (ch3p21)	Gliomas, medulloblastomas Adenomatous colon polyps, adenocarcinoma
VHL	AD	Mutations in <i>VHL</i> gene (ch3p25)	Hemangioblastomas Retinal angiomas, renal cell carcinoma, pheochromocytoma, pancreatic tumors and cysts, endolymphatic sac tumors of the middle ear

^aVarious DNA mismatch repair gene mutations may cause a similar clinical phenotype, also referred to as Turcot syndrome, in which there is a predisposition to nonpolyposis colon cancer and brain tumors.

Abbreviations: AD, autosomal dominant; APC, adenomatous polyposis coli; AR, autosomal recessive; ch, chromosome; NF, neurofibromatosis; PTEN, phosphatase and tensin homologue; TSC, tuberous sclerosis complex; VHL, von Hippel-Lindau.

GRADE II ASTROCYTOMAS These are infiltrative tumors that usually present with seizures in young adults. They appear as nonenhancing tumors with increased T2/FLAIR signal (Fig. 86-1). If feasible, patients should undergo maximal surgical resection, although complete resection is rarely possible because of the invasive nature of the tumor. In patients at higher risk for recurrence (subtotal resection or above the age of 40 years), there is evidence that radiation therapy (RT) followed by PCV (procarbazine, cyclohexylchloroethylnitrosourea [CCNU], and

vincristine) chemotherapy may possibly be of benefit. The tumor transforms to a malignant astrocytoma in most patients, leading to variable survival with a median of ~5–10 years. The minority of grade II astrocytomas without *IDH* mutations have a worse prognosis.

High-Grade Astrocytoma • GRADE III (ANAPLASTIC) ASTROCYTOMA These account for ~15–20% of high-grade astrocytomas. They generally present in the fourth and fifth decades of life as variably enhancing tumors. Treatment is the same as for glioblastoma, consisting of maximal safe surgical resection followed by RT and adjuvant temozolomide alone or RT with concurrent and adjuvant temozolomide.

GRADE IV ASTROCYTOMA (GLIOBLASTOMA) Glioblastoma accounts for the majority of high-grade astrocytomas. Approximately 10% of glioblastoma have *IDH* mutations. These tend to arise from lower-grade tumors (secondary glioblastomas) and have a better prognosis. They are the most common malignant primary brain tumor, with >12,000 cases diagnosed each year in the United States. Patients usually present in the sixth and seventh decades of life with headache, seizures, or focal neurologic deficits. The tumors appear as ring-enhancing masses with central necrosis and surrounding edema (Fig. 86-2). These are highly infiltrative tumors, and the areas of increased T2/FLAIR signal surrounding the main tumor mass contain invading tumor cells. Treatment involves maximal surgical resection followed by partial-field external-beam RT (6000 cGy in thirty 200-cGy fractions) with concomitant temozolomide, followed by 6–12 months of adjuvant temozolomide. With this regimen, median survival is increased to 14.6–18 months compared to only 12 months with RT alone, and 5-year survival is ~10%. Patients whose tumor contains the DNA repair enzyme O⁶-methylguanine-DNA methyltransferase (MGMT) are relatively resistant to temozolomide and have a worse prognosis compared to those whose tumors contain low levels of MGMT as a result of silencing of the MGMT gene by promoter hypermethylation. Implantation of biodegradable polymers containing

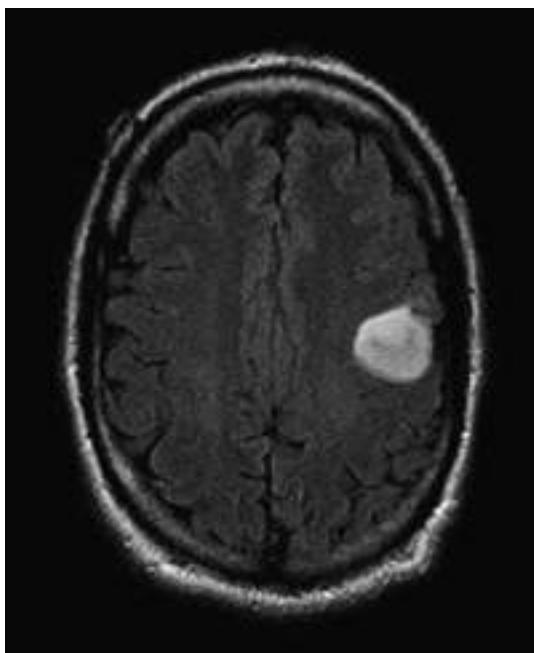


FIGURE 86-1 Fluid-attenuated inversion recovery (FLAIR) MRI of a left frontal low-grade astrocytoma. This lesion did not enhance.

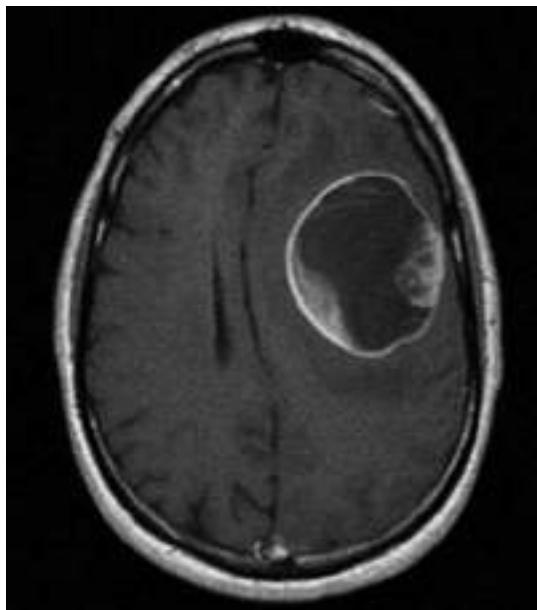


FIGURE 86-2 Postgadolinium T1 MRI of a large cystic left frontal glioblastoma.

carmustine chemotherapy into the tumor bed after resection of the tumor, or addition of tumor treating fields (scalp electrodes delivering low intensity electric currents), produces a modest improvement in survival.

For elderly patients aged >65–70 years, a hypofractionated RT regimen of 40 Gy over 3 weeks with temozolomide is well-tolerated and likely leads to similar outcomes as the 6-week standard RT regimen.

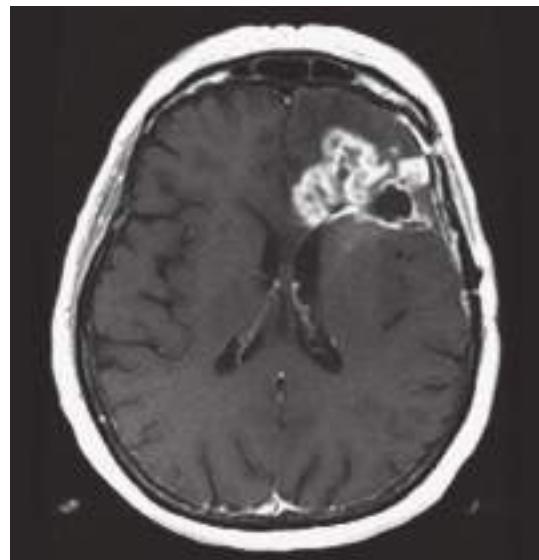
Despite optimal therapy, glioblastomas invariably recur. Treatment options for recurrent disease may include reoperation, carmustine wafers, and alternate chemotherapeutic regimens. Reirradiation is rarely helpful. Bevacizumab, a humanized vascular endothelial growth factor (VEGF) monoclonal antibody, has activity in recurrent glioblastoma, increasing progression-free survival but not overall survival, and reducing peritumoral edema and glucocorticoid use (Fig. 86-3). Treatment decisions for patients with recurrent glioblastoma must be made on an individual basis, taking into consideration such factors as previous therapy, time to relapse, performance status, and quality of life. Whenever feasible, patients with recurrent disease should be enrolled in clinical trials. Novel therapies undergoing evaluation in patients with glioblastoma include targeted molecular agents directed at receptor tyrosine kinases and signal transduction pathways; immunotherapy; oncolytic viruses; antiangiogenic agents; chemotherapeutic agents that cross the blood-brain barrier more effectively than currently available drugs; and infusion of radiolabeled drugs and targeted toxins into the tumor and surrounding brain by means of convection-enhanced delivery.

The most important adverse prognostic factors in patients with glioblastomas are older age, absence of *IDH* mutations, unmethylated MGMT promoter, poor Karnofsky performance status, and unresectable tumor.

Gliosarcomas are a variant of glioblastoma containing both an astrocytic and a sarcomatous component and are treated in the same way as glioblastomas.

OLIGODENDROGLIOMA

Oligodendrogiomas account for ~15–20% of gliomas. They are characterized by codeletion of 1p/19q and usually have *IDH* mutations. Oligodendrogiomas are classified by the WHO into oligodendroglomas (grade II) or anaplastic oligodendroglomas (AOs) (grade III). Oligodendroglomas have distinctive pathologic features such as perinuclear clearing—giving rise to a “fried-egg” appearance—and a reticular pattern of blood vessel growth. Some tumors have both an oligodendroglial as well as an astrocytic component. With molecular testing, it is now clear that almost all these mixed tumors (oligoastrocytomas) are genetically either astrocytomas or oligodendroglomas. As a result,



A

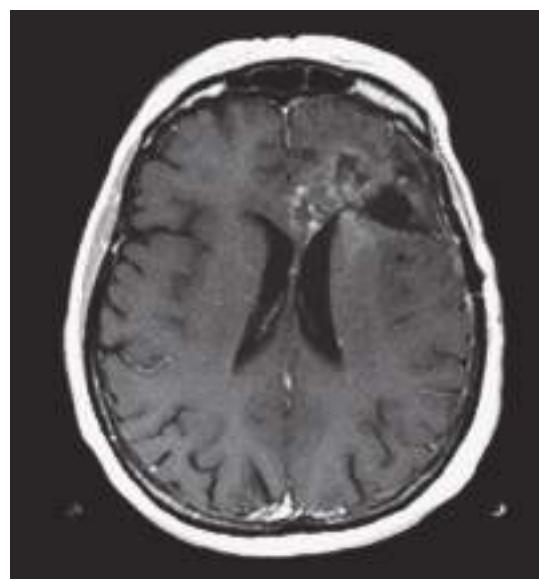


FIGURE 86-3 Postgadolinium T1 MRI of a recurrent glioblastoma before (A) and after (B) administration of bevacizumab. Note the decreased enhancement and mass effect.

the diagnosis of oligoastrocytoma is now rarely made unless molecular testing is not available.

Grade II oligodendroglomas are generally more responsive to therapy and have a better prognosis than pure astrocytic tumors. These tumors present similarly to grade II astrocytomas in young adults. The tumors are nonenhancing and often partially calcified. They should be treated with surgery and, in patients with residual disease or aged >40 years, RT and chemotherapy. Patients with oligodendroglomas have a median survival in excess of 10 years.

AOs present in the fourth and fifth decades as variably enhancing tumors. They are more responsive to therapy than grade III astrocytomas. Treatment involves maximal safe resection followed by RT and PCV or temozolomide chemotherapy. Median survival of patients with AO is in excess of 10 years.

EPENDYMOGENIC TUMORS

Ependymomas are tumors derived from ependymal cells that line the ventricular surface. They account for ~5% of childhood tumors and frequently arise from the wall of the fourth ventricle in the posterior fossa. Although adults can have intracranial ependymomas, they occur more commonly in the spine, especially in the filum terminale of the spinal

cord where they have a myxopapillary histology. Ependymomas that can be completely resected are potentially curable. Partially resected ependymomas will recur and require irradiation. The less common anaplastic ependymoma is more aggressive and is treated with resection and RT; chemotherapy has limited efficacy. Subependymomas are slow-growing benign lesions arising in the wall of ventricles that often do not require treatment.

■ OTHER LESS COMMON GLIOMAS

Gangliogliomas and pleomorphic xanthoastrocytomas occur in young adults. They behave as more indolent forms of grade I gliomas and are usually treated with surgery. Frequently they will have *BRAFV600E* mutations. Brainstem gliomas usually occur in children or young adults. Despite treatment with RT and chemotherapy, the prognosis is poor, with a median survival of only 1 year.

■ PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA

Primary central nervous system lymphoma (PCNSL) is a rare non-Hodgkin lymphoma accounting for <3% of primary brain tumors. For unclear reasons, its incidence is increasing, particularly in immunocompetent, older individuals.

PCNSL in immunocompetent patients is usually a diffuse large B-cell lymphoma. Immunocompromised patients, especially those infected with the human immunodeficiency virus (HIV) or organ transplant recipients, are at risk for PCNSL that is typically large cell with immunoblastic and more aggressive features. Epstein-Barr virus (EBV) plays an important role in the pathogenesis of PCNSL in this population. These patients are usually severely immunocompromised, with CD4 counts of <50/mL.

Immunocompetent patients with PCNSL are older (median 60 years) compared to those with HIV-related PCNSL (median 31 years). PCNSL usually presents as a mass lesion, with neuropsychiatric symptoms, lateralizing signs, or seizures. Ocular and leptomeningeal involvement each occur in 15–20% of patients.

On contrast-enhanced MRI, PCNSL usually appears as a densely enhancing tumor (Fig. 86-4). Immunocompetent patients have solitary lesions more often than immunosuppressed patients. Frequently there is involvement of the basal ganglia, corpus callosum, or periventricular region. Stereotactic biopsy is necessary to obtain a histologic diagnosis. Whenever possible, glucocorticoids should be withheld until after the biopsy has been obtained because they have a cytolytic effect on lymphoma cells and may lead to nondiagnostic tissue. In addition, patients should be tested for HIV, and the extent of disease should be assessed

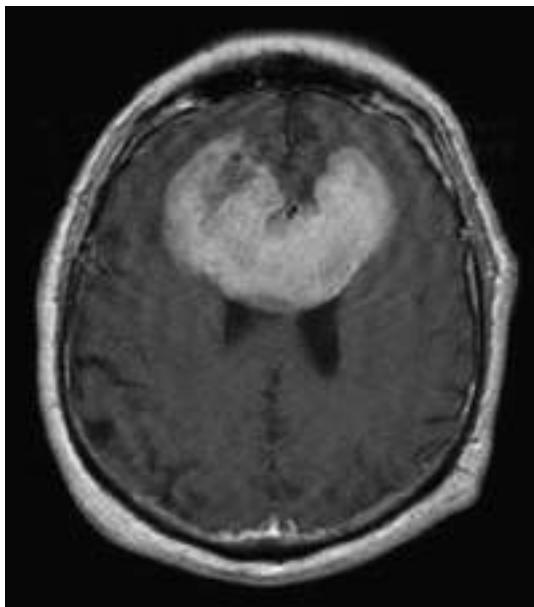


FIGURE 86-4 Postgadolinium T1 MRI demonstrating a large bifrontal primary central nervous system lymphoma (PCNSL). The periventricular location and diffuse enhancement pattern are characteristic of lymphoma.

by performing PET or CT of the body, MRI of the spine, CSF analysis, and slit-lamp examination of the eye. Bone marrow biopsy and testicular ultrasound are occasionally performed.

TREATMENT

Primary Central Nervous System Lymphoma

PCNSL is more sensitive to glucocorticoids, chemotherapy, and RT than other primary brain tumors. Durable complete responses and long-term survival are possible with these treatments. High-dose methotrexate, a folate antagonist that interrupts DNA synthesis, produces response rates ranging from 35 to 80% and median survival of up to 50 months. The combination of methotrexate with other chemotherapeutic agents such as cytarabine increases the response rate to 70–100%. The addition of whole-brain RT to methotrexate-based chemotherapy prolongs progression-free survival but not overall survival, but it is associated with delayed neurotoxicity, especially in patients aged >60 years. As a result, full-dose RT is frequently omitted, but there may be a role for reduced-dose RT. The anti-CD20 monoclonal antibody rituximab has activity in PCNSL and is often incorporated into the chemotherapy regimen. For some patients, high-dose chemotherapy with autologous stem cell rescue may offer the best chance of preventing relapse. At least 50% of patients will eventually develop recurrent disease. Treatment options include RT for patients who have not had prior irradiation, re-treatment with methotrexate, as well as other agents such as temozolomide, rituximab, procarbazine, topotecan, and pemetrexed. High-dose chemotherapy with autologous stem cell rescue may be appropriate in selected patients with relapsed disease.

PCNSL IN IMMUNOCOMPROMISED PATIENTS

PCNSL in immunocompromised patients often produces multiple ring-enhancing lesions that can be difficult to differentiate from metastases or infections such as toxoplasmosis. The diagnosis is usually established by examination of the CSF for cytology and EBV DNA, toxoplasmosis serologic testing, brain PET imaging for hypermetabolism of the lesions which, although nonspecific, can be consistent with tumor, and, if necessary, brain biopsy. Since the advent of highly active antiretroviral drugs, the incidence of HIV-related PCNSL has declined. These patients are preferably treated with high-dose methotrexate-based regimens and initiation of highly active antiretroviral therapy; whole-brain RT is reserved for those who cannot tolerate systemic chemotherapy. In organ transplant recipients, reduction of immunosuppression may improve outcome.

■ MEDULLOBLASTOMAS

Medulloblastomas are the most common malignant brain tumor of childhood, accounting for ~20% of all primary CNS tumors among children. They arise from granule cell progenitors or from multipotent progenitors from the ventricular zone. Approximately 5% of children with medulloblastoma have an inherited syndrome, such as Gorlin, Turcot, or Li-Fraumeni, which predisposes to the development of medulloblastoma. Histologically, medulloblastomas are highly cellular tumors with abundant dark staining, round nuclei, and rosette formation (Homer-Wright rosettes). In the 2016 WHO pathologic classification, they have been divided into four molecular subgroups: (1) WNT-activated (primarily affects children and has the best outcome); (2) SHH-activated (affects adults, infants, and children with the younger patients having the better outcome and adults doing poorly); (3) non-WNT/non-SHH, group 3 (frequently has disseminated CNS disease at diagnosis and has the worst outcome); and (4) non-WNT/non-SHH, group 4 (30% have metastases at diagnosis, but 5-year progression-free survival is 95%). Regardless of subtype, patients present with headache, ataxia, and signs of brainstem involvement. On MRI they appear as densely enhancing tumors in the posterior fossa, sometimes associated with hydrocephalus. Treatment involves maximal surgical resection, craniospinal irradiation, and chemotherapy with agents such as cisplatin, lomustine, cyclophosphamide, and vincristine. Approximately 70% of

patients overall have long-term survival but usually at the cost of significant neurocognitive impairment. A major goal of current research is to improve survival while minimizing long-term complications, and clinical trials are now being designed for specific molecular subgroups.

PINEAL REGION TUMORS

A large number of tumors can arise in the region of the pineal gland. These typically present with headache, visual symptoms, and hydrocephalus. Patients may have Parinaud's syndrome characterized by impaired upgaze and accommodation. Some pineal tumors such as pineocytomas and benign teratomas can be treated by surgical resection. Germinomas respond to irradiation, whereas pineoblastomas and nongerminomatous germ cell tumors require craniospinal radiation and chemotherapy.

EXTRINSIC “BENIGN” TUMORS

MENINGIOMAS

Meningiomas are diagnosed with increasing frequency as more people undergo neuroimaging for various indications. They are now the most common primary brain tumor, accounting for ~35% of the total. Their incidence increases with age. They tend to be more common in women and in patients with neurofibromatosis type 2 (NF2). They also occur more commonly in patients with a past history of cranial irradiation.

Meningiomas arise from the dura mater and are composed of neoplastic meningothelial (arachnoidal cap) cells. They are most commonly located over the cerebral convexities, especially adjacent to the sagittal sinus, but they can also occur in the skull base and along the dorsum of the spinal cord. Meningiomas are classified by the WHO into three histologic grades of increasing aggressiveness: grade I (benign), grade II (atypical), and grade III (malignant).

Many meningiomas are found incidentally following neuroimaging for unrelated reasons. They can also present with headaches, seizures, or focal neurologic deficits. On imaging studies they have a characteristic appearance usually of a densely enhancing extra-axial tumor arising from the dura (Fig. 86-5). Typically they have a dural tail, consisting of thickened, enhanced dura extending like a tail from the mass. The main differential diagnosis of meningioma is a dural metastasis.

If the meningioma is small and asymptomatic, no intervention is necessary and the lesion can be observed with serial MRI studies. Larger, symptomatic lesions should be resected. If complete resection is achieved, the patient is cured. Incompletely resected tumors tend to recur, although the rate of recurrence can be very slow with grade I tumors. Tumors that cannot be resected, or can only be partially removed, may benefit from external-beam RT or stereotactic

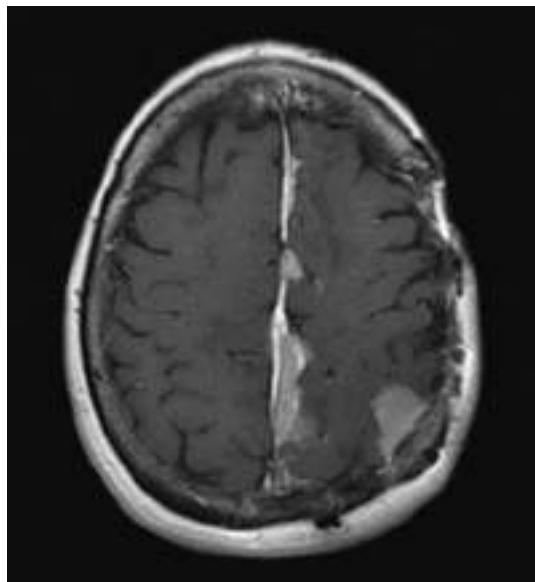


FIGURE 86-5 Postgadolinium T1 MRI demonstrating multiple meningiomas along the falx and left parietal cortex.

radiosurgery (SRS). These treatments may also be helpful in patients whose tumor has recurred after surgery. Hormonal therapy and chemotherapy are currently unproven.

Rarer tumors that resemble meningiomas include hemangiopericytomas and solitary fibrous tumors. Since they share similar molecular alterations, the 2016 WHO classification introduced the combined term solitary fibrous tumor/hemangiopericytoma for this entity. These tumors are treated with surgery and RT but have a higher propensity to recur locally or metastasize systemically.

SCHWANNOMAS

These are generally benign tumors arising from the Schwann cells of cranial and spinal nerve roots. The most common schwannomas, termed *vestibular schwannomas* or *acoustic neuromas*, arise from the vestibular portion of the eighth cranial nerve and account for ~9% of primary brain tumors. Patients with NF2 have a high incidence of vestibular schwannomas that are frequently bilateral. Schwannomas arising from other cranial nerves, such as the trigeminal nerve (cranial nerve V), occur with much lower frequency. Neurofibromatosis type 1 (NF1) is associated with an increased incidence of schwannomas of the spinal nerve roots.

Vestibular schwannomas may be found incidentally on neuroimaging or present with progressive unilateral hearing loss, dizziness, tinnitus, or, less commonly, symptoms resulting from compression of the brainstem and cerebellum. On MRI they appear as densely enhancing lesions, enlarging the internal auditory canal and often extending into the cerebellopontine angle (Fig. 86-6). The differential diagnosis includes meningioma. Very small, asymptomatic lesions can be observed with serial MRIs. Larger lesions should be treated with surgery or SRS. The optimal treatment will depend on the size of the tumor, symptoms, and the patient's preference. In patients with small vestibular schwannomas and relatively intact hearing, early surgical intervention increases the chance of preserving hearing.

PITUITARY TUMORS

These are discussed in detail in [Chap. 373](#).

CRANIOPHARYNGIOMAS

Craniopharyngiomas are rare, usually suprasellar, partially calcified, solid, or mixed solid-cystic benign tumors that arise from remnants of Rathke's pouch. They have a bimodal distribution, occurring predominantly in children but also between the ages of 55 and 65 years.

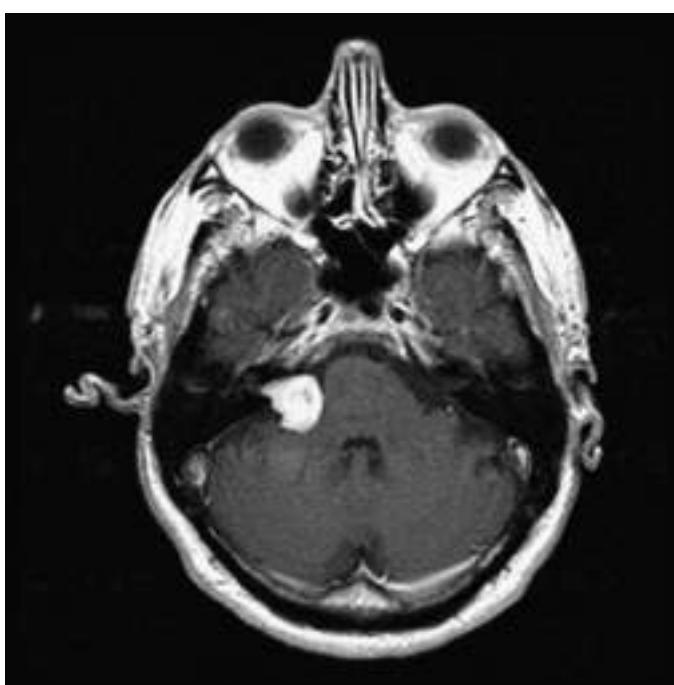


FIGURE 86-6 Postgadolinium MRI of a right vestibular schwannoma. The tumor can be seen to involve the internal auditory canal.

They present with headaches, visual impairment, and impaired growth in children and hypopituitarism in adults. Treatment involves surgery, RT, or a combination of the two.

■ OTHER BENIGN TUMORS

Dysembryoplastic Neuroepithelial Tumors (DNTs) These are benign, supratentorial tumors, usually in the temporal lobe. They typically occur in children and young adults with a long-standing history of seizures. Surgical resection is curative.

Epidermoid Cysts These consist of squamous epithelium surrounding a keratin-filled cyst. They are usually found in the cerebellopontine angle and the intrasellar and suprasellar regions. They may present with headaches, cranial nerve abnormalities, seizures, or hydrocephalus. MRI demonstrates an extra-axial lesion with characteristics that are similar to CSF but have restricted diffusion. Treatment involves surgical resection.

Dermoid Cysts Like epidermoid cysts, dermoid cysts arise from epithelial cells that are retained during closure of the neural tube. They contain both epidermal and dermal structures such as hair follicles, sweat glands, and sebaceous glands. Unlike epidermoid cysts, these tumors usually have a midline location. They occur most frequently in the posterior fossa, especially the vermis, fourth ventricle, and suprasellar cistern. On MRI, dermoid cysts resemble lipomas, demonstrating T1 hyperintensity and variable signal on T2. Symptomatic dermoid cysts can be treated with surgery.

Colloid Cysts These usually arise in the anterior third ventricle and may present with headaches, hydrocephalus, and, very rarely, sudden death. Surgical resection is curative, or a third ventriculostomy may relieve the obstructive hydrocephalus and be sufficient therapy.

NEUROCUTANEOUS SYNDROMES (PHAKOMATOSES)

A number of genetic disorders are characterized by cutaneous lesions and an increased risk of brain tumors. Most of these disorders have an autosomal dominant inheritance with variable penetrance.

■ NEUROFIBROMATOSIS TYPE 1 (NF1) (von RECKLINGHAUSEN'S DISEASE)

NF1 is an autosomal dominant disorder with variable penetrance and an incidence of ~1 in 2600–3000. Approximately one-half of cases are familial; the remainder are caused by new mutations arising in patients with unaffected parents. The *NF1* gene on chromosome 17q11.2 encodes neurofibromin, a guanosine triphosphatase (GTPase)-activating protein (GAP) that modulates signaling through the RAS pathway. Mutations of *NF1* result in a large number of nervous system tumors including neurofibromas, plexiform neurofibromas, optic nerve gliomas, astrocytomas, and meningiomas. In addition to neurofibromas, which appear as multiple, soft, rubbery cutaneous tumors, other cutaneous manifestations of NF1 include café-au-lait spots and axillary freckling. NF1 is also associated with hamartomas of the iris termed Lisch nodules, pheochromocytomas, pseudoarthrosis of the tibia, scoliosis, epilepsy, and mental retardation.

■ NEUROFIBROMATOSIS TYPE 2 (NF2)

NF2 is less common than NF1, with an incidence of 1 in 25,000–40,000. It is an autosomal dominant disorder with full penetrance. As with NF1, approximately one-half of cases arise from new mutations. The *NF2* gene on 22q encodes a cytoskeletal protein, merlin (moesin, ezrin, radixin-like protein) that functions as a tumor suppressor. NF2 is characterized by bilateral vestibular schwannomas in >90% of patients, multiple meningiomas, and spinal ependymomas and astrocytomas. Treatment of bilateral vestibular schwannomas can be challenging because the goal is to preserve hearing for as long as possible. These patients may also have diffuse schwannomatosis that may affect the cranial, spinal, or peripheral nerves; posterior subcapsular lens opacities; and retinal hamartomas.

■ TUBEROUS SCLEROSIS (BOURNEVILLE DISEASE)

This is an autosomal dominant disorder with an incidence of ~1 in 5000–10,000 live births. It is caused by mutations in either the *TSC1* gene, which maps to chromosome 9q34 and encodes a protein termed hamartin, or the *TSC2* gene, which maps to chromosome 16p13.3 and encodes the protein tuberin. Hamartin forms a complex with tuberin, which inhibits cellular signaling through mTOR, and acts as a negative regulator of the cell cycle. Patients with tuberous sclerosis may have seizures, mental retardation, adenoma sebaceum (facial angiofibromas), shagreen patch, hypomelanotic macules, periungual fibromas, renal angiomyolipomas, and cardiac rhabdomyomas. These patients have an increased incidence of subependymal nodules, cortical tubers, and subependymal giant-cell astrocytomas (SEGAs). Patients frequently require anticonvulsants for seizures. SEGAs do not always require therapeutic intervention, but the most effective therapy is with the mTOR inhibitors sirolimus or everolimus, which often decrease seizures as well as SEGAs size.

TUMORS METASTATIC TO THE BRAIN

Brain metastases arise from hematogenous spread and frequently originate from a lung primary or are associated with pulmonary metastases. Most metastases develop at the gray matter–white matter junction in the watershed distribution of the brain where intravascular tumor cells lodge in terminal arterioles. The distribution of metastases in the brain approximates the proportion of blood flow such that ~85% of all metastases are supratentorial and 15% occur in the posterior fossa. The most common sources of brain metastases are lung and breast carcinomas; melanoma has the greatest propensity to metastasize to the brain, being found in 80% of patients at autopsy (Table 86-3). Other tumor types such as ovarian and esophageal carcinoma rarely metastasize to the brain. Prostate and breast cancers also have a propensity to metastasize to the dura and can mimic meningioma. Leptomeningeal metastases are common from hematologic malignancies and also breast and lung cancers. Spinal cord compression primarily arises in patients with prostate and breast cancer, tumors with a strong propensity to metastasize to the axial skeleton.

■ DIAGNOSIS OF METASTASES

Brain metastases are best visualized on MRI, where they usually appear as well-circumscribed lesions (Fig. 86-7). The amount of perilesional edema can be highly variable, with large lesions causing minimal edema and sometimes very small lesions causing extensive edema. Enhancement may be in a ring pattern or diffuse. Occasionally, intracranial metastases will hemorrhage; although melanoma, thyroid, and kidney cancer have the greatest propensity to hemorrhage, the most common cause of a hemorrhagic metastasis is lung cancer because it accounts for the majority of brain metastases. The radiographic appearance of brain metastasis is nonspecific, and similar-appearing lesions can occur with infection including brain abscesses and also with demyelinating lesions, sarcoidosis, radiation necrosis in a previously treated patient, or a primary brain tumor that may be a second malignancy in a patient with systemic cancer. Biopsy is rarely necessary for diagnosis because imaging alone in the appropriate clinical situation usually suffices. However, in

TABLE 86-3 Frequency of Nervous System Metastases by Common Primary Tumors

	BRAIN (%)	LM (%)	ESCC (%)
Lung	41	17	15
Breast	19	57	22
Melanoma	10	12	4
Prostate	1	1	10
GIT	7	—	5
Renal	3	2	7
Lymphoma	<1	10	10
Sarcoma	7	1	9
Other	11	—	18

Abbreviations: ESCC, epidural spinal cord compression; GIT, gastrointestinal tract; LM, leptomeningeal metastases.

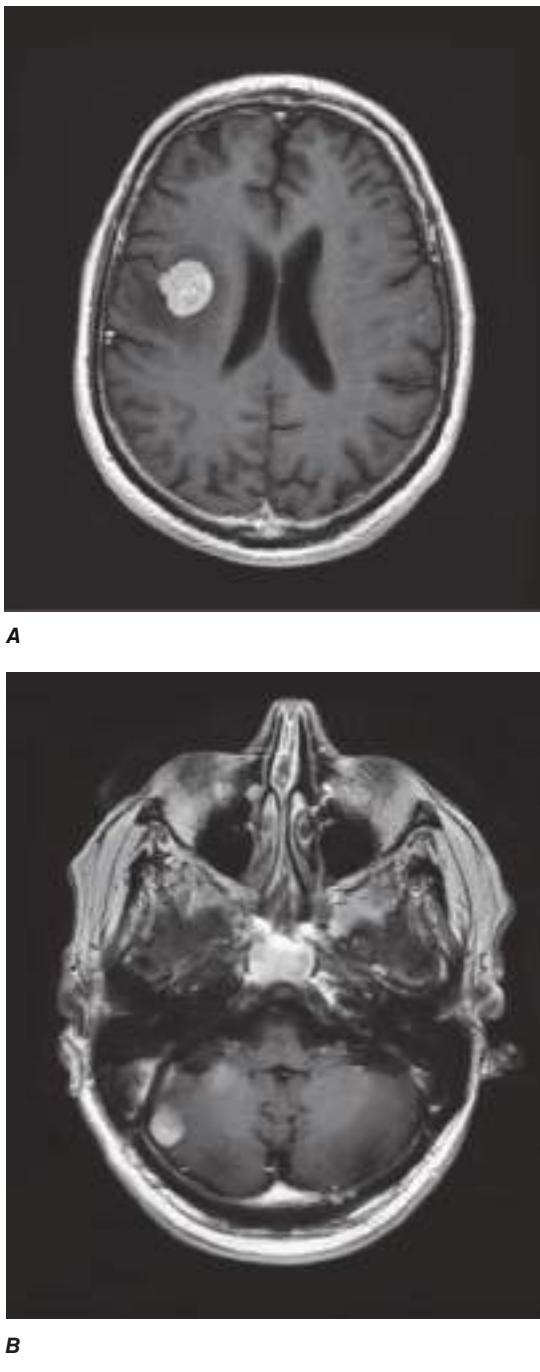


FIGURE 86-7 Postgadolinium T1 MRI of multiple brain metastases from non-small-cell lung cancer involving the right frontal (**A**) and right cerebellar (**B**) hemispheres. Note the diffuse enhancement pattern and absence of central necrosis.

~10% of patients, a systemic cancer may present with a brain metastasis, and if there is not an easily accessible systemic site to biopsy, a brain lesion must be removed for diagnostic purposes.

TREATMENT

Tumors Metastatic to the Brain

DEFINITIVE TREATMENT

The number and location of brain metastases often determine the therapeutic options. The patient's overall condition and current or potential control of systemic disease are also major determinants. Brain metastases are single in approximately one-half of patients and multiple in the other half.

RADIATION THERAPY

The standard treatment for brain metastases has previously been whole-brain radiotherapy (WBRT) usually administered to a total dose of 3000 cGy in 10 fractions. This affords rapid palliation, and ~80% of patients improve with glucocorticoids and RT. However, it is not curative, is associated with neurocognitive toxicity, and produces median survival of only 4–6 months. If feasible, SRS has become the primary radiation oncology approach to brain metastases. It can be delivered through a variety of equally effective techniques including the gamma knife, linear accelerator, proton beam, or CyberKnife, all of which can deliver highly focused doses of RT, usually in a single fraction. SRS can effectively sterilize the visible lesions and afford local disease control in 80–90% of patients. Some patients have been cured of their brain metastases using SRS, whereas this is distinctly rare with WBRT. Traditionally SRS was used only for patients with 1–3 metastases, but recent data suggest that SRS can effectively treat up to 10 lesions. It is, however, confined to lesions of ≤ 3 cm and is most effective in metastases of ≤ 1 cm. The addition of WBRT to SRS improves disease control in the nervous system but does not prolong survival and thus is rarely employed.

SURGERY

Randomized controlled trials have demonstrated that surgical extirpation of a single brain metastasis followed by WBRT is superior to WBRT alone. Removal of two lesions or a single symptomatic mass, particularly if compressing the ventricular system, can also be useful. This is particularly important in patients who have highly radioresistant lesions such as renal carcinoma. Surgical resection can produce rapid amelioration of symptoms, improve control of edema, and result in prolonged survival. WBRT administered after complete resection of a brain metastasis improves disease control but does not prolong survival. Some centers administer focal RT or even SRS to a resected cavity, especially if there is concern that tumor has been left behind.

CHEMOTHERAPY

Chemotherapy is becoming increasingly useful for brain metastases. Metastases from tumor types that are highly chemosensitive, such as germ cell tumors or small-cell lung cancer, may respond to chemotherapeutic regimens chosen according to the underlying malignancy. Increasingly, there are data demonstrating responsiveness of brain metastases to chemotherapy including targeted therapy, such as for patients with lung cancer harboring EGFR mutations that sensitize them to EGFR inhibitors. Immunotherapy may also be effective against those primary tumors that are sensitive to this approach, such as melanoma. Antiangiogenic agents such as bevacizumab are also effective in the treatment of CNS metastases in those primary tumors for which it is approved.

LEPTOMENINGEAL METASTASES

Leptomeningeal metastases are also described as carcinomatous meningitis, meningeal carcinomatosis, or in the case of specific tumors, leukemic or lymphomatous meningitis. Among the hematologic malignancies, acute leukemias most commonly metastasize to the subarachnoid space, followed in frequency by aggressive diffuse lymphomas. Among solid tumors, breast and lung carcinomas and melanoma most frequently spread in this fashion. Tumor cells reach the subarachnoid space via the arterial circulation or occasionally through retrograde flow in venous systems that drain metastases along the bony spine or cranium. In addition, leptomeningeal metastases may develop as a direct consequence of prior brain metastases and occur in almost 40% of patients who have a metastasis resected from the cerebellum.

■ CLINICAL FEATURES

Leptomeningeal metastases are characterized by multilevel symptoms and signs along the neuraxis. Combinations of lumbar and cervical radiculopathies, cranial neuropathies, seizures, confusion, and encephalopathy from hydrocephalus or raised intracranial pressure can be present. Focal deficits such as hemiparesis or aphasia are rarely due

to leptomeningeal metastases unless there is direct brain infiltration. New-onset limb pain in patients with breast cancer, lung cancer, or melanoma should prompt consideration of leptomeningeal spread.

■ LABORATORY AND IMAGING DIAGNOSIS

Leptomeningeal metastases are particularly challenging to diagnose because identification of tumor cells in the subarachnoid compartment may be elusive. MRI can be definitive when there are clear tumor nodules adherent to the cauda equina or spinal cord, enhancing cranial nerves, or subarachnoid enhancement on brain imaging (Fig. 86-8). Imaging is diagnostic in ~75% of patients and is more often positive in patients with solid tumors. Demonstration of tumor cells in the CSF is definitive and often considered the gold standard. However, CSF cytologic examination is positive in only 50% of patients on the first lumbar puncture and still misses 10% after three CSF samples. New

technologies, such as rare cell capture, enhance identification of tumor cells in the CSF. CSF cytologic examination is most useful in hematologic malignancies, especially when combined with flow cytometry to identify a clonal population. Accompanying CSF abnormalities include an elevated protein concentration and an elevated white count; hypoglycorrachia is noted in <25% of patients but is useful when present. Identification of tumor markers may be helpful in some solid tumors.

TREATMENT

Leptomeningeal Metastases

The treatment of leptomeningeal metastasis is palliative because there is no curative therapy. RT to the symptomatically involved areas, such as skull base for cranial neuropathy, can relieve pain and sometimes improve function. Whole-neuraxis RT is avoided because it has significant toxicity with myelosuppression and gastrointestinal irritation as well as limited effectiveness. Systemic chemotherapy with agents that can penetrate the blood-CSF barrier may be helpful. Alternatively, intrathecal chemotherapy can be effective, particularly in hematologic malignancies. This is optimally delivered through an intraventricular cannula (Ommaya reservoir) rather than by lumbar puncture. Few drugs can be delivered safely into the subarachnoid space, and they have a limited spectrum of antitumor activity, perhaps accounting for the relatively poor response to this approach. In addition, impaired CSF flow dynamics can compromise intrathecal drug delivery. Surgery has a limited role in leptomeningeal metastasis; a ventriculoperitoneal shunt can relieve raised intracranial pressure; however, it compromises delivery of chemotherapy into the CSF.



A



B

FIGURE 86-8 Postgadolinium MRI images of extensive leptomeningeal metastases from breast cancer. Nodules along the dorsal surface of the spinal cord (**A**) and cauda equina (**B**) are seen.

EPIDURAL METASTASIS

Epidural metastasis occurs in 3–5% of patients with a systemic malignancy and causes neurologic compromise by compressing the spinal cord or cauda equina. The most common cancers that metastasize to the epidural space are those malignancies that spread to bone, such as breast and prostate. Lymphoma can cause bone involvement and compression, but it can also invade an intervertebral foramen and cause spinal cord compression without bone destruction. The thoracic spine is affected most commonly, followed by the lumbar and then cervical spine.

■ CLINICAL FEATURES

Back pain is the presenting symptom of epidural metastasis in virtually all patients; the pain may precede neurologic findings by weeks or months. The pain is usually exacerbated by lying down; by contrast, arthritic pain is often relieved by recumbency. Leg weakness is seen in ~50% of patients, as is sensory dysfunction. Sphincter problems are present in ~25% of patients at diagnosis.

■ DIAGNOSIS

Diagnosis is established by imaging, preferably with an MRI of the entire spine (Fig. 86-9). Contrast is not required to identify bony or epidural lesions. Any patient with cancer who has severe back pain should undergo an MRI. Plain films, bone scans, or even CT scans may show bone metastases, but only MRI can reliably delineate epidural tumor. For patients unable to have an MRI, CT myelography should be performed to outline the epidural space. The differential diagnosis of epidural tumor includes epidural abscess, acute or chronic hematomas, epidural lipomatosis and rarely, extramedullary hematopoiesis.

TREATMENT

Epidural Metastasis

Epidural metastasis requires immediate treatment. A randomized controlled trial demonstrated the superiority of surgical resection followed by RT compared to RT alone. However, patients must be able to tolerate surgery, and the surgical procedure of choice is

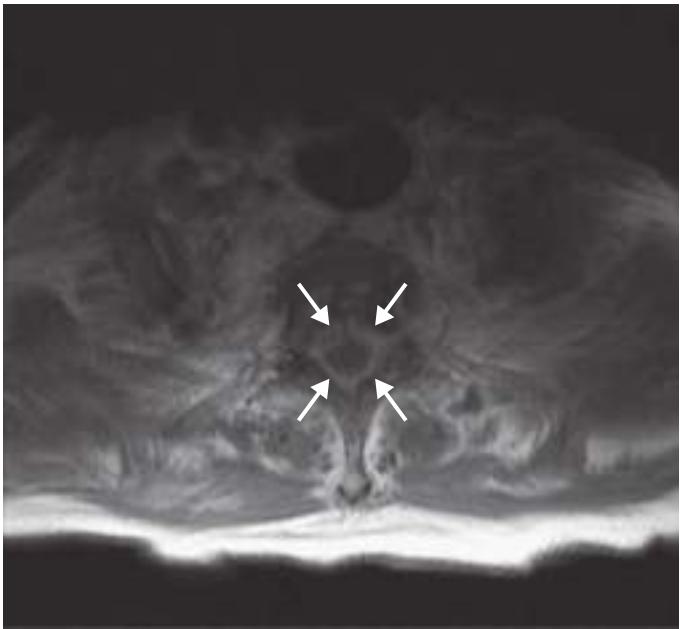


FIGURE 86-9 Postgadolinium T1 MRI showing circumferential epidural tumor around the thoracic spinal cord from esophageal cancer.

a complete removal of the mass, which is typically anterior to the spinal canal, necessitating an extensive approach and resection. Otherwise, RT is the mainstay of treatment and can be used for patients with radiosensitive tumors, such as lymphoma, or for those unable to undergo surgery. SRS is increasingly being used, especially for radioresistant tumor types or for re-irradiation. Chemotherapy is rarely used for epidural metastasis unless the patient has minimal to no neurologic deficit and a highly chemosensitive tumor such as lymphoma or germinoma. Patients generally fare well if treated before there is a severe neurologic deficit. Recovery from paraparesis is better after surgery than with RT alone, but survival is often short due to widespread metastatic tumor.

NEUROLOGIC TOXICITY OF THERAPY

TOXICITY FROM RADIOTHERAPY

RT can cause a variety of toxicities in the CNS. These are usually described based on their relationship in time to the administration of RT: acute (occurring within days of RT), early delayed (months), or late delayed (years). In general, the acute and early delayed syndromes resolve and do not result in persistent deficits, whereas the late delayed toxicities are usually permanent and sometimes progressive.

Acute Toxicity Acute cerebral toxicity may occur during the course of RT to the brain. RT can cause a transient disruption of the blood-brain barrier, resulting in edema and elevated intracranial pressure. This is usually manifest as headache, lethargy, nausea, and vomiting, and can be both prevented and treated with the administration of glucocorticoids. There is no acute RT toxicity that affects the spinal cord.

Early Delayed Toxicity Early delayed toxicity is usually apparent weeks to months after completion of cranial irradiation and is likely due to focal demyelination. Clinically it may be asymptomatic or take the form of worsening or reappearance of a preexisting neurologic deficit. At times a contrast-enhancing lesion can be seen on MRI/CT that can mimic the tumor for which the patient received the RT. For patients with a malignant glioma, this has been described as “pseudoprogression” because it mimics tumor recurrence on MRI but actually represents inflammation and necrotic debris engendered by effective therapy. This is seen with increased frequency when chemotherapy, particularly temozolomide, is given concurrently with RT. Pseudoprogression can resolve on its own or, if very symptomatic, may require resection.

In the spinal cord, early delayed RT toxicity is manifest as a Lhermitte symptom with paresthesias of the limbs or along the spine when the patient flexes the neck. Although frightening, it is benign, resolves on its own, and does not portend more serious problems.

Late Delayed Toxicity Late delayed toxicities are the most serious because they are often irreversible and cause severe neurologic deficits. In the brain, late toxicities can take several forms, the most common of which include radiation necrosis and leukoencephalopathy. Radiation necrosis is a focal mass of necrotic tissue that is contrast enhancing on CT/MRI and may be associated with significant edema. This may appear identical to pseudoprogression but is seen months to years after RT and is always symptomatic. Clinical symptoms and signs include seizures and findings referable to the location of the necrotic mass. The necrosis is caused by the effect of RT on cerebral vasculature with fibrinoid necrosis and occlusion of blood vessels. It can mimic tumor radiographically, but unlike tumor it is typically hypometabolic on a PET scan and has reduced perfusion on perfusion MR sequences. It may require resection for diagnosis and treatment unless it can be managed with glucocorticoids. There are reports of improvement with hyperbaric oxygen or bevacizumab, but symptomatic benefit does not always accompany radiographic improvement.

Leukoencephalopathy is seen most commonly after WBRT as opposed to focal RT. On T2 or FLAIR MR sequences, there is diffusely increased signal seen throughout the hemispheric white matter, often bilaterally and symmetrically. There tends to be a periventricular predominance that may be associated with atrophy and ventricular enlargement. Clinically, patients develop cognitive impairment, a gait disorder, and later urinary incontinence, all of which can progress over time. These symptoms mimic those of normal pressure hydrocephalus, and placement of a ventriculoperitoneal shunt can improve function in some patients but does not reverse the deficits completely. Increased age is a risk factor for leukoencephalopathy but not for radiation necrosis. Necrosis appears to depend on an as yet unidentified predisposition.

Other late neurologic toxicities include endocrine dysfunction if the pituitary or hypothalamus was included in the RT port. An RT-induced neoplasm can occur many years after therapeutic RT for either a prior CNS or a head and neck tumor; accurate diagnosis requires surgical resection or biopsy. In addition, RT causes accelerated atherosclerosis, which can cause stroke either from intracranial vascular disease or carotid plaque from neck irradiation.

The peripheral nervous system is relatively resistant to RT toxicities. Peripheral nerves are rarely affected by RT, but the plexus is more vulnerable. Plexopathy develops more commonly in the brachial than in the lumbosacral distribution. It must be differentiated from tumor progression in the plexus, which is usually visualized by CT/MRI or PET scan demonstrating tumor infiltrating the region. Clinically, tumor progression is usually painful, whereas RT-induced plexopathy is painless. Radiation plexopathy is also more commonly associated with lymphedema of the affected limb. Sensory loss and weakness are seen in both.

TOXICITY FROM CHEMOTHERAPY

Neurotoxicity is second to myelosuppression as the dose-limiting toxicity of chemotherapeutic agents (Table 86-4). Chemotherapy causes peripheral neuropathy from a number of commonly used agents, and the type of neuropathy can vary depending on the drug. Vincristine causes paresthesias but little sensory loss and is associated with motor dysfunction, autonomic impairment (frequently ileus), and, rarely, cranial nerve compromise. Cisplatin causes large fiber sensory loss resulting in sensory ataxia but little cutaneous sensory loss and no weakness. The taxanes also cause a predominantly sensory neuropathy. Agents such as bortezomib and thalidomide also cause neuropathy.

Encephalopathy and seizures are common toxicities from chemotherapeutic drugs. Ifosfamide can cause a severe encephalopathy, which is reversible with discontinuation of the drug and the use of methylene blue for severely affected patients. Fludarabine also causes a severe global encephalopathy that may be permanent. Bevacizumab

Acute encephalopathy (delirium)	Seizures
Methotrexate (high-dose IV, IT)	Methotrexate
Cisplatin	Etoposide (high-dose)
Vincristine	Cisplatin
Asparaginase	Vincristine
Procarbazine	Asparaginase
5-Fluorouracil (\pm levamisole)	Nitrogen mustard
Cytarabine (high-dose)	Carmustine
Nitrosoureas (high-dose or arterial)	Dacarbazine (intraarterial or high-dose)
Ifosfamide	Busulfan (high-dose)
Etoposide (high-dose)	Myelopathy (IT drugs)
Bevacizumab (PRES)	Methotrexate
Chronic encephalopathy (dementia)	Cytarabine
Methotrexate	Thiotepa
Carmustine	Peripheral neuropathy
Cytarabine	Vinca alkaloids
Fludarabine	Cisplatin
Visual loss	Procarbazine
Tamoxifen	Etoposide
Gallium nitrate	Teniposide
Cisplatin	Cytarabine
Fludarabine	Taxanes
Cerebellar dysfunction/ataxia	Suramin
5-Fluorouracil (\pm levamisole)	Bortezomib
Cytarabine	
Procarbazine	

Abbreviations: IT, intrathecal; IV, intravenous; PRES, posterior reversible encephalopathy syndrome.

and other anti-VEGF agents can cause posterior reversible encephalopathy syndrome. Cisplatin can cause hearing loss and less frequently vestibular dysfunction. Immunotherapy with monoclonal antibodies such as ipilimumab or nivolumab can cause an autoimmune hypophysitis, Guillain-Barré syndrome, or an autoimmune encephalitis.

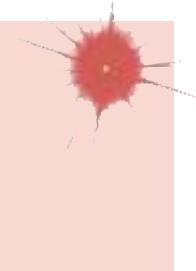
FURTHER READING

- BUCKNER JC et al: Radiation plus procarbazine, CCNU, and Vincristine in low-grade glioma. *N Engl J Med* 374:1344, 2016.
- CANCER GENOME ATLAS RESEARCH NETWORK et al: Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* 372:2481, 2015.
- LOUIS DN et al: The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol* 131:803, 2016.
- MORIKAWA A et al: Characteristics and outcomes of patients with breast cancer with leptomeningeal metastasis. *Clin Breast Cancer* 17:23, 2017.
- OMURO A et al: R-MVP followed by high-dose chemotherapy with TBC and autologous stem-cell transplant for newly diagnosed primary CNS lymphoma. *Blood* 125:1403, 2015.
- PENTSOVA EI et al: Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. *J Clin Oncol* 34:2404, 2016.
- RAMASWAMY V et al: Risk stratification of childhood medulloblastoma in the molecular era: The current consensus. *Acta Neuropathol* 131:821, 2016.
- THIBAULT I et al: Response assessment after stereotactic body radiotherapy for spinal metastasis: A report from the SPIne response assessment in Neuro-Oncology (SPINO) group. *Lancet Oncol* 16:e595, 2015.
- TSAKONAS G et al: Management of brain metastasized non-small cell lung cancer (NSCLC)—From local treatment to new systemic therapies. *Cancer Treat Rev* 54:122, 2017.
- YAMADA Y et al: The impact of histology and delivered dose on local control of spinal metastases treated with stereotactic radiosurgery. *Neurosurg Focus* 42:E6, 2017.

87

Soft Tissue and Bone Sarcomas and Bone Metastases

Shreyaskumar R. Patel



Sarcomas are rare (<1% of all malignancies) mesenchymal neoplasms that arise in bone and soft tissues. These tumors are usually of mesodermal origin, although a few are derived from neuroectoderm, and they are biologically distinct from the more common epithelial malignancies. Sarcomas affect all age groups; 15% are found in children <15 years of age, and 40% occur after age 55 years. Sarcomas are one of the most common solid tumors of childhood and are the fifth most common cause of cancer deaths in children. Sarcomas may be divided into two groups, those derived from bone and those derived from soft tissues.

SOFT TISSUE SARCOMAS

Soft tissues include muscles, tendons, fat, fibrous tissue, synovial tissue, vessels, and nerves. Approximately 60% of soft tissue sarcomas arise in the extremities, with the lower extremities involved three times as often as the upper extremities. Thirty percent arise in the trunk, the retroperitoneum accounting for 40% of all trunk lesions. The remaining 10% arise in the head and neck.

INCIDENCE

Approximately 12,310 new cases of soft tissue sarcomas occurred in the United States in 2016. The annual age-adjusted incidence is 3 per 100,000 population, but the incidence varies with age. Soft tissue sarcomas constitute 0.7% of all cancers in the general population and 6.5% of all cancers in children.

EPIDEMIOLOGY

Malignant transformation of a benign soft tissue tumor is extremely rare, with the exception that malignant peripheral nerve sheath tumors (neurofibrosarcoma, malignant schwannoma) can arise from neurofibromas in patients with neurofibromatosis. Several etiologic factors have been implicated in soft tissue sarcomas.

Environmental Factors Trauma or previous injury is rarely involved, but sarcomas can arise in scar tissue resulting from a prior operation, burn, fracture, or foreign body implantation. Chemical carcinogens such as polycyclic hydrocarbons, asbestos, and dioxin may be involved in the pathogenesis.

Iatrogenic Factors Sarcomas in bone or soft tissues occur in patients who are treated with radiation therapy. The tumor nearly always arises in the irradiated field. The risk increases with time.

Viruses Kaposi's sarcoma (KS) in patients with HIV type 1, classic KS, and KS in HIV-negative homosexual men is caused by human herpesvirus (HHV) 8 (Chap. 190). No other sarcomas are associated with viruses.

Immunologic Factors Congenital or acquired immunodeficiency, including therapeutic immunosuppression, increases the risk of sarcoma.

GENETIC CONSIDERATIONS

Li-Fraumeni syndrome is a familial cancer syndrome in which affected individuals have germline abnormalities of the tumor-suppressor gene *p53* and an increased incidence of soft tissue sarcomas and other malignancies, including breast cancer, osteosarcoma, brain tumor, leukemia, and adrenal carcinoma (Chap. 67). Neurofibromatosis 1 (*NF-1*, peripheral form, von Recklinghausen's disease) is characterized by multiple neurofibromas and café-au-lait spots. Neurofibromas occasionally undergo malignant degeneration to become malignant peripheral nerve sheath tumors. The gene for *NF-1*

is located in the pericentromeric region of chromosome 17 and encodes neurofibromin, a tumor-suppressor protein with guanosine 5'-triphosphate (GTP)ase-activating activity that inhibits Ras function (**Chap. 86**). Germline mutation of the *Rb-1* locus (chromosome 13q14) in patients with inherited retinoblastoma is associated with the development of osteosarcoma in those who survive the retinoblastoma and of soft tissue sarcomas unrelated to radiation therapy. Other soft tissue tumors, including desmoid tumors, lipomas, leiomyomas, neuroblastomas, and paragangliomas, occasionally show a familial predisposition.

Ninety percent of synovial sarcomas contain a characteristic chromosomal translocation $t(X;18)(p11;q11)$ involving a nuclear transcription factor on chromosome 18 called *SYT* and two breakpoints on X. Patients with translocations to the second X breakpoint (*SSX2*) may have longer survival than those with translocations involving *SSX1*.

Insulin-like growth factor (IGF) type II is produced by some sarcomas and may act as an autocrine growth factor and as a motility factor that promotes metastatic spread. IGF-II stimulates growth through IGF-I receptors, but its effects on motility are through different receptors. If secreted in large amounts, IGF-II may produce hypoglycemia (**Chaps. 89 and 399**). A large international sarcoma kindred study including 1162 patients and 6545 Caucasian controls revealed that about half the patients with sarcoma have putatively pathogenic monogenic and polygenic variation in previously reported and new cancer genes, some of them representing therapeutically actionable targets. These patients were diagnosed with sarcoma at an earlier age compared to controls.

CLASSIFICATION

Approximately 20 different groups of sarcomas are recognized on the basis of the pattern of differentiation toward normal tissue. For example, rhabdomyosarcoma shows evidence of skeletal muscle fibers with cross-striations; leiomyosarcomas contain interlacing fascicles of spindle cells resembling smooth muscle; and liposarcomas contain adipocytes. When precise characterization of the group is not possible, the tumors are called *unclassified sarcomas*. All of the primary bone sarcomas can also arise from soft tissues (e.g., extraskeletal osteosarcoma). The entity *malignant fibrous histiocytoma* (MFH) includes many tumors previously classified as fibrosarcomas or as pleomorphic variants of other sarcomas and is characterized by a mixture of spindle (fibrous) cells and round (histiocytic) cells arranged in a storiform pattern with frequent giant cells and areas of pleomorphism. As immunohistochemical suggestion of differentiation, particularly myogenic differentiation, may be found in a significant fraction of these patients, many are now characterized as poorly differentiated leiomyosarcomas, and the terms *undifferentiated pleomorphic sarcoma* (UPS) and *myxofibrosarcoma* are replacing MFH and myxoid MFH.

For purposes of treatment, most soft tissue sarcomas can be considered together. However, some specific tumors have distinct features. For example, *liposarcoma* can have a spectrum of behaviors. Pleomorphic liposarcomas and dedifferentiated liposarcomas behave like other high-grade sarcomas; in contrast, well-differentiated liposarcomas (better termed *atypical lipomatous tumors*) lack metastatic potential, and myxoid liposarcomas metastasize infrequently, but, when they do, they have a predilection for unusual metastatic sites containing fat, such as the retroperitoneum, mediastinum, and subcutaneous tissue. Rhabdomyosarcomas, Ewing's sarcoma, and other small-cell sarcomas tend to be more aggressive and are more responsive to chemotherapy than other soft tissue sarcomas.

Gastrointestinal stromal tumors (GISTs), previously classified as gastrointestinal leiomyosarcomas, are now recognized as a distinct entity within soft tissue sarcomas. Its cell of origin resembles the interstitial cell of Cajal, which controls peristalsis. The majority of malignant GISTs have activating mutations of the *c-kit* gene that result in ligand-independent phosphorylation and activation of the KIT receptor tyrosine kinase, leading to tumorigenesis. Approximately 5–10% of tumors will have a mutation in the platelet-derived growth factor receptor α (*PDGFRA*). GISTs that are wild type for both *KIT* and *PDGFRA* mutations may show mutations in *SDH B, C, or D* and may be driven by the IGF-I pathway.

DIAGNOSIS

The most common presentation is an asymptomatic mass. Mechanical symptoms referable to pressure, traction, or entrapment of nerves or muscles may be present. All new and persistent or growing masses should be biopsied, either by a small incision or by a cutting needle (core-needle biopsy) placed so that it can be encompassed in the subsequent excision without compromising a definitive resection. Lymph node metastases occur in 5%, except in synovial and epithelioid sarcomas, clear-cell sarcoma (melanoma of the soft parts), angiosarcoma, and rhabdomyosarcoma, where nodal spread may be seen in 17%. The pulmonary parenchyma is the most common site of metastases. Exceptions are GISTs, which metastasize to the liver; myxoid liposarcomas, which seek fatty tissue; and clear-cell sarcomas, which may metastasize to bones. Central nervous system metastases are rare, except in alveolar soft part sarcoma.

Radiographic Evaluation Imaging of the primary tumor is best with plain radiographs and magnetic resonance imaging (MRI) for tumors of the extremities or head and neck and by computed tomography (CT) for tumors of the chest, abdomen, or retroperitoneal cavity. A radiograph and CT scan of the chest are important for the detection of lung metastases. Other imaging studies may be indicated, depending on the symptoms, signs, or histology.

STAGING AND PROGNOSIS

The histologic grade, relationship to fascial planes, and size of the primary tumor are the most important prognostic factors. The current American Joint Committee on Cancer (AJCC) staging system is shown in **Table 87-1**. Prognosis is related to the stage. Cure is common in the absence of metastatic disease, but a small number of patients with metastases can also be cured. Historically, most patients with stage IV disease used to die within 12 months, but with availability of multiple lines of treatments, median survival in second-line and beyond ranges from 13 to 14 months, and some patients may live with stable or slowly progressive disease for many years.

TREATMENT

Soft Tissue Sarcomas

AJCC stage I patients are adequately treated with surgery alone. Stage II patients are considered for adjuvant radiation therapy. Stage III patients may benefit from adjuvant chemotherapy. Stage IV patients are managed primarily with chemotherapy, with or without other modalities.

SURGERY

Soft tissue sarcomas tend to grow along fascial planes, with the surrounding soft tissues compressed to form a pseudocapsule that gives the sarcoma the appearance of a well-encapsulated lesion. This is invariably deceptive because “shelling out,” or marginal excision, of such lesions results in a 50–90% probability of local recurrence. Wide excision with a negative margin, incorporating the biopsy site, is the standard surgical procedure for local disease. The adjuvant use of radiation therapy and/or chemotherapy improves the local control rate and permits the use of limb-sparing surgery with a local control rate (85–90%) comparable to that achieved by radical excisions and amputations. Limb-sparing approaches are indicated except when negative margins are not obtainable, when the risks of radiation are prohibitive, or when neurovascular structures are involved so that resection will result in serious functional consequences to the limb.

RADIATION THERAPY

External-beam radiation therapy is an adjuvant to limb-sparing surgery for improved local control. Preoperative radiation therapy allows the use of smaller fields and smaller doses but results in a higher rate of wound complications. Postoperative radiation therapy must be given to larger fields, because the entire surgical bed must be encompassed, and in higher doses to compensate for hypoxia in