

CHAPTER 2

DIAGNOSTIC DECISIONS

CASE: THE CLOT THICKENS

It is a busy Monday at 3 p.m., and Dr Wayne is signing out his patients to you. You both walk over to bed #10 where a young woman looks up from the gurney. He tells you that he picked up the chart of this 25-year-old woman just 15 minutes ago, but the case is an easy one. She has just returned home after an eight-hour plane ride and is now complaining of a “little” shortness of breath and some reproducible chest pain. He goes on to tell you that he ordered a V/Q scan, which has not yet been done. He advises: “If it comes back normal or low probability, send her home. If it comes back intermediate or high, just put her on heparin and admit her as a confirmed PE.” He finishes sign-out and leaves the department to you. You decide to start over with bed #10. *To be continued...*

MAKING THE DIAGNOSIS

In the emergency department, we are expected to diagnose potentially life-threatening illnesses based on scant information in limited amounts of time. Our patients are often unknown to us prior to their presentation and they are not at their best or most cooperative. Inundated with stimuli, distractions, and chaos, we practice in an environment far from the ideal for contemplation. Rarely do we have the luxury of devoting our full attention to any individual decision, but must instead juggle numerous cognitive processes at once.

To compound these difficulties, we often lack the assurance of certainty; we instead treat based on the “most likely.” How we drive our decisions to this point of acceptable likelihood can be a combination of our clinical judgment and diagnostic testing.

Experienced clinicians make an enormous number of decisions in the course of a single shift in the emergency department. Given this density of decisions, we make surprisingly few errors. However, a large number of the errors that lead to poor patient outcome and malpractice suits are cognitive errors of diagnostic decision-making.¹

■ Ill-structured Domain

From the perspective of decision-making, medicine is considered an ill-structured domain; this is in contrast to well-structured fields such as mathematics and the hard sciences. Well-structured domains are defined by rules that translate from one decision to the next.² The problems of medical decisions in general are only compounded in emergency medicine.

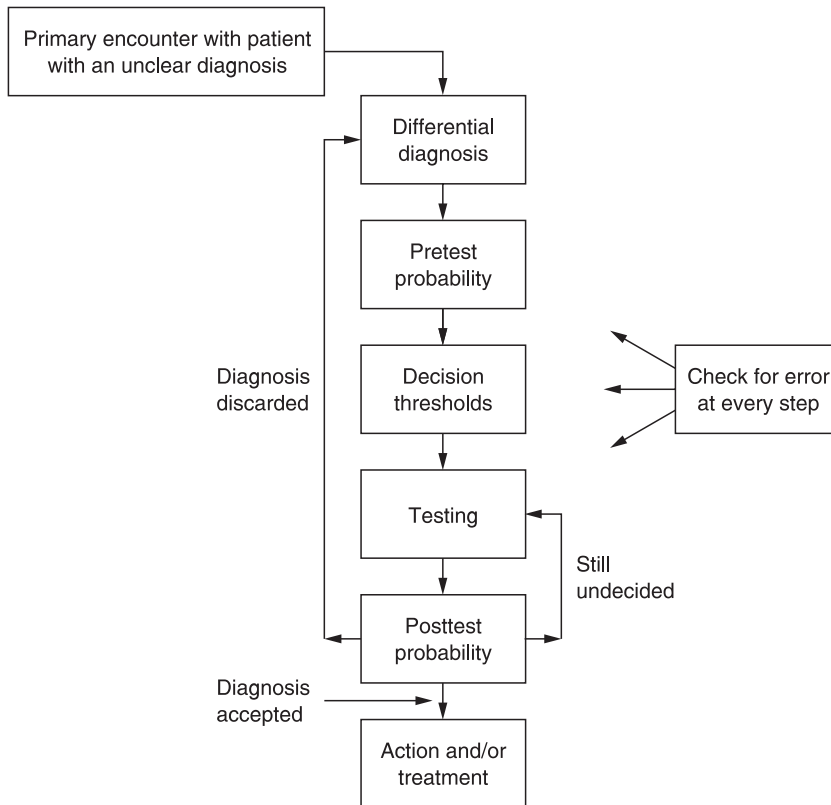
Components of the Ill-structured Domain of Emergency Medicine Decisions

- Complete information is often not available at the onset of the decision-making process. At times, complete information will never become available.
- Problems are dynamic; changes occur even in the midst of the diagnostic process.
- Approaches to problem-solving are often not generalizable to future patients.
- We lack the feedback to know if our decisions were correct; often the final diagnosis is decided only after the patient has left our care.

For all of these reasons, our decisions are more difficult and we require increased vigilance against cognitive pitfalls. By utilizing an approach to diagnostic decisions embracing critical thinking, we can minimize our potential to fall prey to the above errors. In the subsequent sections, we discuss each step of the process of diagnostic decision-making. We will then discuss the potential sources of error inherent in the process (Fig. 2-1).

DIFFERENTIAL DIAGNOSIS

The creation of a differential diagnosis is directly dependent on a clinician's skills and depth of experience. If we do not consider a disease as a possibility, even a full armamentarium of evidence-based medicine may not prevent us from missing the diagnosis.



— FIGURE 2-1 — *The steps of diagnostic decision-making.*

A differential diagnosis is formed based on a history, a focused physical exam, and possibly a limited array of bedside tests, such as fingerstick glucose, 12-lead electrocardiograms, etc. We must critically analyze the data from these assessments and measurements and utilize them carefully to avoid error and bias.

■ Strategies for Forming a Differential

Exhaustive Method

As medical students, we arrived at our differentials via a laborious history, an all-inclusive review of systems, and an extensive physical exam. We then proceeded to list every possible cause of each stated symptom and discovered sign. This might have required a search through a backbreaking

general medicine text looking for all of the possible disease entities consistent with each and every finding. Often we would follow this extensive differential with a suggestion for the scores of tests we should order to narrow down the compendium of possibilities.

We also refer to this strategy as “shotgunning” or the “blunderbuss method.”³ This exhaustive process is *inductive*; i.e. we gather information and then use it to formulate a hypothesis.^{4,5}

Unfortunately, even experienced clinicians, under the rubric of defensive medicine, sometimes utilize the exhaustive method. This tendency may be reinforced during residency training; residents on medical wards frequently order tests only because they know they will be asked about the results on rounds. They may perceive it to be more expedient to simply order the test than to have to argue that they considered that the likelihood of the relevant target disorder was too low to justify testing for it. This leads to costly, unnecessary testing of no benefit to our patients. It also blunts our ability to think analytically about the use of diagnostic testing.

While it may seem counterintuitive, the history and physicals of expert clinicians are actually shorter than those of novices.⁶ This brevity is not the result of carelessness, but instead reflects the ability of the seasoned physician to focus the interview while in the midst of it. The following diagnostic method explains this process.

Hypothetico-deductive Method

As we gather experience, the exhaustive strategy is replaced by a hypothetico-deductive method.^{4,7} The patient’s chief complaint guides our history-taking and systems review, which we then use to focus our physical exam. We are constantly considering and discarding diagnoses even as we are gathering further information to formulate a differential. As the name would indicate, this is a *deductive process*; we form a hypothesis and then we gather data to support or refute it.⁵ In this way, we can create a list of differential diagnoses with a minimum of information. Kassirer termed this strategy the “method of steepest ascent”; a skilled emergency physician can come to the correct diagnosis with an extremely limited data set.⁸

Illness Scripts

The means by which we decide if a diagnosis is plausible is by comparing the current patient’s presentation with our stored illness scripts.⁹ Illness scripts are chunks of information about the presentation of a disease, including symptoms, signs, and the mental picture of patients previously encountered with that disease.⁶ The illness script is a culmination of all of the patients we have seen with the disorder as well as any virtual exposures

we have garnered through readings and case presentations. As our experience grows, we assimilate atypical presentations into the illness scripts. The scripts are stored in our long-term memory; when a patient presents, we move them into our short-term memory. Due to the limits of short-term memory, we generally can consider only three to five scripts at a time to see if they match our current patient's presentation.³ We eliminate the scripts that do not match to allow others into consideration.

The hypothetico-deductive method is much quicker than the exhaustive strategy. We begin forming hypotheses as soon as we pick up the chart. Most presentations can be matched against one or a few illness scripts within one to seven minutes.¹¹ The rapidity with which we form our differential is indicative of our assurance. In other words, when we can form a differential diagnosis quickly, we are more likely to be correct.^{2,6,12–13}

Backwards versus Forward Reasoning

The hypothetico-deductive method is an example of *backward reasoning*. This is something of a misnomer, because backward thinking is not outdated or inferior as the name may suggest, but instead is the quicker and more accurate means of clinical diagnosis.¹⁴ The “backwards” refers to the fact that the situation immediately suggests diagnoses; we then work backwards to see if they fit. The inferior mode of forward reasoning is demonstrated by the exhaustive methods of novices.

Some diagnoses are made within moments of seeing a patient, an EKG, or a radiograph; the method that allows for this instantaneous provisional diagnosis is *pattern matching*.

Pattern Matching

When we see ST-elevations on an EKG, or an ashen, diaphoretic patient clutching his or her chest, we do not engage in extensive cognitive processing. Striking patient appearance or pathognomonic presentations can immediately suggest a diagnosis. However, clinicians experienced with a particular disease may effectively rely on pattern recognition even for atypical presentations. Diagnoses formed by pattern matching occur within seconds of exposure and require little conscious thought.^{2,6}

Pattern matching is operative mostly for visual diagnoses such as rashes, radiographs, and patient appearance. It is also used for haptic pathology such as the pulsating mass in the abdomen of a patient with an abdominal aortic aneurysm. Illness scripts may still play a role as the source of the pattern to which we match the patient presentation. In fact, if we are already considering a diagnosis, we are more likely to be able to appreciate the stimulus which cues the pattern matching.^{15,16}

EXAMPLE

If a patient's triage sheet documents a history of renal failure and five days of missed dialysis, our ability to immediately pattern match the electrocardiogram to the diagnosis of hyperkalemia is augmented.

Recognition-primed Decision Model

Klein first described this model, which emerged from studies of the critical decisions made by firefighters and military personnel.¹⁷ The process allows the clinician to rapidly set diagnostic priorities, recognize and filter diagnostic cues, project what is expected on further examination, and mentally test numerous diagnoses to see if they fit with the patient presentation. It is essentially a distillation of clinical judgment to rapid decision-making. Psychologists describe similar processes as *gestalt*. The recognition-primed decision model integrates the previous three methods we have just discussed. As can be seen in Fig. 2-2, depending on the complexity of the presentation, we shift between the various decision-making strategies.

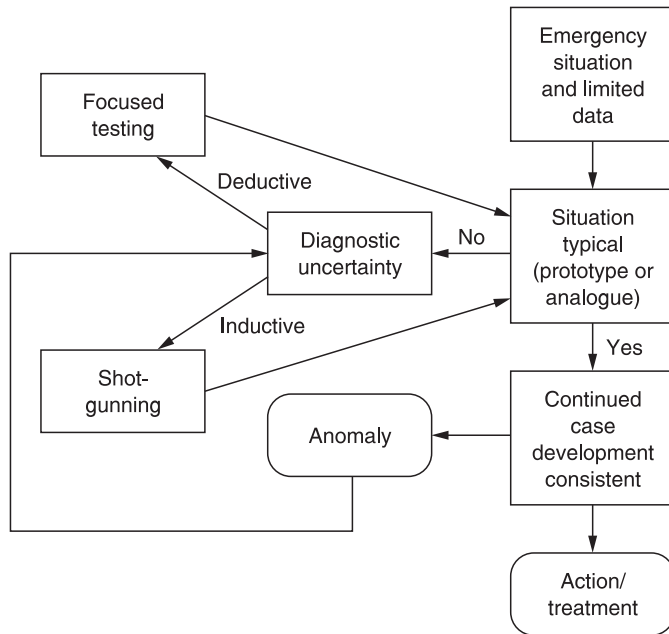
■ Literature Support

Reports of Clinical Manifestations of Disease This type of study assesses the frequency of various signs and symptoms among series of patients independently identified as having a target disease.¹⁸ By using the data from this research, we can decide whether a disease we are considering accounts for all of the important elements of a patient's presentation. If it does not, an alternative or additional diagnosis may need to be entertained.

Differential Diagnosis Lists There are whole textbooks and review articles, which provide a precompiled list of differentials for any given sign or symptom. This form of literature can also be helpful to suggest possibilities that we might not have considered or remembered.^{19,20}

■ Summary

The differential diagnosis provides a list of diagnostic hypotheses. After generating this list, we must systematically evaluate each possibility to determine whether it merits active consideration. We may be confident in quickly dismissing many of the possibilities on the basis of our illness scripts or pattern matching. If we feel obligated to actively consider more



— FIGURE 2-2 — *Recognition primed decision-making.*

than one diagnosis, we need to use validated clinical criteria, laboratory testing or imaging to further narrow the possibilities. The next sections provide a pathway to critically evaluating the possible diagnoses to come to the truth and avoid error.

THE STATISTICS OF LIKELIHOOD

Before we delve into further discussion of the art of diagnostic decision-making, we must digress for a quick review of the mathematics of likelihood.

■ Likelihood

Likelihood is the soul of diagnostic decision-making. It is traditionally expressed in two ways: probability and odds. Each has its uses, so we must be proficient with both.

Probability

Probability is the measurement of likelihood most familiar to us. It is the ratio of one outcome to all outcomes.

$$\text{Probability} = \frac{\text{Particular outcome}}{\text{All outcomes including that particular outcome}}$$

It is usually expressed as a percentage from 0 to 100%. To change the equation above to a percentage, simply multiply by 100. If doing calculations with probability, it should remain as a decimal ranging from 0 to 1.

EXAMPLE

You want to see how many times heads comes up when you flip a coin 100 times. If it comes up 50 times, then you can calculate the probability as:

$$\frac{50 \text{ heads}}{100 \text{ flips}} = 0.5 = 50\%$$

The probability would be 0.5 or 50%.

Odds

Odds are less familiar to physicians, but aficionados of horseracing and Las Vegas excursions are quite used to the terminology. Odds describe the possibility of an event as a *ratio* of one outcome and all of the outcomes that are not that outcome.

$$\text{Odds} = \frac{\text{One outcome}}{\text{All other outcomes except that outcome}}$$

In Las Vegas, the casinos express odds as a number to one. For example, the odds against a ball stopping on the number 19 on the roulette wheel are 35 to 1.* In medicine, we just eliminate the one; we would call those same roulette wheel odds 35.

*The roulette wheel odds are a bit misleading, as the wheel contains the numbers 0 through 36. This would make the actual odds against the number nineteen 36 to 1, but the odds (and therefore the payoff for a winning bet) are listed as 35 to 1 to keep the casino's coffers full. Hence, no one familiar with likelihood plays roulette.

EXAMPLE

If the same coin flip experiment as above is expressed in odds, it would look like:

$$\text{Odds of heads} = \frac{50 \text{ heads}}{50 \text{ not heads (tails)}} = 1.$$

Odds of 1, otherwise known as *even odds*, are the same as a probability of 50%. Any probability less than 50% will be represented by odds between 0 and 1. Probabilities greater than 50% are represented by odds greater than 1. As probabilities get smaller and smaller, they begin to match the odds. For instance, a probability of 10% (0.1) is the same as odds of 0.11. Odds of 20 are the same as a probability of 95%; odds of 100 are equivalent to a probability of 99%. Since odds can go up to infinity, for most decisions we consider odds greater than 100 to be 100% probability.

Converting between odds and probability is quite simple, using these equations:

$$\text{Odds} = \frac{\text{Probability}}{1 - \text{Probability}} \quad \text{Probability} = \frac{\text{Odds}}{1 + \text{Odds}}$$

At this point, you might be questioning why we ever need to think in terms of odds, since probability is more familiar, and often more intuitive. As we will be discussing in subsequent sections, many types of association can only be measured using odds.

PRE-TEST PROBABILITY

Pre-test probability takes the comprehensive list of the differential diagnoses and assigns likelihood to each.

■ Probabilistic Thinking

In the previous section, we discussed the generation of a differential diagnosis; we now must decide the likelihood of each disease on the list. This method of probabilistic thinking will guide the further diagnostic process. We call this likelihood the *pre-test probability*. As the name would indicate, it requires the assignment of probability *before* we perform further diagnostic testing.

It may seem counterintuitive to assign a value to likelihood prior to testing. It evokes the clamoring of “Is that not why we are doing the test in the first place?” Given the presence of perfect diagnostic tests,

it would be a valid complaint. The problem is that very few, if any, of our diagnostic tests are perfect, so we must interpret all of them in light of our pre-test probability. This is an adaptation of the principles of *Bayesian analysis*.

Bayesian Analysis

The reverend Thomas Bayes was an English mathematician in the 1700s. His theorem discussed the alteration of prior probability by new events. After these events, a new post-event probability is created. All of our diagnostic decisions are an adaptation of this theorem. An “event,” in the context of clinical diagnostic reasoning, may be a patient’s answer to a question we ask, information from our physical examination, or the result of a diagnostic test we perform on the patient. We can express the use of Bayes’ theorem in this context with the equation:

Pretest probability × Effects of diagnostic test result = Post-test probability.

The application of this theorem takes what is often an informal process:

I think he may have appendicitis . . .
... so I’ll get a CT scan.
CT shows appendicitis . . .
... so I’ll call the surgeon.

and adds structure, a pathway for critical examination of our thought processes. Mainly, the theorem makes clear that our interpretation of assessments and tests must depend not only on the results, but also on a factor that is entirely independent of the result. We call this the “pre-test” or “pre-assessment” probability of the condition we are considering. This formal decision-making process can appear onerous to perform in our frantic environment, yet we are already performing each step, but perhaps not with the necessary awareness and rigor.

EXAMPLE

You have a patient complaining of acute pain to his right ankle which you note on exam is swollen and painful. Before you had a chance to evaluate the patient, the triage nurse sent a white blood cell (WBC) count. You notice the result of the WBC count is $14 \times 10^3/\text{mL}$. If the patient was asymptomatic until twisting his ankle at a soccer game 30 minutes prior to arrival, your pre-test probability for infection would be extremely low. If, instead, the patient was complaining of a fever, your pre-test probability for an infectious cause would be much higher. How we interpret the WBC count, an imperfect test for

infection, is directly dependent on pre-test probability. If, however, the WBC count was a perfect test for infection,* then pretest probability would be irrelevant and a reevaluation of your soccer player for an infectious source would have to be undertaken. This also illustrates the dangers of sending tests before determining their need and utility.

To use Bayesian analysis to its fullest advantage, we need to estimate our pre-test probability. It is often easier to say that a diagnosis is “unlikely, but possible” than to say your pre-test probability is 10%. The risk stratification of low, moderate, or high probability is often familiar to emergency physicians. This level of pre-test stratification will suffice for most diagnostic decisions. To use these qualitative descriptions requires us to already have familiarity with the situation and the abilities of our diagnostic tests.

Committing to an estimate of pre-test probability allows us to explore the power of new tests and deal with situations we have not thought through in the past. The formal decision-making process we will discuss in subsequent chapters requires a quantitative assessment of pre-test probability. Defining an estimate of pre-test probability also permits subsequent interpretation by other clinicians.

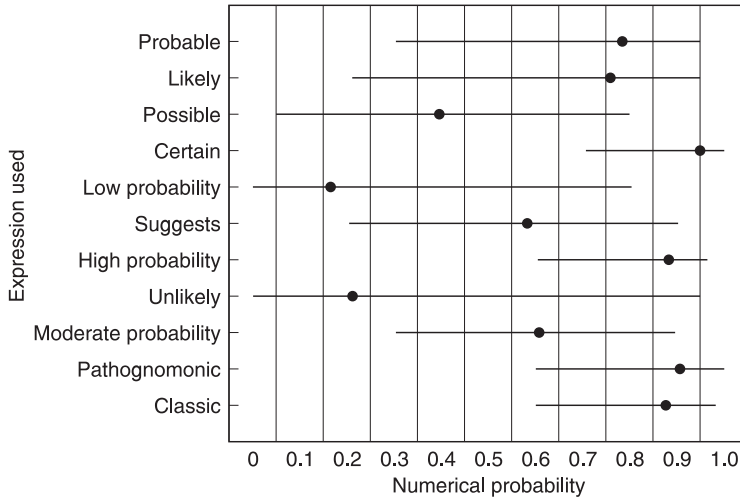
A written notation of moderate pre-test probability is far easier for the physicians on the floor to utilize than jotting down “I sort of suspect he has a dissection.” Even more useful is the commitment to a 35% pre-test probability in the chart.²¹

A study by Bryant illustrates the difficulties of using qualitative words in lieu of numeric probability. In this study, physicians were asked to express numeric equivalents of the words we often use to express pre-test probability.²² The wide ranges in Fig. 2-3 show that the words led to very different assignments of probability amongst the study participants.

The study provides an estimate of the probability ranges to which clinicians frequently assign qualitative estimates of the various descriptive terms:

- Low probability ~10%
- Unlikely ~20%
- Possible ~40%
- Suggestive/moderate ~50%
- Likely/probable ~80%
- Pathognomonic/classic ~90%.

*In this context, a perfect test refers to one that will always be negative in the absence of disease and will always be positive when the patient has the disease. Perfect tests and assessments are exceedingly rare, as we discuss in subsequent sections.



— FIGURE 2-3 — *Ranges of probability associated with qualitative descriptions. The points are the means and the lines are ranges of the physician responses.*

It is not important if the estimate of pre-test probability is off by a small amount, it is only a problem when it is off by orders of magnitude. There will be very little difference if a pre-test probability is estimated at 70% rather than 60%.

EXAMPLE

One of your colleagues from the medicine service stops by the department to give you follow-up on a patient admitted last week with abdominal pain. It turns out the patient had acute intermittent porphyria. In an attempt to avoid missing this diagnosis with future patients, you vow to consider this in the differential of every patient presenting with abdominal pain. Since it is in the differential, you decide to assign a pre-test probability of 10% to the diagnosis in the next patient you see with abdominal pain, despite the absence of any suggestive symptoms or signs. A true estimate of the pre-test probability of this disease in undifferentiated abdominal pain is closer to 0.001%.²³

It is fine to consider rare diagnoses, but this consideration should be tempered with accurate estimations of pre-test likelihood. The evidence for their consideration (from our history, physical, and bedside testing)

should be compelling and they require the use of the threshold approach to decision-making, which we discuss shortly.

If we are considering more than one diagnosis, the combination of pre-test probabilities should total 100%. Assigning a patient a 50% pre-test probability of appendicitis, 50% for gas pain, and 45% for diverticulitis offends the tenets of mathematics and evidence-based medicine alike.

■ Aids to the Estimation of Pre-test Probability

Literature Support

The literature can guide us in the estimation of pre-test probability, thereby removing some of the difficulties of quantitative assignment. Ninety percent of patients in a survey done by Richardson had conditions in which evidence existed in the literature to guide the estimation of pre-test probability.²⁴ Published studies can serve as an aid to the estimation of pre-test probability in three ways.

Published Prevalence

If we can find diagnostic studies with patient populations similar to the patient we are evaluating, we then have a powerful tool for determining pre-test likelihood. By looking at the final number of patients who had gold-standard evidence of a disease, we can obtain a point of departure for estimating the pre-test probability of our individual patient.

EXAMPLE

You have a young patient who presents with the worst headache of her life. You remember reading an article on the use of CT scans to diagnose subarachnoid hemorrhage (SAH). The article had 17% of its patients with the worst headache of their life diagnosed with SAH with lumbar puncture evidence or angiographic evidence.²⁵ Starting with 17% as the average likelihood of these patients, you can consider whether your own patient may have a somewhat greater or smaller individual pre-test probability.

Studies of Differential Diagnosis

This form of literature examines a particular presentation and reports the diseases, which the study patients were eventually proven to have. These studies can give us a rough estimate of pre-test probability.

EXAMPLE

A study is performed to discover the eventual diagnosis of 200 patients presenting with syncope.²⁶ Fifty are found to have a cardiovascular cause, 50 are found to have a neurogenic cause, and the remainder have no etiology found for their loss of consciousness. Based on this study, we can set our pre-test probability for a cardiac cause at approximately 25% if a patient presents with syncope.

Clinical Prediction Rules

The second way the literature can aid pre-test probability assignment is by establishing clinical prediction rules (CPRs) that yield pre-test probability stratification. This form of literature identifies independent signs and symptoms that would indicate the presence of the disease. These signs and symptoms are then validated through additional studies to assure that the components of the CPR accurately predict disease probability. By assigning points to these components, an estimation of pre-test probability is generated.^{27,28} When this form of evidence is available, it provides the highest level of literature support. Clinical prediction rules are discussed more extensively in Part II.

Link to bottom Page 228

EXAMPLE

You have a patient with unilateral leg swelling and recent surgery. An article by Wells allows points assigned to these characteristics to place the patient in a moderate pre-test probability group with a range of 40–50%.²⁹

Using literature aids increases the precision of our pre-test probabilities. Studies show a lack of reproducibility when physicians are asked to evaluate pre-test probabilities using gestalt.³⁰ However, the accuracy of the estimates of *experienced clinicians* in disorders such as pulmonary embolism and acute coronary syndromes is quite good.³¹ The problem with relying on gestalt is that we may not be as experienced with a disorder as we may think, unless we have done a self-comparison of our estimates and a decision rule. For this reason, if a prediction rule or other literature support is available, then it is beneficial to consider the pre-test probability it generates.

We do not need to disregard our gestalt estimate to use these literature-based aids, but instead the two should work in tandem to arrive at the most accurate pre-test probability for the unique patient presentation.

Computer-aided Pre-test Probability

Work is being done on complex Bayesian networks, as well as artificial intelligence to aid diagnostic decision-making. While these technologies are in their early stages, computer aid is currently available for the estimation of pre-test, probably in the form of large patient databases. If the patient's demographic data and symptomatology are cross-referenced with a large database of patients with known disease status, a pre-test probability can be derived.

■ Deciding upon the Pre-test Probability

At first, it may seem daunting to choose an exact number for our estimate of pre-test probability. Often, emergency physicians choose to use only pre-test risk categories. If pre-test probability can be qualitatively split into low, moderate, and high probability, the diagnostic process can still go forward.

EXAMPLE

A reliable practice guideline informs us that we may use an ELISA d-dimer assay in patients we suspect have pulmonary embolism, if we estimate them to be in the low pre-test probability group.³²

If validated clinical guidelines are available, then this strategy is even more viable.

EXAMPLE

A validated strategy for the evaluation of pulmonary embolism offers a path for the work-up of patients placed into low, moderate, or high risk groups.³³ Numeric estimations are unnecessary, because each of the qualitative categories has an associated diagnostic pathway.

For some diseases, no clear literature exists to provide us with *diagnostic strategies* based on qualitative risk stratification. In these situations, we *need* to use numeric estimates. If we master the skills necessary to make these quantitative estimations, we then have the ability to evaluate *any* diagnostic test or process on our own. The methods we use to make this evaluation are the subject of the next section.

DECISION THRESHOLDS

Just as important as the assignment of accurate pre-test probabilities is choosing the probabilities at which we will discard a diagnosis or accept it. We will refer to these probabilities as “decision thresholds.” Pauker and Kassirer first described the use of the threshold approach in diagnostic decision-making.³⁴

In an ideal world, we would accept a diagnosis only with a probability of 100% and would reject it only with a probability of 0%. And we would have cheap, quick, non-invasive tests that could easily achieve this level of certainty in all situations. Such a high level of assurance is rarely achievable in medicine and even less so in the tumultuous environment of the emergency department. In the rare circumstances when it is possible to achieve this degree of certainty, the costs are often prohibitive.

We can often decrease, but rarely completely eliminate, uncertainty regarding whether a particular diagnosis is present or absent. Therefore, for every diagnosis we must factor in the risks of underdiagnosis and overdiagnosis. We must consider the costs, risks, and time expenditures of testing as well.

The concept of decision thresholds is that instead of an absolute assurance, we necessarily seek to drive the likelihood of a diagnosis past a point where we are comfortable either accepting or discarding it. This level will be affected by patient values and concerns; the severity of the potential outcomes; the availability, costs, and invasiveness of relevant diagnostic technology; and many other situational factors.

The best way of representing decision thresholds is graphically. First, we will make a bar showing probabilities from 0 through 100% (Fig. 2-4).

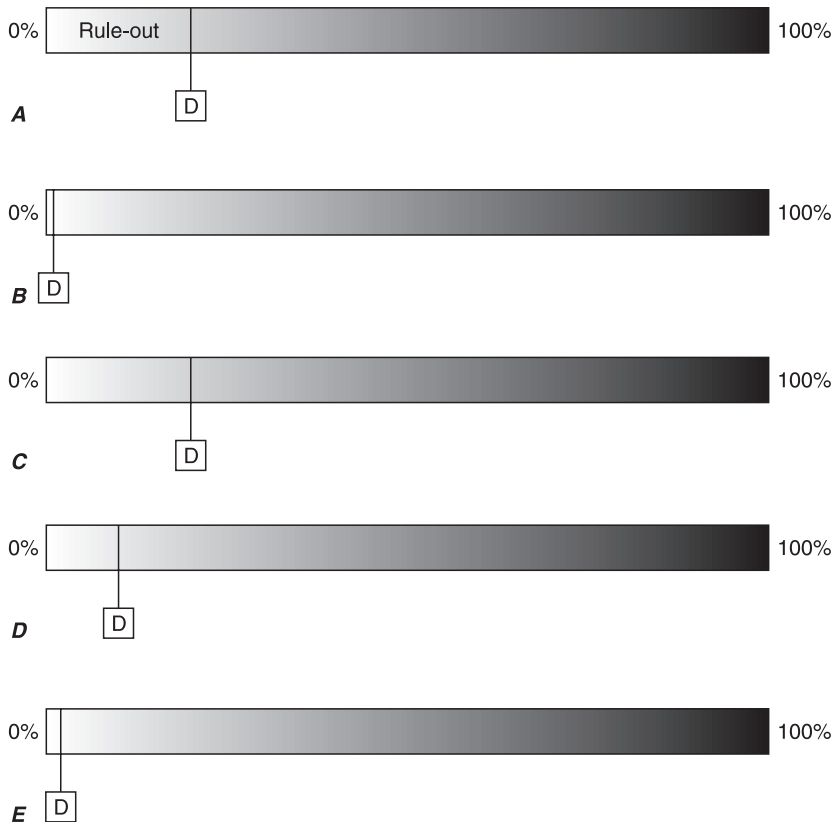
■ Discard Threshold

We can next add a line representing the probability below which we will discard the diagnosis in question. We will call this probability the “discard threshold” (Fig. 2-5A).

In life-threatening diseases or ones associated with significant morbidity, we might place this threshold at 1% or even lower (Fig. 2-5B). For instance, the diagnoses of myocardial infarction, aortic dissection, pulmonary



— FIGURE 2-4 — *Probability bar.*



— FIGURE 2-5 — *A* Discard threshold: probabilities below this level will cause us to discard the diagnosis. *B* Discard threshold of 1% in life-threatening conditions. *C* Discard threshold of 20% for strep pharyngitis in adults. *D* Discard threshold for cholecystitis in a healthy, young patient with suspected cholelithiasis. *E* Discard threshold for the same scenario, but in a less reliable patient.

embolism, and all other life threats, would certainly have a discard threshold of approximately 1%.

In relatively benign diseases in which an underdiagnosis would represent very little damage to the patient, the discard point can be set higher (Fig. 2-5C). For instance, the discard threshold for strep pharyngitis in healthy adults may be as high as 20%, as the risks of missing the diagnosis are extremely small and the disease will usually resolve on its own, even without treatment.

Even given the same disease state, patient characteristics can dictate the set point for the discard threshold.

EXAMPLE

A reliable 30-year-old presents with mild epigastric and right upper-quadrant abdominal pain. Your ultrasound exam indicates the presence of gallstones. You observe the patient for four hours in the emergency department. His pain gets better and he is able to eat one of the mystery meat sandwiches from the refrigerator. We might set the discard threshold at 10% (Fig. 2-5D) for a diagnosis of cholecystitis (as opposed to uncomplicated biliary colic), because if this patient is sent home with good discharge instructions and follow-up at the surgery clinic for probable cholelithiasis, we can expect him to return if he has intractable pain or he develops a fever.

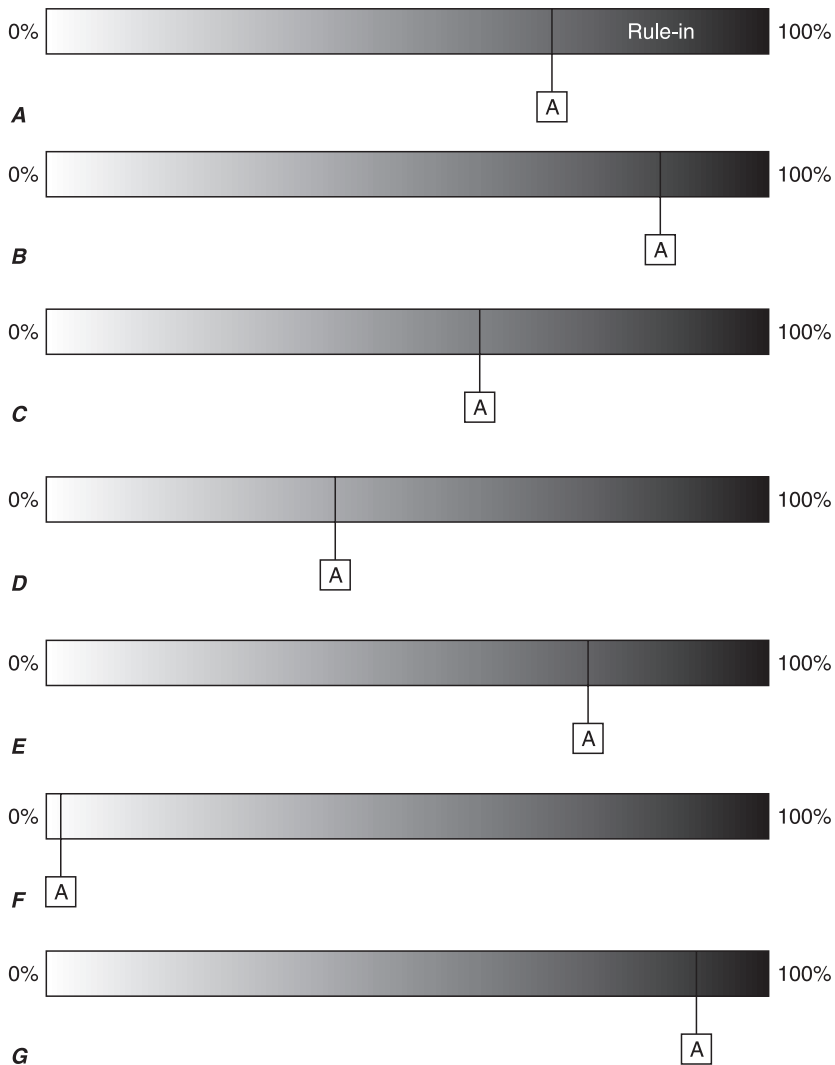
However, if a homeless patient with known substance abuse issues presents with the same scenario, the discard threshold point may have to be lowered (Fig. 2-5E). This patient is less likely to return if his condition declines. Setting his discard threshold at 2% for cholecystitis may be more appropriate because of these factors.

■ Accept Threshold

Next, we need to add a point on the bar above which we will “rule-in” the disease in question (Fig. 2-6A). At any probability above this accept threshold, we will consider the patient to have the diagnosis and, if necessary, treat accordingly (in this discussion, treatment may run the gamut from medications to consultation to admission to simply considering the diagnostic process complete).

The accept threshold is modified by the risks of treatment as well as the risks of overdiagnosis. A patient falsely diagnosed with a pulmonary embolism will suffer both the risks of anticoagulation as well as problems with obtaining health and life insurance. A patient falsely branded with this diagnosis in a medical history will also be potentially misdiagnosed any time he or she presents to a medical professional subsequently with a complaint of chest pain or shortness of breath. Therefore the accept point should be set relatively high in this patient, perhaps at 85% (Fig. 2-6B).

In a patient with epigastric burning, the accept point for peptic ulcer disease can be set much lower, as there is very little risk to falsely overdiagnosing this disease in the emergency department, as long as the patient is given appropriate follow-up (and we are not missing a more serious diagnosis such as MI). The risk of treatment with proton pump inhibitors is quite small and there are no stigmata from this diagnosis. We can therefore pick an accept level of around 60% (Fig. 2-6C).



— FIGURE 2-6 — **A** The accept threshold: above this probability, we will “rule-in” the diagnosis. **B** Accept threshold is set high in diseases in which overdiagnosis or treatment is risky. **C** In diseases in which overdiagnosis or treatment is not dangerous, the accept threshold can be lowered. **D** An accept threshold of 40% for acute coronary syndrome may be established to give heparin. **E** An accept threshold of 75% may be established to more aggressively treat acute coronary syndrome in the emergency department. **F** A low accept threshold for the diagnosis of partial extensor tendon laceration in the emergency department. **G** The surgeon providing follow-up the next day will set the accept threshold for the same diagnosis much higher.

Sometimes, the level of accept threshold will vary depending on which treatments we will give. For instance, in patients presenting with chest pain, the acute coronary syndrome accept threshold for giving aspirin and low-molecular-weight heparin might be 40% (Fig. 2-6D). However, to give that same patient IIb/IIIa inhibitors, our accept threshold may be 75% (Fig. 2-6E), as this therapy is expensive and has increased side-effects.

Similarly, in the initial emergency department visit, we will set accept thresholds for certain diseases far lower than the physician who will provide follow-up.

EXAMPLE

In a patient with a cut from a knife to the dorsum of the first phalanx of the forefinger, our accept threshold for the diagnosis of a partial tendon laceration should be set quite low, perhaps 2%, if there is any question of pain with extension, even in the presence of good strength testing (Fig. 2-6F). If we accept this diagnosis in the emergency department, we will splint the patient and give follow-up with a hand surgeon.

When the hand surgeon reexamines the tendon in her office, she will set the accept threshold for the same diagnosis at a much higher level (Fig. 2-6G), as it will be the difference between weeks of splinting or a simple laceration repair.

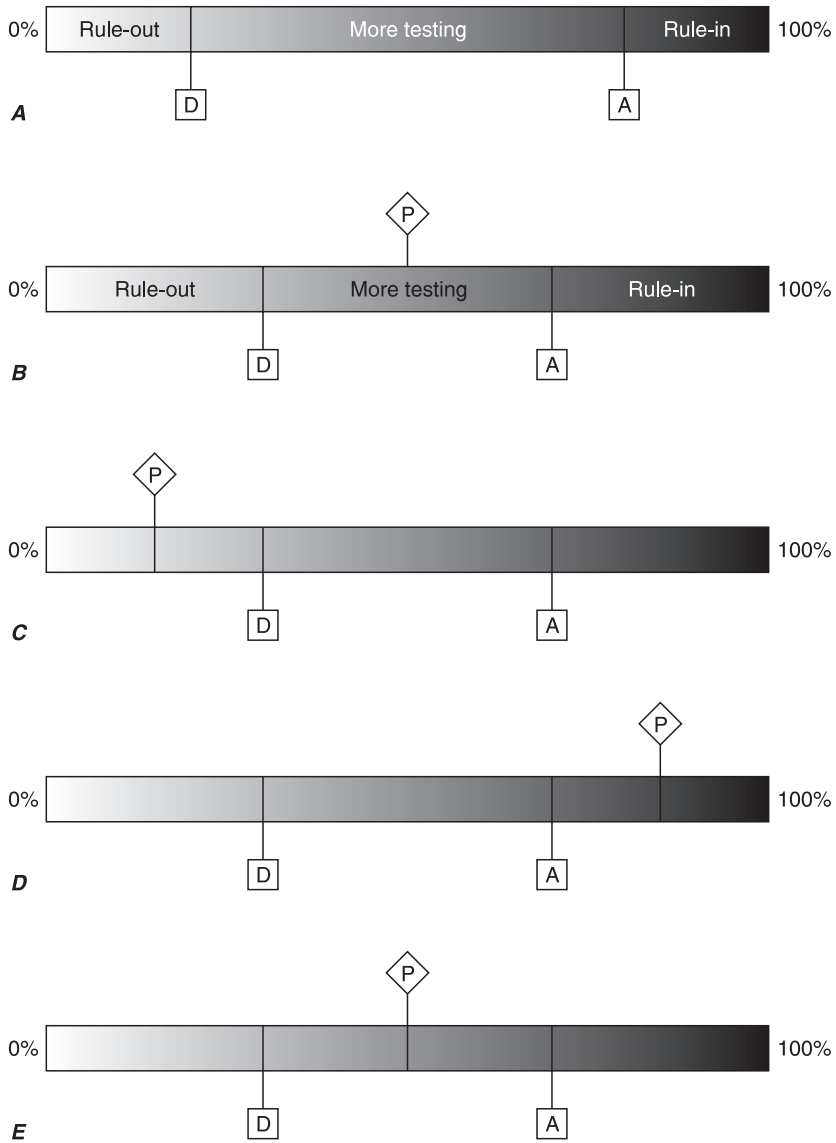
■ Indeterminate Zone

The area between the two thresholds is the indeterminate zone (Fig. 2-7A); more testing will be required for us to make a decision if a probability is in this area. Sometimes we elect to admit patients to the hospital from this zone and let other doctors worry about the final decision.

Pre-test Probability

After plotting the two decision thresholds, we can add our pre-test probability to the bar (Fig. 2-7B).

If the pre-test probability lies below the discard threshold we have chosen (Fig. 2-7C), then further testing for this diagnosis is unnecessary and we should seek another explanation for the patient's complaints. If it lies above the accept threshold (Fig. 2-7D), then further testing may not be necessary and the diagnosis has been made. If it lies between the two thresholds (Fig. 2-7E), then further testing is needed.



— FIGURE 2-7 — *A* The probability bar with both accept and discard thresholds. *B* Pre-test probability can be placed on the bar. *C* Pre-test probability below the discard threshold: no testing needed. *D* Pre-test probability above the accept threshold: no testing needed. *E* Pre-test probability is between the thresholds: additional diagnostic tests are necessary.

We do not advocate the routine drawing of probability bars for each diagnostic decision we make in the emergency department. Often we perform these manipulations mentally, without the need for a sketchpad. We have found that actually drawing the probability bars is an extremely effective technique for teaching house staff and medical students during a shift.

■ Estimation of Thresholds

Often the discard threshold is simple to estimate, being very low in dangerous conditions and higher in diseases without sequelae. The accept threshold is often more difficult to pin down precisely. The key point is that the accept threshold is rarely as critical an issue as the discard threshold. If short-term treatment does not carry risk, then a lower accept threshold will not be a problem. A higher accept threshold may result in unnecessary testing, but most tests are without significant side-effects, though they increase the cost of care. In contrast, setting the discard threshold too high can result in the serious consequences of missed diagnoses.

There are complex formulas that we can use to quantitate the thresholds, but they are unwieldy.⁴⁴ Instead, just as in pre-test probability, we should choose a number that best represents the combination of our clinical estimation and the available evidence and not worry about a few percentage points in either direction.

CHARACTERISTICS OF DIAGNOSTIC TESTS

■ Perfect Diagnostic Test Performance

A perfectly performing test would not have any false positives or false negatives. If the test came back positive, the patient would have the disease and a negative result would only be present in a healthy patient. Furthermore, it would be inexpensive, rapid, and bring no associated risks with its performance.

If such tests existed, we would not need to invest time and effort in the calculation of pre-test probability, the assignment of decision thresholds, or even agonize over a complete differential diagnosis. We could just order the ER ScreenTM, a panel of assays for life threats; our job would be to treat whichever tests returned positive.

We do not have access to many tests even approaching this level of perfection. The ones that are available often exact a high cost to provide us with a high degree of certainty. Therefore, when we order an imperfectly performing diagnostic test, we can only interpret its results based on its characteristics.

■ Testing Vocabulary

Gold Standards

The government used to back all paper money by its value in precious metal; hence the term “gold standard.” Medicine adopted the term to refer to the most accurate means of diagnosing a disease. Often these tests are expensive, time-consuming, invasive, or unsuitable for the clinical environment. When the gold standard is a 6-month follow-up or autopsy, the standard is obviously not feasible in the emergency department.

Even the gold standard test is not necessarily diagnostically perfect, as sometimes these tests have false positives and negatives. It would be better if we could evaluate new diagnostic tests against the Clinical Truth. However, just as in all of life, the truth in medicine is a concept that is wily and difficult to locate. We must instead make do with the best standard available. Hence, a better descriptive term for the gold standard is the *criterion* or *reference standard*, as this does not imply perfection; it just implies that this is the best means of diagnosing a disease currently available to us.

Often the tests we actually perform in the emergency department are surrogates for the criterion standard. Their value was established in well-done (hopefully) comparisons to a *criterion standard*.

The comparison of these surrogate tests to the criterion standard is often plotted on a 2×2 table (Fig. 2-8). The table is traditionally drawn with the known disease status or criterion standard testing on the horizontal and the results of the new test on the vertical. Each of the four boxes on this table will represent one of the following:

- *True positive*: a patient with the disease who tests positive on the new diagnostic test
- *False negative*: a patient with the disease who tests negative on the new diagnostic test
- *True negative*: a patient without the disease who tests negative on the new diagnostic test
- *False positive*: a patient without the disease who tests positive on the new diagnostic test

		Disease	
		Present	Absent
Test result	⊕	True positives	False positives
	⊖	False negatives	True negatives

— FIGURE 2-8 — *The 2 × 2 table you learned to hate in medical school.*

Accuracy

An accurate test will give results that mirror the truth. We can represent the accuracy of a diagnostic test quantitatively by the equation:

$$\text{Accuracy} = \frac{\text{True positives} + \text{True negatives}}{\text{All patients in the study}}.$$

Accuracy attempts to represent the utility of a test using just a single number. This is an oversimplification, because a test is frequently good at ruling-in or ruling-out a disease, but rarely both. The accuracy equation gives little indication of this fact.

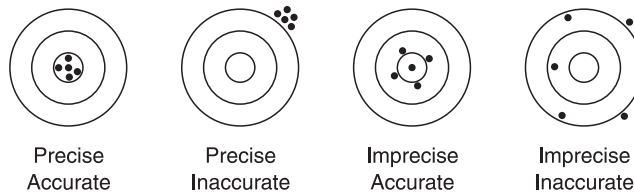
Precision

A precise test will have reproducible results time after time. These results do not necessarily represent the truth; they just need to be consistent (Fig. 2-9). A rifle, which shoots 10 inches to the right of the target every time it is fired is very precise, though not much use for hunting.

■ Interpreting Test Results

When we scan a journal article on diagnostic testing, we may see many different parameters used to evaluate a test.* The problem with the common, traditional test evaluation parameters is that they do not address the meaning of a test result. In the emergency department, we need to know

*The classic parameters are sensitivity and specificity, which we define later.



— FIGURE 2-9 — *Precision and accuracy.*

what to do with the results of the tests we perform on the individual patient we are treating. The ideal test parameter would not only indicate the meaning of a test result, but would also allow us to use it directly in the Bayesian equation we have discussed in the previous section:

Pretest probability \times Effects of diagnostic test result = Post-test probability

Furthermore, we should be able to predict whether any of the test results will allow us to make a change in our clinical suspicion before we even perform the test. If we realize, prior to its performance, that a test will not affect our actions, regardless of whether it is positive or negative, then we can save much time and expense by not using it.

Luckily, we have a way of evaluating test results that fulfills all of these criteria: *likelihood ratios*. Though it is a departure from the typical didactic order, we will discuss likelihood ratios, then segue into the older parameters of sensitivity, specificity, and predictive values.

■ Likelihood Ratios

Likelihood ratios (LRs) allow us to understand the meaning of a test result and use this to alter our pre-test probability. They are the best representation of the accuracy of a test result.

We can define LRs as the ratio of the probability of a test result in patients with disease to the probability of the same test result in patients without disease:

$$\text{LR result} = \frac{\frac{\text{Patients with disease and with result}}{\text{All patients with disease}}}{\frac{\text{Patients without disease and with result}}{\text{All patients without disease}}}$$

The mnemonic for this formula is WOWO, referring to With disease Over WithOut disease.³⁵

- Tests results with likelihood ratios greater than 1 are more likely to occur in patients with the disease than in patients without the disease. LRs above 1 increase our post-test probability of a disease.
- Tests results with likelihood ratios less than 1 are more likely to occur in patients without the disease than in patients with the disease. LRs below 1 decrease our post-test probability of a disease.
- If the likelihood ratio of a test result equals 1, then it is just as likely to occur in a patient with the disease as in a patient without the disease. These test results have no effect on post-test probability and therefore do not help us make a diagnosis.

To reinforce these concepts, consider a likelihood ratio of 10. This means that the result is ten times *more likely* to occur in a patient with the disease than a patient without the disease. This is not bad for a test result, which we would consider “positive”; however, we would certainly prefer an LR of 100 for this purpose.

By the same token, a test result with a likelihood ratio of 0.1 is ten times *less likely* to occur in a patient with the disease than in a patient without the disease. We can consider this result negative, though a likelihood ratio of 0.01 would be even better.

Another way of contemplating likelihood ratios is by their effect on *pre-test odds*. A test result with a likelihood ratio of 20 will yield *posttest odds*, which are twenty times greater than pre-test odds.

EXAMPLE

If our pre-test odds were 4, then a test result with a likelihood ratio of 20 will yield post-test odds of 80.

You may be questioning why we have suddenly gone from the pre-test probabilities of the previous chapter to the irksome pre- and post-test odds. Likelihood ratios can only be directly multiplied with odds, but we will shortly describe how to easily shift back and forth between probabilities and odds. Through these methods, likelihood ratios will allow us to directly modify our pre-test probabilities to yield accurate post-test probabilities.

At first, it may be easier to approach likelihood ratios as they apply to dichotomous test results.

Likelihood Ratio Positive

This value represents the change in pre-test probability caused by a positive test result. It ranges from 1 to infinity and can be represented mathematically as:

$$\text{LR positive} = \frac{\frac{\text{Patients with disease who tested positive}}{\text{All patients with disease}}}{\frac{\text{Patients without disease who tested positive}}{\text{All patients without disease}}}$$

or

$$\text{LR positive} = \frac{\frac{\text{True positives}}{\text{True positives} + \text{False negatives}}}{\frac{\text{False positives}}{\text{False positives} + \text{True negatives}}}$$

Test results having values greater than 5 are moderately useful and those with values greater than 10 have the power to truly alter decision-making. Test results with values from 1 to 5 do very little to alter pre-test probability.

Likelihood Ratio Negative

This value represents the change in pre-test probability caused by a negative test result. It ranges from 0 to 1 and can be expressed mathematically as:

$$\text{LR negative} = \frac{\frac{\text{Patients with disease who tested negative}}{\text{All patients with disease}}}{\frac{\text{Patients without disease who tested negative}}{\text{All patients without disease}}}$$

or

$$\text{LR positive} = \frac{\frac{\text{False negatives}}{\text{True positives} + \text{False negatives}}}{\frac{\text{True negatives}}{\text{False positives} + \text{True negatives}}}$$

Test results with values of less than 0.5 are moderately useful and values less than 0.1 are truly significant and can alter decision-making. Values from 0.5 to 1 do very little to alter the original pre-test probability.

Non-dichotomous Likelihood Ratios

One of the many advantages of likelihood ratios is that they can also easily represent the effects of non-dichotomous test results.

To calculate the likelihood ratio of each test result, we return to our original formula:

$$\text{LR result} = \frac{\frac{\text{Patients with disease and with result}}{\text{All patients with disease}}}{\frac{\text{Patients without disease and with result}}{\text{All patients without disease}}}$$

EXAMPLE

The former study of choice for the initial evaluation of pulmonary embolism was a nucleotide ventilation/perfusion scan (V/Q). Instead of being reported in dichotomous positive/negative values, scans are reported as four results: normal, low probability, intermediate probability, and high probability. We can calculate likelihood ratios for each of these results by using the values provided in the PLOPED study:³⁶

SCAN RESULT	PE	No PE
Normal	5	126
Low probability	39	273
Intermediate probability	105	217
High probability	102	14
Total	251	630

We now have enough information to calculate likelihood ratios for each value:

$$\text{LR normal} = \frac{\frac{5 \text{ (with disease and result)}}{251 \text{ (with disease)}}}{\frac{126 \text{ (without disease and with result)}}{630 \text{ (without disease)}}} = 0.1.$$

In the same way, we can calculate the likelihood ratios for each of the other results:

- LR low probability = 0.4
- LR intermediate probability = 1.2
- LR high probability = 18.3

Based on these likelihood ratios, it is apparent that a high or normal scan can significantly alter our pre-test probability. On the other hand,

the low and intermediate values will not greatly change our pre-test probability. This has led to the reporting of a low or intermediate scan as non-diagnostic. From the perspective of diagnostic decision-making, we gain very little information from either of these two results.

Scalar Likelihood Ratios

Even intervals of scalar values can be reported using likelihood ratios. This avoids the problems created by having just one dichotomous cutoff when interpreting scalar tests.

EXAMPLE

It is a common situation for consultants to ask for a white blood cell (WBC) count when being called to evaluate a patient for appendicitis. The problem with this strategy is that rarely are realistic cutoffs for the results established before ordering the test. Using data from a prior study,^{37,38} Brown and Reeves reported likelihoods for various intervals of the WBC count for patients suspected of appendicitis.³⁹

WBC COUNT ($\times 10^3/\text{mL}$)	LIKELIHOOD RATIO
4–7	0.1
7–9	0.52
9–11	2.8
13–15	1.7
15–17	2.8
17–19	3.5

It can be seen that the WBC count has very little diagnostic effect unless it is below 7 or above 17. This makes the WBC a less than optimal test for appendicitis. If we use an arbitrary cutoff such as >11 instead of interval likelihoods, the test is even less functional.

The power of likelihood ratios is that they simultaneously embrace both the true and false rate for any test result. For instance, the likelihood ratio positive encompasses both the true positive and false positive rate of a test. As we will see shortly, other testing parameters are not capable of this necessary duality.

Ideally, all journals would report likelihood ratios for all studies of diagnostic tests. The literature is unfortunately not yet at this point of sophistication, so we must be familiar with some of the less useful, but more prevalent, test parameters.

■ Sensitivity and Specificity

Sensitivity and specificity were originally developed for use in chemistry research to rate the ability of assays to detect substances.³⁵ Medicine adopted the terms for use in the description of diagnostic tests; like many makeshift measures, they are not ideal for their intended use. The problem is that they measure the likelihood of a test result in patients with the disease status *already known*.³⁵ In the emergency department we need just the opposite; we want to know the likelihood of the patient having the disease once we have the test result. This can only be tangentially answered by sensitivity or specificity.

Sensitivity

A test with perfect sensitivity would be *positive* in every patient *with the disease*.* The classic mnemonic is PID, reminding us that the test is Positive In Disease (Fig. 2-10). Another way of thinking about sensitivity is that a patient with disease will never have a negative result on a perfectly sensitive test.

		Disease	
		Present	Absent
Test result	(+)	True positives	False positives
	(-)	False negatives	True negatives

— FIGURE 2-10 — *Sensitivity represents true positives divided by all patients with disease.*

*In this sense, perfect refers to a sensitivity of 1, or 100%.

We can express sensitivity mathematically as:

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negative}}.$$

Since tests rarely have perfect sensitivity, there will be a number of false negatives; i.e., patients who have the disease but still test negative. The more sensitive the test, the less false negatives there will be. The false negative rate can be expressed as:

$$\text{False negatives} = 1 - \text{Sensitivity}.$$

Negative results on a highly sensitive test are powerful. Sensitive tests are important because they have the ability to aid in ruling-out a disease. Conceptually, a negative result on a sensitive test will sharply decrease the post-test probability.*

This has led to the superior mnemonic of SNOOUT to represent that a SeNsitive test helps to rule OUT disease.⁴⁰

Specificity

A test with perfect specificity would be *negative* in every patient *without the disease*. The classic mnemonic is NIH, reminding us that the test is Negative In Health (Fig. 2-11). Another way to think about specificity is that a patient without the disease will never have a positive result with a perfectly specific test. To express specificity mathematically,

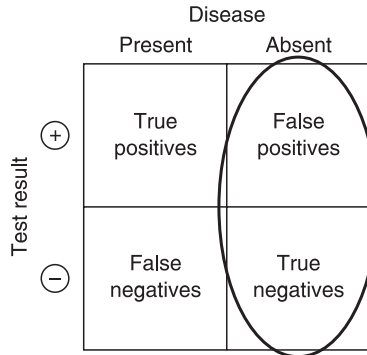
$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negative} + \text{False positives}}.$$

Since tests rarely have a perfect specificity, there will be a number of false positives; i.e., patients who are healthy but test positive. The more specific the test, the less false positives there will be. The false positive rate can be expressed as:

$$\text{False positives} = 1 - \text{Specificity}.$$

Positive results on highly specific tests are important, because they have the ability to aid in ruling-in a disease. Conceptually, a positive result on a

*We say conceptually, because sensitivity is the likelihood of a positive result in the presence of disease, not the likelihood of the absence of disease with a negative result. The latter concept is the definition of the negative predictive value. However, since sensitivity is proportional to the negative predictive value, a negative result on a perfectly sensitive test should rule out the disease. This entire quagmire can be avoided by using likelihood ratios.



— FIGURE 2-11 — *Specificity represents true negatives divided by all patients without the disease.*

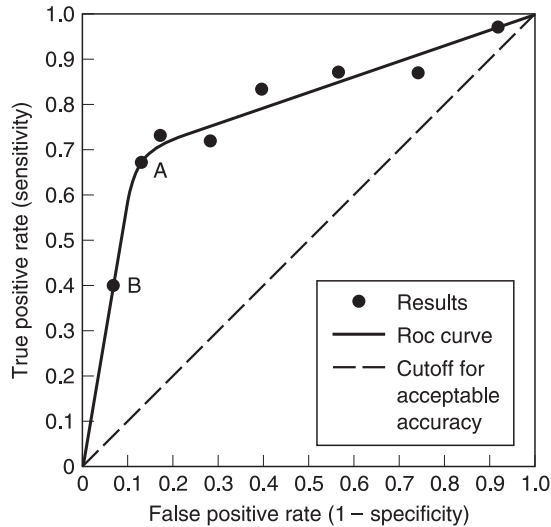
specific test will sharply increase the post-test probability. This concept is represented with the mnemonic SPIN to represent SPecific tests help to rule IN disease.⁴⁰

Often, to increase sensitivity requires a loss of specificity and vice versa. This is because many seemingly dichotomous tests actually generate a wide range of values. The researchers convert these scalar values to a dichotomous result by picking a cutoff for positive and negative. The cutoff for many dichotomous tests is picked from a range of possible points each with a sensitivity/specificity tradeoff.

EXAMPLE

You decide to create a new machine to test for hypoglycemia in diabetics, but instead of reporting the numeric value of the finger-stick sugar, it reports two results: either hypoglycemic or normal. If the machine is set to report hypoglycemia with any sugar level below 100 mg/dL, then it is unlikely that any cases of hypoglycemia will be missed, making the test extremely sensitive. However, at that cutoff point, many patients with normal glucoses will be labeled hypoglycemic; therefore, the test will have low specificity. If the machine was reset to report hypoglycemia below 40 mg/dL, the specificity would increase at the expense of sensitivity.

The sensitivity and specificity will change for any individual value chosen as the cutoff point, as was shown in the preceding example. We rarely have access to any of these sensitivities or specificities other than the one point chosen by study researchers. We can partially overcome



— FIGURE 2-12 — *Receiver operator characteristics curve.*

this problem by obtaining the receiver operator characteristics curve for a test.

Receiver Operator Characteristics (ROC) Curves

These are a graphical representation of true positive rate (sensitivity) on the vertical axis and false positive rate ($1 - \text{specificity}$) on the horizontal. Using these graphs, we can see an estimation of the sensitivity and specificity for the individual values of a scalar test (Fig. 2-12).^{41,42}

Area Under the Curve If we examine the ROC curve, we can appreciate that the greater the area under the curve (AUC), the more accurate the test. If the AUC is 0.5 (represented by the diagonal line in the diagram), then the test is no better than chance alone. If the AUC is 1.0, then the test has perfect accuracy.

Cutoff Values When researchers pick a cutoff value for the positive/negative level of a test, they must first decide whether they wish to maximize sensitivity or specificity. They can then pick a point on the curve, which makes that characteristic highest with minimal false results. Point A on the curve in Fig. 2-12 will maximize sensitivity with the minimal loss of specificity. Point B will give the highest specificity with a minimum of false negatives (sensitivity).

We can also use the ROC curve to observe the likelihood ratios of both individual points on the curve as well as scalar ranges. An article by Choi describes using the slope of various points of the ROC curve to arrive at the appropriate LRs.⁴²

The Dangers of Being Alone

Another serious shortcoming of these sensitivity and specificity tests is that they are often viewed in isolation of each other. The most sensitive test available is useless if it is non-specific. It will help to rule-out a disease if it is negative, but it will be negative so infrequently as to make it clinically useless.

EXAMPLE

A new laboratory test for septic arthritis has a sensitivity of 99.5% and a specificity of 1.8%. You are excited to try the test in the pediatric emergency department, because you have had difficulty easily making this diagnosis in children. You surmise that if the test is negative, you have virtually ruled-out the diagnosis of septic arthritis in your patients. You are frustrated to find that the first twenty patients you send the assay on have positive results. You finally stop using the new assay, because it never helps you make a clinical decision.

A converse of this problem is seen with specific, but insensitive, tests. They may be helpful if they are positive, but they will be negative most of the time.

- If sensitivity + specificity \leq 100, then test = useless!

If we find that the sum of the sensitivity and specificity is less than or equal to 100%, the test is absolutely clinically useless. If we calculate the likelihood ratios of this type of test, by the method we are about to describe, we can prove this to ourselves.

Making Better Use of Sensitivity and Specificity

For the reasons we have mentioned, sensitivity and specificity are not ideal parameters for emergency medicine decision-making. Luckily, we can easily convert these parameters to the far more powerful *likelihood ratios* (positive and negative). Likelihood ratios avoid all of the above problems.

$$\text{LR positive} = \frac{\text{Sensitivity}}{1 - \text{Specificity}};$$

$$\text{LR negative} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}.$$

An astute reader will quickly realize that these formulas are merely a different way of expressing the WOWO formula for likelihood ratios, as mentioned above.

■ Predictive Values

There is another set of parameters which attempt to reconcile the problems of sensitivity and specificity. Predictive values do address the necessity of evaluating a *test result* (like likelihood ratios) as opposed to evaluating a *test* (like sensitivity/specificity). However, predictive values bring problems of their own; they are fixed values that reflect the population included in a particular study and rarely provide the information you need about your own individual patients.

Predictive Value Negative

Given a negative test result, the predictive value negative provides the probability of the patient being healthy. The value is linked to the prevalence of the study population; the predictive value negative *increases* as the prevalence of the disease *decreases*. Mathematically, it can be expressed as:

$$\text{Predictive value negative} = \frac{\text{True negatives}}{\text{True negatives} + \text{False negatives}}.$$

and on the 2×2 table as shown in Fig. 2-13.

		Disease	
		Present	Absent
Test result	(+)	True positives	False positives
	(-)	False negatives	True negatives

— FIGURE 2-13 — *Predictive value negative represents true negatives divided by all negatives.*

Predictive Value Positive

Given a positive test result, the predictive value positive provides the probability of the patient having the disease. The value is inextricably linked to the prevalence of the disease in the study; the predictive value positive *increases* as the prevalence *increases*. It can be expressed in the equation:

$$\text{Predictive value positive} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

and on the 2 × 2 table as shown in Fig. 2-14.

Problems with Predictive Values

Predictive values are linked to the prevalence of the disease in the study from which they were derived; this is problematic. If that study's prevalence matches or approximates the pre-test probability of disease in our patient, then we are able to use these characteristics. If not, then the predictive values may not be accurate to predict disease in our patient. A new pregnancy test that one of the authors is developing demonstrates the extremes to which prevalence can alter predictive values.

EXAMPLE

A new pregnancy test is developed consisting of a piece of loose-leaf paper with the phrase "You are pregnant" written on it in ballpoint pen. You tape this piece of paper to any patient and if it still reads "You are

		Disease	
		Present	Absent
Test result	(+)	True positives	False positives
	(-)	False negatives	True negatives

— FIGURE 2-14 — Predictive value positive represents true positives divided by all positives.

pregnant,” then the test is positive. Obviously, this test has perfect sensitivity (100%), as it will always be positive if taped to a pregnant patient. The specificity will be 0% as the test will never be negative when placed on a non-pregnant patient.

If we performed this test on an Ob/Gyn ward where 95 of the 100 patients on the floor are pregnant, the predictive value positive will be 95% despite the specificity of 0% (Fig. 2-15A). If we took the test to a general medical ward where 95 of the patients are not pregnant and 5 patients are pregnant, the predictive value positive will be 5% despite having an unchanged specificity of 0% (Fig. 2-15B).

Reported predictive values are a less than ideal solution to the problem of appropriately representing the value of a test result. They are inextricably linked to the disease prevalence in the study that reports them and they

	Present	Absent
Test result (+)	95	5
Test result (-)	0	0

$$\text{PVP} = \frac{95}{100} = 95\%$$

A

	Present	Absent
Test result (+)	5	95
Test result (-)	0	0

$$\text{PVP} = \frac{5}{100} = 5\%$$

B

— FIGURE 2-15 — **A** A new pregnancy test applied on an Ob/Gyn floor. **B** The same test applied to a general medical ward.

cannot be directly applied to our patients, unless their pre-test probability exactly matches this prevalence.*

Likelihood ratios give us the information about the test result that really matters; they apply to *any* level of pre-test probability and can be directly inserted into the Bayesian equation.

■ Screening Tests

We do not routinely perform screening tests in the emergency department, though they are an essential part of our public health system. The difference between screening tests and diagnostic tests is that the individuals undergoing screening tests are asymptomatic for the diagnosis and we have no reason to suspect they have a greater risk of disease than the general population. The ideal screening test would be both sensitive and specific, but since most tests cannot be both, screening tests are often quite sensitive with a lower specificity. These less than optimal tests, when positive, would then require a second confirmatory test, which possesses high specificity. An example of this is HIV screening, which starts with a sensitive test, the ELISA. If the ELISA is positive, the extremely specific Western Blot is ordered to confirm the results.

One screening test, which we use routinely and appropriately in the emergency department prior to the administration of medication or radiography, is the urine pregnancy test for women of childbearing age. The urine pregnancy test is an example of an excellent screening exam as it has the power to take a pre-test probability of near zero:

“Doctor, there is no way I can be pregnant. I haven’t been with anyone for two years.”

to a post-test probability of 100%. With a specificity of 100%, a sensitivity of over 98%, a likelihood ratio positive of infinite, and a likelihood ratio negative of 0.02 (it will miss a very early pregnancy which has not produced adequate b-HCG,) it is one of the most powerful tests in emergency medicine.

Unfortunately, we use many inappropriate diagnostic tests as if they were screening tests. To get a white blood cell count as a screen for infection is to guarantee uninterpretable results. The routine coagulation panel for all emergency patients is a gross waste of resources; it screens for

*Even worse, they are misleading even with respect to the original study population. They apply only to the patient whose individual probability of disease is exactly identical to the prevalence within the average of the study population. This may not be true for any individual patient in the entire study!

conditions that we have no reason to suspect. If a patient is receiving anticoagulation or is actively bleeding, then we order a coagulation panel (or better yet, we can break the panel strategy and just order the INR on our patients taking warfarin). Otherwise, there is no reason to support needless and costly standard emergency lab panels.

Playing Testing Jeopardy A practice, which is particularly frustrating, is the ordering of tests without considering their necessity or utility and then trying to determine their value after the results have arrived. This is akin to the American game show Jeopardy, in which an answer is provided and contestants must determine the question. Trying to figure out the meaning of test results that were ordered without indication is not a proper diagnostic strategy.

■ History and Physical Exam

Under the differential diagnosis section, we discussed how a patient's signs and symptoms can suggest diagnoses that we must consider and decide whether or not to evaluate. Our history and physical exam also have the ability to alter the pre-test probability; in this sense they are *diagnostic tests*. In order to make this process objective, we need test characteristics.

While many studies evaluate the test parameters of the components of our history and physical, perhaps the best source for this information is in an ongoing series published in the *Journal of the American Medical Association* (JAMA). The series, entitled The Rational Clinical Exam, takes different diseases and disorders and summarizes the performance of the signs and symptoms conventionally associated with the diagnosis. Particularly relevant to emergency physicians, the *Annals of Emergency Medicine* has begun to publish abstracts and commentaries summarizing the important information from the JAMA series that are relevant to emergency medicine.

The incredible value of this form of evidence is that, in addition to helping us make diagnoses in individual patients, it can allow us to retrain our clinical judgment and *illness scripts*.

EXAMPLE

When diagnosing appendicitis in the emergency department, we have always considered anorexia to be one of the hallmarks of the disorder. When one of our colleagues hands us a copy of the Rational Clinical Exam article on diagnosing appendicitis, we are surprised to see that the likelihood ratios positive and negative (respectively, 1.27 and 0.64) for this symptom are close to one.⁴³ In the future, we stop using anorexia to change our pre-test probability of appendicitis.

■ What are we actually testing?

In order to derive the maximum benefit from the use of a diagnostic test, we must understand what is actually being tested. A pregnancy test does not test for pregnancy, it tests for bHCG; this is why it may test negative during the first two weeks after conception when bHCG levels are too low to detect. If we do not properly understand what a test result means, then we may be led astray from the true diagnosis.

EXAMPLE

We often use the FAST (focused assessment using sonography in trauma) exam to evaluate the abdomen of trauma patients. When we perform this study, we may think we are assessing for intra-abdominal injury. However, the FAST exam does not assess for intra-abdominal injury, it assesses for free intraperitoneal fluid. Furthermore, it cannot detect small amounts of fluid, it requires at the very least a few hundred milliliters of fluid before the test will be positive. If we perform a negative FAST exam on a trauma patient who subsequently has a positive laparotomy, then the study may be branded a false negative. If the intra-abdominal injury was a bowel wall injury with mesenteric hematoma and only 20–30 mL of blood in the abdomen, then the ultrasound was not really a false negative. If, however, the injury was a splenic hematoma with a liter of blood, then the FAST exam was indeed falsely negative.

By the same token, if a FAST exam reveals large amounts of free fluid in a patient with cirrhosis, some would call this a false positive FAST exam. However, it is not really falsely positive, because there was fluid in the belly, it just was not blood.

As we can see, understanding what the test actually examines is vitally important, both to use the results and also to understand the testing characteristics derived from a diagnostic study. If one study uses the presence of fluid while another uses intra-abdominal injury as the test criteria, then the sensitivity and specificity of the FAST exam may be very different in the two studies.

If we understand what a test examines, then we can extrapolate the results to the clinical situation. Free fluid on the FAST exam in a normotensive patient with a grossly enlarged liver and caput medusa may be interpreted as a positive FAST exam that does not necessarily signify an abdominal injury.

APPLICABILITY OF DIAGNOSTIC TESTS

In the previous section, we discussed the process of understanding and using the characteristics of diagnostic tests. The next step is to evaluate whether these characteristics must be altered to account for our individual patient and unique clinical scenario. We call this process of evaluation, and possible adjustment, the *applicability* of evidence.

The evidence we need to determine applicability includes the setting, population, and interpretation of the original study. NEEDLEs can provide these facts in a readily usable form (Fig. 2-16). Using this information, we can decide if the evidence is appropriate to use for diagnostic decision-making in our patient.

Diagnostic Evidence: Elisa d-dimer for Pulmonary Embolism

Benefits/Downsides

- LR + 1.73, LR - 0.11
- No risks, cost minimal

Applicability

- Adult patients in outpatient setting
- Performed in normal hospital laboratories; interpreted by laboratory staff
- Specificity lower in patients >70; sensitivity and specificity lower if symptoms >3 days

Brown MD et al. *Annals Emerg Med* 2002;40(2):133-144

-Valid meta-analysis. (Sens 95% CI 0.88-0.97) (Spec 95% CI 0.36-0.55)

— FIGURE 2-16 — *NEEDLE: Necessary Evidence for Emergency Decisions, Listed and Evaluated.*

■ Applicability

Applicability directly relates to the question “Are our patient and practice setting sufficiently similar to those in the diagnostic study to directly use the study’s test characteristics?” Factors that might cause us to change our interpretation of a test and its applicability are considered below.

Patients

A study of performance of a test in relationship to a disease may not apply to our patient and we therefore may not be able to use the information from such a study to assist in our diagnosis. Factors that may cause a test to be more or less accurate in a given patient include severity of disease, comorbidities, or age.

EXAMPLE

If a patient is pancytopenic, because of myelodysplastic syndrome, we would not think to use his or her white blood cell count to determine the likelihood for appendicitis, even though, ordinarily, a value of less than $7 \times 10^3/\text{mL}$ would have a useful likelihood ratio negative. In the absence of a study that reported WBC counts in pancytopenic patients being considered for appendicitis, a low WBC in this patient should not change our pre-test probability

Evidence for diagnostic tests may not be applicable to whole groups of patients. In this case, we may not even bother to order the test when we are practicing in this setting.

EXAMPLE

During one of your rotations through the surgical intensive care unit (SICU), you have a postoperative patient with sudden-onset shortness of breath. You consider pulmonary embolism in your differential and tell your attending that you want to order a d-dimer. Your attending explains that, while this test is useful in emergency department patients, its applicability is severely diminished in SICU patients. Postoperative patients almost uniformly will test positive regardless of the presence of embolism, therefore this test has very little utility in this setting.

Interpreters

When using a test, we must ask ourselves if its interpretation will be as accurate as in the original study.

EXAMPLE

CT angiogram was found to have very high accuracy for the early diagnosis of ischemic stroke in a well-done study. Fellowship-trained neuroradiologists interpreted the CT scans. At our institution, there are only general radiology residents with little familiarity with CT angiograms of the brain; we would not expect the sensitivities or specificities at our institution to be as high as in the study.

Equipment and Resources

Another part of the applicability of a diagnostic test hinges on the use of the same quality of equipment as in the original study. The accuracy of the test will be less if our diagnostic equipment is inferior.

EXAMPLE

Emergency department ultrasonographers perform an imaginary study demonstrating a likelihood ratio negative of 0.01 for the signs of pericardial tamponade in patients with pericardial fluid. The study was performed using a \$150,000 echocardiography machine with a phased array transducer. It is doubtful that the quality of images using your cut-rate ultrasound machine from the early 1990s will allow the same test characteristics.

Clinical Scenario

Certain clinical scenarios will also allow and necessitate the adjustment of diagnostic test parameters. We can use this variance of testing characteristics in different clinical scenarios to our advantage. Tests that have only middling discriminatory ability can shine in certain circumstances.

EXAMPLE

We know that the FAST exam does not have 100% sensitivity for intra-abdominal injuries or even for the presence of free fluid. However, there are some situations in which we can use the test as if it has a near perfect sensitivity.⁴⁴

If a patient comes to our trauma after *isolated* severe blunt trauma to the abdomen and pelvis, we are placed at a diagnostic nodal point. Should we go to the operating theatre for an exploratory laparotomy or send the patient to angiography for pelvic embolization? To aid our decision, we can perform a FAST exam. If it is negative, then we can conclude that the hypotension is a result of the pelvic fracture and send the patient to the angiography suite. Because the patient is hypotensive, we would expect any intra-abdominal injury significant

enough to cause hypotension to have a large amount of free fluid. If we are able to get good windows with our FAST exam and we do not see free fluid, then we can presume that the bleeding is in the pelvis, even though there may be some free fluid in the belly missed by the ultrasound. We adjust the sensitivity of the FAST exam based on the spectrum of disease of our patient.

We can describe the complex thinking involved in this example:

- We have a hypotensive blunt-trauma patient.
- After a quick primary survey, a negative chest radiograph and subxiphoid transthoracic echocardiogram, we conclude that the patient has isolated trauma to the abdomen and pelvis.
- Therefore, we attribute the hypotension to active bleeding, not to spinal injury, cardiac tamponade, pneumothorax, or other medical causes.
- We deduce that there is no cause for the bleeding that may exist other than pelvic or abdominal after ruling out major extremity bleeding on clinical grounds.
- We decide that if there is not enough blood in the belly to account for the hypotension, then we are going to angiography to stop the pelvic bleeding.
- We consider that an ultrasound of the abdomen, if negative, will have a very low LR for the amount of bleeding that we have decided must exist in order to cause the hypotension. In other words, even though the LR negative for the FAST exam may not be significant enough to rule-out any abdominal bleeding in this patient, we are confident that the sensitivity of US for detecting the amount of blood required to make our patient this hypotensive would be much higher than for the usual minimum threshold for detection, and that the LR for a negative US would therefore be low enough to rule-out the diagnosis of *significant* intra-abdominal bleeding.⁴⁴
- As a result of the negative FAST exam, we conclude that the immediately important bleeding must be retroperitoneal from a severe pelvic fracture and we send the patient to the angiography suite.

This adjustment of test result interpretation to fit the clinical scenario is the hallmark of an adept clinician.

■ Sensitivity Analysis

The adjustment of a test to make it applicable to our patient and clinical scenario is often more of an art than a science. In some circumstances (e.g., the WBC in the myelodysplastic patient), we will simply decide we cannot use the test.

In other circumstances, semi-quantitative adjustment is possible through a process called *sensitivity analysis*.⁴⁰ This process involves deciding how much loss of accuracy we can accept and still be able to use the test to make a decision. If the test result would normally allow us to make a decision with a wide margin of certainty, but our applicability assessment causes us to downgrade the accuracy, we may still use the test. If, however, a test result barely would allow us to cross a decision threshold, we may no longer be sure that a test with diminished applicability will allow us to make the decision. The next section will deal with the determination of this margin of error for our decision-making.

POST-TEST PROBABILITY

With an understanding of the power and pitfalls of diagnostic tests, we can see how they alter our pre-test probability of disease. The most efficient method utilizes likelihood ratios, as they allow direct calculations of post-test probability.

■ Obtaining Post-test Probabilities

We started with the equation:

$$\text{Pretest probability} \times \text{Effects of diagnostic test} = \text{Post-test probability.}$$

We can now alter the equation to include the use of likelihood ratios:

$$\text{Pretest odds} \times \text{Likelihood ratio} = \text{Post-test odds.}$$

Unfortunately, likelihood ratios can only be directly multiplied with pre-test odds and not pre-test probability.

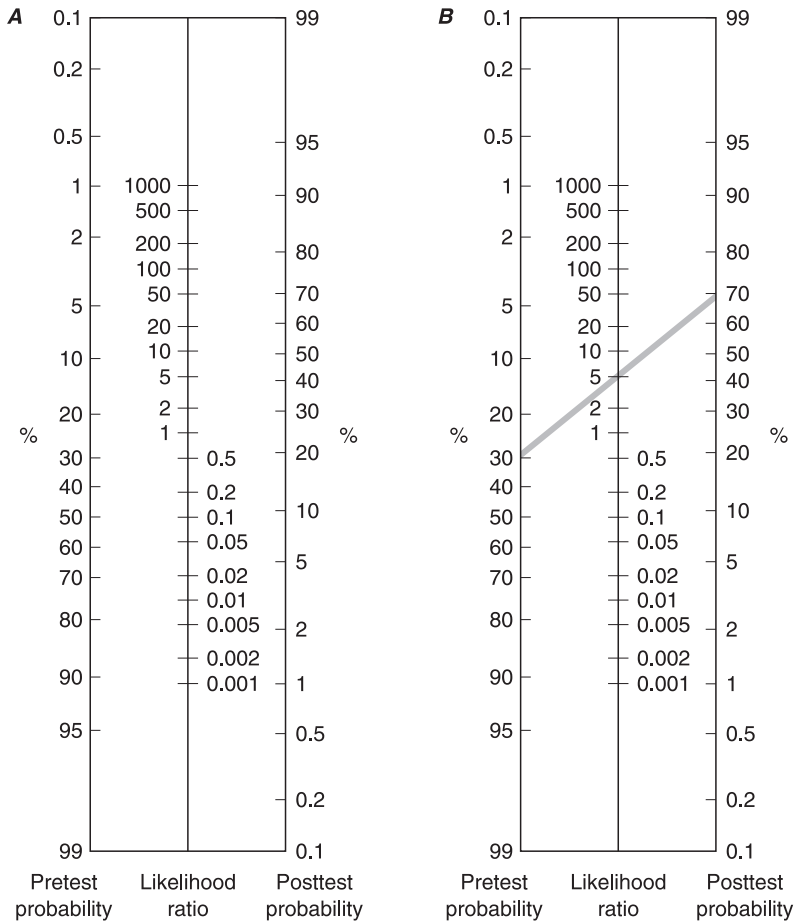
To convert pre-test probability to odds, use the formula:

$$\text{Odds} = \frac{\text{Probability}}{1 - \text{Probability}}.$$

To convert post-test odds back to post-test probability, use the formula:

$$\text{Probability} = \frac{\text{Odds}}{1 + \text{Odds}}.$$

If this seems like a lot of calculating, there is a much easier solution; a nomogram exists that takes pre-test probability and LR and gives the post-test probability (Fig. 2-17A).⁴⁵ Using the nomogram involves simply



— FIGURE 2-17 — **A,B** Nomograms for calculating post-test probability. (Adapted from Fagan, 1975⁴⁵)

drawing a line through the pre-test probability and likelihood ratio of the test result to yield the post-test probability.

EXAMPLE

If our pre-test probability is 30% and the LR for our positive result is 5, then our post-test probability is 70% (Fig. 2-17B).

In addition to the nomogram, there is software available for PDAs (personal digital assistants) that will calculate post-test probability as well as

convert sensitivity/specificity to likelihood ratios. A free version can be found at www.cebm.utoronto.ca. There are also online calculators for post-test probability, which allow us to avoid calculating odds: <http://araw.mede.uic.edu/cgi-bin/testcalc.pl>.

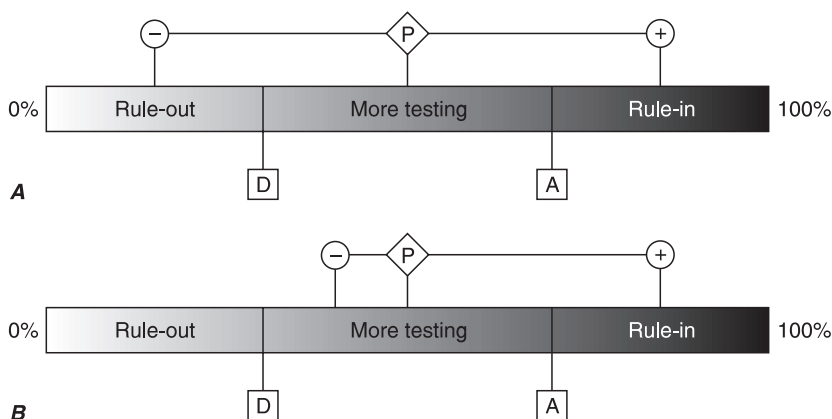
■ Plot the Post-test Probabilities *before* Performing the Test

Before even performing a test, we can plot the post-test probabilities of its results on our decision bar, using the likelihood ratios of each result. In Fig. 2-18A, the post-test probabilities of each test result cross the decision thresholds. If we perform this test, it will allow us to make a decision no matter what its results. If this is the case, it is a fortuitous situation; however, a test will often only allow the crossing of one decision threshold (Fig. 2-18B). If you plot the post-test probabilities and realize that neither result will cause a crossing of a decision threshold, do not perform the test.

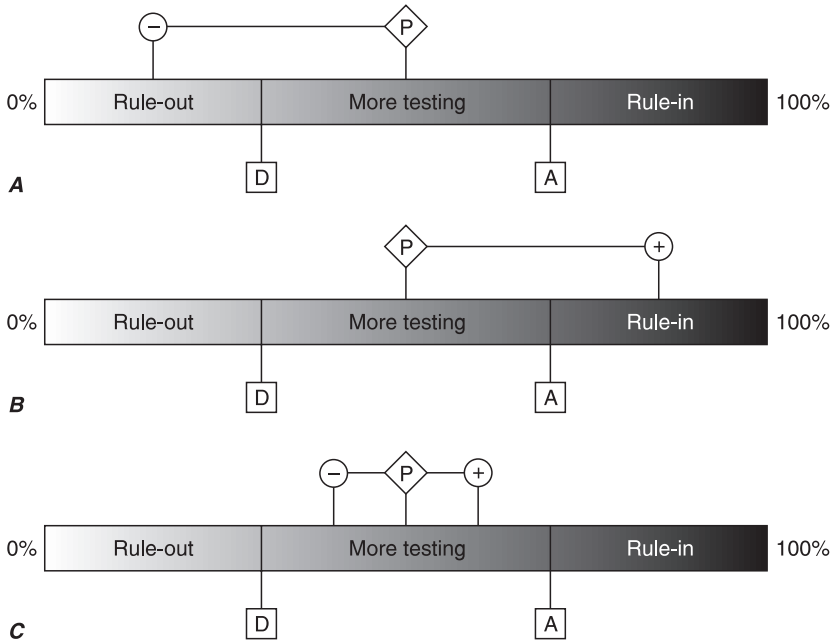
If a test has a result that will allow the crossing of at least one decision threshold, then it can help make the diagnosis.

■ Perform Testing

After the results return, we can see where on the decision bar the post-test probability lies. If we are below the discard threshold (Fig. 2-19A), then we must seek another diagnosis. If we are above the accept threshold (Fig. 2-19B), then we have “ruled-in” the diagnosis and should take action.



— FIGURE 2-18 — **A** Post-test probabilities plotted before testing. **B** Only one of the test results crosses a decision threshold.



— FIGURE 2-19 — **A** The post-test probability is below the discard threshold. **B** The post-test probability crosses the accept threshold. **C** The post-test probability is still in the indeterminate zone.

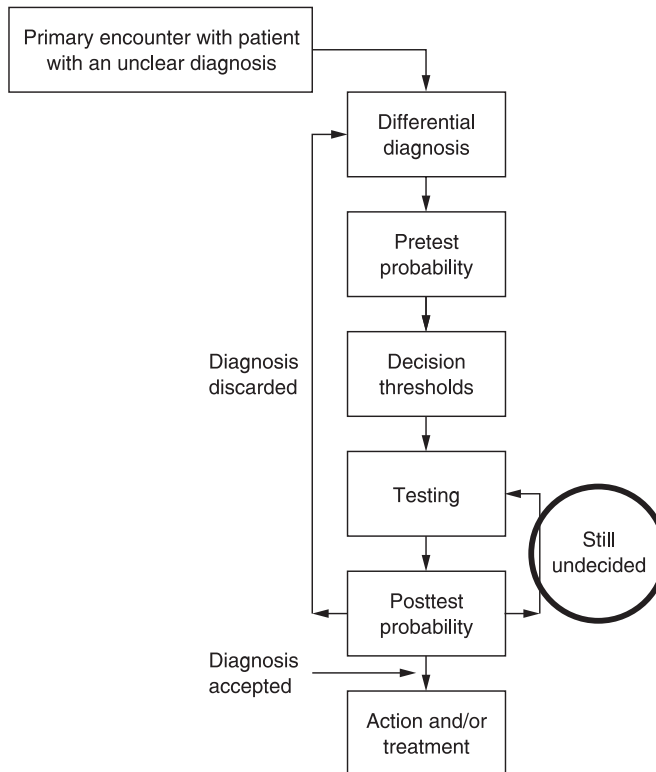
This action may be to administer a medication, admit the patient, call a consultant, or simply give the patient appropriate discharge instructions and a follow-up appointment.

If the post-test probability still lies within the indeterminate zone (Fig. 2-19C), then we still have not made or eliminated the diagnosis and further testing is needed. The venue for this continued testing depends on the clinical situation. It can take place in the emergency department, in the hospital, or as an outpatient.

The diagnostic testing we have already performed may not have been a waste of time. The post-test probability after the first round of testing becomes the new pre-test probability for any further tests we order (Fig. 2-20).

SOURCES OF ERROR

In emergency medicine, we need to make rapid, accurate diagnoses. A good deal of emergency medical cognition is performed subconsciously, almost automatically. The impetus for many of our diagnostic decisions



— FIGURE 2-20 — *If our post-test probability has not allowed us to discard or accept the diagnosis, then we must perform further testing. We then use the generated post-test probability as our new pre-test probability for this additional testing.*

comes from behind the scenes. If we have integrated flawed reasoning into our diagnostic methodology, we are prone to make errors. By recognizing this potential and systematically eliminating sources of error, we can free our decision-making process from the potential pitfalls.

■ Diagnostic Errors

Error in medicine has received an enormous amount of attention in recent years. Kassirer has delineated the types of error that occur during the process of diagnostic decision-making.^{46,47}

- *Faulty triggering:* The signs and symptoms do not lead us to a consideration of the diagnosis.

- *Faulty context formulation:* We assign an improper diagnosis because of the setting or other external factors.
- *Errors in gathering or processing information:* We make an error because we have performed a poor history, conducted a flawed exam, or ordered the wrong test.
- *Faulty interpretation of clinical testing:* We draw improper conclusions from test results, leading to misdiagnosis.
- *Errors in verification:* We prematurely discard or accept a diagnosis before a sufficiently thorough evaluation
- *No-fault errors:* We miss the proper diagnosis, but in retrospect, no clear cognitive missteps can be found.

We can simplify this schema even further by dividing the cognitive missteps of diagnostic decision-making into errors of information gathering and errors of information integration.⁴⁸

Errors pertaining to the gathering and acquisition of information are known as *slips* and *lapses*.

■ Slips and Lapses

These forms of error involve the gathering of false data; by virtue of this erroneous information, our diagnostic pathways are compromised. Many slips and lapses are caused by the frequent interruptions and multiple simultaneous stimuli characteristic of the environment of a typical emergency department.

Slips

If we are distracted, our inattention can cause us to misread or falsely receive information. We may look at the number 1.1, but see 11; if this number is the white blood cell count in an oncology patient, the results of this error can be serious. When looking at a chest radiograph, if we are simultaneously signing an order sheet, a subtle pneumothorax can easily be missed.

Cognitive drift is the cause of many slip-type errors. Optimally, we process one task or thought pattern at a time. If we want to check labs, but the computerized information system takes 45 seconds to start-up, we naturally might use this time to think of other things than the values we wanted to check. When the computer system does eventually start, we may be distracted. Informatics experts therefore view sub-second computer information system response times as essential to avoid error due to cognitive drift.⁴⁹

Lapses

Lapses are errors of memory. Just as in slips, distraction can play a role in causing us to forget information we normally can recall readily. Often these lapses deal with skills and tasks we normally would perform automatically. If in the course of the preparations for suturing an arm laceration, we look up at the full chart rack, we may forget to perform our standard search to diagnose foreign bodies. Ordinarily, we have a set order for our preparations: first we inject anesthetic, then we clean the wound, then the search for foreign bodies, and finally we suture. However, a distraction can cause us to deviate from this routine and forget our normal process, leading to a missed diagnosis.

It is also a lapse when we place drops of urine on a bedside pregnancy test and read the results as soon as the indicator lines appear. We are forgetting the necessity of waiting the proscribed period of time for the reagents to mix with the urine. This lapse can lead to a false diagnosis of a patient's pregnancy status.

■ Mistakes

Slips and lapses cause us to bring *false data* into our diagnostic decision-making process. If the data we assimilate is correct, but we process it erroneously, cognitive psychologists would refer to this error type as a *mistake*.

While slips and lapses were errors of information gathering, mistakes are errors of information processing and utilization. The difficulty of using information to make diagnoses in the emergency department is the rapidity with which we must make decisions. One factor which contributes to our ability to make decisions in the rapid manner necessary in our practice, is the use of heuristics.⁵⁰

Heuristics

Croskerry provided a definition of heuristics as a

“Cognitive process that simplifies clinical decision-making operations, describing the everyday intuitive decisions that emergency physicians make without resorting to formal decision analysis.”⁵¹

While heuristics can aid in the urgent formulation of a diagnosis, these rules of thumb can also lead to error. Non-beneficial heuristics are often referred to as “biases,” though the ambiguity of this term has lead Croskerry to call them a *cognitive disposition to respond (CDR)*.^{7,50}

Comprehensive lists of heuristics that may affect the thinking of emergency physicians can be found in the literature.^{7,52}

Ultimately, the use of heuristics is unavoidable. McDonald advises physicians to become aware of the heuristics used in decision-making and to consider an ongoing process of pruning and re-examining such criteria.⁵³ Awareness of potentially flawed or failed heuristics can prevent them from negatively influencing our decisions. Relevant clinical evidence can be a powerful adjunct to rejecting or revising erroneous heuristics. What follows is a discussion of the potential errors inherent in each phase of diagnostic decision-making.

■ Errors in the Creation of a Differential Diagnosis

If our illness scripts are well-developed and extensive, we are able to accurately and quickly develop a differential. If inexperience or an anomalous patient presentation forces us into a state of diagnostic uncertainty, we are more prone to error. It is in this state that we are most subject to biasing heuristics.

ROWS: Rule-out Worst Scenario

It is our nature as emergency physicians to assume the worst. Asking the question “What is going to kill this patient?” is one of the foundations of emergency medicine decision-making. We always search for the most serious or life-threatening explanation for a patient’s presentation and then attempt to “rule-out” these dangerous diagnoses. In most circumstances, this heuristic is beneficial and essential given the unique nature of our practice. If taken too far, it can lead to needless work-ups for benign conditions.⁷

Availability Bias

Diseases that are easier to remember will spring to mind readily; diseases that are difficult to remember are often not considered. This heuristic can cause us to leave infrequently encountered disorders off the list of differential diagnoses.^{50,54}

Confirmation Bias

This heuristic deals with the momentum that occurs when a clinician latches on to one diagnosis. Often, we will accept new evidence only if it confirms this working diagnosis and ignore the information if it does not. This is a form of “anchoring”; it is human nature to regard positive, confirmatory evidence more highly than negative information.⁵⁵ When we set our stakes on one diagnosis, other potential entities can be missed.

Occam's Razor

William of Occam, a medieval philosopher, postulated the principle of parsimony; i.e., “Plurality should not be posited without necessity.” This principle demands the search for the simplest explanation for any course of events. When found, that simplest explanation will likely be the correct one. To state the theory another way, “One should not increase, beyond what is necessary, the number of entities required to explain anything.”⁵⁶ While it is quite satisfying to postulate a unifying diagnosis for all aspects of a patient's presentation, we must keep in mind that two or more problems might have caused the patient to arrive in the emergency department. While beneficial as a general mindset, if only singular explanations are considered, we will miss diagnoses.^{7,50}

Search Satisficing

Similar to confirmation bias above, this source of error stems from the lack of consideration of other diagnoses once we are positive about one diagnosis. A Nobel Prize winning economist, Herbert Simons, first described the concept; the term is a combination of satisfying and sufficing.⁵⁷ This heuristic causes us to abandon the search for further fractures on a radiograph once one fracture is found. In the same vein, if after exploration a foreign body is discovered in a laceration, there is a strong urge to cease the search for additional foreign bodies.^{7,17} This bias is a form of premature closure, a cessation of the search prior to finding *all* of the correct diagnoses.

Prior Extensive Work-ups

When we deal with a patient with multiple presentations and/or an extensive work-up preceding the current visit, it can predispose us to error. This bias causes a narrowing of the differential in such patients as they have already been worked up the “yin-yang.”⁷ The danger lies in the possibility that such patients may have a new condition or a missed diagnosis despite their prior exhaustive work-up. “Frequent fliers” also fall into this category; this bias causes us to miss the intracranial bleed on the inebriated gentleman who has visited the emergency department every night for the past three years.

■ Errors in Assigning Pre-test Probability

The process of assigning pre-test probabilities is not free of cognitive pitfalls. Just as heuristics affected differential diagnosis, they can lead to errors in the estimation of pre-test probability. Though there are many forms of

bias that can cause false estimation of pre-test probability, the following are common in emergency medicine.⁷

Sampling Bias

Limited or skewed exposure to the diseases that cause various presentations can lead to an over-representation of rare diseases or an under-representation of common ones. If this biased exposure is applied to a different population, then it can foil pre-test probability estimates.⁵⁰

EXAMPLE

A medical student has worked on the wards at a referral center for pulmonary hypertension patients. Three of his first five patients with shortness of breath had this disease as the etiology of their dyspnea. When he rotates to the emergency department and presents his sixth patient with shortness of breath to his attending, he naturally places the pre-test probability of pulmonary hypertension at 60%. In the general emergency population with dyspnea, the prevalence of pulmonary hypertension is actually much lower than 60%.

This heuristic also pertains to diseases that receive extensive journal coverage. Some rare conditions receive literature coverage out of proportion to their prevalence; this can lead to faulty estimation of pre-test probability.⁵⁸

Saliency Bias

When assigning pre-test probabilities, certain diseases come to mind more quickly than others do. For instance, we may have seen a striking example of the disease on our previous shift; this form of the bias is motivated by a recency effect. Novel clinical features of a disease may make it more striking than others with equal prevalence. If a previous misdiagnosis has led to legal action or embarrassment, a clinician will often vow consciously or subconsciously to never miss this diagnosis again. The *saliency* of these diagnoses may lead us to assign a higher than accurate pre-test probability. This source of error is similar to the availability bias we discussed under differential diagnosis.⁵²

[Link to Page 62](#)

EXAMPLE

A patient presenting with isolated substernal chest pain wound up dying in the department from a thoracic aortic dissection during one of your shifts. From that point on, you assign high pre-test probabilities

to the diagnosis of dissection in all of your chest pain patients. This falsely elevated pre-test probability requires you to send all of these patients for CT angiograms. Without this bias, in a majority of these patients your history and physical would allow you to discard the diagnosis of dissection without testing.

Equal Weighting of Clinical Characteristics

This form of bias assigns equal value to clinical characteristics despite the varying significance of these characteristics. Examining articles on the clinical manifestations of disease can aid in the elimination of this negative heuristic.

EXAMPLE

A patient presents with fever and headache, but no Kernig's sign or Brudzinski's sign. If we give the same weight to the absence of the latter two signs as to the presence of the first two, we would calculate a falsely low pre-test probability for the diagnosis of meningitis. The presence of fever and headache are much more predictive of meningitis than the absence of the latter two signs.⁵⁹

Avoidance of this form of bias relies on the integration of the frequency of the clinical manifestations of disease into our illness scripts.

Base Rate Neglect

This bias causes the assumption that all possibilities in the differential list have equal pre-test probability despite varying prevalence. This can be a byproduct of the ROWS heuristic causing us to assign the same probability to all life-threatening diagnoses as we would to less dangerous, but more common, diagnoses.⁷

[*Link to Page 62*](#)

EXAMPLE

A patient comes to the emergency department with crushing chest pain. You quickly make a differential consisting of acute coronary syndrome, pneumothorax, pericarditis, pulmonary embolism, and Boerhave's rupture of the esophagus. If you calculated a pre-test probability of 20% for each of the disorders, you would be ignoring the fact that Boerhave's is much less common than the other diagnoses.

Novices also commonly fall prey to this form of error, as it is often easier to create a differential than to accurately assign pre-test probabilities to each

of its constituents. Assigning an equal value to each part of the differential is an easy but flawed way of dealing with this difficulty.

Gamblers' Fallacy

If an uninformed gambler sees the roulette wheel stop on red five times in a row, he assumes that the bet on black for the next spin is a “sure thing.” In decision-making, this bias causes the erroneous assumption that if we diagnose a number of patients with a serious condition, it is less likely for the next patient to have this same serious condition.² This misconception of chance can result in a false decrease in our estimation of pre-test probability.⁵⁰

EXAMPLE

During your shift, you have three patients in a row present with chest pain who then have enzyme evidence of myocardial infarction. If a fourth patient comes in with chest pain, this bias would cause the underestimation of pre-test probability of myocardial infarction as the etiology of his chest pain.

Prototypical Error

This bias causes the underestimation of pre-test probability when a presentation of a disease is not the one described in the textbooks. It is particularly dangerous when *atypical* presentation is far more common than the *classic* presentation. This failed heuristic is most prevalent during the nascent stages of our development of illness scripts. As we progress through our career, our illness scripts embrace a wide range of disease presentations. Inevitably, we begin to understand that the classic presentation is very infrequently the same as the *typical* presentation.

EXAMPLE

A 70-year-old diabetic woman presents with epigastric pain and shortness of breath. This patient's pre-test probability for ischemic heart disease should be quite high regardless of the fact that she does not have retrosternal chest pain.

■ Errors in the Decision to Perform Diagnostic Tests

As we discussed in the prior sections, testing without diagnostic uncertainty is not only unnecessary, but can lead to false results. Cognitive biases can lead to performing tests we do not need and can obscure the diagnostic process.

Commission Bias

In this context, commission bias refers to the performing of diagnostic testing even if the pre-test probability lies above the accept threshold or below the discard threshold. We often feel the need to *do something*. This bias is operative when we do not feel that our clinical reasoning is a tangible *something* or confidence in our decision-making is lacking. In other situations, it will be the patient directly requesting an objective test as opposed to the supposedly subjective judgment of the clinician. Most patients are quite satisfied with an explanation of the thought process that led us not to need a test, if we take the time to delineate our reasoning. If we still feel the need to order a test, instead of one that would simply confirm what we already have decided, the more worthwhile test would be one that has the potential to alter our diagnosis.

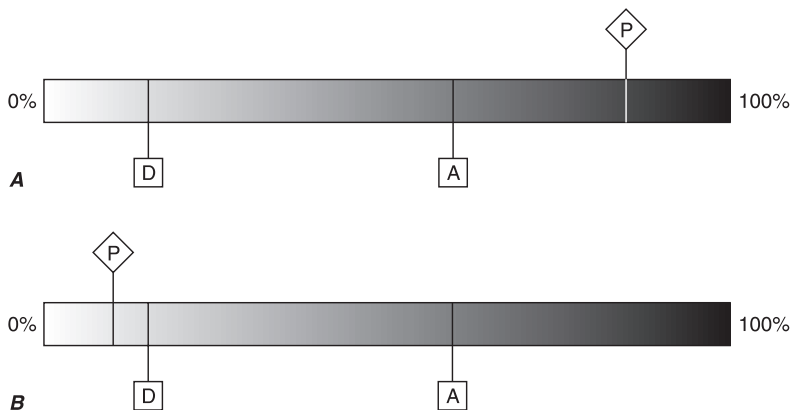
EXAMPLE

You have two 5-year-old patients complaining of a sore throat. The first has anterior cervical lymphadenopathy, exudate on his tonsils, and a low-grade fever. He is not complaining of a runny nose or a cough. You set your pre-test probability at 85% for strep pharyngitis. You set your discard threshold at 15% as missing this diagnosis is unlikely to result in harmful sequelae.⁶⁰ You set your accept threshold at 60% as the risks of over-treatment and over-diagnosis are quite small when considering your patient (Fig. 2-21A).

You decide that you will treat the patient with penicillin without any further testing, but the patient's mother pleads with you to do that new rapid strep test. You realize that the rapid strep test, if positive, will do nothing to change your plan to treat. A negative rapid test will not lower your pre-test probability enough to keep you from treating the patient based on your clinical exam.* You explain this to the mother and she accepts your reasoning.

Your second patient keeps wiping his runny nose with his sleeve. As you examine his throat, you notice only a mild redness. He has no palpable lymph nodes and is without fever. You set his pre-test probability for strep at 10% with the same decision thresholds (Fig. 2-21B). You tell the mother that you are comfortable sending the patient home without antibiotics and follow-up with his pediatrician in two days to ensure the resolution of what is likely a viral illness. The mother pleads with you to do a strep test. As you think about the test in this situation,

*Rapid strep test: sensitivity = 80%, specificity = 99%, LR+ = 80, LR- = 0.2. The likelihood ratio negative would not drive your post-test probability to below your accept threshold. In a later section, the mathematics necessary to do these calculations will be discussed.



— FIGURE 2-21 — **A** The pre-test probability crosses the accept threshold: no testing is necessary. **B** The pre-test probability is below the discard threshold: no testing should be necessary.

you realize that a negative result will do very little to change your thinking, but a positive result would actually make you strongly consider “Strep.” Pharyngitis as the false positive rate of this test is low. You perform the rapid test; the patient’s mother feels quite relieved to hear it is negative.

If we do order a test, even when the pre-test probability would allow action without testing, we should only use tests that can change our decision to treat or discard a diagnosis.

Entrapment Bias

If halfway through a movie we realize that we are not enjoying ourselves, we have two choices. We can stay to the end or we can get up and leave. The rationale often raised for the first choice is that, since we have already spent the money for the ticket, we might as well stay through until the end, as we cannot get our investment back. Of course, this choice guarantees that we waste additional time, which could have been used for other entertainment. This misconception is an example of entrapment bias. This bias is pervasive in medicine as well; if we start a work-up and then realize it is not needed, we do not need to continue the work-up.

EXAMPLE

One of your residents is presenting a 25-year-old male with chest pain. He details his thorough history and physical and tells you that ordinarily

he would send the patient home now, with primary care follow-up. However, a set of cardiac enzymes was sent before he got to the patient. He tells you that now we will have to admit the patient, because one set of enzymes is not enough to rule-out a myocardial infarction, and since the work-up has already begun, it must be completed. You inform him that since he did not feel the evaluation was necessary before sending the enzymes, the work-up does not need to be completed. You document your thought process in the chart and send the patient home.

■ Errors in the Interpretation of Test Results

How we determine what a test result means to the diagnosis of our patient is also error-prone.

Forced Dichotomy

This bias refers to a desire to see results as positive or negative even when the true result may be indeterminate. Many clinicians find it easier to think of tests in this manner, but as we have discussed above, even tests that are normally reported as positive or negative actually are scalar with predetermined cutoffs. The bias is especially operative with tests that do not easily fit into a dichotomous means of reporting, such as radiological studies.

EXAMPLE

You send a 28-year-old female for a CT scan of the abdomen and pelvis for severe right lower-quadrant abdominal pain. After she returns, you pull the radiology report up on the computer. You quickly scroll down to the bottom of the report, which reads: “No evidence of appendicitis noted.” You interpret this as a negative and send the patient home. She comes back eight hours later with generalized abdominal pain and a rigid abdomen. The operative report after the completion of her surgery describes a perforated appendix with peritonitis.

You return to the original CT report, which has the line: “Appendix not visualized; no contrast in cecum.” In retrospect, this CT result is not positive, nor is it negative; the patient needed further diagnostic testing in order to determine her diagnosis.

One possible means of avoiding this bias is the assignment of individual likelihood ratios to clinically indeterminate results. This strategy was successfully used for the low and intermediate results of a V/Q scan as we mentioned in the prior section.

At times, even the proper assignment of likelihood ratios does not prevent this bias from causing poor decisions. At times, a test branded positive or negative will cause acceptance of a diagnosis regardless of the actual test result characteristics.

EXAMPLE

One of your residents is evaluating a patient with shortness of breath. Due to the patient's history and physical exam, your resident assigns a low pre-test probability to this patient for the diagnosis of pulmonary embolism. He appropriately orders a d-dimer (Elisa LR + = 1.73, LR – = 0.11). When the result comes back positive, he decides to admit the patient and start heparin. Even though the results of the test were positive, the likelihood ratio of this result does almost nothing to change the previously low assessment of pre-test probability.

■ Avoiding Cognitive Error

Biases such as the above are integrated into our thinking processes. Though sometimes useful, they can lead to missed diagnoses. Research in cognitive psychology indicates that biases can be unlearned by awareness and deliberate analysis of our behavior. One form of this process of cognitive debiasing is provided by metacognition.⁵¹

Metacognition

This branch of cognitive psychology deals with the awareness and understanding of our own thoughts.⁶¹ Metacognition allows us to analyze our behavior and search for mental processes that predispose us to error.^{17,51,62} Experience brings the ability to know what we do not know and makes us aware of when we are not processing information well. Metacognition is difficult for the novice, both because of a deficiency in clinical knowledge as well as a lack of past mistakes to draw upon.¹⁷

Cognitive Forcing Strategies

Often, an awareness of factors that predispose us to error can aid in their avoidance. Some biases are so entrenched in our thought processes that, even with the awareness of metacognition, we are still prone to error. There are strategies that we can use to lower the potential risks even further, such as cognitive forcing strategies.

This process evaluates consistent cognitive errors and creates solutions, which force the avoidance of these errors.⁵¹ An example is a small sign

placed in the physician charting area during the autumn and winter months reading: “Is this carbon monoxide poisoning?” Similarly, cognitive forcing is operative when we enter the complaint of chest pain radiating to the back into an electronic charting system causing a small box to pop up on the screen. The box has a message urging the physician to consider thoracic aortic dissection. In both examples, the physician is forced to consider adding a diagnosis to the differential. Even if in the vast majority of cases the physician would have already considered the diagnosis prior to the cognitive forcing, the combination of exhaustion, a busy department, and the juggling of numerous simultaneous tasks can lead to occasional misses. These forcing strategies attempt to prevent missed diagnoses even when our decision-making process is clouded.

EXAMPLE

When a patient presents in extremis with crushing substernal chest pain, we must rapidly make the diagnosis and treat with numerous medications. Realizing that this is a situation where contemplative critical thinking is difficult for the nurses and physicians, the director of the emergency department decided to develop a cognitive forcing strategy with the help of the pharmacy. All bottles of nitroglycerin have a sticker placed on top of the cap reading “Pt on VIAGRA?” This action was motivated by the knowledge that, in the chaos of treating a sick patient, it is easy to forget the severe hypotension that can be caused when nitroglycerin is administered to patients taking sildenafil and similar agents.

Algorithms

Diagnostic algorithms, such as Advanced Cardiac Life Support (ACLS) and Advanced Trauma Life Support (ATLS), provide another form of debiasing. If a patient presents with pulseless electrical activity (PEA), the ACLS course teaches the eight most common diagnoses along with mnemonics to aid in memory. The difficulty with this form of debiasing is that these algorithms are only as effective as a clinician’s ability to remember them.

It is interesting that reluctance to adopt an algorithm can itself serve as a cognitive forcing strategy. ATLS advocates a film of the pelvis on all trauma patients; this is against the practice of many emergency physicians, when the patient has undergone only mild trauma. When we consciously discard the algorithm’s recommendations, it still forces us to consider the possibility of pelvic trauma. If we decide not to obtain the radiograph, we know we are flouting the surgeons’ standard of care and may apply increased diligence to assure that there is no pelvic injury.

■ Summary

Diagnostic error is a looming threat during every shift in the emergency department. By carefully screening our thinking for bias and false assumptions, we have an opportunity to reduce these errors. In particularly error-prone circumstances, cognitive forcing strategies can further limit the potential for error and harm. Algorithms and literature support, in the form of clinical guidelines and prediction rules, give us a springboard to developing our own ideal diagnostic plans.

PUTTING IT ALL TOGETHER

We have presented a structured approach to evidence-based diagnostic decision-making in emergency medicine. This approach brings evidence and quantitative information to bear on areas that traditionally involve purely “clinical” reasoning. While the process may seem daunting and unwieldy on first inspection, this approach of critical thinking quickly becomes intuitive and fully integrated into practice. We are convinced that you will ultimately find that the pain of acquiring these skills is compensated many fold by virtue of having become a smarter, more insightful, and also quicker emergency physician.

The alternative is to make non-critical diagnostic decisions. It may seem clinically intuitive that a patient with a negative CT scan of the head does not have a subarachnoid hemorrhage, until we miss one. It may make sense clinically that a patient with a normal WBC does not have a surgical abdomen, until she returns with a perforation. These two misjudgments were probably pointed out to us as problematic during the nascent stages of our training, but what of the scores of slightly flawed assumptions we do not even know exist within our biased thinking?

These methods do not require the abandonment of clinical judgment. In fact, as we have discussed, clinical judgment is essential to all parts of the process of evidence-based decision-making. Further, we build our clinical judgment on experience, and the foundation of experience is often the mistakes we have made. Evidence-based diagnostic decision-making offers another path to developing our clinical judgment by allowing the examination of seemingly logical diagnostic plans before we make a mistake.

More than anything else, this strategy forces us to think through our plans. This process of critical evaluation allows us to avert errors and avoid causing harm to our patients.

CASE: THE CLOT THICKENS

It is a busy Monday at 3 p.m., and Dr Wayne is signing out his patients to you. You both walk over to bed #10 where a young woman looks up from the gurney. He tells you that he picked up the chart of this 25-year-old woman just 15 minutes ago, but the case is an easy one. She has just returned home after an eight-hour plane ride and is now complaining of a “little” shortness of breath and some reproducible chest pain. He goes on to tell you that he ordered a V/Q scan, which has not yet been done. He advises: “If it comes back normal or low probability, send her home. If it comes back intermediate or high, just put her on heparin and admit her as a confirmed PE.” He finishes sign-out and leaves the department to you. You decide to start over with bed #10.

After a brief history and physical exam, you gather the following information.

Clinical Exam

Chief Complaint I took a plane flight yesterday and now my chest hurts.

History of Present Illness The patient is a 25-year-old female, who flew from California back to New York yesterday. This morning, around 10 a.m., she experienced the sudden onset of right-sided chest pain described as constant, respirophasic, without radiation or alteration with changed position. She is also complaining of mild dyspnea.

Past Medical History None.

Past Surgical History None.

Family History Paternal grandfather died of myocardial infarction at age 78.

Social History No tobacco or drug use, including cocaine.

Medications Oral contraceptives.

Allergies None.

Physical Exam Well-kempt 25 y/o female appearing her age. Temperature 99.2°F; pulse 96/min; BP 110/72 mmHg; respiration 24/min; SaO₂ 98% on room air.

- HEENT normal
- Neck: No jugular venous distention, trachea is midline.

- Lungs: Clear to auscultation bilaterally. Mild tenderness right chest wall which reproduces the patient's pain.
- Heart: S1 S2, regular rate and rhythm and no murmurs, rubs, or gallops.
- Abdomen: Benign.
- Extremities: Normal exam. No calf tenderness, masses, or cords.

You send Cindy, the medical student rotating in the department, in to interview the patient as well. After Cindy's interview, the two of you discuss the diagnostic plan.

Differential Diagnosis

When you ask Cindy her differential diagnosis, she proceeds to do an extremely thorough job of listing every known cause of chest pain that has ever plagued a human being. Her list has little connection to the patient's age, risk factors, or circumstances.

You then relay to her your own list: Pattern matching brought up the diagnosis of pulmonary embolism as soon as you read the chief complaint on the nurse's triage note. Upon interviewing the patient, you mentally reviewed many of the diagnoses on Cindy's list but discounted almost all of them except spontaneous pneumothorax, costochondritis, or musculo-skeletal pain.

Pre-test Probability

Pulmonary embolism (PE) remains at the top of your list of differential diagnoses. You know of quite a few different clinical prediction rules in the literature, which can aid in the calculation of pre-test probability for pulmonary embolism. You are most familiar with the Wells' criteria and decide to use them in this case.⁶³ You know this clinical prediction rule has been validated and has proven to be accurate in the emergency department population.

Your patient scores 3 points for pulmonary embolism being more likely than any alternative diagnosis; she has none of the other Wells' criteria (Fig. 2-22). A score of 3 points translates into a moderate pre-test probability, or quantitatively, a 21% pre-test probability.

Though they are not part of the Wells' criteria, you decide to adjust the pre-test probability due to the use of oral contraceptives and the plane flight. This alteration of pre-test probability in light of the patient's unique circumstances is based on your clinical experience. You decide to raise the pre-test probability of pulmonary embolism to 30%; you write this number in the chart.

Wells' Criteria for PE

Criteria	Points
Suspected DVT	3.0
An alternative diagnosis is less likely than PE	3.0
Heart rate >100 beats/min	1.5
Immobilization or surgery in previous 4 weeks	1.5
Previous DVT/PE	1.5
Hemoptysis	1.0
Malignancy (on treatment or treated within past 6 months)	1.0

Score	Mean probability of PE	Risk
<2	3.6%	Low
2-6	20.5%	Medium
>6	66.7%	High

— FIGURE 2-22 — *Wells' criteria for estimating quantifying the pre-test probability of pulmonary embolism.*⁶³

You do not know of any decision aids for the assignment of probabilities for the other diagnoses on the list, so you use your clinical judgment. You assign a pre-test probability of 5% to spontaneous pneumothorax and 65% to costochondritis or musculoskeletal pain. The ROWS heuristic leads you to pursue the diagnosis of pulmonary embolism first, even though it is not the most likely diagnosis based on pre-test probability.

Decision Thresholds

Next, you pick your discard and accept thresholds for pulmonary embolism. Missing a clot can be life-threatening, not because of the current situation, but because of the risk of a subsequent larger embolism. Falsely diagnosing a pulmonary embolism is also problematic. If the patient is erroneously given this diagnosis, it will expose her to the risks of anticoagulation, affect insurance premiums, and forever brand future medical interactions.

You decide to set the discard threshold at 2% and the accept threshold at 80%. To illustrate for Cindy what these numbers mean, you decide to draw the threshold diagrams for your work-up of pulmonary embolism.

You ask Cindy how she would go about working up this patient for pulmonary embolism. She tells you she wants a chest x-ray, an arterial blood gas, and an EKG.

Chest Radiograph You tell Cindy that the chest x-ray very rarely will help with the diagnosis of pulmonary embolism, but it may be useful to evaluate other conditions. You decide that you will get a chest x-ray as it will allow you to drive the probability of spontaneous pneumothorax below the discard threshold. You could have easily argued that, in the course of further studies for pulmonary embolism, the chest would be imaged and therefore hold off on the chest x-ray.

Electrocardiogram The EKG is also rarely useful in modifying the pre-test probability of PE. If acute coronary syndrome was in the differential, it would certainly be worth getting an EKG to evaluate for this possibility as well as other cardiac disorders, but in this patient, this is not a consideration given the absence of any risk factors. Supposedly, pathognomonic signs such as S1 Q3 T3 are neither sensitive, nor specific for pulmonary embolism.⁶⁴



Arterial Blood Gas (ABG) As a field, we are moving towards a consensus that, aside from assessing a patient after intubation, this test is not useful in the emergency department. When we look at the data from PIOPED, the test has little utility in the workup of pulmonary embolism.⁶⁵

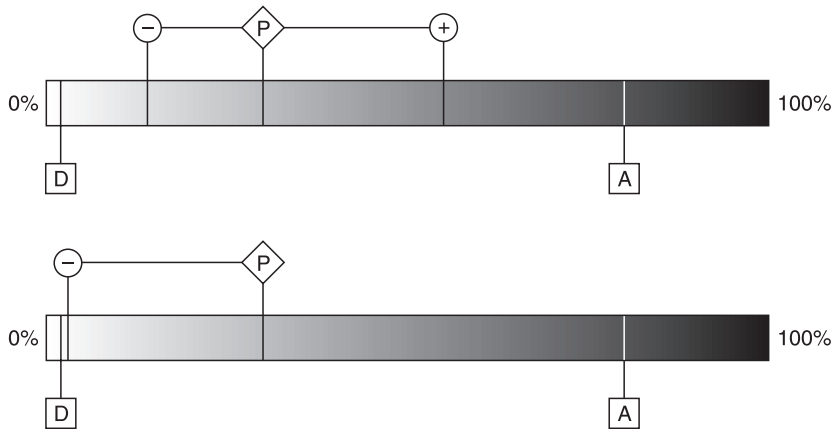
The ABG is at its most sensitive when the criteria for an abnormal test are a $\text{PaCO}_2 < 35$, a $\text{PaO}_2 < 80$, or an Aa gradient > 20 . The LR negative using this criterion is 0.9.⁶⁵ This likelihood ratio has no ability to change our pre-test probability.

While you continue to discuss further diagnostic workup, you ask the nurse to take the patient over to get her x-ray, if her urine pregnancy test is negative.

You continue to barrage Cindy with questions about other diagnostic tests for pulmonary embolism. Digging deep into her first two years of medical school, she suggests a d-dimer and Dopplers of the legs to look for DVT.

D-Dimer This test has revolutionized the work-up of low-risk patients presenting to the emergency department with suspected DVT or PE. Unfortunately, the institution you are working in has a first-generation latex

d-dimer as its only assay. The characteristics of this test are a likelihood ratio positive of 2.9 and negative of 0.39.⁶⁶ We see that it would do little to alter our pre-test probability. If we had an ELISA d-dimer, a negative result would have a more significant effect. The likelihood ratio negative of an ELISA assay is 0.06. You elect not to order the d-dimer assay.



Dopplers of the Legs A positive DVT in a patient with chest symptoms virtually rules-in the diagnosis of pulmonary embolism. However, in a patient without leg symptoms or signs, the diagnostic yield of this exam is quite low. A negative ultrasound exam for DVT would do little to alter our pre-test probability for pulmonary embolism.

You tell Cindy that since a conventional pulmonary angiogram is not available at your hospital, there are two remaining choices: V/Q scan or CT angiogram.

Ventilation/Perfusion (V/Q) Scan We have discussed the likelihood ratios of the interval results of the V/Q scan in the prior sections.

RESULTS OF V/Q SCAN	LIKELIHOOD RATIO
Normal	0.1
Low	0.4
Intermediate	1.2
High	18.3

From these values, we can see that only a high-probability scan will have the ability to allow us to stop the diagnostic process without further testing.

RESULTS OF V/Q SCAN	POST-TEST PROBABILITY
Normal	4%
Low	15%
Intermediate	34%
High	89%

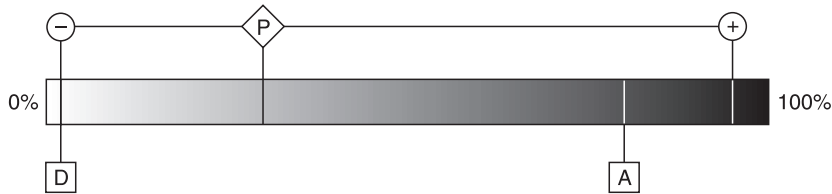
The most common results of a V/Q scan are low or moderate probability.⁶⁷ Even a normal result will not drive our post-test probability below the discard threshold.

CT Angiogram of the Chest A year ago, you only ordered CT angiograms for pulmonary embolism when your patients had severe underlying lung disease. You worried about the poor sensitivities for subsegmental clots and questioned the ability of your radiology department to interpret these scans. Two things have changed in your institution in the past few months. All scans are now performed on a new multidetector machine capable of 1-mm cuts and imaging of the entire pulmonary vasculature during a single breath. In addition, an outside group with fellowship training in CT imaging of the body now reads all CT angiograms of the chest.

These two factors have made multidetector spiral CT angiogram your first test of choice for pulmonary embolism. The most recent studies of this level of technology give a LR negative of approximately 0.05 and LR positive of 45.^{68,69} These levels of accuracy are only achievable if poor quality scans, which do not achieve good visualization of subsegmental branches, are read as inconclusive and not as negative.

Not only are the testing characteristics of CT superior, but there is actual patient-oriented outcome data for this test. The risk of recurrent venous thromboembolic disease (PE or DVTs) in patients sent home after a negative CT angiogram and without anticoagulation is 0.5–1% at 3 months.⁶⁹⁻⁷¹ These data are as compelling as the excellent likelihood ratios. Your viewpoint is supported by the clinical policy of the British Thoracic Society, which accepts a negative new-generation CT angiogram as sufficient to rule out the diagnosis of PE (Fig. 2-23).⁷²

If your institution had older machines or a radiology department inexperienced with this exam, then the likelihood ratios would have to be



— FIGURE 2-23 — *The CT angiogram has the ability to both rule-in and rule-out the diagnosis of pulmonary embolism in your patient.*

adjusted. In this case, the *test characteristics* in the above-mentioned studies would not be *applicable* to your patient.

Testing

You put the order for CT scan in the computer. The patient has returned from x-ray; you pull the radiograph up on the point-of-care radiology system. The chest film appears completely normal. This does not affect the probability of pulmonary embolism, but you discard the diagnosis of spontaneous pneumothorax.

Cindy asks if we need renal function laboratory tests before the CT scan. You tell her that, in the absence of any risk factors such as preexisting problems or a history of diabetes, the pre-test probability for renal disease is almost nil.⁷³

Post-test Probability

Your patient returns from radiology; a few minutes later, you get a call from the attending radiologist. He first discusses the technical quality of the scan. This is of vital importance as, if no clot was visualized on a poor quality scan, then the result is not negative, it is indeterminate. This scan was of excellent quality, with good timing of the dye bolus and visualization of the entire pulmonary vascular tree. No clot was visualized, nor was any additional lung or chest pathology noted.

Action

You explain the results to your patient with Cindy listening in. You tell them both that a negative CT scan is sufficient to drive the post-test probability of PE below the discard threshold. You believe she is suffering from a musculoskeletal condition and tell her to buy and take ibuprofen for the pain and inflammation.

You also explain that complete diagnoses are often difficult in the emergency department. You want her to follow-up with her private medical doctor in the next few days for a check-up.

