

# 1

## Basic Principles of Diagnostic Test Use and Interpretation\*

---

*Diana Nicoll, MD, PhD, and William M. Detmer, MD*

The clinician's main task is to make reasoned decisions about patient care despite imperfect clinical information and uncertainty about clinical outcomes. While data elicited from the history and physical examination are often sufficient for making a diagnosis or as a guide to therapy, more information may be required. In these situations, clinicians often turn to diagnostic tests for help.

### 1. BENEFITS; COSTS AND RISKS

When used appropriately, diagnostic tests can be of great assistance to the clinician. Tests can be helpful for **screening**, ie, to identify risk factors for disease and to detect occult disease in asymptomatic persons. Identification of risk factors may allow early intervention to prevent disease occurrence, and early detection of occult disease may reduce disease morbidity and mortality through early treatment. Optimal screening tests meet the criteria listed in Table 1–1.

Tests can also be helpful for **diagnosis**, ie, to help establish or exclude the presence of disease in symptomatic persons. Some tests assist in early diagnosis after onset of symptoms and signs; others assist in differential diagnosis of various possible diseases. Still others help determine the stage or activity of disease.

Finally, tests can be helpful in **patient management**. Tests can help to: (1) evaluate the severity of disease, (2) estimate prognosis, (3) monitor the course of disease (progression, stability, or resolution), (4) detect disease recurrence, (5) select drugs and adjust dosages, and (6) select and adjust therapy.

---

\*Chapter modified, with permission, from Tierney LM Jr, McPhee SJ, Papadakis MA (editors): *Current Medical Diagnosis & Treatment 1997*. Appleton & Lange, 1997.

**TABLE 1-1. CRITERIA FOR USE OF SCREENING PROCEDURES.**

<p>Characteristics of population</p> <ol style="list-style-type: none"> <li>1. Sufficiently high prevalence of disease.</li> <li>2. Likely to be compliant with subsequent tests and treatments.</li> </ol> <p>Characteristics of disease</p> <ol style="list-style-type: none"> <li>1. Significant morbidity and mortality.</li> <li>2. Effective and acceptable treatment available.</li> <li>3. Presymptomatic period detectable.</li> <li>4. Improved outcome from early treatment.</li> </ol> <p>Characteristics of test</p> <ol style="list-style-type: none"> <li>1. Good sensitivity and specificity.</li> <li>2. Low cost and risk.</li> <li>3. Confirmatory test available and practical.</li> </ol>
--

Diagnostic tests are not without disadvantages, however. First, diagnostic tests can be expensive. An individual test such as MRI of the head can cost more than \$1400.00. Diagnostic tests as a whole account for approximately one-fifth of health-care expenditures in the USA. Second, some tests carry a risk of morbidity or mortality. For instance, intravenous contrast material used in some CT scans leads to death from anaphylaxis in approximately one in 30,000 examinations. Third, some diagnostic tests cause discomfort to patients. The discomfort from tests such as sigmoidoscopy or barium enema, for example, will deter some patients from completing a diagnostic workup. Fourth, the result of a diagnostic test often has implications for further care in that a test result may mandate further testing or frequent follow-up. For instance, a patient with a falsely positive fecal occult blood test may incur significant cost, risk, and discomfort during follow-up sigmoidoscopy, barium enema, or colonoscopy. Last, classifying a healthy patient as diseased based on a falsely positive diagnostic test can cause psychologic distress and may lead to risks from unnecessary therapy.

Therefore, when ordering diagnostic tests, clinicians should weigh these potential costs and disadvantages against the potential benefits.

## 2. PREPARATION FOR DIAGNOSTIC TESTS

Factors affecting both the patient and the specimen are important.

### Patient Preparation

The preparation of the patient is important for certain tests. For example, a fasting state is needed for optimal glucose and triglyceride

measurements. Controlled conditions are frequently needed for endocrinology testing. For instance, posture and sodium intake must be strictly controlled when measuring renin and aldosterone levels. Strenuous exercise should be avoided before obtaining some tests such as creatine kinase, since vigorous muscle activity can lead to falsely abnormal results.

### Specimen Collection

Careful attention must be paid to patient identification and specimen labeling. Knowing when the specimen was collected may be important. For instance, aminoglycoside levels cannot be interpreted appropriately without knowing whether the specimen was drawn just before (“trough” level) or just after (“peak level”) drug administration. Drug levels cannot be interpreted if they are drawn during the drug’s distribution phase (eg, digoxin levels drawn during the first 6 hours after an oral dose). Substances that have a circadian variation (eg, cortisol) can be properly interpreted only with knowledge of the time of day the sample was drawn.

During specimen collection, other principles should be remembered. Specimens should not be drawn above an intravenous line, as this may contaminate the sample with intravenous fluid. Excessive tourniquet time will lead to hemoconcentration and increased concentration of protein-bound substances such as calcium. Lysis of cells during collection of a blood specimen will result in spuriously increased serum levels of substances concentrated in cells (eg, lactate dehydrogenase and potassium). Certain test specimens may require special handling or storage (eg, blood gas specimens). Delay in delivery of specimens to the laboratory can result in ongoing cellular metabolism and therefore spurious results for some studies (eg, low blood glucose).

## 3. TEST CHARACTERISTICS

Table 1–2 lists the characteristics of all useful diagnostic tests. Most of the principles detailed below can be applied not only to diagnostic tests but also to historical facts and physical examination findings.

### Accuracy

The accuracy of a laboratory test is its correspondence with the true value. An inaccurate test is one that differs from the true value even though the results may be reproducible (Figures 1–1A and 1–1B). In

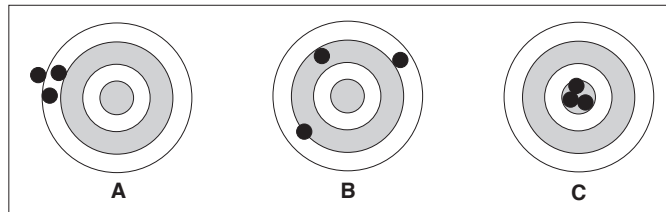
TABLE 1-2. PROPERTIES OF USEFUL DIAGNOSTIC TESTS.

1. Test methodology has been described in detail so that it can be accurately and reliably reproduced.
2. Test accuracy and precision have been determined.
3. The reference range has been established appropriately.
4. Sensitivity and specificity have been reliably established by comparison with a gold standard. The evaluation has used a range of patients, including those who have different but commonly confused disorders and those with a spectrum of mild and severe, treated and untreated disease. The patient selection process has been adequately described so that results will not be generalized inappropriately.
5. Independent contribution to overall performance of a test panel has been confirmed if a test is advocated as part of a panel of tests.

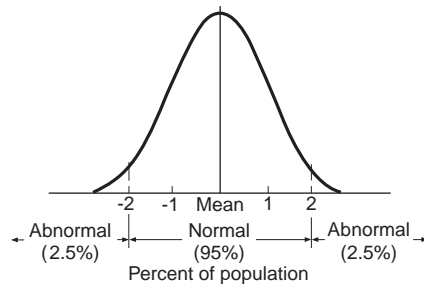
the clinical laboratory, accuracy of tests is maximized by calibrating laboratory equipment with reference material and by participation in external quality control programs.

### Precision

Test precision is a measure of a test's reproducibility when repeated on the same sample. An imprecise test is one that yields widely varying results on repeated measurements (Figure 1-1B). The precision of diagnostic tests, which is monitored in clinical laboratories by using control material, must be good enough to distinguish clinically relevant changes in a patient's status from the analytic variability of the test. For instance, the manual white blood cell differential count is not precise enough to detect important changes in the distribution of cell



**Figure 1-1.** Relationship between accuracy and precision in diagnostic tests. The center of the target represents the true test result. Figure (A) represents a diagnostic test which is precise but inaccurate; on repeated measurement, the test yields very similar results, but all results are far from the true value. Figure (B) shows a test which is imprecise and inaccurate; repeated measurement yields widely different results, and the results are far from the true value. Figure (C) shows an ideal test, one that is both precise and accurate.



**Figure 1-2.** The reference range is usually defined as within 2 standard deviations (SD) of the mean test result (shown as -2 and 2) in a small population of healthy volunteers. Note that in this example, test results are normally distributed; however, many biologic substances will have distributions that are skewed.

types, because it is calculated by subjective evaluation of a small sample (100 cells). Repeated measurements by different technicians on the same sample result in widely different results. Automated differential counts are more precise because they are obtained from machines that use objective physical characteristics to differentiate a much larger sample (10,000 cells).

### Reference Range

Patient test results are interpreted by comparing them with published reference ranges. These ranges are method- and laboratory-specific. In practice, reference ranges often represent test results found in 95% of a small population presumed to be healthy; by definition, 5% of healthy patients will have a positive (abnormal) test (Figure 1-2). As a result, slightly abnormal results should be interpreted in a critical fashion: they may be either truly abnormal or falsely abnormal. The practitioner should be aware also that the more tests ordered, the greater the chance for obtaining a falsely abnormal result. For instance, a healthy person subjected to 20 independent tests has a 64% probability of having at least one abnormal test result (Table 1-3).

It is important to consider also whether published reference ranges are appropriate for the patient being evaluated, since some ranges depend on age, sex, weight, diet, time of day, activity status, or posture. For instance, the reference ranges for hemoglobin concentration are age- and sex-dependent.

TABLE 1-3. RELATIONSHIP BETWEEN THE NUMBER OF TESTS AND THE PROBABILITY THAT A HEALTHY PERSON WILL HAVE ONE OR MORE ABNORMAL RESULTS.

Number of Tests	Probability That One or More Tests Will Be Abnormal
1	5%
6	26%
12	46%
20	64%

### Interfering Factors

The results of diagnostic tests can be altered by external factors, such as ingestion of drugs; and internal factors, such as abnormal physiologic states. External interferences can affect test results in vivo or in vitro. In vivo, alcohol increases  $\gamma$ -glutamyl transpeptidase, and diuretics can affect sodium and potassium concentrations. Cigarette smoking can induce hepatic enzymes and thus reduce levels of substances such as theophylline that are metabolized by the liver. In vitro, cephalosporins may produce spurious serum creatinine levels due to interference with a common laboratory method of creatinine measurement. Internal interferences result when an abnormal physiologic state interferes with the measurement of a test. As an example, patients with gross lipemia may have spuriously low serum sodium levels if the test methodology used includes a step where serum is diluted before sodium is measured. Because of the potential for test interference, clinicians should be wary of unexpected test results and should investigate reasons other than disease that may explain abnormal results, including laboratory error.

### Sensitivity and Specificity

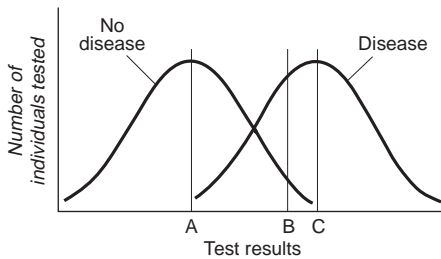
Clinicians should use measures of test performance such as sensitivity and specificity to judge the quality of a diagnostic test for a particular disease. Test **sensitivity** is the likelihood that a diseased patient has a positive test. If all patients with a given disease have a positive test (ie, no diseased patients have negative tests), then the test sensitivity is 100%. A test with high sensitivity is useful to exclude a diagnosis because a highly sensitive test will render few results that are falsely negative. To exclude infection with the AIDS virus, for instance, a clinician might choose a highly sensitive test such as the HIV antibody test.

A test's **specificity** is the likelihood that a healthy patient has a negative test. If all patients who do not have a given disease have negative tests (ie, no healthy patients have positive tests), then the test specificity is 100%. A test with high specificity is useful to confirm a diagnosis, because a highly specific test will have few results that are falsely positive. For instance, to make the diagnosis of gouty arthritis, a clinician might choose a highly specific test, such as the presence of negatively birefringent needle-shaped crystals within leukocytes on microscopic evaluation of joint fluid (Figure 2–7).

To determine test sensitivity and specificity for a particular disease, the test must be compared against a “gold standard,” a procedure that defines the true disease state of the patient. For instance, the sensitivity and specificity of the ventilation/perfusion scan for pulmonary embolus is obtained by comparing the results of scans with the gold standard, pulmonary arteriography. However, for many disease states (eg, pancreatitis), such a gold standard either does not exist or is very difficult or expensive to perform. Therefore, reliable estimates of test sensitivity and specificity are sometimes difficult to obtain.

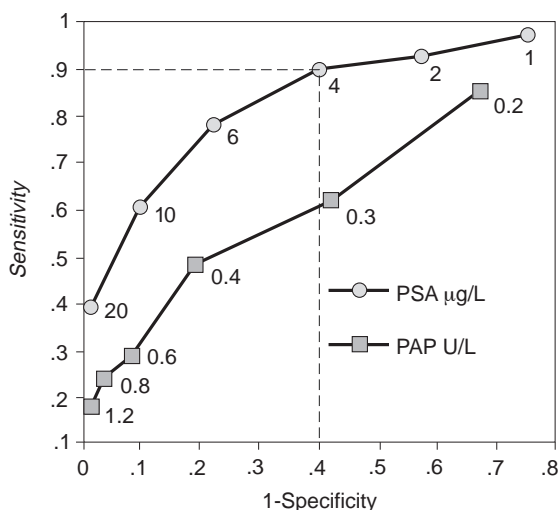
Test sensitivity and specificity depend on the reference range used, ie, the cutoff point above which a test is interpreted as abnormal (Figure 1–3). If the cutoff is modified, sensitivity will be enhanced at the expense of specificity or vice versa.

Sensitivity and specificity values can also be affected by the population from which these values are derived. For instance, many



**Figure 1–3.** Hypothetical distribution of test results for healthy and diseased individuals. The position of the “cutoff point” between “normal” and “abnormal” (or “negative” and “positive”) test results determines the test’s sensitivity and specificity. If point (A) is the cutoff point, the test would have 100% sensitivity but low specificity. If point (C) is the cutoff point, the test would have 100% specificity but low sensitivity. For most tests, the cutoff point (point B) is determined by the reference range, ie, the range of test results that are within 2 SD of the mean (point B). In some situations, the cutoff is altered to enhance either sensitivity or specificity.

1



**Figure 1-4.** Receiver operator characteristic (ROC) curves for prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP) in the diagnosis of prostate cancer. For all cutoff values, PSA has higher sensitivity and specificity; therefore, it is a better test based on these performance characteristics. (Modified and reproduced, with permission, from Nicoll D et al: Routine acid phosphatase testing for screening and monitoring prostate cancer no longer justified. *Clin Chem* 1993;39:2540.)

diagnostic tests are evaluated first using patients who have severe disease and control groups who are young and well. Compared with the general population, this study group will have more results that are truly positive (because patients have more advanced disease) and more results that are truly negative (because the control group is healthy). Thus, test sensitivity and specificity will be higher than would be expected in the general population, where more of a spectrum of health and disease is found. Clinicians should be aware of this **spectrum bias** when generalizing test results to their own practice.

The performance of two tests can be compared by plotting the sensitivity and (1 minus the specificity) of each test at various reference range cutoff values. The resulting receiver operator characteristic (ROC) curve will often show which test is better; a clearly superior test will have an ROC curve that always lies above and to the left of the inferior test curve. For instance, Figure 1-4 shows the ROC curves for prostate-specific antigen (PSA) and prostatic acid phosphatase

(PAP) in the diagnosis of prostate cancer. PSA is a superior test because it has higher sensitivity and specificity for all cutoff values.

1

#### 4. USE OF TESTS IN DIAGNOSIS AND MANAGEMENT

The value of a test in a particular clinical situation depends not only on the test's sensitivity and specificity but also on the probability that the patient has the disease before the test result is known (**pretest probability**). The results of a valuable test will substantially change the probability that the patient has the disease (**posttest probability**). Figure 1–5 shows how sensitivity, specificity, and posttest probability can be calculated from test results on patients also evaluated using the gold standard by first filling in a 2 × 2 table using the definitions of true-positive, true-negative, false-positive, and false-negative tests.

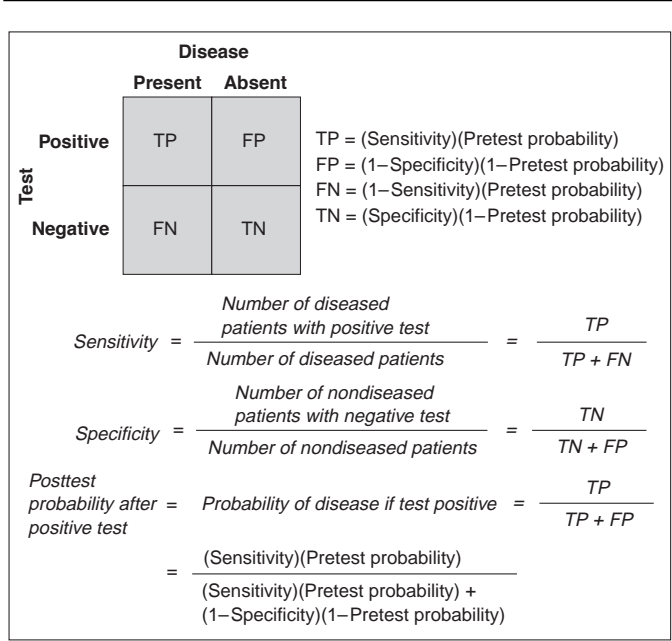


Figure 1–5. Calculation of sensitivity, specificity, and probability of disease after a positive test (posttest probability). (TP, true positive; FP, false positive; FN, false negative; TN, true negative.)

TABLE 1-4. INFLUENCE OF PRETEST PROBABILITY ON THE POSTTEST PROBABILITY OF DISEASE WHEN A TEST WITH 90% SENSITIVITY AND 90% SPECIFICITY IS USED.

Pretest Probability	Posttest Probability
0.01	0.08
0.50	0.90
0.99	0.999

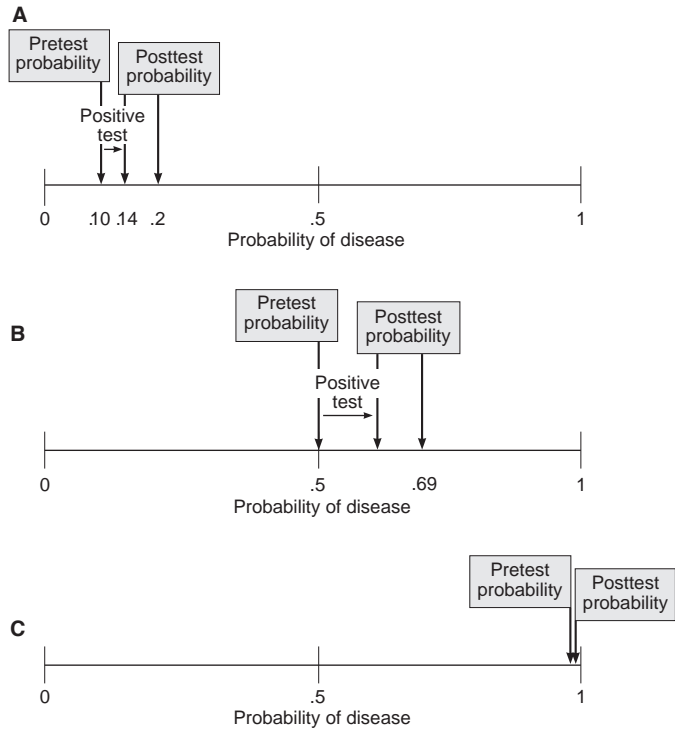
The posttest probability can be calculated directly from the known sensitivity and specificity and the estimated pretest probability (or disease prevalence).

The pretest probability of disease has a profound effect on the posttest probability. As demonstrated in Table 1-4, when a test with 90% sensitivity and specificity is used, the posttest probability can vary from 1% to 99% depending on the pretest probability of disease. Furthermore, as the pretest probability of disease decreases, it becomes less likely that someone with a positive test actually has the disease and more likely that the result represents a false positive.

As an example, suppose the clinician wishes to calculate the posttest probability of prostate cancer using the PSA test with a cutoff of 4 ng/mL. Using the data shown in Figure 1-4, sensitivity is 90% and specificity 60%. The clinician estimates the pretest probability of disease given all evidence and then calculates the posttest probability using the approach shown in Figure 1-5. The pretest probability that an otherwise healthy 50-year-old man has prostate cancer is equal to the prevalence of prostate cancer in that age group (probability = 10%) and the posttest probability is only 20%—ie, even though the test is positive, there is still an 80% chance that the patient does not have prostate cancer (Figure 1-6A). If the clinician finds a prostate nodule on rectal examination, the pretest probability of prostate cancer is 50% and the posttest probability using the same test is 69% (Figure 1-6B). Finally, if the clinician estimates the pretest probability to be 98% based on a prostate nodule, bone pain, and lytic lesions on spine x-rays, the posttest probability using PSA is 99% (Figure 1-6C). This example illustrates that pretest probability has a profound effect on posttest probability and that tests provide more information when the diagnosis is truly uncertain (pretest probability about 50%) than when the diagnosis is either unlikely or nearly certain.

## 5. ODDS-LIKELIHOOD RATIOS

Another way to calculate the posttest probability of disease is to use the odds-likelihood approach. This method may be easier for clinicians to



**Figure 1-6.** Effect of pretest probability and test sensitivity and specificity on the posttest probability of disease. (See text for explanation.)

use. Sensitivity and specificity are combined into an entity called the likelihood ratio (LR).

$$LR = \frac{\text{Probability of result in diseased persons}}{\text{Probability of result in nondiseased persons}}$$

Every test has two likelihood ratios, one corresponding to a positive test (LR<sup>+</sup>) and one corresponding to a negative test (LR<sup>-</sup>):

1

$$\begin{aligned} \text{LR}^+ &= \frac{\text{Probability that test is positive in diseased persons}}{\text{Probability that test is positive in nondiseased persons}} \\ &= \frac{\text{Sensitivity}}{1 - \text{specificity}} \end{aligned}$$

$$\begin{aligned} \text{LR}^- &= \frac{\text{Probability that test is negative in diseased persons}}{\text{Probability that test is negative in nondiseased persons}} \\ &= \frac{1 - \text{sensitivity}}{\text{Specificity}} \end{aligned}$$

Lists of likelihood ratios can be found in some textbooks, journal articles, and computer programs (see Table 1–5 for sample values). Likelihood ratios can be used to make quick estimates of the usefulness of a contemplated diagnostic test in a particular situation. The simplest method for calculating posttest probability from pretest probability and likelihood ratio is to use a nomogram (Figure 1–7). The clinician places a straightedge through the points that represent the pretest probability and the likelihood ratio and then notes where the straightedge crosses the posttest probability line.

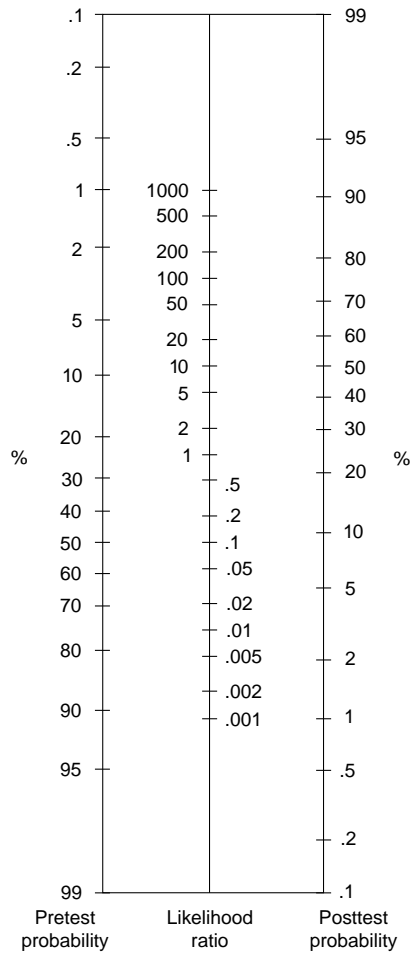
A more formal way of calculating posttest probabilities uses the likelihood ratio as follows:

$$\text{Pretest odds} \times \text{Likelihood ratio} = \text{Posttest odds}$$

To use this formulation, probabilities must be converted to odds, where the odds of having a disease are expressed as the chance of having the disease divided by the chance of not having the disease. For instance, a probability of 0.75 is the same as 3:1 odds (Figure 1–8).

To estimate the potential benefit of a diagnostic test, the clinician first estimates the pretest odds of disease given all available clinical information and then multiplies the pretest odds by the positive and negative likelihood ratios. The results are the **posttest odds**, or the odds that the patient has the disease if the test is positive or negative. To obtain the posttest probability, the odds are converted to a probability (Figure 1–8).

For example, if the clinician believes that the patient has a 60% chance of having a myocardial infarction (pretest odds of 3:2) and the



1

**Figure 1-7.** Nomogram for determining posttest probability from pretest probability and likelihood ratios. To figure the posttest probability, place a straightedge between the pretest probability and the likelihood ratio for the particular test. The posttest probability will be where the straightedge crosses the posttest probability line. (Adapted and reproduced, with permission, from Fagan TJ: Nomogram for Bayes's theorem. *N Engl J Med* 1975;293:257.)

1

TABLE 1-5. LIKELIHOOD RATIOS FOR A SAMPLE OF DIAGNOSTIC TESTS.

Test	Disease	LR <sup>+</sup>	LR <sup>-</sup>
Carcinoembryonic antigen	Dukes A colon cancer	1.6	0.87
Creatine kinase MB	Myocardial infarction	32	0.05
Free thyroxine index	Hyperthyroidism	6.8	0.06
Ferritin	Iron deficiency anemia	85	0.15
Antinuclear antibody	SLE	4.5	0.13

creatine kinase MB test is positive (LR<sup>+</sup> = 32), then the posttest odds of having a myocardial infarction are

$$\frac{3}{2} \times 32 = \frac{96}{2} \quad \text{or} \quad 48 : 1 \text{ odds} \left( \frac{48/1}{48/1 + 1} = \frac{48}{48 + 1} = 98\% \text{ probability} \right)$$

If the CKMB test is negative (LR<sup>-</sup> = 0.05), then the posttest odds of having a myocardial infarction are

$$\frac{3}{2} \times 0.05 = \frac{0.15}{2} \text{ odds} \left( \frac{0.15/2}{0.15/2 + 1} = \frac{0.15}{0.15 + 2} = 7\% \text{ probability} \right)$$

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

Example: If probability = 0.75, then

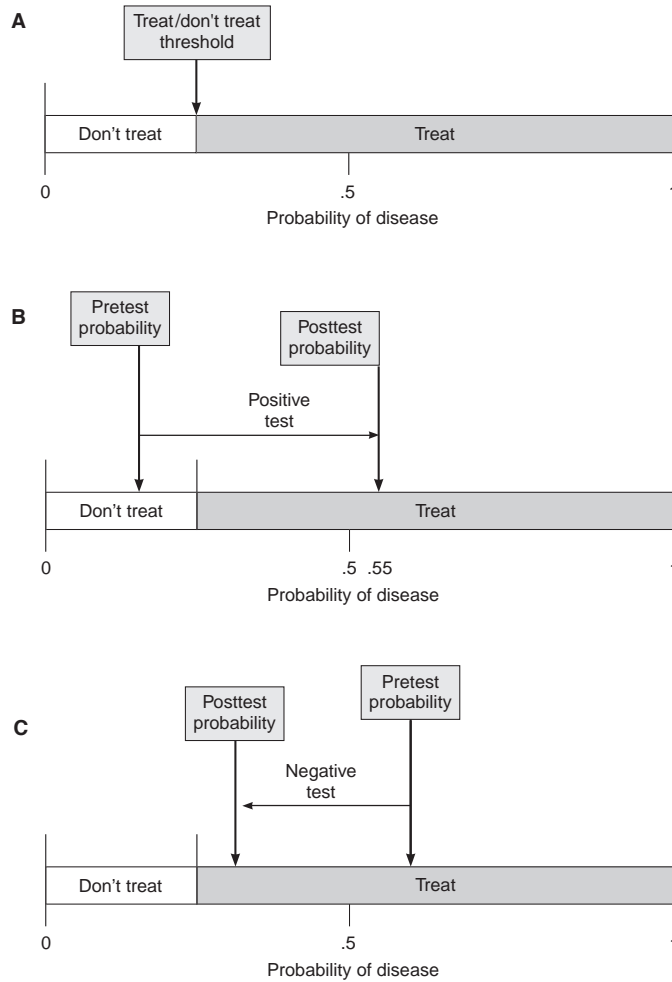
$$\text{Odds} = \frac{0.75}{1 - 0.75} = \frac{0.75}{0.25} = \frac{3}{1} = 3:1$$

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

Example: If odds = 3:1, then

$$\text{Probability} = \frac{3/1}{3/1 + 1} = \frac{3}{3 + 1} = 0.75$$

Figure 1-8. Formulas for converting between probability and odds.



**Figure 1-9.** Threshold approach applied to test ordering. If the contemplated test will not change patient management, the test should not be ordered. (See text for explanation.)

## Sequential Testing

1

To this point, the impact of only one test on the probability of disease has been discussed, whereas during most diagnostic experiences, clinicians obtain clinical information in a sequential fashion. To calculate the posttest odds after three tests, for example, the clinician might estimate the pretest odds and use the appropriate likelihood ratio for each test:

$$\text{Pretest odds} \times LR_1 \times LR_2 \times LR_3 = \text{Posttest odds}$$

When using this approach, however, the clinician should be aware of a major assumption: the chosen tests or findings must be conditionally independent. (For instance, with liver cell damage, the aspartate aminotransferase [AST] and alanine aminotransferase [ALT] enzymes may be released by the same process and are thus not conditionally independent.) If conditionally dependent tests are used in this sequential approach, an overestimation of posttest probability will result.

## Threshold Approach to Decision Making

When contemplating the use of a test in a management decision, the clinician should first decide on the treatment threshold and then assess whether the test result would shift the probability of disease across this treatment threshold. For example, a clinician might decide to treat with antibiotics if the probability of streptococcal pharyngitis in a patient with a sore throat is greater than 25% (Figure 1-9A). If, after reviewing evidence from the history and physical examination, the clinician estimates the pretest probability of strep throat to be 15%, then a diagnostic test such as throat culture ( $LR^+ = 7$ ) would be useful only if a positive test would shift the probability of disease above 25%. Use of the nomogram shown in Figure 1-7 indicates that the posttest probability would be 55% (Figure 1-9B), and ordering the test is thus justified as it affects patient management. On the other hand, if the history and physical examination had suggested that the pretest probability of strep throat were 60%, the throat culture ( $LR^- = 0.33$ ) would be indicated only if a negative test would lower the disease probability below 25%. In this case, using the nomogram shown in Figure 1-7, the posttest probability after a negative test would be 33% (Figure 1-9C). Therefore, ordering the throat culture would not be justified.

In summary, tests should be ordered only if they will affect patient management.