Over the past decade, many clinicians have recognized the need to acquire an expanded clinical vocabulary and conceptual framework, with which to interpret and act upon clinical information from the patient's history, the physical examination, and the diagnostic laboratory. This expanded vocabulary acknowledges the probabilistic nature of clinical events that occur consequent to diagnostic and therapeutic strategies and thereby provides clinicians with the tools for more meaningful communication with colleagues and patients. In essence, this approach addresses the uncertainty inherent in the information used to make diagnostic and therapeutic decisions, and its use permits a more explicit approach to making decisions concerning patient management.

As the medical student progresses through the clinical years and beyond, he or she undoubtedly will hear statements couched in the "old vocabulary." Statements such as "In our hands complications with this test are rare," "splenomegaly is often present in this disorder," and "if the exercise test is positive, then this patient most likely has coronary disease" are vague and leave room for variable interpretation. Terms such as "rare," "often," and "most likely" have different meanings for different listeners and therefore are limiting if one is attempting to communicate precisely. Perhaps students also will hear statements couched in the "new" clinical vocabulary. Statements such as "The risk of death from this procedure has been reported in several series to range from 1 to 5%," "splenomegaly is reportedly present in about 40 to 50% of patients with this disease," and "if the exercise test is positive, then this patient most likely has coronary disease" are used to replace the vague terminology. This chapter is an introduction to this new vocabulary using a clinical vignette about a patient with a common health problem, chest pain. We believe the concepts inherent in this expanded framework and vocabulary are useful tools in the everyday practice of medicine.

Clinical Decision Making

Vignette

Ms. Jones, a single, 42-year-old associate editor for a large publishing house, is treated by Dr. Burns for mild hypertension. On this particular visit, Ms. Jones indicates that she has been experiencing episodes of "burning and pressure type" chest pain 2 or 3 times per week, lasting for about 30 to 40 minutes, which usually are relieved by Maalox. The episodes have occurred at work generally about an hour after lunch. She had ascribed these to tension related to her job but became concerned last week when they occurred during her aerobics class. Each time the pain occurred about 15 minutes into the class and went away with rest about 10 minutes. Ms. Jones admits that although she knows that such symptoms could be a sign of heart trouble, she cannot believe that this could be causing the problem because she is in good shape physically, has never smoked, and does not have a family history of heart disease. In fact, she participated in the class yesterday and did not experience any discomfort.

Upon hearing this, Dr. Burns knows that Ms. Jones is concerned about the possibility of heart disease as a cause for her symptoms and is fearful of discovering that anything might be wrong. After a more complete description of the chest pain (radiation, location, quality, duration, and description of other episodes) and a directed physical examination, Dr. Burns will need to decide on an optimal course of action for Ms. Jones and make a recommendation to her.

Coronary artery disease (CAD) is a major health problem in the United States, and chest pain is a common medical complaint encountered by internists and other physicians who provide care for adults. Accordingly, Dr. Burns has devoted special emphasis to this subject in the time he has available for reading and in his other continuing education activities. He knows that a number of workers have collected data from patients with chest pain that might be angina pectoris (classically, chest pain due to myocardial ischemia occurring on a consistent basis with physical activity and relieved by rest) and have categorized the pattern of pain into several subsets that correlate with the probability of finding serious coronary artery disease by coronary arteriography.

Dr. Burns considers the following subsets: typical or classic angina (pain that occurs consistently with exercise and is relieved by rest within 20 minutes); atypical chest pain (pain that has some features of angina but does not have a consistent relationship with activity and relief by rest); and noncardiac chest pain (i.e., pain that has no features resembling angina). Such data have been derived from clinical databases where historical information has been collected from a large number of patients, typically from a few hundred to several thousand. To arrive at a plausible estimate of Ms. Jones's likelihood of coronary artery disease on the basis of her history and the results of his physical examination, Dr. Burns refers to an article by Drs. Diamond and Forrester (1979) entitled "Analysis of Probability as an Aid in the Clinical Diagnosis of Coronary Artery Disease." Diamond and Forrester related the probability of a person with a complaint of chest pain having coronary artery disease to age, sex, and pain pattern. A modification of their tabular data as it appeared in that article is reproduced in Table 6.1.

Probability

Dr. Burns interprets Ms. Jones's pain pattern as atypical angina, mainly because of the inconsistent relationship with physical activity. Extrapolating from Table 6.1, he notes that a reasonable estimate for the likelihood that a woman of
Ms. Jones's age with atypical angina has significant coronary artery disease is about 13%. To arrive at a specific estimate for Ms. Jones, he decides to revise this figure downward to 10% because she is at the younger end of the age interval 40 to 49 and lacks other risk factors besides mild hypertension. Dr. Burns recognizes his uncertainty about this estimate and considers this likelihood to lie within a range of 5 to 20%, although his best guess about Ms. Jones's likelihood of CAD is 10%.

Dr. Burns now must decide whether he should recommend a noninvasive test such as the ECG treadmill stress test or an invasive procedure such as coronary arteriography or simply do nothing. Although other exercise tests are available, such as the stress thallium test, Dr. Burns considers the treadmill test the simplest to perform and the preferred noninvasive test for Ms. Jones.

The coronary arteriogram has been considered the definitive test for determining the presence and severity of CAD and would provide the best information about the anatomic state of Ms. Jones's coronary arteries. It has served as the "gold standard" against which the accuracy of other diagnostic tests for CAD are measured. An arteriogram would require a one-day hospitalization, is invasive and is associated with a mortality risk of about 1/1000, or with morbidity such as myocardial infarction (2/1000), arterial complications (5–8/1000) and ventricular fibrillation (5/1000) (Franch et al., 1982). On the other hand, the stress test is safer (risk of mortality on the order of 1/10,000), can be done in Dr. Burns's office, and would allow a revised determination of Ms. Jones's likelihood of having serious CAD (Clarke and Bruce, 1979).

A patient undergoing an ECG stress test walks on a moving treadmill that is increased in speed at 3-minute intervals. The test is continued until the patient either experiences signs or symptoms of cardiac dysfunction or achieves 85% of his or her predicted maximal heart rate for a specified period of time. In addition to the above symptoms, a specific electrocardiographic feature is looked for: depression of the ST segment. Figure 6.1 shows a picture of the normal ECG, and a change in the ST segment that would be associated with myocardial ischemia. This segment may become depressed to a variable degree; 0.5, 1.0, 1.5, or 2 mm or more. As the criterion (i.e., the "cutoff point") for an abnormal ST segment depression changes, the likelihood of making errors in interpretation also changes.

**Predictive Value, Sensitivity, and Specificity**

Dr. Burns now quickly goes through a series of calculations to determine what Ms. Jones's likelihood of CAD would be depending on the result of the exercise test. This is called the "posttest" likelihood, or alternatively, the "predictive value" of the test result. Dr. Burns uses a hand-held programmable calculator to perform these calculations quickly and automatically, but for purposes of this discussion, we will describe the process of determining the posttest likelihood as Dr. Burns would if he were doing this by hand. Some important terms and the notation used to represent them are explained in Table 6.2.

Three parameters must be known or estimated in order to derive Ms. Jones's likelihood of CAD after being tested on the treadmill. These are the sensitivity and specificity of the treadmill test and the prior probability of CAD in Ms. Jones. Recall that Dr. Burns estimated the latter as between 5 and 20% because of Ms. Jones's age and pain pattern. The sensitivity and specificity of a diagnostic test are determined by comparing the test results in an appropriate sample of individuals against a gold standard (more will be

![Figure 6.1](image-url)

**Figure 6.1**
Electrocardiographic patterns. The top and middle tracings show examples of ST segment depression. A normal tracing of the ST segment is shown below for comparison.

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Nonanginal chest pain</th>
<th>Atypical angina</th>
<th>Typical angina</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>30–39</td>
<td>5.2</td>
<td>0.8</td>
<td>21.8</td>
</tr>
<tr>
<td>40–49</td>
<td>14.1</td>
<td>2.8</td>
<td>46.1</td>
</tr>
<tr>
<td>50–59</td>
<td>21.5</td>
<td>8.4</td>
<td>58.9</td>
</tr>
<tr>
<td>60–69</td>
<td>28.1</td>
<td>18.6</td>
<td>67.1</td>
</tr>
</tbody>
</table>

Table 6.2
Terms and Notation Used to Calculate Predictive Value

<table>
<thead>
<tr>
<th>Term</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of disease: The physician's estimate of the patient's likelihood of having the disease before a diagnostic test is performed. Also termed the prevalence of disease, or prior probability of disease.</td>
<td>( p(D^+) )</td>
</tr>
<tr>
<td>Probability of not having disease: 1 - ( p(D^+) ).</td>
<td>( p(D^-) )</td>
</tr>
<tr>
<td>Test sensitivity: The probability that a test result will be positive among patients with the disease.</td>
<td>( p(T^+</td>
</tr>
<tr>
<td>Test specificity: The probability that a test result will be negative among patients without the disease.</td>
<td>( p(T^-</td>
</tr>
<tr>
<td>Predictive value positive: The probability that a patient has the disease given a positive test result.</td>
<td>( p(D^+</td>
</tr>
<tr>
<td>Predictive value negative: The probability that a patient does not have the disease given a negative test result.</td>
<td>( p(D^-</td>
</tr>
</tbody>
</table>

\*The "|" bar indicates conditional probability; the parameter on the left of the bar being conditional on the right parameter. In this example it is the probability of being test positive in the group that has disease.

Table 6.3
Methods for Calculating Predictive Values

A. General Form

<table>
<thead>
<tr>
<th>Stress test result</th>
<th>Coronary artery disease*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Present</td>
</tr>
<tr>
<td>TP = a</td>
<td>FP = b</td>
</tr>
<tr>
<td>Negative</td>
<td>FN = c</td>
</tr>
</tbody>
</table>

\*determined by arteriography.

PVP = \( \frac{a}{a+b} \).

PVN = \( \frac{c}{c+d} \).

B. Specific Example

CA Disease

<table>
<thead>
<tr>
<th>Stress test result</th>
<th>Coronary artery disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Positive</td>
<td>4</td>
</tr>
<tr>
<td>Negative</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Given sensitivity = .75.

Given specificity = .80.

PVP = \( \frac{(0.75)(0.05)}{(0.75)(0.05) + (1 - 0.80)(0.95)} = 17\% \)

PVN = \( \frac{(0.80)(0.05) + (0.80)(0.95) + (1 - 0.75)(0.05)}{(0.80)(0.05) + (0.80)(0.95) + (1 - 0.75)(0.05)} = 98\% \)

If this probability were .20, then:

PVP = \( \frac{(0.75)(0.20) + (0.75)(0.20) + (1 - 0.80)(0.80)}{(0.75)(0.20) + (0.75)(0.20) + (1 - 0.80)(0.80)} = 48\% \)

PVN = \( \frac{(0.80)(0.20) + (0.80)(0.20) + (1 - 0.75)(0.80)}{(0.80)(0.20) + (0.80)(0.20) + (1 - 0.75)(0.80)} = 93\% \)

An alternative method for calculating predictive values is shown in Table 6.3. Table 6.3A indicates that PVP is derived by the following: true positive rate/(true positive rate + false positive rate). Similarly, the PVN is obtained by: true negative rate/(true negative rate + false negative rate). As indicated in Table 6.3B, out of every 100 patients, 5 will have disease (since Dr. Burns has estimated Ms. Jones's likelihood at 5%). Now the sensitivity and specificity can be used to derive the numbers of patients falling into individual cells. Of these 5 with CAD, 4 will be test positive (sensitivity = .75). Of the 95 people without CAD, 76 will be test negative (specificity = .80). The predictive values will be:

PVP = \( \frac{4}{23} = 0.17 = 17\% \)

PVN = \( \frac{76}{77} = 0.98 = 98\% \)

An essential point to note is that the predictive values change as the prior estimate of Ms. Jones's probability of having CAD changes. Thus, for a physician to interpret a result of a "less than perfect" test meaningfully requires estimation of the patient's likelihood of disease before the test is performed. Figure 6.2 illustrates this principle using the ECG stress test by relating the predictive value positive and negative of an exercise test to a changing prior probability of CAD at a constant test sensitivity of .75 and a specificity of .80. (These and the other figures reproduced here were quickly graphed by Dr. Burns using his desktop personal computer and the appropriate software program.) Table 6.4 gives the PVP and PVN in tabular form as the prevalence of CAD increases.
Making the Decision

By using the exercise test result to revise Ms. Jones's probability of CAD, Dr. Burns can now decide if the PVN would be such that he and Ms. Jones would feel reassured enough not to pursue a coronary arteriogram and conversely, if the PVP would be high enough to recommend pursuing a coronary arteriogram. In arriving at these "threshold probabilities" at which the choice of treatment options would be affected, Dr. Burns must analyze his beliefs about the natural history of significant CAD, the efficacy and disadvantages of the alternative treatments for this problem, and Ms. Jones's attitudes about various health outcomes. This is no small task and requires the ability to evaluate results of clinical research, strong communication skills, and experience.

Using diagnostic information to revise clinical probabilities is only a first step in the process of clinical decision making. Completing this process requires integrating this revised probability estimate with probabilistic information regarding the different treatment options and the patient's attitudes toward the treatments. Acquiring skills in formal decision analysis will permit the clinician to follow through with the probabilistic approach to diagnosis described here.

Characteristics of Diagnostic Tests

In calculating Ms. Jones's likelihood of CAD based on stress test results, Dr. Burns needed to use the sensitivity and specificity of this test. He obtained these values from published reports of stress test results in persons known to have or to be free of CAD according to cardiac catheterization.

He understood that these test characteristics express how good or how poor a test is in predicting the presence of disease. His understanding of how these characteristics are derived helped him decide whether or not to use the stress test.

Most diagnostic tests are indirect measures of disease and thus must be standardized against some "best estimate" of disease, an estimate often referred to as the "gold standard."

For the stress test, the gold standard has been coronary arteriography. The validity of a diagnostic test is determined by comparing test results with the gold standard results, both performed in an appropriate number of selected patients. When agreement does not occur, the diagnostic test is said to be falsely positive or falsely negative. Thus the sensitivity and specificity of a diagnostic test describe the ability of the test to correctly classify a patient when compared with a gold standard.

One additional feature that Dr. Burns notes is the test "cutoff" point. Clinicians generally desire that diagnostic test results be reported as either abnormal or normal (i.e., categorical). This desire for categorizing results presents a problem, since measurements of biologic variables usually form a continuum of values. Continuous values can be made categorical by selecting a point (cutoff point) along the continuum and assigning all values on one side of this point to the abnormal category and those on the other side of the point to the normal category. The sensitivity and specificity of a test are determined by where the cutoff point is selected. This is true because test values between diseased and non-diseased populations usually overlap. If the cutoff point is chosen such that the test has high sensitivity (high true positive rate), then the specificity (true negative rate) usually is lowered.

This unavoidable tradeoff is important to the clinician, since different cutoff points will produce different test characteristics. Thus the clinician will obtain different predictive values based on the particular cutoff point used. In our example with the exercise stress test, the cutoff point can be set at 0.5, 1.0, 1.5, or 2 mm of ST depression. The effect of using different cutoff points is seen in Figure 6.3. If the cutoff point is chosen as an ST segment depression of >2.5 mm, the test has a higher PVP at any given pretest probability than if less stringent cutoff points were chosen. This occurs because there will be fewer false positive tests. However, it should be remembered that there also will be a greater number of false negative test results. This is the price paid for choosing a more stringent criterion for an abnormal result.
of disease are calculated by dividing the prior prevalence by \((1 - 0.8)\). As indicated in this table, the prior odds that the likelihood ratio for the stress test is 3.75 (0.75 divided by 1-specificity). This is the same as dividing the sensitivity of a test by one minus the prior prevalence. The posttest odds are easily converted to posttest probability by using the odds in the numerator divided by 1 plus the odds (i.e., \(0.94/1.94 = 0.49\), or 49%).

Because the likelihood ratio incorporates both the sensitivity and specificity of a test, its value will vary depending on the cutoff point chosen. The different likelihood ratios for a particular test as the cutoff point changes may be derived from the receiver-operating-characteristic (ROC) curve for the particular test under consideration. This curve is a plot of the sensitivity versus 1 minus the specificity for a test as the cutoff point is changed. Using his personal computer, Dr. Burns can easily plot this curve with data obtained from the laboratory. In minutes he can produce a plot such as that seen in Figure 6.4, a hypothetical ROC curve for the stress test. Any pair of coordinates taken from this curve may be used to calculate the likelihood ratio and thus, along with the prior odds of disease, may be used to calculate the posttest odds of disease. It is clear from this curve that both the false positive and false negative rates are low if a cutoff point of 1.6 or 1.8 is chosen. ROC curves may also be used to compare the efficiency of different diagnostic tests used for a particular disease. Hanley and McNeil (1982) give an excellent discussion of this use of ROC curves.

A test of very high specificity, for example, would give a relatively large likelihood ratio even if the sensitivity of the test were low. This occurs because 1-specificity, if a small value, will give a relatively large likelihood ratio. This may be seen in Table 6.6 where the cutoff point has been set to give a specificity of 95% and a sensitivity of 60%. In this case, the likelihood ratio is 8.66 and, as can be seen by multiplying this value times the pretest odds, will give a much higher posttest odds than in the data in Table 6.4. Unfortunately, something must be lost for obtaining a higher likelihood ratio, and in this case the false negative rate is 40% (1 minus 0.60).

Caveats

When the student actually goes to the literature to learn the performance characteristics of signs and symptoms and other
Figure 6.4
Effect of choosing different cutoff points for normal versus ab-
normal results on the likelihood that an individual patient will have
a particular disease. The likelihood ratios for a particular test with
different cutoff points are derived from the receiver-operating-
characteristic (ROC) curve for the test, which is a plot of the sen-
sitivity versus 1 minus the specificity at the various cutoff points.

Second, the clinician still needs to incorporate the prob-
abilistic information gained from the history, physical ex-
amination, and laboratory tests into patient management
decisions. In many situations these decisions are complex,
and for selected clinical situations we advocate using deci-
sion trees as a tool to enable one to maintain an explicit
approach to patient management. It makes little sense to
use an explicit approach to diagnosis and then suddenly
hide the assumptions and steps in reasoning that went into
making a therapeutic management decision.

### Table 6.6
Likelihood Ratio Calculation of Posttest Odds

<table>
<thead>
<tr>
<th>Pretest prevalence of disease</th>
<th>Pretest odds of disease</th>
<th>Posttest odds of disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.01:1</td>
<td>.06:1</td>
</tr>
<tr>
<td>.05</td>
<td>.05:1</td>
<td>.32:1</td>
</tr>
<tr>
<td>.10</td>
<td>.11:1</td>
<td>.67:1</td>
</tr>
<tr>
<td>.15</td>
<td>.18:1</td>
<td>1.06:1</td>
</tr>
<tr>
<td>.20</td>
<td>.25:1</td>
<td>1.50:1</td>
</tr>
<tr>
<td>.25</td>
<td>.33:1</td>
<td>2.00:1</td>
</tr>
<tr>
<td>.30</td>
<td>.43:1</td>
<td>2.57:1</td>
</tr>
<tr>
<td>.35</td>
<td>.54:1</td>
<td>3.25:1</td>
</tr>
<tr>
<td>.40</td>
<td>.67:1</td>
<td>4.00:1</td>
</tr>
<tr>
<td>.45</td>
<td>.82:1</td>
<td>4.91:1</td>
</tr>
<tr>
<td>.50</td>
<td>1.00:1</td>
<td>6.00:1</td>
</tr>
<tr>
<td>.55</td>
<td>1.22:1</td>
<td>7.33:1</td>
</tr>
</tbody>
</table>

Test sensitivity = .60.
Test specificity = .90.
Likelihood ratio = TP/FP = .60/1 - .90 = 6.00.

### References

Clarke JL, Bruce RA. Exercise testing. In: Cohn PF, ed. Diagnosis
and therapy of coronary artery disease. Boston: Little, Brown,
1979.

Department of Clinical Epidemiology and Biostatistics, McMaster
University Health Sciences Center. How to read clinical jour-
nals: II. To learn about a diagnostic test. Can Med Assoc J

Diamond GA, Forrester JS. Analysis of probability as an aid in the
1979;300:1350–58.

Franch RH, King SB, Douglass JS, Jr. Technique of cardiac cath-
ereterization. In: Hurst W, ed. The heart. 5th ed. New York:

Goldschlager N. Use of the treadmill test in the diagnosis of cor-
onary artery disease in patients with chest pain. Ann Intern

Griner PF, Mayeski RJ, Mushlin AI, Greenland P. Selection and
interpretation of diagnostic tests and procedures: principles and

Hanley JA, McNeil BJ. The meaning and use of the area under a
receiver operating characteristic curve (ROC). Radiology

Ransohoff DF, Feinstein AR. Problems of spectrum and bias in
evaluating the efficacy of diagnostic tests. N Engl J Med

Rifkin RD, Hood WB, Jr. Bayesian analysis of electrocardiographic

Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia:
W.B. Saunders, 1980.