CME CREDIT

Accuracy and predictive values in clinical decision-making

MARK J. EISENBERG, MD, MPH

SUMMARY In clinical practice, the accuracy and predictive values of a diagnostic test may differ substantially from values cited in published reports, owing to a lower prevalence of most diseases in clinical populations than in study populations. To correct this problem, published assessments of diagnostic tests should standardize accuracy and predictive values to account for disease prevalence.

KEY POINTS The accuracy of a test varies directly with the prevalence of the disease in question, and the upper and lower bounds of accuracy are determined by the test's sensitivity and specificity. When disease prevalence equals 50%, a test's accuracy is exactly midway between its sensitivity and specificity. If a test's sensitivity and specificity have the same value, its accuracy will also equal this value, regardless of disease prevalence. A test's positive and negative predictive values are also strongly affected by disease prevalence. Positive predictive values are high when disease prevalence is high, and they are low when disease prevalence is low. Negative predictive values have an inverse relationship.

INDEX TERMS: PREDICTIVE VALUE OF TESTS; PREVALENCE; DECISION MAKING
CLEVE CLIN J MED 1995; 62:311–316

From the Department of Cardiology, The Cleveland Clinic Foundation.
Address reprint requests to M.J.E., Division of Clinical Epidemiology, Jewish General Hospital, 3755 Côte Ste-Catherine Rd., Montreal, Quebec, H3T 1E2, Canada.
Supported by an Institutional National Research Service Award, HL 07192, Training Program in Heart and Vascular Diseases, NHLBI, Bethesda, Maryland.

THE ACCURACY and positive and negative predictive values of diagnostic tests in clinical practice may differ from those cited in published reports, because the prevalence of most diseases is lower in clinical populations than in study populations.[1–4] How then are clinicians to assess the performance of diagnostic tests? This article suggests ways researchers could end the confusion.

DEFINING THE TERMS

Accuracy, sometimes known as *diagnostic accuracy*, is a global measure of the value of a diagnostic test. It is the proportion of patients correctly identified as either having or not having the disease in question (*Table 1*). In contrast, predictive values provide more specific information. The *positive predictive value* is the proportion of patients with a positive test result who have the disease; the *negative predictive value* is the proportion of patients with a negative test result who do not have the disease.

Although the accuracy and predictive values of a test vary with the prevalence of disease in the

TABLE 1
ASSESSING THE PERFORMANCE OF A DIAGNOSTIC TEST

Compiling the results

	Disease present	**Disease absent**	
Positive test result	True positive (TP)	False positive (FP)	TP+FP
Negative test result	False negative (FN)	True negative (TN)	FN+TN
	TP+FN	FP+TN	Total

Calculating the performance

Sensitivity = TP/(TP + FN)
Specificity = TN/(FP + TN)
Accuracy = (TP + TN)/Total
Disease prevalence = (TP + FN)/Total
Positive predictive value = TP/(TP + FP)
Negative predictive value = TN/(FN + TN)

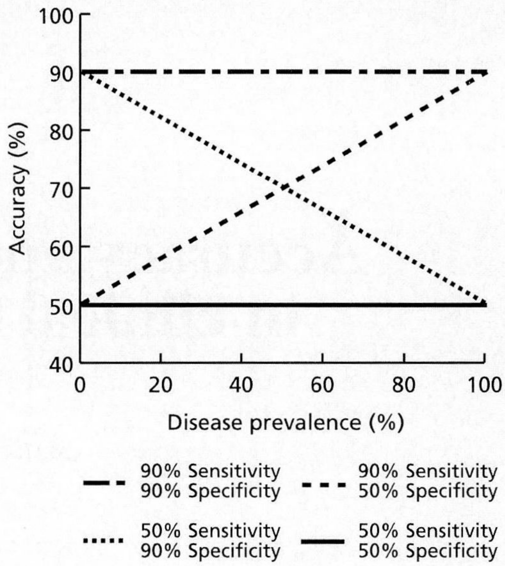

FIGURE 1. Relationships among accuracy and sensitivity, specificity, and disease prevalence. At a disease prevalence of 100%, accuracy is equal to test sensitivity. At a disease prevalence of 0%, accuracy is equal to test specificity. Between these two extremes, accuracy varies directly (linearly) with disease prevalence. When a test's sensitivity and specificity have the same value, accuracy will also equal this value and it will not vary with changes in disease prevalence.

population in which the test is used, sensitivity and specificity are relatively stable.[1–4] *Sensitivity* is the proportion of patients with a disease who have a positive test result, and *specificity* is the proportion of patients without a disease who have a negative test result (*Table 1*). In most situations, the sensitivity and specificity of a diagnostic test vary very little with differences in disease prevalence.[1–4] However, a test may perform differently in populations with different distributions of disease severity.[5] For example, treadmill stress testing will have a different sensitivity and specificity in a group of patients with severe triple-vessel coronary artery disease than in a group of patients with mild single-vessel disease.[6–8]

Because sensitivity and specificity are relatively stable test characteristics, many clinicians consider them the best measures of the value of a diagnostic test. However, when employing a diagnostic test, clinicians are faced with positive or negative test results. Since predictive values directly measure the reliability of positive and negative test results, they are often of greater clinical value than sensitivity and specificity.

Because accuracy and predictive values may vary considerably with disease prevalence, reports that do not state the disease prevalence in the population in which these test characteristics were measured may be misleading. For example, a test may be reported as having a high sensitivity, specificity, accuracy, and positive predictive value, but a marginal negative predictive value. Close examination of the study population, however, often discloses that the proportion of patients with the disease is much higher than in most clinical settings. A more realistic prevalence of disease may yield a significantly different accuracy, a dramatically lower positive predictive value, and a higher negative predictive value.[9–11] Consequently, to properly assess the clinical value of a diagnostic test, it is important to know its accuracy and positive and negative predictive values in different patient populations.

ACCURACY

The accuracy of a test varies directly with disease prevalence, and the upper and lower bounds of accuracy are determined by the test's sensitivity and specificity (*Figure 1*).[10] In a population with a disease prevalence of 100%, the accuracy of a test equals its sensitivity; in a population with a disease prevalence of 0%, accuracy equals specificity. Between the bounds determined by sensitivity and specificity, accuracy varies directly (linearly) with disease prevalence. Thus, disease prevalence determines the trade-off in importance of sensitivity and

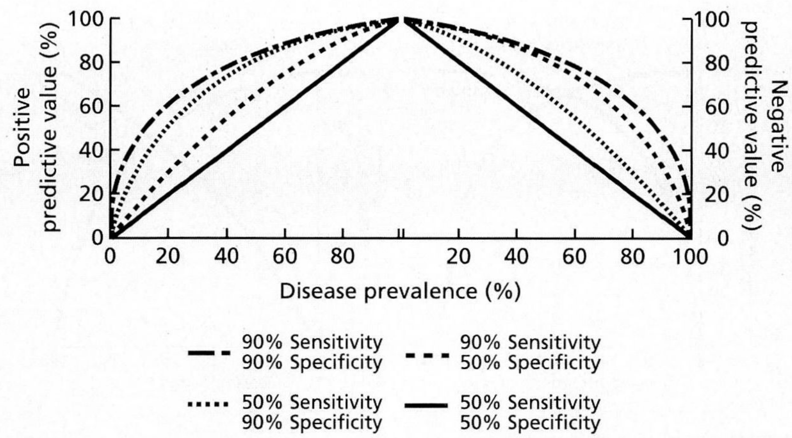

FIGURE 2. Relationships among predictive values and sensitivity, specificity, and disease prevalence. Disease prevalence is an important determinant of positive and negative predictive values. Positive predictive values are high when disease prevalence is high, and they are low when disease prevalence is low. Negative predictive values have an inverse relationship. In contrast, positive and negative predictive values both increase and decrease in tandem with increases and decreases in sensitivity and specificity. However, sensitivity has a greater influence on negative predictive values, and specificity has a greater influence on positive predictive values. This relationship is demonstrated by the areas under two curves: test A, with a sensitivity of 90% and a specificity of 50%; and test B, with a sensitivity of 50% and a specificity of 90%. Test A (high sensitivity) has higher negative predictive values than test B, and test B (high specificity) has higher positive predictive values than test A.

specificity. If disease prevalence equals 50%, a test's accuracy is exactly midway between its sensitivity and specificity. If a test's sensitivity is the same as its specificity, (eg, sensitivity = 90%, specificity = 90%), its accuracy also has this value, regardless of disease prevalence.

For example, the accuracy of a test with a sensitivity of 90% and a specificity of 50% can range between 90% and 50% (*Figure 1*). If the prevalence of disease is close to 0%, the test's accuracy is close to 50%, because the few patients who have the disease (almost all of whom are correctly identified owing to the high sensitivity of the test) are far outnumbered by the patients who do not have the disease (many of whom are incorrectly identified as having the disease owing to the low specificity of the test). However, in a population with a disease prevalence close to 100%, the test's accuracy is close to 90%, because the few patients who do not have the disease (many of whom are incorrectly identified owing to the low specificity of the test) are far outnumbered by the patients who have the disease (most of whom are correctly identified owing to the high sensitivity of the test). At a disease prevalence of 50%, the test's accuracy is 70%, exactly midway between the values for sensitivity and specificity.

Although frequently reported, accuracy is often a poor measure of a diagnostic test.[12] Suppose a test has a sensitivity of 0% and a specificity of 90%. In a population with a high prevalence of disease, the test's accuracy will be close to 0%, but in a population with a low disease prevalence (common in many clinical settings), its accuracy will be close to 90%. Although this hypothetical test cannot identify any patients with the disease in question (sensitivity = 0%), it may still be reported as having an accuracy of 90%. Thus, the accuracy of a given test may vary widely in different populations. Consequently, this test characteristic may be a very misleading measure of the value of a diagnostic test.

PREDICTIVE VALUES

Like accuracy, predictive values may vary substantially with disease prevalence.[1–4,12] The positive predictive value of a test is high if disease prevalence is high, and low if disease prevalence is low; negative predictive values have an inverse relationship (*Figure 2*). Diagrams that simultaneously demonstrate positive and negative predictive values at different disease prevalences can provide clinicians with a good idea of the value of a positive or a negative test result.

Consider a test having a sensitivity of 90% and a specificity of 90% (*Figure 3*). If a patient is at high risk for a disease (eg, if the disease prevalence is greater than 50% in patients with similar clinical characteristics), the clinician can be confident that a positive test result indicates the presence of disease, as the positive predictive values range from 90% to 100%. A negative test result for this same patient, however, does not guarantee that the disease is not present, as the negative predictive value ranges from

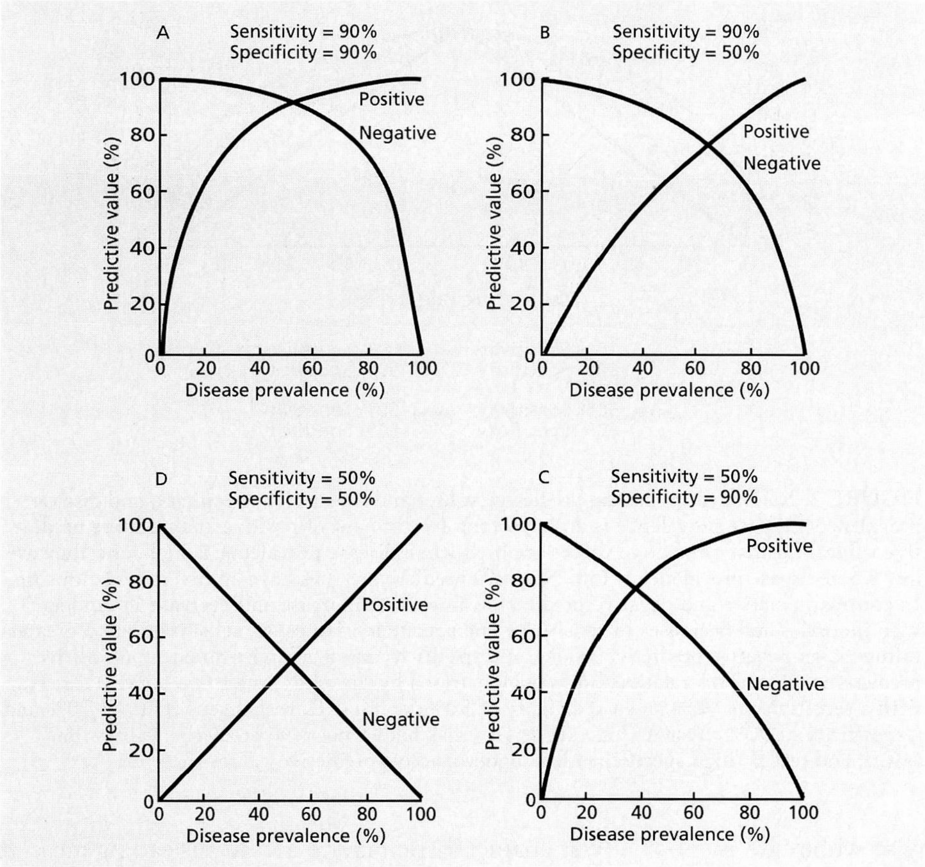

FIGURE 3. Trade-off in positive and negative predictive values. Positive and negative predictive values vary inversely. When test sensitivity equals test specificity (A and D), positive and negative predictive values are equal at a disease prevalence of 50%. At disease prevalences < 50%, negative predictive values are always higher than positive predictive values, and at disease prevalences > 50%, positive predictive values are always higher than negative predictive values. When test sensitivity is greater than test specificity (B), positive and negative predictive values are equal at a disease prevalence > 50%. If sensitivity is less than specificity (C), positive and negative predictive values are equal at a disease prevalence < 50%.

90% to 0%. In contrast, if a patient is at low risk for the disease (eg, disease prevalence < 20%), the clinician can be confident that a negative test result indicates the patient does not have the disease (the negative predictive values range from 97% to 100%). However, a positive test result for this patient is not very reassuring, because the positive predictive value may range from 69% to 0%.

Although predictive values are strongly influenced by disease prevalence, they are also affected by sensitivity and specificity. Both positive and negative predictive values vary in tandem with sensitivity and specificity (*Figure 2*). However, sensitivity has greater influence on negative predictive values, and specificity has greater influence on positive predictive values (*Figure 3*).[2]

In most clinical situations, the disease prevalence is less than 50%. Therefore, in clinical populations, negative predictive values are almost always in the acceptable range, but positive predictive values are often low (*Figures 2* and *3*). In a patient population with a disease prevalence of 10%, for example, a test with a sensitivity and specificity of 50% has a negative predictive value of 90%. Surprisingly, a test with a sensitivity and specificity of 90% would have a similar negative predictive value (99%) in the same patient population. However, neither test will have a positive predictive value greater than 50%. In order to optimize positive predictive values in low-risk patient populations (ie, with a disease prevalence < 50%), clinicians should be particularly concerned about test specificity, as the positive predictive value will always be better with a test that has a high specificity (*Figures 2* and *3*).

A test with a sensitivity of 90% and a specificity of 50% has the same overall diagnostic ability as a test with a sensitivity of 50% and a specificity of 90%, but their diagnostic abilities at particular disease prevalences may differ markedly (*Figure 4*). In a population with a disease prevalence of 20%, a test with a sensitivity of 90% and a specificity of 50% has a positive predictive value of 31%, a negative predictive value of 95%, and an accuracy of 58%. In the same population, a test with a sensitivity of 50% and a specificity of 90% has a positive predictive value of 56%, a negative predictive value of 89%, and an accuracy of 82%. Thus, at a disease prevalence of 20%, the two tests have similar nega-

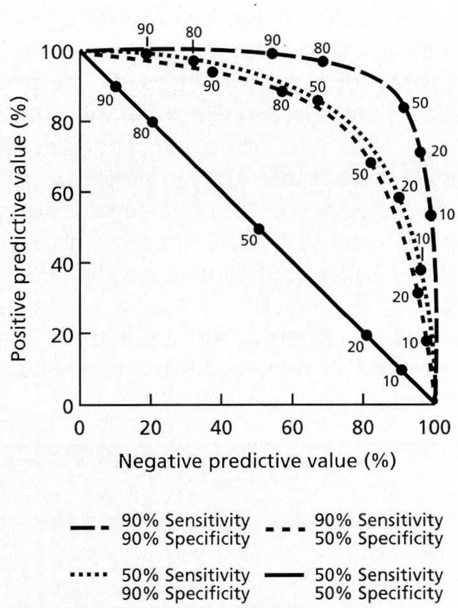

FIGURE 4. Comparison of the diagnostic abilities of different tests by graphing positive vs negative predictive values. Numbers on curves refer to disease prevalences. Tests with high sensitivities and specificities have greater discriminating abilities than tests with low sensitivities and specificities. If two tests have different sensitivities and specificities but their sums are equal (eg, sensitivity + specificity = 140 = 90 + 50 or 50 + 90), the overall discriminating abilities of the two tests will be equal. Despite identical overall discriminating abilities, at particular disease prevalences, the two tests may have markedly different positive and negative predictive values. The test with the higher specificity will perform better at low disease prevalences, and the test with the higher sensitivity will perform better at high disease prevalences.

tive predictive values, but the test with the higher specificity has a higher positive predictive value and accuracy. Since disease prevalence is much less than 50% in most clinical settings, test specificity may be crucial, while test sensitivity may be of lesser importance.

STANDARDIZING ACCURACY AND PREDICTIVE VALUES

Because accuracy and predictive values may vary substantially with disease prevalence and because diagnostic tests are often used in clinical settings in which disease prevalence is much less than in reported study populations, the current practice of reporting accuracy and predictive values without reference to disease prevalence is often confusing and, in many cases, misleading.

TABLE 2
ACCURACY AND PREDICTIVE VALUES OF A DIAGNOSTIC TEST WITH A SENSITIVITY OF 50% AND A SPECIFICITY OF 90%

Disease prevalence (%)	Accuracy (%)	Positive predictive value (%)	Negative predictive value (%)
0	90	0	100
10	86	36	94
20	82	56	88
30	78	68	81
40	74	77	72
42*	73	78	71
50	70	83	64
60	66	88	55
70	62	92	44
80	58	95	31
90	54	98	17
100	50	100	0

*The sensitivity and specificity of the diagnostic test were determined in a study population with a disease prevalence of 42%.

The clinical value of diagnostic tests would be clarified if the reporting of accuracy and predictive values were standardized. Any of three methods could be employed. The simplest method is to state the disease prevalence along with the accuracy and predictive values. For example, a test may be found to have a sensitivity of 50% and a specificity of 90% in a study population with a disease prevalence of 42%. The accuracy and the predictive values for this test could be reported in the following format: ACC_{42} = 73%, PPV_{42} = 78%, and NPV_{42} = 71%; in which ACC is the accuracy, PPV is the positive predictive value, and NPV is the negative predictive value.[1] This format emphasizes that the reported accuracy and predictive values apply to the diagnostic test only when it is used in a patient population with a disease prevalence of 42%. A clinician whose patients have a lower disease prevalence will know that this same diagnostic test will have an accuracy between 73% and 90% (upper bound determined by test specificity), a positive predictive value lower than 78% (perhaps substantially lower), and a negative predictive value higher than 71%.

A second method is to calculate the test's accuracy and predictive values in a standard population with a disease prevalence of 50%, regardless of the actual prevalence of disease in the study population. In the example cited above, ACC_{50} = 70%, PPV_{50} = 83%, and NPV_{50} = 64%. Like the first method, this format highlights disease prevalence, and it lets clinicians estimate accuracies and predictive values applicable to their individual clinical settings. This method also ensures comparability of accuracy and predictive values among published reports, and it provides a reasonable trade-off between high and low values for accuracy and predictive values. The disadvantage of this method is that most patient populations have disease prevalences much less than 50%, so these standardized values are not very realistic for most clinicians.

A third possibility is to report a range of accuracies and predictive values corresponding to the test's performance in patient populations with different disease prevalences. Values could be presented in either tabular (*Table 2*) or graphic form (*Figures 1* and 3). Although the most cumbersome, this method provides the most information. In addition to providing the same information as the other two methods, it presents accuracies and predictive values that apply to a variety of clinical settings. This method allows physicians to estimate more closely accuracies and predictive values for their patient populations, and it also illustrates important trends in these values with changes in disease prevalence.

CONCLUSION

Many studies of diagnostic tests are performed in populations with high disease prevalences. In these situations, the test's accuracy is unusually dependent on the sensitivity, positive predictive values are high, and negative predictive values are low. In most clinical situations, however, disease prevalence is low, and these same diagnostic tests may have different accuracies (because they become less dependent on sensitivity and more dependent on specificity), lower positive predictive values, and higher negative predictive values.

Since accuracy and predictive values may vary substantially in different populations, the current method of reporting these test characteristics without reference to disease prevalence may be misleading. Consequently, a standard format is needed for the presentation of accuracy and predictive values. Any of three formats could be used. The first states the disease prevalence in the population being studied, the second refers to a hypothetical, standard population with a disease prevalence of 50%, and the third provides a range of accuracies and predictive values calculated for different disease prevalences. A standardized format for the reporting of accuracies and predictive values would reduce the confusion that currently surrounds these test characteristics and provide a clearer understanding of the true value of a diagnostic test.

ACKNOWLEDGMENT

The author is indebted to Robert Jimison for preparation of the figures.

REFERENCES

1. **Patterson RE, Horowitz SF.** Importance of epidemiology and biostatistics in deciding clinical strategies for using diagnostic tests: a simplified approach using examples from coronary artery disease. J Am Coll Cardiol 1989; **13:**1653–1665.
2. **Sox HC Jr.** Probability theory in the use of diagnostic tests: an introduction to critical study of the literature. Ann Intern Med 1986; **104:**60–66.
3. **Griner PF, Mayewski RJ, Mushlin AI, Greenland P.** Selection and interpretation of diagnostic tests and procedures: principles and applications. Ann Intern Med 1981; **94:**557–600.
4. **McNeil BJ, Keeler E, Adelstein SJ.** Primer on certain elements of medical decision making. N Engl J Med 1975; **293:**211–215.
5. **Ransohoff DF, Feinstein AR.** Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; **299:**926–930.
6. **Hlatky MA, Pryor DB, Harrell FE Jr, Califf RM, Mark DB, Rosati RA.** Factors affecting sensitivity and specificity of exercise electrocardiography: multivariable analysis. Am J Med 1984; **77:**64–71.
7. **Weiner DA, Ryan TJ, McCabe CH, et al.** Exercise stress testing: correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the Coronary Artery Surgery Study (CASS). N Engl J Med 1979; **301:**230–235.
8. **Pryor DB, Harrell FE Jr, Lee KL, Califf RM, Rosati RA.** Estimating the likelihood of significant coronary artery disease. Am J Med 1983; **75:**771–780.
9. **Martin TW, Seaworth JF, Johns JP, Pupa LE, Condos WR.** Comparison of adenosine, dipyridamole, and dobutamine in stress echocardiography. Ann Intern Med 1992; **116:**190–196.
10. **Eisenberg MJ, Pilote L.** Pharmacologic stress echocardiography [letter]. Ann Intern Med 1992; **117:**168–169.
11. **Eisenberg MJ, Schiller NB.** Bayes' theorem and the echocardiographic diagnosis of cardiac tamponade. Am J Cardiol 1991; **68:**1242–1244.
12. **Beytas EM, Debatin JF, Blinder RA.** Accuracy and predictive value as measures of imaging test performance. Invest Radiol 1992; **27:**374–378.