Simple Statistical Measures for Diagnostic Accuracy Assessment

Jayawant N. Mandrekar, PhD

Abstract: The aim of diagnostic medicine research is to estimate and compare the accuracy of diagnostic tests to provide reliable information about a patient's disease status and thereby influencing patient care. When developing screening tools, researchers evaluate the discriminating power of the screening test by using simple measures such as the sensitivity and specificity of the test, as well as the positive and negative predictive values. In this brief report, we discuss these simple statistical measures that are used to quantify the diagnostic ability of a test.

Key Words: Sensitivity, Specificity, Predictive values, Prevalence.

The purpose of a diagnostic test is to classify or predict the presence or absence of a disease. The clinical performance of a diagnostic test is based on its ability to correctly classify subjects into relevant subgroups. As new diagnostic tests are introduced, it is important to evaluate the quality of the classification obtained from this new test in comparison with existing tests or the gold standard.

To illustrate the simple measures of diagnostic accuracy, consider an example where results from a diagnostic test such as roentgenogram or computer tomographic scan and the true disease or condition of the patient is known (Table 1). The accuracy of any test is measured by comparing the results from a diagnostic test (positive or negative) with the true disease using a gold standard (presence or absence) (see Table 1).[1]

In this brief report, we first introduce the different methods used to quantify the diagnostic ability of a test, namely sensitivity, specificity, likelihood ratio (LR), positive predictive value, and negative predictive value. We then discuss the impact of disease prevalence on these measures.

SENSITIVITY AND SPECIFICITY

The two basic measures of quantifying the diagnostic accuracy of a test are the sensitivity and specificity.[1,2] Sensitivity is the ability of a test to detect the disease when it is truly present, whereas specificity is the probability of a test to exclude the disease status in patients who do not have the disease. Thus, sensitivity is given by the ratio of true positives/(true positives + false negatives), and specificity is given by the ratio of true negatives/(true negative + false positives). In the example given in Table 1, the sensitivity is 90% (270/300) and specificity is 60% (60/100).

In describing a diagnostic test, one needs to report both sensitivity and specificity because they are inherently linked in that as the value of one increases, the value of the other decreases. These are also dependent on the patient characteristics and the disease spectrum, for example, bigger tumors are easier to detect than smaller benign lesions. In practice, investigators aim for maximizing both sensitivities and specificities. Given the financial and emotional implications associated with a disease process, higher sensitivity is often considered desirable in a diagnostic setting, and higher specificity is desirable in a screening set up.

POSITIVE AND NEGATIVE PREDICTIVE VALUES

Clinically, it is always important to know how good the test is at predicting the true disease status given the findings from new proposed test. This is captured by the predictive values. Positive predictive value (PPV) is the probability that a patient has the disease given that the test results are positive, and the negative predictive value (NPV) is the probability that a patient does not have the disease given that the test results are indeed negative. PPV is therefore given by the ratio of true positives/(true positives + false positives), and NPV is given by the ratio of true negatives/(true negatives + false negatives). In the example given in Table 1, PPV is 87% (270/310) and NPV is 67% (60/90).

PREVALENCE AND ITS IMPACT

Prevalence is defined as the prior probability of the disease before the test is carried out. Therefore, by definition, both sensitivity and specificity can be applied to other populations that have different prevalence rates. The predictive values on the other hand are dependent on the prevalence of the disease being tested.[2,3] As described in Table 1, the disease prevalence in that case is 75% (300/400). Instead, consider a slightly different scenario in which the disease prevalence is 50% (200/400) (Table 2).

The sensitivity and specificity are still 90% (180/200), and 60% (120/200). However, the PPV and NPV are 69%

TABLE 1. Test Results by Disease Status with Disease Prevalence of 75%

Test Results	Disease Status (Gold Standard)		Total
	Present	Absent	
Positive	True positive (270)	False positive (40)	310
Negative	False negative (30)	True negative (60)	90
Total	300	100	400

TABLE 2. Test Results by Disease Status with Disease Prevalence of 50%

Test Results	Disease Status (Gold Standard)		Total
	Present	Absent	
Positive	True positive (180)	False positive (80)	260
Negative	False negative (20)	True negative (120)	140
Total	200	200	400

(180/260) and 86% (120/140), respectively. Consider another scenario where the disease prevalence is only 5%. In this case, the PPV and NPV would be 11% and 99%, respectively. Thus, the rarer the prevalence of the disease, the more certain one can be that a negative test result indeed means that there is no disease (i.e., higher NPV). Similarly, the rarer the prevalence of the disease, the less certain one can be that a positive test result indicates the presence of a disease (i.e., lower PPV). PPV and NPV estimates obtained from one study can therefore not be applied universally without information on prevalence. In addition, as can be seen by the examples above, the number of people who are diagnosed as false positives increases in cases of low prevalence, even though the sensitivity and the specificity remain high.

LIKELIHOOD RATIO

LR is given by the ratio of the probability of the test result among patients who truly had the disease to the probability of the same test among patients who do not have the disease. In other words, the LR is really the ratio of sensitivity to (100 − Specificity). Therefore, it is independent of prevalence of the disease. The LR for both examples 1 and 2 is 2.25 (90/[100 − 60]).

The magnitude of the LR informs about the certainty of a positive diagnosis. As a general guideline, a value of LR = 1 indicates that the test result is equally likely in patients with and without the disease, values of LR >1 indicate that the test result is more likely in patients with the disease and values of LR <1 indicate that the test result is more likely in patients without the disease.[2]

SUMMARY

Studies designed to measure the performance of diagnostic tests are important for patient care and health care costs. Attention must be given to include proper representation of patients with the disease or condition of interest along with healthy participants to ensure that the study results are generalizable to the population of interest. Extrapolation of results obtained from one study to other populations requires a good understanding of the underlying prevalence and its impact on the estimates of PPV and NPV.

REFERENCES

1. Altman DG, Bland JM. Statistics notes: diagnostic tests 1: sensitivity and specificity. *Br Med J* 1994a;308:1552.
2. Zhou XH, Obuchowski NA, Obuchowski DM. Statistical Methods in Diagnostic Medicine. New York: John Wiley and Sons, 2002.
3. Altman DG, Bland JM. Statistics notes: diagnostic tests 2: predictive values. *Br Med J* 1994b;309:102.