ERIC Notebook

Second Edition

Assessment of Diagnostic and Screening Tests

Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

Diagnostic and screening tests are used to detect the presence or severity of disease in individuals. Clinicians rely on these tests to make decisions in treating patients. Thus, it is important to assess the performance of diagnostic and screening tests before they are adopted in a clinical setting. The performance of any new diagnostic or screening test is assessed by comparing actual test results to the patient's true disease status (as assessed by a *gold standard*). The four measures used to evaluate a new test are the *sensitivity*, *specificity,* and *positive and negative predictive values.*

The gold standard

Gold standard is a term for the most definitive diagnostic procedure, e.g. microscopic examination of a tissue specimen, or the best available laboratory test, e.g. serum antibodies to HIV. Sometimes it can refer to a comprehensive clinical evaluation, e.g. clinical assessment of arthritis. Gold standard procedures can often be costly, invasive, and/or uncomfortable. New tests that are less invasive and less expensive are compared against gold standards to assess the new test's accuracy. A test which detects a marker in the blood for prostate cancer may not be as sensitive as taking a biopsy from the prostate itself, but the discomfort of biopsy may make the blood assay a better alternative.

Calculating the test results

A table like the one below is used to group individuals into one of four disease-test categories.

	Truth (Gold Standard)	
Results of test	Disease Present	Disease Absent
Positive Test	a = True Positive	b = False Positive
Negative Test	c = False Negative	d = True Negative

Prevalence of the disease =

$$\frac{\text{True positives}}{\text{Total population}} = \frac{(a+c)}{a+b+c+d)}$$

Sensitivity vs. specificity

Sensitivity and specificity are measures that assess the validity of diagnostic and screening tests. These measures reflect how well the test is

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH detecting the disease and classifying individuals into disease and non-disease groups.

Sensitivity (Se or Sn) describes how well the test detects disease in all who truly have disease, or the percent of diseased individuals who have positive test results

Specificity (Sp), describes how well the test is detecting non-diseased individuals as truly not having the disease, or the percent of non-diseased individuals who have

$$Se = \frac{True-positives}{True-positivies+False-negatives} \text{ x } 100$$

$$Se = \frac{a}{(a+c)} \text{ x } 100$$

$$Sp = \frac{True-negatives}{True-negativies+False-positives} \text{ x } 100$$

$$Sp = \frac{d}{(b+d)} \text{ x } 100$$

A highly sensitive test means that a large percent of people who have disease are classified correctly as having the disease. A highly specific test means that a large percent of individuals without disease are classified correctly as not having disease. An ideal test would be both highly sensitive and highly specific, where disease would be detected in 100% of those who truly have disease (100% sensitivity), and disease would be ruled out in 100% of those who are truly disease-free (100% specificity).

Example

If a test is 95% sensitive and 98% specific, then 5% of the diseased individuals will have negative test results (the test is incorrectly classifying 5% of the diseased individuals), and 2% of the disease-free individuals tested will have positive test results (the test is incorrectly classifying 2% of the disease-free individuals).

False-positives and false-negatives

A *false positive* is an individual who is incorrectly diagnosed as a case when, in fact, they do not have the disease. A *false negative* is an individual who is incorrectly diagnosed as a non-case, when in fact the person does have the disease.

100% - % sensitivity = % false negatives

100% - % specificity = % false positives

Positive and negative predictive values

The *positive predictive value* (PPV) is the percent of positive tests that are truly positive. The *negative predictive value* (NPV) is the percent of negative tests that are truly negative.

Like sensitivity and specificity, PPV and NPV also show how well the test is classifying individuals into disease and non-disease groups, but the denominator for PPV is the total number of persons who test positive (a+b), while that for NPV is the total number who test negative (c+d). A test with a high PPV value means that there is only a small percent of false-positives within all the individuals with positive test results. A test with a high NPV value means that there is only a small percent of false-negatives within all the individuals with negative test results.

$$PPV = \frac{True\ Positives}{Positive\ Tests} \text{ x } 100$$

$$PPV = \frac{a}{(a+b)} \text{ x } 100$$

$$NPV = \frac{True\ Negatives}{Negative\ Tests}\ x100$$

$$NPV = \frac{d}{(c+d\ )} \text{ x } 100$$

Example

A certain test, e.g. a stress ECG, which has a PPV of 90% and a NPV of 95% is used to screen 5,000 people for coronary heart disease. Forty percent of the individuals (2,000 people) have positive test results and 60% (3,000 people) have negative test results. But the gold standard for CHD found that 1,800 of those who tested positive (90% of 2,000) truly have CHD, and 2,850 of those who tested negative (95% of 3,000) are truly non-cases.

Pros and cons of specificity and sensitivity

Ideally, an investigator would prefer a diagnostic test that is both 100% sensitive and 100% specific. However, this scenario rarely occurs. It is important in clinical decision-making to know the sensitivity and specificity of the test you are conducting and to weigh the pros and cons of using tests with different levels of sensitivity and specificity. For instance, if a disease is not life threatening if left untreated, the costs of treatment are high, and invasive surgery is required, then a very specific diagnostic test is preferred over a more sensitive test. If the disease under study is life threatening if left untreated, and the survival rate is improved with immediate treatment, then the sensitivity of a diagnostic test is of greater importance than its specificity.

Example

Schizophrenia has a low prevalence in the U.S. at around 1%. A new diagnostic test that is 99% sensitive and 99% specific is used to screen 10,000 patients for schizophrenia. Of those 10,000, we would expect 100 to truly be suffering from schizophrenia, or 1% of our population. Of those 100, 99 (99% of 100) would have positive test results. Of the 9,900 who are truly without disease, 9801 (99% of 9,900) would be classified as disease-free. However, there would be 99 (9,900-9801) false positives. This test would give 198 (99+99) positive test results. Therefore, even with a test that is 99% sensitive and 99% specific, the PPV would only be 50% (99/198).

The prevalence affects the predictive values

The prevalence of a disease affects the PPV and NPV values. If a disease has a low prevalence and the test being used to assess disease in individuals is not 100% sensitive or 100% specific, as will most likely be the case, then false-positives may overwhelm the positive test results.

Terminology

False positive: An individual who is incorrectly diagnosed as having disease

False negative: An individual who is incorrectly diagnosed as not having disease

Gold standard: The most definitive diagnostic procedure to detect disease

Negative predictive value: The percent of negative tests that are truly negative

Positive predictive value: The percent of positive tests that are truly positive

Sensitivity: The percent of diseased individuals who have positive test results

Specificity: The percent of non-diseased individuals who have negative test results

Practice Questions.

A test is used to screen people for hepatitis B. The sensitivity of the test is 95% and the specificity of the test is 90%. Assume that the total number of persons being tested for hepatitis B is 50,000. Assume that the true prevalence of hepatitis B in the population is 100 per 50,000.

	Disease present	Disease absent
Positive test	a= true positive	b= false positive
Negative test	c= false negative	d= true negative

1) Calculate the number of true positives

2) Calculate the number of false positives

References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

The University of North Carolina at Chapel Hill, Department of Epidemiology Courses: Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, and Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-2013.

Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH, Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP) to promote the strategic growth of the epidemiologic capacity of the DVA.

Answers to Practice Questions

	Disease present	Disease absent
Positive test	a= true positive	b= false positive
Negative test	c= false negative	d= true negative

a+ b+ c+ d= 50,000
a+c= 100 (100 people have Hepatitis B)
b+d = Total – (a+c) = 50,000-100=49900

1) Calculate the number of true positives
Sensitivity = 95% = a / (a+c) a= (0.95)(a+c)= 0.95*100= 95 true positives

A sensitivity of 95% means that the test detects as positive 95% of the diseased individuals.

2) Calculate the number of false positives
Specificity = 90%= d / (d+b)
d=(0.90)(d+b) = 0.90*49900)=44910
b= (b+d)-d= (49900-44910)=4990 false positives

A specificity of 90% means that the test detects as negative 90% of the non-diseased individuals.