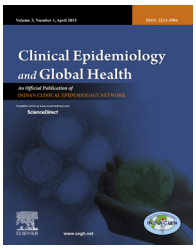
Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cegh



Continuing Education

Evaluation of diagnostic tests

Rashmi Kumar *

Professor, Department of Pediatrics, King George's Medical University, Lucknow, India

ARTICLE INFO

Article history:
Received 17 April 2015
Accepted 1 December 2015
Available online 22 December 2015

Keywords:
Diagnostic tests
Evaluation
Sensitivity
Specificity
Likelihood ratio

ABSTRACT

Measures of test efficacy are its repeatability or reliability and its validity. Validity is computed by comparing it against an older 'gold standard' test, which is supposed to unequivocally give the diagnosis. Sensitivity and specificity are inherent properties of a test. Predictive values give the probability of disease or no disease when the test result is known and vary with the prevalence of disease in the setting it is used. Receiver operator characteristic curves are used to determine the best cut off and also to choose between tests with numerical values. Likelihood ratios give the odds of disease at a particular test result.

One of the most important tasks, a clinician has to perform when faced with a patient is that of reaching a diagnosis. Most astute clinicians use a thorough knowledge of literature along with a judicious use of diagnostic tests, good judgment, and a ready approach to organize the information. While newer diagnostic tests are continually coming into use, not much was said till recently about assessment of the test itself. The science of Clinical Epidemiology addresses the diagnostic process and the interpretation and evaluation of diagnostic data, both clinical and laboratory.

1. Measures of test efficacy

Most diagnostic testing concerns measurements. A good test is easy, inexpensive, safe and easily available. One of the attributes of a good test is its *reliability or repeatability*. Repeat measurements are likely to vary even in the same subjects. These differences increase the "noise" around any measurement and should be minimized. Validity of a test is an expression of the degree to which it is supposed to measure, and how well it discriminates between diseased and non-diseased.

1.1. Probability of a diagnosis

Whenever we consider a diagnosis, we talk about probabilities. On the basis of the history and clinical examination, we already have an idea of the probability of a diagnosis (prior probability). This prior or pretest probability depends on clinical judgment and prevalence in the clinical setting in which the patient is seen. Above a certain level of probability, we should treat the patient and below a certain level we would not. In between these two levels, we would take the help of diagnostic tests. These tests are expected to move us along the scale of probability into either the "treat" or "not treat" zones (post-test probability) (Fig. 1).

How much the test affects probability of the diagnosis will depend in how good or discriminatory the test is.

* Tel.: +91 9415408777; fax: +91 522 2257243.
E-mail address: rashmik2005@gmail.com

http://dx.doi.org/10.1016/j.cegh.2015.12.001

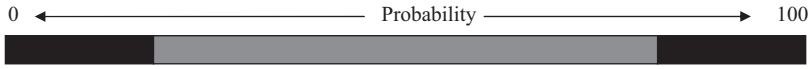


Fig. 1 – Probability of a diagnosis.

1.2. *The gold standard*

To study the validity of a test, we need to compare it with an older, more established test – the 'gold standard'. The gold standard test may be more difficult, risky, expensive or not easily available but it should unequivocally tell us whether the disease is present or absent. The gold standard result is equated with presence or absence of the disease in question. Sometimes an ideal gold standard is not available and a combination of tests, follow-up data or response to treatment, etc. are taken as the gold standard. For example, to study the accuracy of a new test for enteric fever, a combination of the Widal test and blood culture (either one or both) may be taken as the gold standard. The new test being evaluated and the gold standard test are done uniformly in all subjects in a blinded fashion. This would give us 4 groups of subjects as shown in the 2 × 2 table (Fig. 2).

1.3. *Sensitivity and specificity*

Two of the attributes of a test are its sensitivity and specificity. Sensitivity is the ability of a test to pick out those patients, who really have the disease. It is synonymous with positivity in disease (PiD) rates or true positivity. Specificity is the ability of a test to confirm that the disease is absent when it is truly absent and is synonymous with negative in health (NiH) rate or true negativity. Sensitivity and specificity can be derived by comparing the test with the ''real answer'' about whether the disease is present or absent in a simple 2 × 2 table given.

$$\text{Sensitivity} = a/a + c; \qquad \text{Specificity} = d/b + d$$

Sensitivity and specificity are attributes of a test. Some tests are more sensitive, while others more specific. Very often there is a trade off between sensitivity and specificity. Sensitive tests are more useful for ruling out a diagnosis, if they are negative. They are used early in the diagnostic process such as in screening. A sensitive test is also used for dangerous but treatable conditions when it is important not to miss a single patient. A specific test is more useful for ruling in a diagnosis, if it is positive and therefore more useful for final diagnosis. They are used when a false diagnosis or label would cause alarm. What degree of sensitivity/specificity one chooses also depends on the clinical situation and the trade offs involved. For example, one would like a highly sensitive test, when one is considering a diagnosis of bacterial meningitis as the condition is fatal if left untreated, and one would rather err on the side of treating them. Before conveying a diagnosis of say, malignancy one would want to be very sure and therefore use a very specific test.

	Gold Standard +ve	Gold Standard –ve
Test +ve	**a**	**b**
Test –ve	**c**	**d**

Fig. 2 – 2 × 2 table. *a* – patients with disease and a positive test; *b* – patients without disease but a positive test; *c* – patients with disease but a negative test; *d* patients without disease and a negative test.

2. Predictive values

Sensitivity and specificity are good for describing a test but in a clinical situation, the clinician has the test result and wants to know whether the disease is present or not. For this purpose, he needs to know its predictive value.

Positive predictive value (PPV) is the proportion of those with a positive test, who actually have the disease, i.e. PPV = true positive/total positives. From the 2 × 2 table above, PPV = $a/a + b$.

Negative predictive value (NPV) is the proportion of those, who test negative and are actually non-diseased, i.e. NPV = true negatives/total negatives = $d/c + d$.

Predictive values are thus calculated horizontally.

It is important to understand, however, that predictive value of a positive or negative test is not constant but changes with the prevalence of the disease in the situation, in which the test is used. Predictive values can be increased or decreased by choosing the type of patients to be tested. For example, PPV of amniotic fluid alpha fetoprotein (AFP) for neural tube defect can be increased by doing it in mothers with high serum AFP. This in effect means increasing the prevalence or pretest probability of the condition in the tested population. PPV will also decrease greatly, if the test is used for screening in the community. PPV also equals the post-test probability mentioned above.

3. Tests with numerical values

The above discussion was centered on tests, which gave a positive or a negative answer, for example, biopsy findings of Reid Sterberg cells in Hodgkin's disease or Western Blot assay in HIV. Many laboratory tests give the answer in numerical values on a continuous scale. It is often up to the clinician to decide the cut-off he chooses, above or below which the measurement is considered abnormal. Once a cut-off point is chosen, the continuous data are converted into dichotomous positive or negative, and the sensitivity and specificity at that cut-off can be calculated as above. One can also calculate sensitivity and specificity at a range of cut-off values. Usually