

Objective assessment of diagnostic tests validity: a short review for clinicians and other mortals. Part I

Nermin N. Salkić

University Clinical Center Tuzla
Department of Gastroenterology
Bosnia and Herzegovina

Corresponding author:
Nermin Salkić
Miroslava Krleža 15/26
75000 Tuzla,
Bosnia and Herzegovina
nsmermin@gmail.com

Received: 30 November 2008
Accepted: 23 December 2008

The whole point of a diagnostic test is to use it to make a diagnosis, thus the obvious need is to know how accurately a particular diagnostic test detects patients with or without a disease. In order to know it, a clinician or a researcher should have a basic understanding of the principles of objective appraisal of diagnostic tests. In this short review, the author presents the most common biostatistical methodology for assessment of the validity of diagnostic tests. Definitions and interpretations of sensitivity, specificity, positive and negative predictive values are also provided together with how they are calculated.

Key words: Sensitivity, Specificity, Positive predictive value, Negative predictive value.

Introduction

The very first time I encountered the concept of sensitivity, specificity, positive and negative predictive value, in the end I had a huge question mark above my head. Those somewhat vague parameters, which have an important role in describing the diagnostic performance of a particular diagnostic test, seemed to have a logic that is beyond my grasp.

However, later, when I became much more interested in biostatistics, I realized the importance of measures deriving from a work of a British mathematician and Presbyterian minister Thomas Bayes (1702–1761) which is known today as Bayes' theorem(1).

Why is it so important? The most common use of this approach will be when your

patient has an abnormal lab test result and you wonder "What does this really mean?" or in other words, "How likely it is that this patient really has the disease in question?" Bayes' theorem allows us to evaluate the diagnostic performance of each particular test and also to compare several of them.

A two-by-two table

The best way to understand this concept is by example. Having a piece of paper, pen and calculator nearby, while reading this text is advisable.

Usually, a diagnostic test is validated by comparison against an established gold standard in an appropriate group of subjects. In order to make the analysis, we need

to make something called a *2-by-2 table*, which is displayed in a Table 1. The Table is self-explanatory, but be sure always to label the table with the test results on the left side and the disease status on the top as shown.

Table 1 A two-by-two table notation for expressing the results of validation study for diagnostic or screening test

	Disease present	Disease absent
Test positive	True positives (TP)	False positives (FP)
Test negative	False negatives (FN)	True negatives (TN)

Now, let us set the stage for our hypothetical problem. Liver biopsy is currently considered to be the gold standard in the assessment of the presence and degree of liver fibrosis in various liver diseases such as viral hepatitis etc. (2). However, it is associated with the possibility of severe complications and serious discomfort for the patient (3). Therefore, our hypothetical researchers decided to evaluate a non-invasive marker of liver fibrosis comparing it with the gold standard (liver biopsy).

The researchers recruited 189 patients. After performing liver biopsy, 43 of them had liver fibrosis, while 146 did not. On the other hand, after performing a non-invasive test for liver fibrosis, 61 patients were positive for the presence of liver fibrosis, while 128 of them were negative. Now, let us make a 2-by-2 table from this data (Table 2).

Table 2 Two by two table showing the results of a validation study of non-invasive liver fibrosis test against the gold standard

	Liver biopsy positive	Liver biopsy negative	Total
Test positive	43 (TP)	18 (FP)	61
Test negative	0 (FN)	128 (TN)	128
Total	43	146	189

TP-true positive; TN-true negative; FP-false positive; FN-false negative

Sensitivity

The sensitivity or true positive rate (positive for disease) answers the question “How good is this test in picking up people who have the condition?” In other words sensitivity is the probability of a positive test among those who actually do have the condition (4, 5). It is calculated by using the formula

$$\text{Sensitivity} = TP / (TP + FN)$$

Sensitivity for our diagnostic test would be $43 / (43 + 0) = 1 = 100\%$. So, one could say that this non-invasive test for liver fibrosis will detect all patients that actually have it. The higher the value of sensitivity, the higher the proportion of those with the actual condition among those that test positive.

Specificity

The specificity or true negative rate (negative in health) answers the question “How good is this test at correctly excluding people without the condition?” Specificity is the probability of a negative test in those without the condition. (4, 5) The higher the specificity, the higher the proportion of those without the actual condition among those that test negative. Specificity is calculated by using the formula

$$\text{Specificity} = TN / (FP + TN)$$

Therefore, in our example the specificity would be $128 / (18 + 128) = 0.88 = 88\%$ which means that our test will correctly classify 88% of those that actually do not have liver fibrosis. Yet, there will still be 12% of those who will test positive despite the fact that they do not have liver fibrosis.

Pre-test and post-test probabilities

Sensitivity and specificity are one way to evaluate the diagnostic ability of a test. Since they are dealing with probabilities before ac-

tually performing a test they are also called pre-test probabilities. From a clinical point of view, all we have is the result of a test, so clinicians are much more interested in knowing what proportion of patients with an abnormal test are truly abnormal or vice versa – the proportion of patients with a normal test who do not have the condition (6). These questions are answered by using so called, post-test probabilities and positive and negative predictive values, since we are dealing with numbers after actually performing a test.

Pre-test probabilities are of great use when we have several diagnostic tests at our disposal and we need to select the one with the best chance of detecting the condition. On the other hand, the whole point of a diagnostic test is to use it to make a diagnosis, so we need to know the probability that the test will give a correct diagnosis; sensitivity and specificity do not provide us with this type of information. (6) Predictive values however, do have one important limitation: they are measures that we calculate from a defined population with a defined prevalence. If we change the prevalence, the predictive values also change, therefore they do not necessarily apply to another population. Post-test probabilities do not have this particular limitation.

So, what are positive and negative predictive values and how do we calculate them?

Positive predictive value

Positive predictive value (PPV) answers the question “In group of patients with positive test, what is the proportion of those with the condition?” or in other words, it is the proportion of patients with positive test results that are correctly diagnosed (4, 6). The higher the PPV, the higher is our certainty that patient with a positive test really has the condition. It is calculated by using the formula

$$PPV = TP / (TP + FP),$$

or in our hypothetical research $PPV = 43 / (43 + 18) = 0.70 = 70\%$. This practically means that when we have a patient tests positive for liver fibrosis by our non-invasive test, he will actually have a 70% probability of really having liver fibrosis. This number tells us that our test is not particularly reliable in detecting the presence of the disease.

Negative predictive value

The negative predictive value (NPV) answers the question “In the group of patients with negative test results, what is the proportion of those without the condition?” This is the proportion of patients with negative test results who are correctly diagnosed (4, 6). The higher the NPV, the higher is our certainty that a patient with negative test results does not have the condition. Negative predictive value is calculated by using the formula

$$NPV = TN / (FN + TN),$$

or in our case $NPV = 128 / (128 + 0) = 1 = 100\%$. This is an excellent NPV which means that we can be fairly certain that our patient with a negative non-invasive liver test really does not have liver fibrosis. There are particular clinical situations when it is important to know that our patient does not have the condition. These are settings where tests with high NPV have their significance, regardless of their PPV.

Sensitivity, specificity, PPV and NPV are looking at a one side of the coin. But, what if we want to look at both sides of coin, or in other words, if we want to assess the overall accuracy of a test, taking into account true positive and true negative cases? Also, the positive and negative predictive values depend crucially on prevalence; when we change the prevalence, PPV and NPV change also. How do we avoid this problem?

These are issues to be addressed in the next part of our series.

References

1. Anonymous. Wikipedia: Thomas Bayes. 2008 [cited November 16th, 2008]; Available from: http://en.wikipedia.org/wiki/Thomas_Bayes
2. Lok AS, McMahon BJ. Chronic hepatitis B. *Hepatology* 2001;34:1225-41.
3. Cadranet JF, Rufat P, Degos F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). *Hepatology* 2000;32:477-81.
4. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ* 1997;315:540-3.
5. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994;308:1552.
6. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994;309:102.