Understanding sensitivity, specificity and predictive values

7 authors, including:

Miodrag Stojanović
University of Niš
48 PUBLICATIONS   **312** CITATIONS

SEE PROFILE

Marija Anđelković Apostolović
School of Medicine, University of Nis , Serbia
20 PUBLICATIONS   **46** CITATIONS

SEE PROFILE

Dijana Stojanovic
University of Niš
34 PUBLICATIONS   **188** CITATIONS

SEE PROFILE

Aleksandra M. Ignjatović
University of Niš
48 PUBLICATIONS   **210** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

PhD thesys of the first author View project

ERA EDTA congres View project

Understanding sensitivity, specificity and predictive values

Razumevanje osetljivosti, specifičnosti i prognostičkih vrednosti

Miodrag Stojanović*, Marija Apostolović*, Dijana Stojanović*, Zoran Milošević*, Aleksandra Toplaović*, Vesna Mitić Lakušić[†], Mlađan Golubović[‡]

*Faculty of Medicine, University of Niš, Niš, Serbia; [†]Clinical Center Podgorica, Podgorica, Montenegro; [‡]Clinical center Niš, Niš, Serbia

Introduction

Modern medicine and medical science have experienced a change in knowledge and dramatic increase in technology. Nowadays "Hi-Tech" tests are widespread and doctors can hear a statement: "This new test is very reliable, it is the most sensitive test at the market, you should use it", practically every day. The basic idea for using diagnostic tests aims to easier making the diagnosis of a disease and so enable appropriate treatment. How to know if a test is useful? Will a test point out to a doctor whether a patient is ill or not?

Due to the fact that the doctors and medical research workers are not so common with the statistical methods, the aim of this paper was to explain the basics of sensitivity, specificity and predictive values in a very simple way, using simple logic formulas, and presenting the samples from medical practice.

Diagnostic test and the gold standard

Diagnostic tests are all the tests that physician can use in the process of making the diagnosis of a particular disease. It is a procedure performed to confirm or determine the presence of a disease in a person suspected of having the disease, usually following the report on signs and symptoms, or based on the results of other medical tests [1].

The most accurate test for determining a disease is a "gold standard". Since it represents the best of the existing tests, we may consider the "gold standard" as a currently preferred method for diagnosing a specific disease [2]. It is often invasive or expensive; therefore some other diagnostic test may be used instead. Hence, a newly designed test has to be initially validated by comparing its results with a gold standard due to establish the exact health status of a person.

Some of the known test examples are: uriscreen for urinary tract infection, blood pressure for hypertension, pap smear for cervical carcinoma, mammography for breast cancer, prostate specific antigen (PSA) for prostate cancer, fecal occult blood for colon cancer, ocular pressure for glaucoma, colonoscopy to find early cancers and potentially cancerous polyps, ultrasound for thyroid cancers, nuclear medicine techniques to examine a lymphoma, measuring blood sugar for diabetes mellitus, taking a complete blood count for bacterial infection [3]. Gold standard may be arbitrary and may change. When a new test is under consideration for using in practice it should be good enough to replace the gold standard for some particular disease, otherwise should be discharged or used as a preliminary test. The doctor's goal is to realize which test and testing strategies are best for making the correct diagnosis.

Validity

Validity is the capability of a test to point out which people have a disease and which do not. It is the test accuracy, or the extent to which a test is able to measure what should be measured [4]. Validity is estimated by two objective measures: sensitivity and specificity [5].

The ideal test should correctly identify all tested people with or without disease with 100% of accuracy, which is practically impossible. Traditionally, to help understanding sensitivity, specificity and predictive values, the best method for explanation is based on 2 × 2 contingency table. Suppose a population of 1,000 people, 100 of them have a disease X, 900 do not have the disease according to the gold standard results (Table 1). Here, the rows represent the screening test results and the columns the true condition of a person according to the gold standard. Screening test is used to identify 180 people with the disease (a + b).

Table 1

The table 2 × 2 (for diagnostic test results)

Screening test results	Disease X		Total (n)
	present (n)	absent (n)	
Positive	80 (a)	100 (b)	180 (a + b)
Negative	20 (c)	800 (d)	820 (c + d)
Total	100 (a + c)	900 (b + d)	1000 (a + b + c + d)

a – true positive (sick people correctly diagnosed as sick);b – false positive (healthy people wrongly diagnosed as sick); c – false negative (sick people wrongly diagnosed as healthy); d – true negative (healthy people correctly diagnosed as healthy).

Sensitivity

Sensitivity of a clinical test represents test ability to correctly identify people with illness (a) within all people with illness (a + c). It is a proportion of people with disease who positive, expressed in percentages. Sensitivity as a fixed test characteristic provides a true positive rate [5, 6].

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives}$$

If we apply screening test to our hypothetical population and receive that 80 of the 100 people with disease X test positive, than the sensitivity of this test is 80/100 or 80% (Table 1). A test with 80% of sensitivity detects 80% of true ill patients, while 20% (false negative) will not be detected.

Specificity

The specificity of a clinical test represents test ability to correctly identify people without illness (d) within all people free from illness (b + d). It is a proportion of people without disease who test negative. Specificity is also a fixed characteristic of the test and represents true negative rate [7, 8].

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives}$$

This hypothetical population (Table 1) demonstrates that 800 of the 900 people without disease X test negative, so specificity of this screening test is 800/900 or 89%. Practically, this test detects 89% of healthy people but 11% will be assumed as false positive.

For a test to be accurate, both sensitivity and specificity should be high. When measuring sensitivity, we only calculate those people with disease. High sensitive test detects a high percentage of positive cases while missing few. Also, a negative result would suggest the absence of disease according to test with high sensitivity. On the contrary, specificity highlights negative test results. A highly specific test is good for detection of a disease if a person tests positive, likewise it does not falsely diagnose disease when none is present.

It is worthy to mention that specificity and sensitivity of a quantitative test depend on a cut-off value. This is the value which determines the limit between positive and negative test results. In a situation when the cut-off is reduced, most people with the disease will be correctly identified, but at the same time the number of false positives will be increased. Raising the cut-off value will show more false negatives, but will reduce the number of false positives [9, 10]. Practically, sensitivity and specificity are inversely proportional, as one increases the other decreases and *vice versa*.

The lack of adequate education about interpreting test results (especially relevant to tests that may have minor and insignificant findings) may lead to misdiagnosing [11]. It is known that in two cases if test result doesn't correspond to the real situation an error has occurred. Therefore it is essential to pay attention to false positive and false negative test results. These errors are closely associated with the terms of type I and type II errors in hypothesis testing. In hypothesis testing "null hypothesis" matches to the natural state (in our situation people who are free from disease). As opposed to the null hypothesis there is an "alternative hypothesis" which corresponds to the ill people. Type I error, also known as "α" error appears when we reject null hypothesis which is actually correct [12, 13]. Type I error corresponds to the false positive results. Type II error or "β" error appears when accepting null hypothesis, when actually it is not really true [12, 13]. This error corresponds to false negative test results. It depends on a situation in which a false result is more undesirable. Minimizing false positives and false negatives at the same time maximizes sensitivity and specificity. Generally, it is not benign to tell someone after testing that he has a serious disease (false positive) when he does not really have the same (HIV for example). Moreover, it is inexcusable to overlook a disease when it really exists (false negative).

Sensitivity and specificity do not depend on the disease prevalence [14]. They are conditional on the patient either having or not having a disease and represent the power of a diagnostic test to discriminate between those with and without disease. When a patient has a positive test result, does it actually mean that he/she has disease or not? Sensitivity and specificity cannot answer such a question, thus it is worthy to know predictive values.

Predictive values

The real questions to be answered are the following: "What is the probability that a person with a positive test results will have the disease? Also if a person has a negative test, what is the likelihood that he is healthy?" These questions refer to what's called the "predictive values". Therefore, the mission of the clinician is to determine the likelihood of a disease present given a positive test (positive predictive value – PPV), or the likelihood of a disease absent given a negative test (negative predictive value – NPV).

Positive predictive value

The positive predictive value or precision rate is defined as a proportion of people with a positive test result (a) who are actually ill (a + b) [5, 8]. It is calculated by the formula:

$$Positive\ predictive\ value = \frac{True\ positives}{True\ positives + False\ positives}$$

In our population from 180 (a + b) people who test positive, 80 (a) of them actually have a disease X. PPV in this case is 80/180 or 44%. This practically means that a person who tests positive has a 44% likelihood of having a disease X. Less false positives (higher specificity) signifies a higher PPV in the observed population.

examined (Table 2). Since this is the same test, sensitivity and specificity remain the same 90% and 80%, respectively. According to this new scenario, PPV is 450/550, or 82%, whereas the NPV is 400/450, or 89%. In the case when we repeat the same test to some other population with different disease prevalence, we will notice that the PPV increases with the increasing disease prevalence while the NPV decreases in the same situation (Table 3). Though the sensitivity and specificity remain the same, the PPV has changed remarkably.

Table 2

Calculation of predictive values at 50% disease prevalence

Test	Disease present (n)	Disease absent (n)	Total (n)
Positive	450	100	550
Negative	50	400	450
Total	500	500	1000

Prevalence = 500/1000 = 0.5; sensitivity = 450/ 500 = 0.9; specificity = 400/500 = 0.8; PPV = 450/550 = 0.82; NPV = 400/450 = 0.89; PPV – positive predictive value; NPV – negative predictive value.

Table 3

Relationship between disease prevalence and predictive values for a test of 90% sensitivity and 80% specificity

Prevalence (%)	Positive predictive value (%)	Negative predictive value (%)
5	0.19	0.99
10	0.44	0.97
50	0.82	0.89
70	0.90	0.77
90	0.96	0.47

Negative predictive value

The negative predictive value is defined as a proportion of people with a negative test result (d) who actually do not have disease (b + c) [5, 8]. The formula for this measure is:

$$Negative\ predictive\ value = \frac{True\ negatives}{True\ negatives + False\ negatives}$$

We can see from our example (Table 1) that out of 820 people disease free, 800 people test negative. That basically means, a person who tests negative has 800/820 = 0.97 or 97% likelihood of not having a disease. NPP of enormous 99% indicates that a negative screening test actually dismisses disease. On the other hand, a very low PPV (only 44%) indicates that every second ill patient is diagnosed wrongly according to this test. A highly sensitive test (small number of false negatives) will have a higher NPV in the observed population.

In clinical practice, the value of a test result for an individual patient depends on the prevalence of tested population (the proportion of the population that has the disease at a given time) [15, 16]. We should expect that for any given population, as the prevalence of a disease increases, the test PPV will also increase and *vice versa*, as disease prevalence in the population being tested decreases, the PPV of that test will also decrease, while the NPV will increase. To help you understand these relationships we will illustrate this through the following example. Imagine that we now apply the same screening test we used before (Table 1) to another population where the disease prevalence is 50%, respectively 500 sick patients out of 1,000

The utility of predictive values is limited because the statistics is determined by sensitivity and specificity of a test, as well as with the prevalence of disease which can vary [17]. In general, specificity has more impact on a positive predictive value in the case of low disease prevalence [18]. Constancy of sensitivity and specificity are an important feature of a test when using in similar patients and similar settings. Predictive values although associated with sensitivity and specificity will change with the prevalence of target disease (Figure 1).
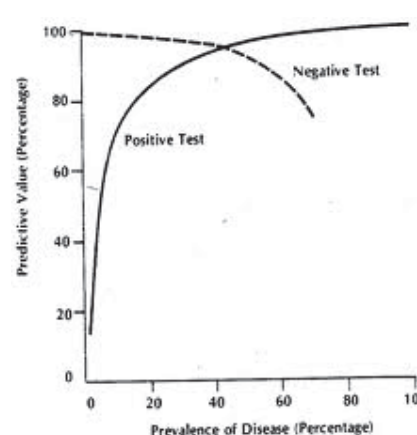

Fig. 1 – Disease prevalence and predictive values

illustrates the effect of the disease prevalence on the PPV and the NPV. Decreasing the disease prevalence increases the number of false-positive test results, while increasing the disease prevalence decreases the number of false-negative test results.

It is vital to note while comparing two populations that predictive values are only significant if the disease prevalence is the same in both populations. They are used along with the specificity and sensitivity when the prevalence in the target population is known. In the case of unknown prevalence, the sensitivity and specificity are the primary measurement used to evaluate the accuracy of a test [19].

Conclusion

To enable interpretation of diagnostic test accuracy it is necessary to understand the concepts of sensitivity, specificity and predictive values. These calculations require a design of the two-by-two table. Sensitivity and specificity of a test have limited clinical usefulness of the disease likelihood in an individual patient. Although sensitivity and specificity are not under the influence of the disease prevalence they can be affected by the differences in disease characteristics. In everyday clinical practice knowing the predictive values is more significant for measuring diagnostic accuracy. So, next time you hear about a new reliable test, ask yourselves what gold standard is performed, what is the disease prevalence, and most important of all, what are the sensitivity, specificity, positive and negative predictive values? If the answers are satisfactory, you can recommend this test for use.

R E F E R E N C E S

1. *Al-Gwaiz LA, Babay HH.* The diagnostic value of absolute neutrophil count, band count and morphologic changes of neutrophils in predicting bacterial infections. Med Princ Pract 2007; 16(5): 344−7.
2. *Spiegelman D, Schneeweiss S, McDermott A.* Measurement error correction for logistic regression models with an "alloyed gold standard".Am J Epidemiol 1996; 145(2): 184−96.
3. *Kanchanaraksa S.* Evaluation of Diagnostic and Screening Tests: Validity and Reliability. Baltimore: Johns Hopkins University; 2012. [cited 2012 March 5]. Available from: http://ocw.jhsph.edu/courses/fundepi/PDFs/Lecture11.pdf
4. *Singh AS.* Statistical Research design and basic techniques in epidemiological research. Eur J Appl Sci 2012; 4(1): 27−35.
5. *Singh A, Masuku M.* Understanding and applications of test characteristics and basics inferential statistics in hypothesis testing. Eur J Appl Sci 2012; 4(2): 90−7.
6. *Jaeschke R, Guyatt G, Sackett DL.* Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1994; 271(5): 389−91.
7. *Parikh R, Mathai A, Parikh S, Chandra SG, Thomas R.* Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008; 56(1): 45−50.
8. *Jaeschke R, Guyatt GH, Sackett DL.* Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994; 271(9): 703−7.
9. *Spitalnic S.* Test properties I: Sensitivity, Specificity, and Predictve values. Hosp Physician 2004; 40(9): 27−31.
10. Module 2. Screening Sensitivity, Specificity, and Predictive Values of Diagnostic and Screening Tests. Available from: www.medschool.lsuhsc.edu/module2.screening
11. *Ray MM.* Introduction to epidemiology. Sudbury, US: Jones and Bartlett Publishers, Inc; 2010.
12. *Devashish S, Yadav UB, Sharma P.* The concept of sensitivity and specificity in relation to two types of errors and its application in medical research. J Reliabil Statistical Studies 2009; 2(2): 53−8.
13. *Milosevic Z, Bogdanovic D.* Hypothesis testing and estimation of population parameters based on a sample. In: *Bogdanovic MZ*, editor. Statistics and informatics for medical science. Nis: Galaksija;. 2012. p. 84−5. (Serbian)
14. *Kelly H, Bull A, Russo P, McBryde E.* Estimating sensitivity and specificity from positive predictive value, negative predictive value and prevalence: application to surveillance systems for hospital-acquired infections. J Hosp Infect 2008; 69(2): 164−8.
15. *Altman DG, Bland JM.* Statistics notes: Diagnostic tests 2: Predictive values. BMJ 1994; 309: 102.
16. *Akobeng AK.* Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatr 2007; 96(3): 338−41.
17. *Collier J, Huebscher R.* Sensitivity, specificity, positive and negative predictive values: diagnosing purple mange. J Am Acad Nurse Pract 2010; 22(4): 205−9.
18. *Chu K.* An introduction to sensitivity, specificity, predictive values and likelihood ratios. Emerg Med 1999; 11(3): 175−81.
19. *Rothman KJ, Greenland S.* Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998.