 DOI: 10.2478/v10011-007-0039-z

UDK 577.1 : 61 ISSN 1452-8258

JMB 26: 309–313, 2007

Professional paper
Stručni rad

SIMPLE STATISTICS IN DIAGNOSTIC TESTS

JEDNOSTAVNA STATISTIKA U DIJAGNOSTIČKIM TESTOVIMA

Diler Aslan[1], Sverre Sandberg[2]

[1]Pamukkale University School of Medicine, Department of Biochemistry, Denizli, Turkey
[2]Lab. Clin. Biochem., Haukeland University Hospital and NOKLUS, Bergen, Norway

Summary: Diagnostic performance of a laboratory test is one of the key elements in decision making on diagnosis, screening, monitoring, risk assessment and prognosis of diseases. Sensitivity, specificity, likelihood ratios, diagnostic odds ratios and receiver operating characteristic curves are the measures of diagnostic accuracy of a test. The pretest probability of a disease or a target condition can be enhanced by the use of these measures, and hence the decision is made with the posttest probability. These measures are also used for analysis and critical appraisal of literature for finding the best evidence in the five-step model of evidence-based medicine approach, as well as for integrating the research results into clinical usage. In this context, the specialists in laboratory medicine should assess the diagnostic performance of a laboratory test as well as its analytical performance in order to take part in the management of health care services and health care resources. The aim of this review is to summarize the simple Statistics in diagnostic tests.

Keywords: simple statistics, diagnostic accuracy measures, diagnostic tests

Kratak sadržaj: Dijagnostičke performanse laboratorijskog testa jedan su od ključnih elemenata u procesu odlučivanja o dijagnozi, skriningu, praćenju, proceni rizika i prognozi bolesti. Mere dijagnostičke tačnosti jednog testa su osetljivost, specifičnost, odnosi verovatnoće, odnosi dijagnostičkih šansi i »ROC« krive. Verovatnoća bolesti ili medicinskog stanja pre testa može se uz upotrebu ovih metoda bolje proceniti, tako da se odluka donosi prema proceni verovatnoće posle testa. Te mere takođe se koriste za analizu i kritičko razmatranje literature u cilju pronalaženja najboljih dokaza u okviru modela iz oblasti medicine zasnovane na dokazima koji sadrži pet koraka, kao i za uključivanje rezultata istraživanja u kliničku upotrebu. Shodno tome, specijalisti laboratorijske medicine treba da procene dijagnostičke performanse kako bi učestvovali u upravljanju službama javnog zdravstva i resursima zdravstvene nege. Cilj ovog pregleda je da ukratko izloži primenu jednostavnih statističkih metoda u proceni dijagnostičkih testova.

Ključne reči: jednostavna statistika, mere dijagnostičke tačnosti, dijagnostički testovi

Introduction

The diagnostic accuracy of a test is one of the factors which are taken into account in the decision making for a disease. The clinicians should estimate the pretest probability before ordering a test. The prior /pretest probability may simply be the prevalence of a particular disease in the patient population. After studying the history and the physical examination, the prevalence is adjusted, and we get the prior or pretest probability.

Which test can be ordered for revising the probability of a disease is decided according to the knowledge of the diagnostic accuracy or effectiveness, risk and cost-effectiveness of the test.

In this review, the statistical method to determine the diagnostic accuracy of a test including some concepts and calculations of measures of diagnostic accuracy, such as sensitivity, specificity, likelihood ratios, diagnostic odds ratios, and receiver operating characteristics curves, have been summarized.

Measures of Diagnostic Accuracy

The third step in the five-step model of the Evidence-Based Medicine (EBM) approach (*1. Ask an answerable clinical question, 2. Search the literature, 3. Analyze and critically appraise the literature for the best evidence, 4. Make a decision by integrating the evidence with the clinical expertise and the patient's*

Address for correspondence:

Prof. Dr. Diler Aslan
Pamukkale University School of Medicine
Department of Biochemistry, Denizli, Turkey

value, *5. evaluate the performance*), is »critically appraising the evidence for its validity and relevance«. The evidence is appraised according to the type of question created (aetiology, diagnosis, prognosis, therapy, cost-effectiveness, and quality of life) in the first step. If the question is how to diagnose a disease, the articles on the diagnostic accuracy studies will be searched for and critically appraised (1–4).

Appraising can be made by asking three questions: 1) Is the study valid? 2) Are the results important? 3) Can the results be applied in the patient care? (2).

The second and third questions can be answered by evaluation of the diagnostic performance of the test.

The diagnostic performance of a test can be evaluated in several ways by using the measures of diagnostic accuracy. The measures of diagnostic accuracy are sensitivity and specificity, likelihood ratios, diagnostic odds ratio and receiver operating characteristic (ROC) curves.

Sensitivity and specificity values of a test are used in the calculations (LR's, DOR) and construction (ROC Curve).

Sensitivity and Specificity

Sensitivity (Sn) of a test is its capability to make a correct (positive) diagnosis in patients who have the disease in question. Thus it is called true positive (TP) rate, or probability of true positive results in patients who have the disease [Pr (T^+/D^+]. Specificity (Sp) of a test is its capability to make correct (negative) diagnosis in patients who do not have the disease. Thus it is called true negative (TN) rate, or probability of true negatives in patients who do not have the disease [Pr (T^-/D^-)].

Sn and Sp of a diagnostic test are determined by administering the test to two groups: 1) a group of patients known to have the disease or condition, 2) a group of patients known not to have the disease or condition according to the reference test (or gold standard).

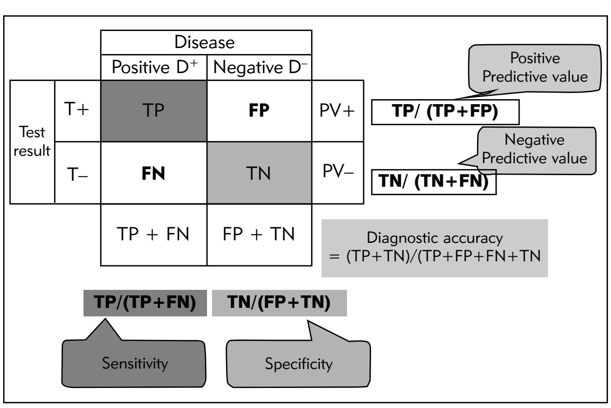

Figure 1 The 2 × 2 table method.

The Sn is calculated as the proportion (or percentage) of patients known to have the disease whose test is positive; Sp is the proportion (or percentage) of patients known to be free of the disease whose test is negative. The Sn and Sp can be calculated by using the 2 × 2 table method (*Figure 1*).

Using Sensitivity and Specificity (SnNout, SpPin)

If a test has high Sn or high Sp, they can be helpful for ruling out or ruling in the diagnosis, respectively (2).

High sensitivity expresses that the test is unlikely to produce false negative results. Then, if the test is negative, it rules out the diagnosis. This can be remembered by the mnemonic **SnNout**.

High specificity expresses that the test is unlikely to produce false positive results. Then, if the test is positive, it rules in the diagnosis. This can be remembered by the mnemonic **SpPin**. Sensitivity and specificity are in principle independent of the prevalence of the disease, but that is not always so. In a low prevalence situation, the sensitivity will usually be lower than in a high prevalence situation, whereas for the specificity the opposite is true.

Revising Prior Probability

Predictive values of the test

The predictive value of a positive test (PV+ or PPV) expresses the probability that a patient has the disease, given that the test result is positive. The negative predictive value of a test (PV- or NPV) expresses the probability that a patient does not have the disease, given that the test result is negative. PPV or post-test probability and NPV can be calculated by using the 2 × 2 Table Method (*Figure 1*). They are conditional probabilities, and are affected by the prevalence of disease in the patient population being studied (Equation 1 and 2).

$$\text{PPV} = \frac{(\text{sensitivity})\ (\text{prevalence})}{(\text{sensitivity})\ (\text{prevalence}) + (1 - \text{specificity})\ (1 - \text{prevalence})} \quad (1)$$

$$\text{NPV} = \frac{(\text{specificity})\ (1 - \text{prevalence})}{(\text{specificity})\ (1 - \text{prevalence}) + (1 - \text{sensitivity})\ (\text{prevalence})} \quad (2)$$

Likelihood Ratios

Likelihood ratios state how many times more likely particular test results are in patients with disease than in those without the disease. Thus, a positive test has one likelihood ratio and a negative test another. They are calculated from sensitivity and specificity of the test (1–5):

$$\text{Likelihood Ratio for a positive test (LR+)} = \frac{\text{odds of a positive test with the disease}}{\text{odds of a positive test without the disease}} \quad (3)$$

hence

$$\text{LR+} = \frac{\text{True positive rate}}{\text{False positive rate}} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (4)$$

hence

$$\text{Likelihood Ratio for a negative test (LR–)} = \frac{\text{odds of a negative test with the disease}}{\text{odds of a negative test without the disease}} \quad (5)$$

$$\text{LR–} = \frac{\text{True negative rate}}{\text{False negative rate}} = \frac{\text{Specificity}}{1 - \text{Sensitivity}} \quad (6)$$

Likelihood ratio is an alternative method for calculating post-test probabilities, and work with the prior odds rather than prior probabilities. Likelihood ratios allow computations of pretest probabilities using Bayes' Theorem and the Equation 8.

$$\text{Posterior Probability} = \frac{\text{Posttest odds}}{1 + \text{Posttest odds}} \quad (7)$$

$$\text{Posttest odds} = \text{Pretest odds} \times \text{LR} \quad (8)$$

$$\text{Pretest probability} = \frac{\text{Prior probability}}{1 - \text{prior probability}} \quad (9)$$

By the use of Nomogram (*Figure 2*) established by Fagan (1975), the posterior probability can be obtained without performing extra calculation.

As seen, there are several advantages of LRs compared to sensitivity and specificity. LRs give direct information about the power of a test. A LR of 1 has no influence on the pretest probability. Test results with LRs greater than 10 or less than 0.1 are generally clinically useful, but again the usefulness will in some degree depend on the pretest probability (1–5).

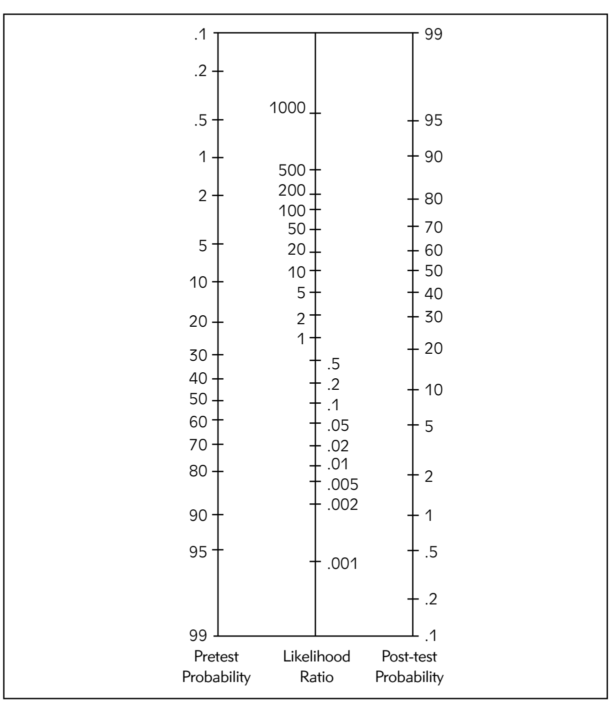

Figure 2 Nomogram.

Diagnostic Odds Ratio

Diagnostic Odds Ratio (DOR) expresses the odds of the positive test results in patients with the disease compared with the odds of the positive test results in those without the disease. With the combination of the equations 3 and 5, the equation is:

$$\text{DOR} = \frac{\text{LR+}}{\text{LR–}} \quad (10)$$

The diagnostic odds ratio summarizes the diagnostic accuracy of a test as a single value.

DORs cannot be translated into useful measures in the interpretation of a test result for an individual patient.

Receiver Operating Characteristic (ROC) Curve

When the test results are measured on a continuum (ratio scale), sensitivity and specificity depend upon where the cutoff is set between »negative« and »positive« results. ROC Curves are used to show the pattern of sensitivities and specificities observed when the performance of a diagnostic test is evaluated at different cutoff points. Thus, the threshold giving the highest sensitivity and specificity can be determined.

ROC Curves were developed in the communications field as a way to display signal-to-noise ratios. The true positives can be thought to be the correct signals from a diagnostic test and false positives to be noise. The ROC curve is a plot of Sn (or true positive rate) versus the (1-Sp) (or false positive rate) (*Figure 3*).

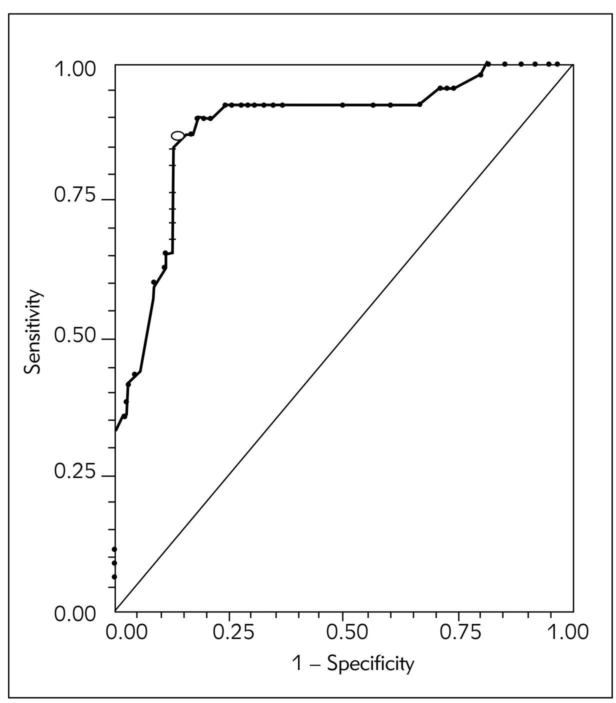

Figure 3 Receiver operating characteristic curve.

The closer a ROC curve is to the upper left-hand corner of the graph, the better is the test, because the true positive rate is 1 and the false positive rate is 0.

The ROC curves are a useful graphical method for comparing two or more diagnostic tests. The significance of the difference between the diagnostic accuracy of two tests can be determined by calculating the area under each ROC Curve.

Equations for Confidence Intervals

Confidence Intervals (CI's) provide useful information about estimates. Studies of diagnostic tests should report both point estimates and CI's (1–6).

Diagnostic accuracy is usually given as proportions and 95% Confidence limits for proportions is given by observed proportion ±1.96 × standard error of the observed proportion

or

$$p \pm 1.96 \times \sqrt{\frac{p\,(1-p)}{n}} \qquad (11)$$

P = proportion (such as Sn, Sp)

Table I Equations for 95% Confidence limits.

Measure of Diagnostic Accuracy	95% Confidence Limits
Sensitivity	$Sn \pm 1.96 \times \sqrt{\frac{Sn(1-Sn)}{TP+FN}}$ (12)
Specificity	$Sp \pm 1.96 \times \sqrt{\frac{Sp(1-Sp)}{FP+TN}}$ (13)
PPV	$PPV \pm 1.96\sqrt{Var\,(PPV)^{*}}$ (14)
NPV	$NPV \pm 1.96\sqrt{Var\,(NPV)^{**}}$ (15)
OR	$\exp\left[\ln(OR) \pm 1.96\sqrt{\frac{1}{TP}+\frac{1}{FP}+\frac{1}{FN}+\frac{1}{TN}}\right]$ (16)
LR+	$\left[\left(e^{\pm 1.96\sqrt{\frac{(1-Sens)}{(a+c)Sens}+\frac{Spec}{(b+d)(1-Spec)}}}\right)LR+ve\right]$ (17)
LR–	$\left[\left(e^{\pm 1.96\sqrt{\frac{Sens}{(a+c)(1-Sens)}+\frac{(1-Spec)}{(b+d)Spec}}}\right)LR-ve\right]$ (18)

$$^{*}Var(PPV) = \frac{[p.(1-Sp).(1-p)]^2.\frac{Sn.(1-Sn)}{TP+FN}\ [p\cdot Sn.1-Sp)\cdot(1-p)]^2\cdot\frac{Sp.(1-Sp)}{FP+TN}}{[Sn.p+(1-Sp).(1-p)]^4}$$

$$^{**}Var(NPV) = \frac{[Sp.(1-p).p]^2.\ \frac{Sn.(1-Sn)}{TP+FN_1}\ 1-\cdot Sn)\cdot(1-p)\cdot p]^2\cdot\ \frac{Sp.(1-Sp)}{FP+TN}}{[(1-Sp).p+Sp.(1-p)]^4}$$

Specific and Global, Unconditional and Conditional Measures of Diagnostic Accuracy

Measures of diagnostic accuracy or test performance can be classified into four groups: 1) Global, unconditional, 2) Global, conditional, 3) Specific, unconditional, 4) Specific, conditional (1).

The class of a measure of diagnostic accuracy should be considered when the diagnostic accuracy of a test is assessed, because the correct decision depends upon whether it is global or specific and unconditional or conditional.

Global measure is the overall test performance in a single figure (DOR, ROC Curve, and accuracy). Specific measures are influenced by the choice of thresholds that separate test positives from negatives (Sn, Sp, LR+, LR–, PPV, and NPV). The measures are also classified according to the influence of the prevalence of the disease. Those under the influence of the prevalence are the conditional measures (PPV, NPV, Accuracy), and those that in principle are not under the influence are unconditional measures (Sn, Sp, LR+, LR–, DOR and ROC curve). Remember, however, that in some situations prevalence will also influence these measures.

Table II The types of measures of the diagnostic accuracy of a test.

Measure	Global or specific	Unconditional or conditional
Sensitivity Specificity LR+ LR–	Specific	Unconditional
DOR ROC Curve	Global	
PPV NPV	Specific	Conditional
Accuracy or efficiency	Global	

Example Study

Castello-Boerrigter LC, et al. Amino-Terminal Pro-B-Type Natriuretic peptide and B-Type Natriuretic Peptide in the General Community: Determination and Detection of Left Ventricular Dysfunction, JACC, 2007.

1. Calculate sensitivity and specificity for NT-proBNP and BNP Assays, fill in the appropriate cell.

		LVD EF≤40		
		D+	D–	
NT-proBNP	≥228	32	256	
	<228	5	1576	
		37	1832	N = 1869

		LVD EF≤40		
		D+	D–	
BNP	≥66	30	346	
	<66	7	1486	
		37	1832	N = 1869

2. Calculate the values measures of diagnostic accuracy shown in the table by using the equations above, and fill the blank cells.

NT-proBNP Population n= 1869		BNP Population n= 1869	
LVD EF≤40 n=37		LVD EF≤40 n=37	
Cutpoint (pg/mL)	228	Cutpoint (pg/mL)	66
Sn (%)		Sn (%)	
Sp (%)		Sp (%)	
PPV (%)		PPV (%)	
NPV (%)		NPV (%)	
Pretest Pr (%)		Pretest Pr (%)	
LR+		LR+	
LR-		LR-	
DOR		DOR	
Pretest odds		Pretest odds	
For +ve results		**For +ve results**	
Posttest odds		Posttest odds	
Postest Pr (%)		Postest Pr (%)	
For -ve results		**For -ve results**	
Posttest odds		Posttest odds	
Posttest Pr (%)		Posttest Pr (%)	

References

1. Battaglia M, Bucher H, Egger M, Grossenbacher F, Minder C, Pewsner D. The Bayes Library of Diagnostic Studies and Reviews, Berne, February. 2002 (http://www.medepi.net/meta/guidelines/BAYES_Library.pdf).
2. Heneghan C, Badenoch D. Evidence-Based Medicine Toolkit, BMJ Books; 2Rev Ed edition, London, UK. 2006.
3. Price CP, Christenson RH. Evidence-based Laboratory Medicine From Principles to Outcomes, AACCPress, Washington DC. USA. 2003.
4. Dawson B, Trapp RC. Basic & Clinical Biostatistics, McGraw-Hill Comp. NY, USA. 2001.
5. Jones R, Payne B. Clinical Investigation and Statistics in Laboratory Medicine, ACB Venture Publications, Association of Clinical Biochemists London, UK. 1997.
6. Mercaldo ND, Zhou XH, and Lau KF. »Confidence Intervals for Predictive Values Using Data from a Case Control Study« UW Biostatistics Working Paper Series. Working Paper 271. 2005 (http://www.bepress.com/uwbiostat/ paper271).

Received: August 15, 2007
Accepted: September 10, 2007