International Journal of Cardiovascular Sciences. 2016;29(3):218-222

REVIEW ARTICLE

Diagnostic Accuracy Measures in Cardiovascular Research

Leonardo Silva Roever Borges

Universidade Federal de Uberlândia, Uberlândia, MG – Brazil

Abstract

Diagnostic accuracy is the ability of a test to discriminate between the target condition and health; it can be quantitated using diagnostic accuracy measures such as sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios, the area under the ROC curve, odds ratio of the diagnosis and Youden index. The diagnostic accuracy measures are related to the different aspects of the diagnostic procedure. Some measures are used to evaluate the discriminative property of the test, while others are used to assess its predictive ability. Diagnostic accuracy measures are not fixed indicators of a test performance; some are very sensitive to disease prevalence, while others are sensitive to disease spectrum and definition. This review described the definitions and characteristics of diagnostic accuracy measures used in cardiovascular research.

Introduction

The accuracy of a diagnostic test shows how this test correctly discriminates two conditions of interest: health and disease. This discriminative capacity can be quantified through diagnostic accuracy measures: sensitivity and specificity, Positive and Negative predictive values (PPV and NPV), Positive and Negative Likelihood Ratio (PLR and NLR), area under the ROC curve (AUC), the Youden index and the diagnostic odds ratio (DOR).[1-14]

Keywords

Diagnosis; Data Accuracy; Evidence-Based Medicine; Clinical Study.

Sensitivity and specificity

A perfect diagnostic test has the potential to adequately discriminate between patients with and without disease. Unfortunately, diagnostic procedures can only make a partial distinction between individuals with present or absent disease. The values of a diagnostic test that are greater than or equal to the cutoff indicate the presence of disease, while values below the cutoff rule out the disease.[15]

The values above the cutoff are not always indicative of a disease, as healthy individuals may also sometimes have higher values. Such elevated values of a given parameter of interest are called false-positive (FP) values. On the other hand, values below the cutoff are mainly found in individuals without the disease, but some individuals with the disease may also show them, being called false-negative (FN) values.[15]

The cutoff value divides the population of assessed individuals with and without the disease into four subgroups, considering the values of the parameters of interest:

- A – True positive (TP): patients with the disease and the value of a parameter of interest greater than or equal to the cutoff.
- B – FP: patients without the disease and the value of a parameter of interest greater than or equal to the cutoff.
- C – True-negative (TN): patients without the disease and the value of a parameter of interest below the cutoff.
- D – FN: patients with the disease and the value of a parameter of interest below the cutoff.

The method for performing the calculation of diagnostic accuracy is performed using a 2x2 table, with groups of individuals divided according to the gold

Mailing Address: Leonardo Silva Roever Borges
Rua Rafael Rinaldi, 431, Martins. Postal Code: 38400-384. Uberlândia, MG – Brazil
E-mail: leonardoroever@hotmail.com

DOI: 10.5935/2359-4802.20160030

Manuscript received June 22, 2016; revised manuscript June 27, 2016; accepted July 10, 2016.

standard or reference method in columns and categories, according to the test result (Table 1).[15]

Sensitivity is expressed in percentage and defined as the probability of obtaining a positive result in patients with the disease (TP / TP + FN). Specificity is defined as the probability of obtaining a negative result in individuals without the disease (TN / TN + FP).[15]

Neither sensitivity nor specificity is influenced by the disease prevalence. This means that the results of a study can be easily transferred to another environment with different prevalence of the disease in the population. However, sensitivity and specificity may vary widely depending on the disease spectrum in the studied group.[15]

SnNOut is used to indicate when a sign/test or symptom = has a negative (N) result in a highly sensitive test (Sn), which excludes the diagnosis (Out). SpPIn indicates when a sign/test or symptom has a positive (P) result in a highly specific test (Sp), confirming the diagnosis (In).[11,12]

The pretest odds are the chance of an individual having the target disease before the test is performed. The post-test odds are the chance that a patient has the target disease after the test is performed. The pretest probability (prevalence) is the probability that an individual has the target disease before the test is performed and the post-test probability is the probability that an individual with a specific test result has the target disease.[11,12]

Predictive values

The PPV defines the probability of having the disease of interest in an individual with a positive result.

Table 1
Diagnostic accuracy measures

		Target disorder		Total
		Present	Absent	
Diagnostic test result	Positive	A(TP)	B(FP)	a+b
	Negative	C(FN)	D(TN)	c+d
Total		a+c	b+d	a+b+c+d

Sensitivity= a/(a+c)
Specificity = d/(b+d)
PLR = sensitivity/(1-specificity)
NLR = (1-sensitivity)/specificity
Positive predictive value = a/(a+b)
Negative predictive value = d/(c+d)
Pretest probability (prevalence) = (a+c)/(a+b+c+d)
Odds = probability/(1- probability)
Pretest Odds = prevalence/(1-prevalência)
Post-test Odds = pretest odds x LR
Probability = odds/ (odds + 1)
Post-test probability = post-test odds /(post-test odds +1)
Accuracy = (PV + NV)/(PV + FP + NV + FN)
TPF - total of positive tests in individuals with the disease
FPF - total of positive tests in individuals without the disease
Diagnostic Odds Ratio: PLR/NLR = (TP/FN)/(FP/TN)
DE = (TP + TN/ (TP + TN + FP + FN)
Youden index: (sensitivity + specificity) - 1

TP: true-positive; FP: false positive; FN: false-negative; TN: true-negative; PLR: positive likelihood ratio; NLR: negative likelihood ratio; TPF: true-positive fraction; FPF: false-positive fraction; DE: diagnostic efficacy.

The PPV represents a proportion of patients with a positive test result in a total of individuals with positive results (TP / TP + FP).[15,16]

The NPV describes the probability of not having a disease in an individual with a negative result. The NPV is defined as a proportion of individuals without the disease with a negative result in a total of individuals with negative results (TN / TN + FN).[15,16]

Unlike sensitivity and specificity, PPV and NPV are largely dependent on the prevalence of disease in the studied population. Therefore, the predictive value of a study should not be transferred to another population with different disease prevalence. Prevalence affects PPV and NPV differently. The PPV is increases, while the NPV decreases with the increase in disease prevalence in the population.[15,16]

Likelihood ratio

LR is a very useful measure in diagnostic accuracy, being defined as the ratio of the expected test result in individuals with a certain disease to individuals without the disease. Simply put, the LR says how much more likely it is to have the particular test result in individuals with the disease than in those without it. When both likelihoods are equal, such test is of no value and its LR = 1. The LR for tests with positive results (PLR) indicate how much more likely a test with a positive result will occur in individuals with the disease, compared to those without the disease (PLR = sensitivity / (1-specificity). PLR is generally greater than 1, because it is more likely that a positive test result will occur in individuals with the disease than in subjects without the disease. The greater the PLR, the more the test is indicative of disease. Good diagnostic tests have a PLR > 10 and its positive result brings a significant contribution to diagnosis.

A NLR is the ratio of the likelihood of a negative result occurs in individuals with the disease to the likelihood that the same result occurs in individuals without the disease (LR = (1-sensitivity) / specificity). The NLR shows how much less likely a negative test result will occur in a patient than in an individual without the disease.

The NLR is generally less than 1, because it is less likely that the negative test result will occur in individuals with than in individuals without disease. Good diagnostic tests have NLR < 0.1.

Both sensitivity and specificity are used to calculate the likelihood rate and it is evident that neither PLR nor NLR depend on disease prevalence in the assessed groups.

Therefore, the likelihood ratios from a study are applicable to any other clinical setting, as long as disease definition is not changed. If the disease definition varies, none of the measured and calculated options should be applied to another clinical context.

The ROC curve

There is a pair of diagnostic sensitivity and specificity for each of the individual's cutoffs. To construct a Receiver Operating Characteristic (ROC) Curve chart, one uses these pairs of values in the chart as 1-specificity on the x axis and sensitivity on the y axis (Figure 1).[15,17]

The shape of a ROC curve and the AUC helps to verify the discriminative power of a test. The closer the curve is located to the top left corner and the larger the ROC AUC, the better the test is to discriminate between patients and non-patients. The AUC may have any value between zero and 1 and that is a test quality indicator. A perfect diagnostic test has an AUC = 1.[17]

We consider a non-discriminating test when it has an AUC of 0.5. Usually it can be said that the ratio between the AUC and the diagnostic accuracy applies as described in Table 2.[7]

AUC is a global measure of diagnostic accuracy. It discloses nothing about the patient's parameters, such as sensitivity and specificity. By comparing the two areas under the ROC curves, one can estimate which one of the two tests is more adequate to differentiate health from disease or any other two conditions of interest.[17]

Diagnostic odds ratio

The diagnostic odds ratio (DOR) is also a global measure of diagnostic accuracy, used to generate the estimate the discrimination power of diagnostic procedures and to compare diagnostic accuracies between two or more diagnostic tests. The DOR of a test is the ratio of the odds of positivity in patients with the disease, compared to the odds in individuals without the disease [DOR = (TP/FN) / (FP/TN)].[18]

The DOR depends significantly on the sensitivity and specificity of a test. In a test with high specificity and low sensitivity, the FP and FN rates have a high DOR.

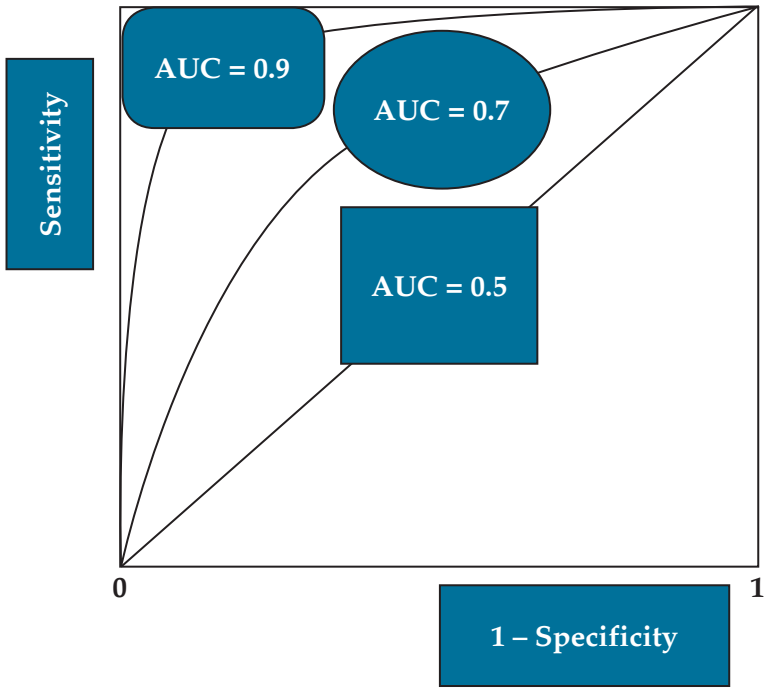

Figure 1
ROC curve depiction.

Table 2
Association between the area under the ROC curve and diagnostic accuracy

Area	Diagnostic accuracy
0.9 - 1.0	Excellent
0.8 - 0.9	Very good
0.7 - 0.8	Good
0.6 - 0.7	Fair
0.5 - 0.6	Poor
< 0.5	Do not use the test

With the same test sensitivity, the DOR increases with increasing test specificity.[18]

Diagnostic efficacy

The diagnostic efficacy (ED) is a global measure of diagnostic accuracy, expressed as a proportion of individuals correctly classified among all [DE = (TP + TN / (TP + TN + FP + FN)]. It is affected by the prevalence of the disease. With the same sensitivity and specificity, the diagnostic accuracy of a certain disease increases as the prevalence of the disease decreases.[15]

Youden index

The Youden index is one of the oldest diagnostic accuracy measures. This is a test performance measure. It is used to assess the overall discriminating power of a diagnostic procedure and to compare that test with other tests. The Youden index is calculated by subtracting

1 from the sum of sensitivity and specificity of the test and it is not expressed as a percentage, but as part of a whole number: (sensitivity + specificity) – 1.[19]

For a test with poor diagnostic accuracy, the Youden index is zero, whereas in a perfect test, the Youden index is equal to 1. The Youden index is not sensitive to differences in test sensitivity and specificity, which is its main disadvantage. The Youden index is not affected by the prevalence of the disease, but by disease spectrum, as well as the specificity of sensitivity, LR and DOR. [19]

Conclusion

Decision-making in cardiovascular practice is often based on complex, however, incomplete evidence. The accuracy measures of a diagnostic test represent a tool to improve cardiovascular decision-making and patient care.

Author contributions

Conception and design of the research:Borges LSR. Acquisition of data:Borges LSR. Analysis and interpretation of the data: Borges LSR. Writing of the manuscript:Borges LSR. Critical revision of the manuscript for intellectual content: Borges LSR.

Potential Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Sources of Funding

There were no external funding sources for this study.

Study Association

This study is not associated with any thesis or dissertation work.

References

1. Sackett DL, Haynes RB. The architecture of diagnostic research. BMJ. 2002;324(7336):539-41.
2. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ. 2002;324(7335):477-80.
3. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. BMJ 2002;324(7338):669-71.
4. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ. 2002;324(7339):729-32. Erratum in: BMJ. 2006;333(7575):944.
5. Winkens R, Dinant GJ. Evidence base of clinical diagnosis: rational, cost effective use of investigations in clinical practice. BMJ. 2002;324(7340):783.
6. Irwing L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120(8):667-76.
7. Bruns DE. The STARD initiative and the reporting of studies of diagnostic accuracy. Clin Chem. 2003;49(1):19-20.
8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem. 2003;49(1):7-18.
9. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. 2003;49(1):1-6.
10. Rennie D. Improving reports of studies of diagnostic tests: the STARD initiative. JAMA. 2003;289(1):89-90.
11. Guyatt G, Rennie D, Meade MO, Cook DJ. (eds.). Users' guide to the medical literature: a manual for evidence-based clinical practice. 2nd ed. New York: McGraw Hill Medical; JAMA evidence; 2008.
12. Guyatt G, Rennie D, Meade MO, Cook DJ. (eds.). Users' guide to the medical literature: a manual for evidence-based clinical practice. 3rd ed. New York: McGraw Hill Medical; JAMA evidence; 2014.
13. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. BMJ. 2002;324(7338):669-71.
14. Raslich MA. Markert RJ, Stutes SA. Selecting and interpreting diagnostic tests. Biochemia Medica. 2007;17(2):151-60.
15. Eusebi P. Diagnostic accuracy measures. Cerebrovasc Dis. 2013;36(4):267-72.
16. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ. 2004;329(7458):168-9.
17. Obuchowski NA, Lieber ML, Wians FH Jr. ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin Chem. 2004;50(7):1118-25.
18. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003;56(11):1129-35.
19. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-35.