Evaluation and application of diagnostic accuracy in clinical decision-making

Prabhaker Mishra[1], Chandra Mani Pandey[1], Uttam Singh[1], Sachin Singh Yadav[2], Vishal Sharma[3]

[1]Department of Biostatistics & Health Informatics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India.
[2]Department of Community Medicine, MM Institute of Medical Sciences and Research, Mullana, Ambala, Haryana, India.
[3]Department of Community Medicine, Army College of Medical Sciences, New Delhi, India.
Correspondence to: Prabhaker Mishra, E-mail: mishrapk79@gmail.com

Received February 09, 2016. Accepted February 25, 2016

Abstract

A diagnostic test is any kind of medical test performed to determine the presence or absence of a disease when a subject shows signs or symptoms of the disease. Measures of diagnostic accuracy tell us about the ability of a test to discriminate between disease and health. There are many measures used to show the accuracy of the diagnostic test including sensitivity, specificity, etc. We calculate these measures in 2×2 table, comparing with gold standard and it is possible only when result of the gold standard as well as new test is dichotomous nature. Many diagnostic test results have continuous scale and for such variable a series of cutoff points for disease and corresponding measures of accuracy can be calculated using receiver operating characteristics curve. In this article, we have discussed the various measures of diagnostic accuracy and their computational methods commonly used to assess the performance of a diagnostic test.

KEY WORDS: Sensitivity, specificity, overall accuracy, ROC curve, area under the curve

Introduction

A diagnostic test is used to determine the presence or absence of a disease when a subject shows signs or symptoms of the disease. A screening test identifies asymptomatic individuals who may have the disease. The diagnostic test is performed after a positive screening test to establish a definitive diagnosis.[1] Example there are many diagnostic test used to diagnose the disease: pap smear for cervical dysplasia or cervical cancer, fasting blood cholesterol for heart disease, fasting blood sugar for diabetes, etc.[1] Diagnostic accuracy relates to the ability of a test to discriminate between the disease and non-disease. This discriminative potential can be quantified by the measures of diagnostic accuracy such as true positive rate (sensitivity), true negative rate (specificity), positive predictive value (PPV), negative predictive value (NPV), likelihood ratio positive (LR+), likelihood ratio negative (LR−), area under the ROC curve, Youden's index, and diagnostic odds ratio (DOR). Different measures of diagnostic accuracy relate to the different aspects of diagnostic procedure, example some measures are used to assess the discrimination property of the test, others are used to assess its predictive ability.[2] In this article, we have discussed the measures of accuracy of diagnostic test as well as concept and application of ROC curve.

Access this article online	
Website: http://www.ijmsph.com	Quick Response Code:
DOI: 10.5455/ijmsph.2016.09022016401	

Measures of Diagnostic Accuracy Used

In the medical sciences, there are many measures used to calculate indices related to diagnostic accuracy are: sensitivity, specificity, false negative rate, false positive rate, overall accuracy, PPV, NPV, LR+, and LR−. Above discussed measures can be computed in 2×2 table by comparing gold standard. In some situation, our test variable is in continuous scale and computation by 2×2 table is not possible. To overcome of this problem, we used ROC curve analysis to find out a series of cutoff value of this continuous variable with corresponding measures of diagnostic accuracy.[3] Youden's index is a measure of diagnostic accuracy used to calculate using sensitivity and specificity of the test. DOR is also an important measure of a test to find out the chances of getting correct diagnosis compared to wrong diagnosis.

Discussion about Measures of Diagnostic Accuracy

In this article, we have discussed the measures of accuracy of the diagnostic test, with an hypothetical example: In a study of 100 suspected persons of tuberculosis (TB), 40 persons have been diagnosed as TB patients whereas rest 60 were identified as free of TB disease (healthy) using "diagnostic test A." Same persons were again investigated using another "diagnostic test B" which predicted that of 40 TB patients, only 30 have TB disease and in 60 TB free persons, only 40 were free from TB. If "diagnostic test A" is a gold standard, then accuracy of the diagnostic test B could be calculated using 2×2 table given as follows [Table 1].

Gold Standard Test

The best available method for establishing the presence or absence of the target condition (disease). It is often invasive or expensive.[3] Using 2×2 tables, we compare the performance of the new test (usually less expensive) with the gold standard test. In the Table 1, "Test A" is gold standard whereas "Test B" is the new test that accuracy has to be evaluated.

Sensitivity (True Positive Rate)

The ability of a test to identify disease in a person, when person has disease.[4] In other words, it is proportion of correctly identified disease cases out of true disease cases. In the Table 1, diagnostic test A (gold standard) identified 40 disease cases (a+c) whereas of these disease cases, test B could be correctly identified as disease were 30 (a). So sensitivity of test B is a/(a+c) = 30/40 = 0.75.

False Negative Rate

The lack of ability of a test to identify disease in a person, when person has disease. In other words, it is proportion of missed disease cases, out of true disease cases. In the Table 1, diagnostic test A (gold standard) identified 40 disease cases (a+c) whereas out of these disease cases, test B identified 30 disease cases and missed disease cases were 10(c). So false negative rate of test B is c/(a+c) = 10/40 = 0.25.

Ideally for a good diagnostic test, false negative should be zero. For any diagnostic test, sum of sensitivity and false negative rate is always equal to 1.

Specificity (True Negative Rate)

The ability of a test to correctly identify a person as disease free, when person has no disease.[4] In other words, it is proportion of correctly identified disease free persons out of true disease free persons. In the Table 1, diagnostic test A (gold standard) identified 60 disease free persons (b+d) whereas out of these disease free persons, test B, correctly identified as disease free people were 40 (d). So specificity of test B is d/(b+d) = 40/60 = 0.67.

False Positive Rate

The lack of ability of a test to identify non-disease person, when person does not has disease. In other words, it is proportion of incorrectly identified a person has disease, when that person does not have disease. In the Table 1, diagnostic test A (gold standard) identified 60 disease free persons (b+d) whereas out of these disease free persons, test B correctly identified as disease free were 40 and rest 20 (b) persons were wrongly diagnosed as disease. So false positive rate of test B is b/(b+d) = 20/60 = 0.33.

Ideally for a good diagnostic test, false positive should be zero. For any diagnostic test, sum of specificity and false positive rate is always equal to 1.

Positive Predictive Value (PPV)

The proportion of correctly identify disease persons from out of total disease persons predicted by the test itself. In the Table 1, diagnostic test B predicted 50 disease cases (a+b) whereas out of these prediction, correct prediction were 30(a). So PPV of test B is a/(a+b) = 30/50 = 0.60.

Negative Predictive Value (NPV)

The proportion of correctly identify disease free persons from out of total disease free persons predicted by the test itself. In the Table 1, diagnostic test B predicted 50 disease free persons (c+d) while out of these prediction, correct prediction was 40(d). So NPV of test B is d/(c+d) = 40/50 = 0.80.

Ideally for a good diagnostic test, each of PPV and NPV should be 1 (or 100%) and it is possible only when false positive rate and false negative rate are zero.

In a situation, when 2×2 table is not given and only sensitivity, specificity, false positive rate, false negative rate and prevalence of disease are known. We can compute PPV and NPV of the test using following formula:

$$\text{PPV} = \frac{(\text{Sensitivity} \times \text{p})}{(\text{Sensitivity} \times p) + (\text{False positive} \times q)}$$

$$\text{NPV} = \frac{(\text{Specificity} \times q)}{(\text{False negative} \times p) + (\text{Specificity} \times q)}$$

where p = prevalence, $q = 1-p$.

Overall Accuracy

Diagnostic accuracy (effectiveness) of a new test is a proportion of sum of the correctly identified disease cases and disease free persons with total number of persons available for screening.[5] In the Table 1, correct predictions by test B were 30(a) and 40(d) among all subjects 100(n). So overall accuracy of test B is (a+d)/n = (30+40)/100 = 70/100 = 0.70

Ideally for a good diagnostic test, overall accuracy should be 1 (or 100%) and it is possible only when false negative and false positive are zero.

Table 1: Showing comparative prediction of disease (present/absent) by two diagnostic tests

		Diagnostic test "A"		
		Present	Absent	Total
Diagnostic test "B"	Present	30 (a)	20 (b)	50 (a+b)
	Absent	10 (c)	40 (d)	50 (c+d)
	Total	40 (a+c)	60 (b+d)	100 (n)

n = a+b+c+d, "Test A" is gold standard, "Test B" accuracy to be measured

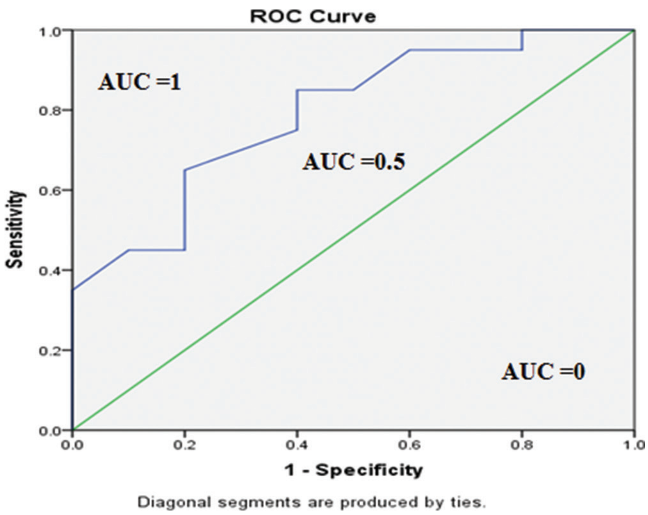

Figure 1: ROC curve plot between false positive rate and sensitivity at different cutoff value of a continuous test variable.

Likelihood Ratio Positive (LR+)

Probability of an individual with the disease having the positive test result (sensitivity) w.r.t. probability of an individual without the disease having the positive test result (false positive). In other words, likelihood ratio positive (LR+) is ratio of sensitivity with false positive rate. In the Table 1, LR+ is sensitivity/false positive rate = 0.75/0.33 = 2.27.

Likelihood Ratio Negative (LR–)

Probability of an individual with the disease having a negative test result (false negative) w.r.t. probability of an individual without the disease having a negative test result (specificity). In other words, likelihood ratio negative (LR–) is ratio of false negative rate with specificity. In the Table 1, LR– is false negative rate/specificity = 0.25/0.67 = 0.37.

For a good diagnostic test, LR+ should be >1 and LR– should be <1. The higher LR+, the test is more indicative of a sensitive for disease and their positive result has a significant contribution to the diagnosis. The lower LR–, test is more significant contribution of specificity of the test.

Youden's Index

Youden's index is one of the measures for diagnostic accuracy. It is also a global measure of a test performance, used for the evaluation of overall discriminative power of a diagnostic procedure and for comparison of this test with other tests.[6] Youden's index is calculated using given formula.

Youden index = [(Sensitivity + Specificity) – 1].

where sensitivity and specificity expressed in proportion (not in percentage). For a test with poor diagnostic accuracy, Youden's index is equals 0 and in a perfect test, Youden's index is equal to 1. Youden's index is not sensitive for differences in the sensitivity and specificity of the test, which is its main disadvantage. Example, a test with sensitivity 0.8 and specificity 0.6 has the same Youden's index (0.4) as a test with sensitivity 0.7 and specificity 0.7.

Diagnostic odds ratio (DOR)

DOR is also one measure for diagnostic accuracy, used for estimation of discriminative power of a diagnostic test and also for the comparison of diagnostic accuracy between two or more diagnostic tests. DOR of a test is the ratio of the two odds, that is, odd of the disease in subjects in positive group (odd of the true positive with false positive) w.r.t. Odd of the disease in subjects in negative group (odd of the false negative with true negative).[7] In the Table 1,

Odd of the true positive with false positive = a/b = 30/20
Odd of the false negative with true negative = c/d = 10/40
DOR: (a/b) / (c/d) = ad/bc = 30×40/20×10 = 6
For a useful test, DOR should be >1.

Receiver Operating Characteristic (ROC) curve

ROC curves are used in medicine to determine all possible cutoff value for a clinical variable (continuous scale).[8] To construct a ROC curve, minimum required data are continuous variable for which cutoff value to be obtained and corresponding binary outcome of the disease (like present/absent). Through the ROC curve analysis, we could get all possible cutoff points of the continuous scale variable with corresponding diagnostic accuracy. In ROC curve [Figure 1], we plot a graph between false positive rate (1-specificity) on the *x*-axis and sensitivity on the *y*-axis.[9] The shape of a ROC curve and the area under the curve (AUC) helps us to estimate the discriminative power of a test. The AUC can have any value between 0 and 1 and it is a good indicator of the goodness of the test. The larger AUC is located to upper-left hand corner where sensitivity = 1, specificity = 1, and test showing the best discriminating power between diseased and non-diseased. If test sensitivity is equally proportion to false positive, test is called useless (i.e. sensitivity = false positive rate), that is curve is at diagonal line. Test is called just opposite discriminating power between disease/non-disease, when ROC curve is at right bottom corner (false negative rate = 1, false positive rate = 1). A perfect diagnostic test has an AUC = 1.0 and just opposite discrimination, AUC = 0, whereas a non-discrimination test has an AUC = 0.5.

Conclusion

Diagnostic accuracy is a method to evaluate as well as to check validity of a diagnostic test. It is important part of the medicine where we check the presence/absence of a particular disease before providing its treatment. Through a correct diagnosis, we can ensure that all the patients would be identified as disease while healthy persons would be identified as non-disease. In case of wrong diagnosis, results are to missing disease as well as falsely diagnosis of disease. To correctly identifying the disease, Sensitivity, LR+ and PPV should be maximum and false negative rate minimum. To correctly identifying the non-disease from non-disease persons, specificity and NPV should be maximum whereas LR– and false positive rate minimum. ROC curve which is visual representation of the accuracy of the diagnostic test is compare using its AUC. For a good test, AUC should be 1 whereas for useless test as well as just opposite discrimination of disease/non-disease, AUC are 0.5 and 0.0, respectively. Recent developments in ROC methodology provide a wide range of statistical tools to evaluate the diagnostic accuracy of biomarkers. As technologies advance, there is a critical window of time to harness and direct development of new diagnostics to benefit patients. The goal is not just to create more tests, but to develop rapid, reliable, accurate, simple tests that will reduce time to a diagnosis and truly improve the quality of care and patient outcomes while reducing health-care costs.[10]

References

1. Kanchanaraksa S. *Evaluation of Diagnostic and Screening Tests: Validity and Reliability*. Baltimore: John Hopkins University, 2008.
2. Šimundić AM. Measures of diagnostic accuracy: basic definitions. Med Biol Sci 2008;22(4):61–5.
3. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. British Med J 2002;324(7338):669.
4. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56:45–50.
5. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of “overall accuracy” to evaluate the validity of screening or diagnostic tests. J Gen Intern Med 2004;19(5 Pt 1):460–5.
6. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J 2005;47:458–72.
7. 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.
8. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143(1):29–36.
9. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol 2004;5(1):11–8.
10. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 2011;48(4):277–87.

How to cite this article: Mishra P, Pandey CM, Singh U, Yadav SS, Sharma V. Evaluation and application of diagnostic accuracy in clinical decision-making. Int J Med Sci Public Health 2016;5:2190-2193

Source of Support: Nil, **Conflict of Interest:** None declared.