Understanding diagnostic tests – Part 3: Receiver operating characteristic curves

Rakesh Aggarwal, Priya Ranganathan[1]

Department of Gastroenterology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, [1]Department of Anaesthesiology, Tata Memorial Centre, Mumbai, Maharashtra, India

Abstract

In the previous two articles in this series on biostatistics, we examined the properties of diagnostic tests and various measures of their performance in clinical practice. These performance measures vary according to the cutoff used to distinguish the diseased and the healthy. We conclude the series on diagnostic tests by looking at receiver operating characteristic curves, a technique to assess the performance of a test across several different cutoffs, and discuss how to determine an optimum cutoff.

Keywords: Biostatistics, receiver operating characteristic curve, sensitivity, specificity

Address for correspondence: Dr. Priya Ranganathan, Department of Anaesthesiology, Tata Memorial Centre, Ernest Borges Road, Parel, Mumbai - 400 012, Maharashtra, India.
E-mail: drpriyaranganathan@gmail.com

INTRODUCTION

In two previous articles in this series,[1,2] we discussed some of the properties of diagnostic tests. The sensitivity and specificity of a test inform us about the likelihood of a positive or a negative result, given that the disease of interest is present or absent, whereas positive and negative predictive values tell us about the probability of presence or absence of the disease, given that a test's result is positive or negative.[1] The latter values are heavily influenced by the prevalence of disease in the population being tested and are more relevant to clinicians.[1] The positive and negative likelihood ratios, another way of looking at diagnostic tests, represent the probability that someone with the disease has a particular test result as compared to someone without the disease.[2] A test with a higher positive likelihood ratio and a lower negative likelihood ratio is better at discriminating between those with and without disease.

All the attributes of diagnostic tests discussed in the previous articles depend on the cutoff value used to define the presence or absence of disease. However, the cutoffs are not cast in stone, and it is not infrequent for different cutoffs to be used to define disease or health. This change can markedly affect the performance characteristics of the test. In this third and final article on diagnostic tests, we look at another way of assessing a diagnostic test, namely the receiver operating characteristic (ROC) curve, which looks at the performance of the test over a range of cutoffs.

HOW ARE RECEIVER OPERATING CHARACTERISTIC CURVES PLOTTED?

An ROC curve is constructed by plotting sensitivity (proportion of cases having positive test or the proportion of cases correctly identified as having disease or "true positives"/"all cases") against "1 − specificity" (i.e., the proportion of

Access this article online	
Quick Response Code:	**Website:** www.picronline.org
	DOI: 10.4103/picr.PICR_87_18

How to cite this article: Aggarwal R, Ranganathan P. Understanding diagnostic tests – Part 3: Receiver operating characteristic curves. Perspect Clin Res 2018;9:145-8.

controls having positive test or proportion of controls incorrectly classified as having disease or "false-positives"/"all controls"), for each possible cutoff score. By convention, sensitivity (or the true-positive rate) is plotted along the "y" axis, whereas "1 − specificity" (or the false-positive rate) is plotted along the "x" axis. The ROC curve thus provides a graphical representation of the proportion of patients with the disease of interest correctly identified as positive against the proportion of healthy subjects incorrectly identified as positive for each cutoff score.

Let us, as an example, think of a test which can have values of 0–14, with higher values more likely to indicate disease and lower values indicating health. This test is administered to 40 persons each with and without the disease of interest, whose test results are shown in Figure 1. One could now use different cutoffs (e.g., 0.5, 1.5, 2.5..., 12.5, 13.5) to define the test result as positive or negative. The number of persons with or without disease who test positive or negative would vary according to the cutoff used [Table 1]. A lower cutoff would lead to more patients with disease being picked up correctly but a higher proportion of false-positives among healthy persons. On the other hand, a higher cutoff would miss some persons with disease but would lead to fewer false-positives. Using these numbers, one can easily calculate sensitivity and "1 − specificity" for each cutoff [Table 1]. If one plots these values, one obtains a curved line which is referred to as the ROC curve [Figure 2].

INTERPRETING RECEIVER OPERATING CHARACTERISTIC CURVES

A test with good performance would be expected to correctly diagnose nearly all the cases, i.e., to have a high sensitivity.
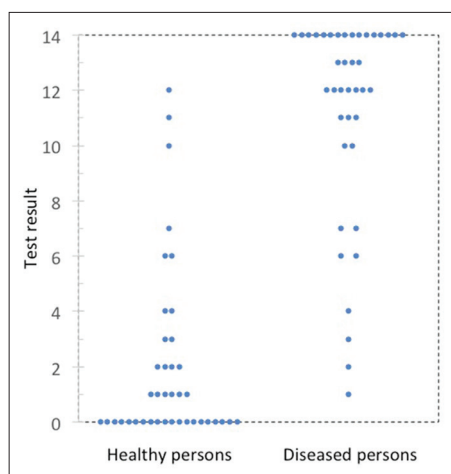

Figure 1: A hypothetical test with possible test result values of 0–14 is offered to forty persons known to have disease and forty healthy persons. The number of persons in each group with each possible test result is shown. In general, higher values are more likely in diseased persons than in healthy persons

Further, it would be expected to correctly diagnose nearly all the controls, i.e., have a very low false-positive rate (or a low "1 − specificity"). For such a test, the points on the ROC curve for cutoffs that provide good discrimination between persons with and without the disease would be expected to lie close to the top-left corner of the plot [Figure 3] (curve A). In fact, for a perfect test which accurately diagnoses all the cases and controls, sensitivity and specificity would both be 1.0 and "1 − specificity" would be zero. The ROC curve for such a test would rise vertically from the origin to the left top corner of the box and then run horizontally across to the right. By comparison, a test with a larger number of false-positive or negative tests would not reach as close to the left upper corner [Figure 3] (curve B). It is customary to draw a diagonal line on the ROC curve extending from left lower end (sensitivity = 0 and false-positivity rate = 0) to right upper end (sensitivity = 1.0 and false-positivity rate = 1.0) of the box in which the ROC is drawn. For all points on such a line [Figure 2] (line C), the values of sensitivity and false-positivity rate are identical. This line represents a hypothetical test for which, using any cut-off, positive results are as frequent in cases as in controls, i.e., the test does not discriminate at all between persons with and without the disease. Such a test would have no clinical use.

AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE

ROC curves also permit a numeric assessment of the overall performance of diagnostic tests. This is done by estimating the area under (i.e., to the right of and beneath) the curve and is expressed as a proportion of total area of
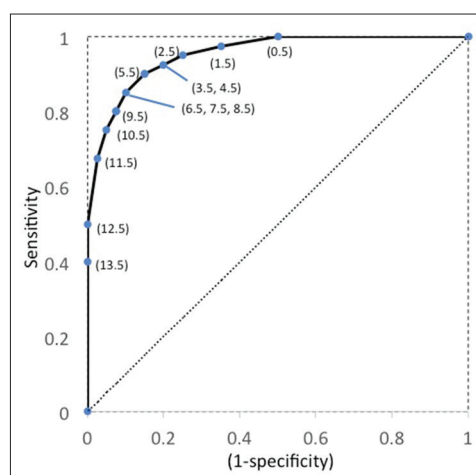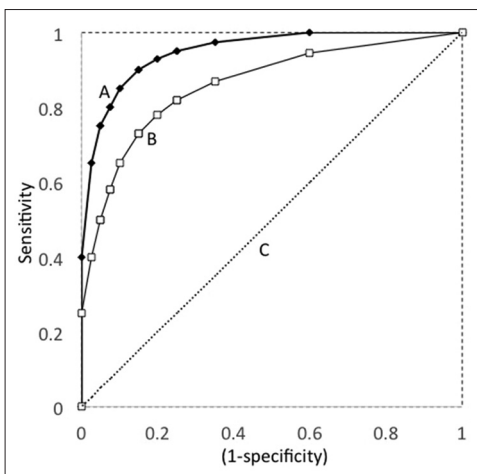

Figure 2: Receiver operating characteristic (ROC) curve for hypothetical data shown in Figure 1. From the data in Figure 1, sensitivity and false-positivity (=1 − specificity) rates were calculated for various possible cutoffs [Table 1]. A plot of these values yielded this ROC curve. The values in parentheses represent the cut-off value(s) that each point on the curve corresponds to. The dotted diagonal line represents a test that does not discriminate at all between those with and without disease (see text for details)

Table 1: Number of persons who are correctly classified as having disease (true positives; among 40 diseased persons) or not having disease (true negatives; among 40 healthy persons) using different cutoffs

Cutoff	Number of persons correctly classified among		Sensitivity "C"="A"/40	Specificity "D"="B"/40	1-specificity "E"=1-D
	Persons with disease (true positives) "A"	Persons without disease (true negatives) "B"			
0.5	40	20	1.000	0.500	0.500
1.5	39	26	0.975	0.650	0.350
2.5	38	30	0.950	0.750	0.250
3.5	37	32	0.925	0.800	0.200
4.5	36	34	0.900	0.850	0.150
5.5	36	34	0.900	0.850	0.150
6.5	34	36	0.850	0.900	0.100
7.5	32	37	0. 800	0.925	0.075
8.5	32	37	0.800	0.925	0.075
9.5	32	37	0.800	0.925	0.075
10.5	30	38	0.750	0.950	0.050
11.5	27	39	0.675	0.975	0.025
12.5	20	40	0.500	1.000	0
13.5	16	40	0.400	1.000	0

The numbers in columns marked "A" and "B" can be used to calculate sensitivity and specificity (columns marked "C" and "D"). The latter can then be used to calculate "1−specificity (column "E")". Values of "C" and "E" are then used to draw the receiver operator characteristic curve

Figure 3: Comparison of performance of tests using receiver operating characteristic (ROC) curves. A test with ROC curve which is located closer to the left upper corner (e.g., curve "A") has a better discrimination ability than a test with a curve that is located farther from this corner (e.g., curve "B"). The former would also have a higher value of area under curve, which is a quantitative measure of a test's performance. The diagonal line (line "C"; with area under curve = 0.50) represents a test with no discriminating ability. An ideal test would be expected to have an area under ROC curve value of 1.0

the square in which the curve is drawn. A test with higher sensitivity and specificity would reach closer to the left upper corner and hence would have a higher area under the curve. This measure can also be used to compare the performance of two different tests for the diagnosis of a particular disease. Thus, a test with larger area under the ROC curve is preferred over another test with smaller area under the curve [e.g., in Figure 2, the test with ROC curve A would be preferred over that with ROC curve B]. A test with area under the curve value of 0.5 (e.g., curve C) has no diagnostic value, as discussed above. For an ideal test, area under the ROC curve would be expected to be 1.0.

CHOOSING THE CUTOFF VALUE FOR A TEST

ROC curve is also helpful in deciding the optimum cutoff for a test. One possible cutoff could be one which is least likely to lead to misclassification, i.e., is likely to have the least number of false-positives and false-negatives taken together. This is represented by the point on the ROC curve that has the least distance from the top-left corner of the box. For instance, in Figure 2, the point nearest to the top-left corner is the one for the cutoff of 5.5, suggesting that this may be the optimal cutoff to differentiate persons with disease from those without disease. This point, as compared to other possible cutoffs, has the minimum value for $(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$. A simpler and more commonly used alternative is the use of cutoff with the maximum sum of sensitivity and specificity. It is calculated as the cutoff with maximum value of Youden's index, which is defined as (sensitivity + specificity − 1). Its values can vary between −1.0 and 1.0, and higher values indicate a test cutoff with higher discriminative ability.

However, these apply only if misclassification in either direction is given equal weightage. In clinical situations, the importance of a false-negative test is often different from that of a false-positive test. If one wishes the test to have a high sensitivity at the cost of some loss of specificity, one can choose as cutoff, a point where the curve becomes horizontal (e.g., in Figure 2, one could decide to use 1.5 or 2.5 as the cutoff). Alternatively, if one prefers a test with higher specificity with some loss of sensitivity, one could choose a point where the curve stops being vertical (e.g., in Figure 2, using 11.5 or 12.5 as the cutoff). For instance, for an assay for hepatitis

B surface antigen (HBsAg) in serum, one could use a lower cutoff value when the test is done for screening of donated blood in blood banks than when it is used to test blood from patients attending a clinic. In the former situation, we wish to detect blood units with the minutest amounts of HBsAg so that these can be excluded from the blood supply system (i.e., we wish to minimize the risk of transfusion-related infection even at the cost of discarding some blood units that contain no or so little virus that these cannot transmit infection). Therefore, we prefer a lower cutoff, with greater sensitivity at the cost of some loss of specificity. On the other hand, in the clinic situation, we wish to be certain that anyone with a positive test actually has the infection; any false-positive test in this situation would cause unwarranted psychological stress to the person and further costly testing and treatment. Thus, in this situation, we use a higher cutoff, preferring specificity over sensitivity.

SUGGESTED READING

The readers may want to read a study by Oh and Bae who assessed the effect of use of different cutoff levels of an antigen in the serum for detecting recurrent disease in women treated for cervical cancer undergoing posttreatment surveillance on the test's sensitivity and specificity.[3] Further, they used these data to create an ROC curve, calculated the area under this curve, and determined the optimal cutoff using Youden's index.

It may be pertinent to point out here that a lower cutoff may be preferred when this blood test is used for surveillance, as in this study; in this situation, one would prefer a higher sensitivity (fewer false-negatives) even at the cost of some loss of specificity (more false-positives). By comparison, for the use of this blood test as a confirmatory test, a higher cutoff with higher specificity (fewer false-positives) may be preferred, even though that would be associated with a loss of sensitivity (i.e., a larger number of false-negatives).

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests – Part 1. Perspect Clin Res 2018;9:40-3.
2. Ranganathan P, Aggarwal R. Understanding the properties of diagnostic tests – Part 2: Likelihood ratios. Perspect Clin Res 2018;9:99-102.
3. Oh J, Bae JY. Optimal cutoff level of serum squamous cell carcinoma antigen to detect recurrent cervical squamous cell carcinoma during post-treatment surveillance. Obstet Gynecol Sci 2018;61:337-43.S