Determining the Area under the ROC Curve for a Binary Diagnostic Test

SCOTT B. CANTOR, PhD, MICHAEL W. KATTAN, PhD

The authors provide a simple calculation for the unbiased estimation of the area under the ROC curve for a binary diagnostic test or a continuously valued test result that is effectively used in a binary way. The formula described can be used to interpret the discriminative ability of a diagnostic test. **Key words:** ROC curve analysis, binary diagnostic test; discriminative ability. **(Med Decis Making 2000;20:468–470)**

Receiver operating characteristic (ROC) curve analysis is used to determine the discriminative ability of a diagnostic test.[1] A common metric for determining the level of discriminative ability is the area under the ROC curve (AUC), which ranges from 0.5 for a test with no discriminative ability at all (i.e., the test is as good as tossing a coin to determine whether or not the test result should be classified as "positive") to 1.0 for a test with perfect discrimination.

Some diagnostic tests are binary, such that only one cutoff point is available, i.e., only one pair of values for sensitivity and specificity exists. In addition, some diagnostic tests theoretically have more than one cutoff point but are effectively employed as binary. When used, they are reported in the literature with only one set of values for sensitivity and specificity. We report a way to determine the area under the ROC curve for a binary diagnostic test where only one cutoff point or one paired value of sensitivity and specificity is used.

Assume the diagnostic test has one pair of values for sensitivity and specificity, designated SENS and SPEC, respectively. The ROC curve for a binary (positive–negative) test consists of two line segments (see figure 1). The first line segment connects the origin, the point in the bottom left corner of the figure that classifies all samples as "negative" (sensitivity = 0; specificity = 1), with the point representing the binary test (sensitivity = SENS; specificity = SPEC). The first line segment has slope equal to the likelihood ratio for a "positive test result," i.e., slope = positive likelihood ratio = $P(T+|D+)/P(T+|D-)$.

The second line segment connects the point representing the binary test with the point located in the top right corner of the figure that classifies all samples as "positive" (sensitivity = 1; specificity = 0). The second line segment has slope equal to the likelihood ratio for a "negative test result," i.e., slope = negative likelihood ratio = $P(T-|D+)/P(T-|D-)$.*

Divide the region under the ROC curve into three sections: triangle A, rectangle B, and triangle C. The sum of the areas of the individual sections constitutes the area under the ROC curve. The formula for the area of a triangle is $A = \frac{1}{2} \cdot \text{base} \cdot \text{height}$. The formula for the area of a rectangle is $A = \text{base} \cdot \text{height}$.

Triangle A has legs of length (1 − SPEC) and SENS.

Triangle C has legs of length SPEC and (1 − SENS).

Rectangle B has sides of length SPEC and SENS.

Received March 28, 2000, from The University of Texas M. D. Anderson Cancer Center, Department of Health Services Research, Houston, Texas (SBC); and Memorial Sloan-Kettering Cancer Center, Department of Biostatistics and Epidemiology, New York, New York (MWK). Revision accepted for publication June 6, 2000. Supported in part by a grant from the National Cancer Institute CA82710 (SBC).

Address correspondence and reprint requests to Dr. Cantor: The University of Texas M. D. Anderson Cancer Center, Department of Health Services Research, 1515 Holcombe Boulevard, Box 40, Houston, TX 77030-4095; telephone: (713) 745-8552; fax: (713) 792-8700; e-mail: ⟨sbcantor@mdanderson.org⟩.

*If we have a poor diagnostic test, so that both line segments have slopes close to 1.0, the ratio of net costs of treating nondiseased individuals compared with the net benefits of treating diseased individuals (the C/B ratio) of the clinical situation may identify the optimum cutoff point as either a test with perfect specificity and no sensitivity or a test with perfect sensitivity and no specificity. This may happen because the slopes of the two line segments do not straddle the critical slope $R = (1 - p[D])/p[D] \cdot (C/B)$, where p[D] is the prevalence of disease. See Cantor et al. for details.[2]

Therefore,

$$\text{Area of triangle A} = \tfrac{1}{2}\cdot(1 - \text{SPEC})\cdot\text{SENS}$$
$$\text{Area of triangle C} = \tfrac{1}{2}\cdot\text{SPEC}\cdot(1 - \text{SENS})$$
$$\text{Area of rectangle B} = \text{SPEC}\cdot\text{SENS}$$

The area under the ROC curve is the sum of the areas of regions A, B, C.

$$\begin{aligned}\text{Sum of A, B, C} &= \tfrac{1}{2}\cdot\text{SENS} - \tfrac{1}{2}\cdot\text{SPEC}\cdot\text{SENS}\\ &\quad + \tfrac{1}{2}\cdot\text{SPEC} - \tfrac{1}{2}\cdot\text{SPEC}\cdot\text{SENS}\\ &\quad + \text{SPEC}\cdot\text{SENS}\\ &= \frac{\text{SENS} + \text{SPEC}}{2}\end{aligned}$$

Thus, the area under the ROC curve for a diagnostic test with a single operating point is equal to ½·(SENS + SPEC).†

Example 1. The AUC for any point on the chance line (line of no discriminative ability) in a ROC diagram is 0.50. The line is such that SENS = 1 − SPEC.

$$\text{AUC} = \frac{\text{SENS} + \text{SPEC}}{2} = \frac{(1 - \text{SPEC}) + \text{SPEC}}{2} = 0.5$$

Example 2. A perfect test has sensitivity and specificity equal to 1; this test also has area under the ROC curve equal to 1.0.

$$\text{AUC} = \frac{\text{SENS} + \text{SPEC}}{2} = \frac{1 + 1}{2} = 1.0$$

Example 3. In fact, all binary tests with their ROC points on a given 45-degree line have the same area under the ROC curve. These tests all have the same Youden index, as defined by SENS + SPEC − 1 = Y, where Y is a constant between 0 and 1 for a useful (but not perfect) diagnostic test.[3] Thus, in this special case, the area under curve can be computed as follows:

$$\text{AUC} = \frac{\text{SENS} + \text{SPEC}}{2} = \frac{(Y + 1)}{2}$$

†Another geometric proof is plausible. In this alternative construction, one divides the region under the ROC curve in figure 1 into two triangles by drawing a line from the point representing the binary diagnostic test to the bottom right corner of the plot. The triangle on the left has a base of length 1 and height of SENS; the triangle on the right also has a base of length 1 but a height of SPEC. Thus, the triangles have areas of ½·SENS and 1/2·SPEC respectively, and the total area of the curve is the sum of the two triangular regions or ½·(SENS + SPEC), the same result as presented in the text.

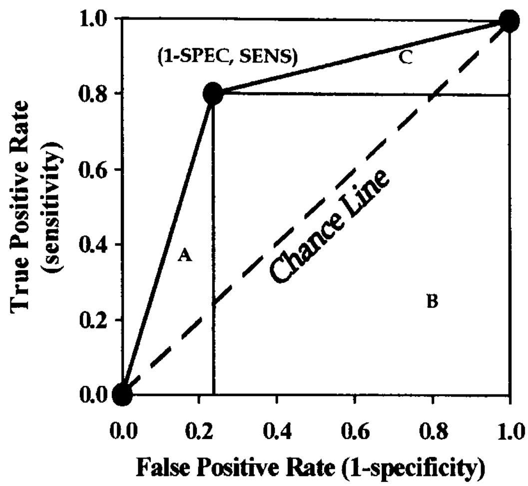

FIGURE 1. Plot of receiver operating characteristic (ROC) curve with only one paired value of sensitivity and specificity.

Example 4. A diagnostic test of good accuracy might have sensitivity equal to 0.8 and specificity equal to 0.8. By the above formula, the area under the receiver operating characteristic curve will also be 0.8.

$$\text{AUC} = \frac{\text{SENS} + \text{SPEC}}{2} = \frac{0.8 + 0.8}{2} = 0.8$$

Discussion

The simple calculation provided here yields an unbiased estimate of the area under the ROC curve for a binary diagnostic test or a continuously valued test result that is effectively used in a binary fashion. An example of the latter is serum prostate-specific antigen (PSA) level in the decision to biopsy for prostate cancer.[4] While PSA takes on continuous values with a minimum of 0, many urologists simply recommend biopsy when the PSA is above 4.0 ng/mL. Our method would not be suitable for judging PSA when used as a continuous diagnostic test but would be appropriate for judging its ability to detect cancer when used in this binary fashion.

Gaussian methods are available for approximating the entire ROC curve when only one cutoff point is given. The methods used for determining sample size for studies of the ROC curve area when the literature provides only one pair of estimates for sensitivity and specificity[5] can be extended by creating a family of plausible ROC curves[6]; the area under the ROC curve can then be calculated using nonparametric methods.[7] Clearly, the method presented in the present paper underestimates the more "fuller," concave curve that would be available if all the cutoff points had been reported. However, the method described in the present paper evaluates the discriminative ability of a test in its simplest dichotomous version.

We have computed the area under the ROC curve for a single operating point based on a geometric construction. The usefulness of this computation might be based on the interpretation of the area under the curve. In other words, students who are beginning to learn about ROC curve analysis typically ask, "What is the area under the ROC curve for a good diagnostic test?" Or more typically, "The area under the ROC curve for a diagnostic test I wish to evaluate is 0.85. Is this a great, good, or fair diagnostic test?"

Students and evaluators of diagnostic tests can now point to the simple formula described in this paper to interpret the discriminative ability of a diagnostic test. Weinstein and Fineberg,[8] in their discussion of ROC curve analysis, present a figure representing three diagnostic tests. These tests have approximate areas under the curve of 0.8, 0.65, and 0.5, indicating "good," "fair," and "poor" discriminative abilities. Biggerstaff[9] extends the geometric approach we used by illustrating innovative methods for the comparison of rival binary diagnostic tests. Certainly, students may benefit from graphic presentations of areas under the ROC curve.

However, students may have a better sense of the relative accuracy of a diagnostic test based upon the values of sensitivity and specificity. If, for example, a student feels that a diagnostic test that has sensitivity and specificity equal to 0.8 is a "good" diagnostic test, then that student should interpret an ROC curve with area under the curve equal to 0.8 to be of "good" discriminative ability. Similarly, if students understand that a diagnostic test with sensitivity and specificity equal to 0.95 is of "excellent" discriminative ability, then those students should consider an area under the ROC curve of 0.95 to be "excellent" as well.

The authors thank the reviewers for their contributions that enhanced the clarity of the manuscript.

References

1. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283–98.
2. Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. J Clin Epidemiol. 1999;52:885–92.
3. Hilden J, Glasziou P. Regret graphs, diagnostic uncertainty and Youden's index. Stat Med. 1996;15:969–86.
4. Slawin KM, Kattan MW, Scher HI, Scardino PT. Prostate-specific antigen in screening, prognosis of the primary tumor and monitoring of recurrent disease. In: Perry MC (ed). 2000 Spring Educational Book. Alexandria, VA: American Society of Clinical Oncology, 2000:596–608.
5. Obuchowski NA. Sample size calculations in studies of test accuracy. Stat Meth Med Res. 1998;7:371–92.
6. Dorfman DD, Alf E Jr. Maximum-likelihood estimation of parameters of signal detection theory and determination of confidence intervals—rating-method data. J Math Psychol. 1969; 6:487–96.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve. Radiology. 1982;143: 29–36.
8. Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia, PA: W. B. Saunders, 1980.
9. Biggerstaff BJ. Comparing diagnostic tests: a simple graphic using likelihood ratios. Stat Med. 2000;19:649–63.

Correction

An error appeared in the article, Development of a pediatric multiple organ dysfunction score: use of two strategies (Leteurtre et al., Med Decis Making 1999;19:399–410). On page 407, in the right column, line 9 states that "No death occurred in patients with PEMOD or PELOD scores of 0, whereas the PICU mortality rate was 100% for patients with PEMOD score >50 or PELOD score >26." The last part of the sentence should be changed as follows: "whereas the PICU mortality rate was 100% for patients with PEMOD score $\geq$19 or PELOD score $\geq$50."