Radiology

Nancy A. Obuchowski, PhD

Index terms:
Diagnostic radiology
Statistical analysis

Published online
10.1148/radiol.2291010898
Radiology 2003; 229:3–8

Abbreviations:
FPR = false-positive rate
MS = multiple sclerosis
ROC = receiver operating characteristic

[1] From the Department of Biostatistics and Epidemiology/Wb4, Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH 44195. Received May 8, 2001; revision requested June 11; revision received August 1; accepted August 2. **Address correspondence to** the author (e-mail: *nobuchow@bio.ri.ccf.org*).

Receiver Operating Characteristic Curves and Their Use in Radiology[1]

Sensitivity and specificity are the basic measures of accuracy of a diagnostic test; however, they depend on the cut point used to define "positive" and "negative" test results. As the cut point shifts, sensitivity and specificity shift. The receiver operating characteristic (ROC) curve is a plot of the sensitivity of a test versus its false-positive rate for all possible cut points. The advantages of the ROC curve as a means of defining the accuracy of a test, construction of the ROC, and identification of the optimal cut point on the ROC curve are discussed. Several summary measures of the accuracy of a test, including the commonly used percentage of correct diagnoses and area under the ROC curve, are described and compared. Two examples of ROC curve application in radiologic research are presented.

Sensitivity and specificity are the basic measures of the accuracy of a diagnostic test. They describe the abilities of a test to enable one to correctly diagnose disease when disease is actually present and to correctly rule out disease when it is truly absent. The accuracy of a test is measured by comparing the results of the test to the true disease status of the patient. We determine the true disease status with the reference standard procedure.

Consider as an example the test results of 100 patients who have undergone mammography (Table 1). According to biopsy results and/or 2-year follow-up results (ie, the reference standard procedures), 50 patients actually have a malignant lesion and 50 patients do not. If these 100 test results were from 100 asymptomatic women without a personal history of breast cancer, then we might define a positive test result as any that represents a "suspicious" or "malignant" finding and a negative test result as any that represents a "normal," "benign," or "probably benign" finding. We have used a cut point for defining positive and negative test results. The cut point is located between the suspicious and probably benign findings. The estimated sensitivity with this cut point is (18 + 20)/50 = 0.76, and the specificity is (15 + 3 + 18)/50 = 0.72.

Alternatively, if these 100 test results were from 100 asymptomatic women with a personal history of breast cancer, then we might use a different cut point, such that a positive test result represents a probably benign, suspicious, or malignant finding and a negative test result represents a normal or benign finding. The estimates of sensitivity and specificity would change (ie, they would now be 0.96 and 0.36, respectively).

Important point: Sensitivity and specificity depend on the cut point used to define positive and negative test results. As the cut point shifts, the sensitivity increases while the specificity decreases, or vice versa.

COMBINED MEASURES OF SENSITIVITY AND SPECIFICITY

It is often useful to summarize the accuracy of a test by using a single number; for example, when comparing two diagnostic tests, it is easier to compare a single number than to compare both the sensitivity and the specificity values of the tests. There are several such summary measures; I will describe a popular but easily misinterpreted one that is usually referred to simply as accuracy. Using the second cut point in Table 1, we can compute accuracy as the percentage of correct diagnoses in the entire sample—that is, (48 + 18)/100 = 0.66, or 66%. The strength of this measure of accuracy is its simple computa-

TABLE 1
Results from Mammography Study with 100 Patients

Cut Point and Reference Standard Result	Radiologist's Interpretation					Total
	Normal	Benign	Probably Benign	Suspicious	Malignant	
Cut point 1*						
Reference standard result						
Cancer present	2	0	10	18†	20†	50
Cancer absent	15	3	18	13‡	1‡	50
Cut point 2*						
Reference standard result						
Cancer present	2	0	10†	18†	20†	50
Cancer absent	15	3	18‡	13‡	1‡	50

Note.—Data are numbers of patients with the given result in a fictitious study of mammography in which 50 patients had a malignant lesion and 50 did not.

* For cut point 1, a positive result is defined as a test score of suspicious or malignant; for cut point 2, a positive result is defined as a test score of probably benign, suspicious, or malignant.

† Test results considered true-positive (for estimating sensitivity) with this cut point.

‡ Test results considered false-positive (for estimating the false-positive rate [FPR] or specificity) with this cut point.

TABLE 2
Effect of Prevalence on Accuracy

Reference Standard Result	Radiologist's Interpretation					Total
	Normal	Benign	Probably Benign	Suspicious	Malignant	
Cancer present	2	0	10	18	20	50
Cancer absent	285	57	342	247	19	950

Note.—Data are numbers of patients with the given result in a fictitious study of mammography with 1,000 patients. This data set represents a modification of the data set in Table 1 so that the prevalence of cancer is 5%. When cut point 2 (described in the note to Table 1) is used with this data set, the estimated sensitivity ([10 + 18 + 20]/50 = 0.96) and specificity ([285 + 57]/950 = 0.36) are the same as with the data set in Table 1. However, one commonly used estimate of overall accuracy is the percentage of correct diagnoses in the sample. With this data set it is 39% ([10 + 18 + 20 + 285 + 57]/1,000 = 0.39), which is not the same as with the data set in Table 1.

tion. It has several limitations, however: Its magnitude varies as the prevalence of disease varies in the sample, it is calculated on the basis of only one cut point, and false-positive and false-negative results are treated as if they are equally undesirable. As an illustration of the first limitation, note that in Table 2 the prevalence of disease is 5% instead of the 50% in Table 1. The sensitivity and specificity values are the same in Tables 1 and 2, yet the estimated accuracy value in Table 2 drops to (48 + 342)/1,000 = 0.39, or 39%.

Important point: A measure of test accuracy is needed that combines sensitivity and specificity but does not depend on the prevalence of disease.

RECEIVER OPERATING CHARACTERISTIC CURVE

In 1971, Lusted (1) described how receiver operating characteristic (ROC) curves could be used to assess the accuracy of a test. An ROC curve is a plot of test sensitivity (plotted on the y axis) versus its FPR (or 1 − specificity) (plotted on the x axis). Each point on the graph is generated by using a different cut point. The set of data points generated from the different cut points is the empirical ROC curve. We use lines to connect the points from all the possible cut points. The resulting curve illustrates how sensitivity and the FPR vary together.

Figure 1 illustrates the empirical ROC curve for the mammography example. Since in our example there are five categories for the test results, we can compute four cut points for the ROC curve. The two endpoints on the ROC curve are 0,0 and 1,1 for FPR, sensitivity. The points labeled 1 and 2 on the curve correspond to the first and second cut points, respectively, that are defined in the note to Table 1. Estimations of the other points are provided in Table 3.

The ROC plot has many advantages over single measurements of sensitivity and specificity (2). The scales of the curve—that is, sensitivity and FPR—are the basic measures of accuracy and are easily read from the plot; the values of the cut points are often labeled on the curve as well. Unlike the measure of accuracy defined in the previous section (ie, the percentage of correct diagnoses), the ROC curve displays all possible cut points. Because sensitivity and specificity are independent of disease prevalence, so too is the ROC curve. The curve does not depend on the scale of the test results (ie, we can alter the test results by adding or subtracting a constant or taking the logarithm or square root without any change to the ROC curve) (3). Lastly, the ROC curve enables a direct visual comparison of two or more tests on a common set of scales at all possible cut points.

It is often convenient to make some assumptions about the distribution of the test results and then to draw the ROC curve on the basis of the assumed distribution (ie, assumed model). The resulting curve is called the fitted or smooth ROC curve. The fitted curve for the mammography study is plotted in Figure 1; it was constructed from a binormal distribution (ie, two normal distributions: one for the test results of patients without breast cancer and another for test results of patients with breast cancer) (Fig 2). The binormal distribution is the most

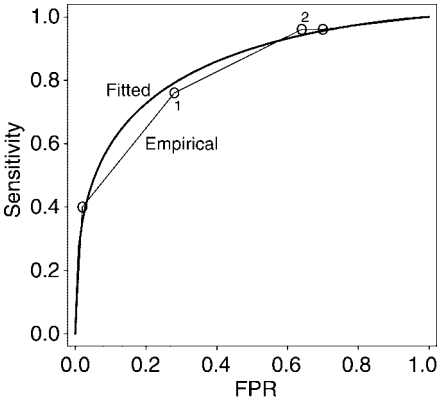

Figure 1. Graph of the empirical and fitted ROC curves for the mammography study. The points on the empirical curve are marked with open circles and are estimated in Table 3. The points labeled *1* and *2* on the curve correspond to the first and second cut points, respectively, that are defined in the note to Table 1.

TABLE 3
Construction of Empirical ROC Curve for Mammography Study

Cut Point	Sensitivity*	FPR†
Between normal and benign	0.96 (48/50)	0.70 (35/50)
Between benign and probably benign	0.96 (48/50)	0.64 (32/50)
Between probably benign and suspicious	0.76 (38/50)	0.28 (14/50)
Between suspicious and malignant	0.40 (20/50)	0.02 (1/50)

Note.—These data represent estimations of the points on the empirical ROC curve marked with open circles and depicted in Figure 1. The ROC curve in Figure 1 was constructed on the basis of the data in Table 1, with sensitivity and the FPR estimated at each possible cut point.
* Data in parentheses are those used to calculate the sensitivity value.
† Data in parentheses are those used to calculate the FPR (or 1 − specificity) value.

commonly used distribution for estimating the smooth ROC curve. There are computer programs (for example, *www-radiology.uchicago.edu/sections/roc/software.cgi*) for estimating the smooth ROC curve on the basis of the binormal distribution; these programs make use of a statistical method called maximum likelihood estimation.

An ROC curve can be constructed from objective measurements of a test (eg, serum glucose level as a test for diabetes), objective evaluation of image features (eg, the computed tomographic [CT] attenuation coefficient of a renal mass relative to normal kidney), or subjective diagnostic interpretations (eg, the five-category Breast Imaging Reporting and Data System scale used for mammographic interpretation) (5). The only requirement is that the measurements or interpretations can be meaningfully ranked in magnitude. With objective measurements the cut point is explicit, so one can choose from an infinite number of cut points along the continuum of the test results. For diagnostic tests whose results are interpreted subjectively, the cut points are implicit or latent in that they only exist in the mind of the observer (6). Furthermore, it is assumed that each observer has his or her own set of cut points.

The term *receiver operating characteristic curve* comes from the idea that, given the curve, we, the receivers of the information, can use (or operate at) any point on the curve by using the appropriate cut point. The clinical application determines which cut point is used. For example, for evaluating women with a personal history of breast cancer, we need a cut point with good sensitivity (eg, cut point 2 in Table 1), even if the FPR is high. For evaluating women without a personal history of breast cancer, we require a lower FPR. For each application the optimal cut point (2,7) can be determined by finding the sensitivity and specificity pair that maximizes the function sensitivity − m(1 − specificity), where m is the slope of the ROC curve as follows:

$$m = \frac{\text{Prob}_{\text{Norm}}}{\text{Prob}_{\text{Dis}}} \times \frac{(C_{\text{FP}} - C_{\text{TN}})}{(C_{\text{FN}} - C_{\text{TP}})},$$

$\text{Prob}_{\text{Norm}}$ is the probability that the patient's condition is normal before the test is performed, Prob_{Dis} is the probability that the patient has the disease before the test is performed, C_{FP} is the cost (ie, the financial cost and/or health "cost") of a false-positive result, C_{TN} is the cost of a true-negative result, C_{FN} is the cost of a false-negative result, and C_{TP} is the cost of a true-positive result.

MEASURES OF ACCURACY BASED ON THE ROC CURVE

One of the most popular measures of the accuracy of a diagnostic test is the area under the ROC curve. The ROC curve area can take on values between 0.0 and 1.0. A ROC curve with an area of 1.0 is shown in Figure 3. A test with an area under the ROC curve of 1.0 is perfectly accurate because the sensitivity is 1.0 when the FPR is 0.0. In contrast, a test with an area of 0.0 is perfectly inaccurate. That is, all patients with disease are incorrectly given negative test results and all patients without disease are incorrectly given positive test results. With such a test it would be better to convert it into a test with perfect accuracy by reversing the interpretation of the test results. The practical lower bound for the ROC curve area is then 0.5. The line segment from 0,0 to 1,1 has an area of 0.5; it is called the chance diagonal (Fig 3). If we relied purely on guessing to distinguish patients with from patients without disease, then the ROC curve would be expected to fall along this diagonal line. Diagnostic tests with ROC curve areas greater than 0.5 have at least some ability to discriminate between patients with and those without disease. The closer the ROC curve area is to 1.0, the better the diagnostic test. One method (8) of estimating the area under the empirical ROC curve is described and illustrated in the Appendix. There are other methods (9,10) of estimating the area under the empirical ROC curve and its variance; all of these methods rely on nonparametric statistical methods.

The ROC curve area has several interpretations: *(a)* the average value of sensitivity for all possible values of specificity, *(b)* the average value of specificity for all possible values of sensitivity (11,12), and *(c)* the probability that a randomly selected patient with disease has a test result that indicates greater suspicion than a randomly chosen patient without disease (9).

In Figure 1 the area under the empirical ROC curve for mammography is 0.82; that is, if we select two patients at random—one with breast cancer and one without—the probability is 0.82 that the patient with breast cancer will have a more suspicious mammographic result. The area under the fitted curve is slightly larger at 0.84. When the number of cut points is small, the area under the empirical ROC curve is usually smaller than the area under the fitted curve.

The ROC curve area is a good summary measure of test accuracy because it does not depend on the prevalence of disease or the cut points used to form the curve.

However, once a test has been shown to distinguish patients with disease from those without disease well, the performance of the test for particular applications (eg, diagnosis, screening) must be evaluated. At this stage, we may be interested in only a small portion of the ROC curve. Furthermore, the ROC curve area may be misleading when one is comparing the accuracies of two tests. Figure 4 illustrates the ROC curves of two tests with equal area. At the clinically important FPR range (for example, 0.0–0.2), however, the curves are different: ROC curve A demonstrates higher sensitivity than does ROC curve B. Whenever the ROC curves of two tests cross (regardless of whether or not their areas are equal), it means that the test with superior accuracy (ie, higher sensitivity) depends on the FPR range; a global measure of accuracy, such as the ROC curve area, is not helpful here.

Important point: There are situations where we need a more refined measure of diagnostic test accuracy than the area under the ROC curve.

One alternative is to use the ROC curve to estimate sensitivity at a fixed FPR (or, as appropriate, we could use the FPR at a fixed sensitivity). As an example, in Figure 1 the sensitivity at a fixed FPR of 0.10 is 0.60. This measure of accuracy allows us to focus on the portion of the ROC curve that is of clinical relevance.

Another alternative measure of accuracy is the partial area under the ROC curve. It is defined as the area between two FPRs, e_1 and e_2 (or, as appropriate, the area between two false-negative rates). If $e_1 = 0$ and $e_2 = 1$, then the area under the entire ROC curve is specified. If $e_1 = e_2$, then the sensitivity at a fixed FPR is given. The partial area measure is thus a "compromise" between the entire ROC curve area and the sensitivity at a fixed FPR.

To interpret the partial area, we must consider its maximum possible value. The maximum area is equal to the width of the interval—that is, $e_2 - e_1$ (13). McClish (13) and Jiang et al (14) recommend standardizing the partial area by dividing it by its maximum value. Jiang et al (14) refer to this standardized partial area as the partial area index. The partial area index is interpreted as the average sensitivity for the range of FPRs examined (or the average FPR for the range of sensitivities examined). As an example, in Figure 1, the partial area in the FPR range of 0.00–0.20 is 0.112; the partial area index is 0.56. In other words, when the FPR is between 0.00 and 0.20, the average sensitivity is 0.56.

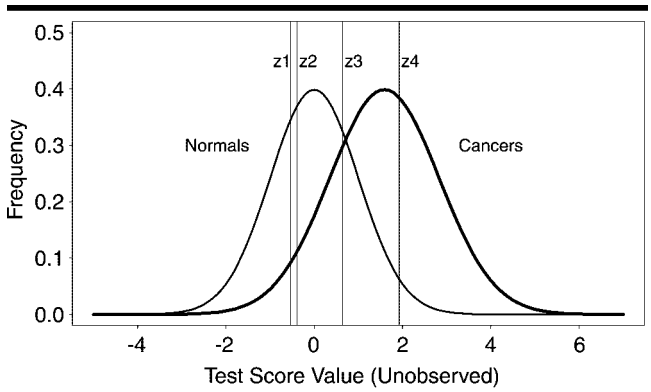

Figure 2. Graph shows the binormal distribution that best fits the mammography study data. By convention, the distribution of unobserved variables for the patients without cancer is centered at zero (ie, $\mu_1 = 0$) with variance (σ_1^2) equal to 1. For these data, the center of the distribution of the unobserved variables for the patients with cancer is estimated to be 1.59 (ie, $\mu_2 = 1.59$) with variance (σ_2^2) estimated to be 1.54. The binormal distribution can be described by its two parameters (4), a and b, as $a = (\mu_1 - \mu_2)/\sigma_2$ and $b = \sigma_1/\sigma_2$. The four cut points z_1, z_2, z_3, and z_4 define the five categories of test results. That is, a variable with a value below the point defined by z_1 indicates a normal result; a variable with a value between z_1 and z_2, a benign result; a variable with a value between z_2 and z_3, a probably benign result; a variable with a value between z_3 and z_4, a suspicious result; and a variable with a value above the point defined by z_4, a malignant result. Note that the binormal variables exist only in the mind of the reader (ie, they are unobserved). When the reader applies the cut points z_1, z_2, z_3, and z_4 to the unobserved variables, we obtain the observed five categories of test results.

EXAMPLES OF ROC CURVES IN RADIOLOGY

There are many examples of the application of ROC curves in radiologic research. I present two examples here. The first example illustrates the comparison of two diagnostic tests and the identification of a useful cut point. The second example describes a multireader study of the differences in diagnostic accuracy of two tests and differences in reader performance.

The first example is the study of Mushlin et al (15) of the accuracy of magnetic resonance (MR) imaging for detecting multiple sclerosis (MS). Three hundred three patients suspected of having MS underwent MR imaging and CT of the head. The images were read separately by two neuroradiologists without knowledge of the clinical course of or final diagnosis given to the patients. The images were scored as definitely showing MS, probably showing MS, possibly showing MS, probably not showing MS, or definitely not showing MS. The reference standard consisted of results of a review of the clinical findings by a panel of MS experts, results of follow-up for at least 6 months, and results of other diagnostic tests; the results of CT and MR imaging were not included to avoid bias.

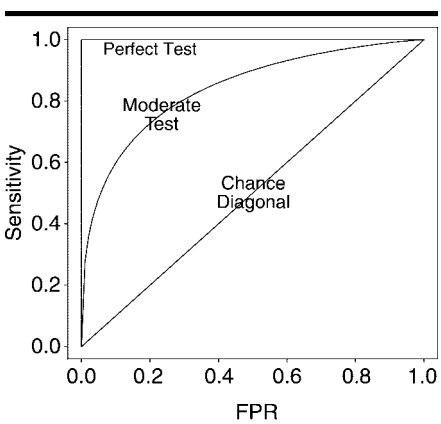

Figure 3. Graph shows comparison of three ROC curves. A perfect test has an area under the ROC curve of 1.0. The chance diagonal has an ROC area of 0.5. Tests with some discriminating ability have ROC areas between these two extremes.

The estimated ROC curve area for MR imaging was 0.82, indicating a good, but not definitive, test. In contrast, the estimated ROC curve area of CT was only 0.52; this estimated area was not significantly different from 0.50, indicating that CT results were no more accurate than guessing for diagnosing MS. The authors concluded that a "definite MS"

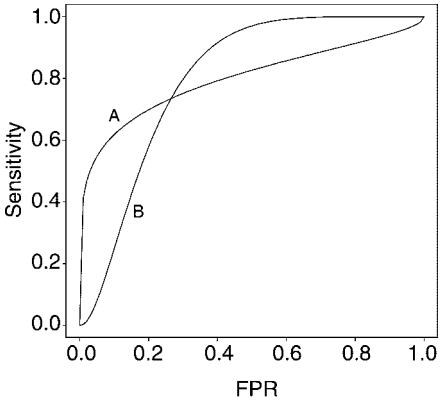

Figure 4. Graph shows two crossing ROC curves. The ROC areas of the two tests are the same at 0.80; however, for the clinically important range (ie, an FPR of less than 0.20), test *A* is preferable to test *B*.

Standard of Reference Result:	"Normal"	"Suspicious"	"Diseased"	Total
Disease present	0	4	6	10
Disease absent	7	3	2	12

Step 1: Identify pairs and assign scores to each pair

Possible Pairings			No. of Such Pairings	Score
Patient with Disease	vs	Patient without Disease		
"Normal"	vs	"normal"	0	0.5
"Normal"	vs	"suspicious"	0	0.0
"Normal"	vs	"diseased"	0	0.0
"Suspicious"	vs	"normal"	$4 \times 7 = 28$	1.0
"Suspicious"	vs	"suspicious"	$4 \times 3 = 12$	0.5
"Suspicious"	vs	"diseased"	$4 \times 2 = 8$	0.0
"Diseased"	vs	"normal"	$6 \times 7 = 42$	1.0
"Diseased"	vs	"suspicious"	$6 \times 3 = 18$	1.0
"Diseased"	vs	"diseased"	$6 \times 2 = 12$	0.5

Step 2: Sum the M × N scores. This is written below as (no. of pairings times score):
$(28 \times 1.0) + (12 \times 0.5) + (8 \times 0) + (42 \times 1.0) + (18 \times 1.0) + (12 \times 0.5) = 100$.

Step 3: Divide the sum by the total number of pairs (ie, M × N). This gives the estimated area under the empirical ROC curve: $100 / (10 \times 12) = 0.833$.

Figure A1. Fictitious data set and example of how to calculate the area under the empirical ROC curve.

reading at MR imaging essentially establishes the diagnosis of MS (MR images in only two of 140 patients without MS were scored as definitely showing MS, for an FPR of 1%). However, a normal MR imaging result does not conclusively exclude the diagnosis of MS (MR images in 35 of 163 patients with MS were scored as definitely not showing MS, for a false-negative rate of 21%).

In the second example, Iinuma et al (16) compared the accuracy of conventional radiography and digital radiography for the diagnosis of gastric cancers. One hundred twelve patients suspected of having gastric cancer underwent conventional radiography, and 113 different patients with similar symptoms and characteristics underwent digital radiography. Six readers interpreted the images from all 225 patients; the readers were blinded to the clinical details of the patients. The images were scored with a six-category scale, in which a score of 1 indicated that cancer was definitely absent; a score of 2, cancer was probably absent; a score of 3, cancer was possibly absent; a score of 4, cancer was possibly present; a score of 5, cancer was probably present; and a score of 6, cancer was definitely present. The diagnostic standard consisted of the findings of a consensus panel of three radiologists (not the same individuals as the six readers) who examined the patients and were told of the findings of other tests, such as endoscopy and histopathologic examination after biopsy.

The ROC curve areas of the six readers were all higher with digital radiography than with conventional radiography; the average ROC curve areas with digital and conventional radiography were 0.93 and 0.80, respectively. By plotting the fitted ROC curve areas of each of the six readers, the authors determined that for five of the six readers, digital radiography resulted in higher sensitivity for all FPRs; for the sixth reader, digital radiography resulted in considerably higher sensitivity only at a low FPR.

In summary, the ROC curve has many advantages as a measure of the accuracy of a diagnostic test: *(a)* It includes all possible cut points, *(b)* it shows the relationship between the sensitivity of a test and its specificity, *(c)* it is not affected by the prevalence of disease, and *(d)* from it we can compute several useful summary measures of test accuracy (eg, ROC curve area, partial area). The ROC curve alone cannot provide us with the optimal cut point for a particular clinical application; however, given information about the pretest probability of disease and the relative costs of diagnostic test errors, we can find the optimal cut point on the ROC curve. There are many study design issues (eg, patient and reader selection, verification and diagnostic standard bias) that need to be considered when one is conducting and interpreting the results of a study of diagnostic test accuracy. Many of these issues will be covered in a future article.

APPENDIX

The area under the empirical ROC curve can be estimated as follows: First, consider every possible pairing of patients with disease and patients without disease. Give each pair a score of 1.0 if the test result for the patient with disease is higher (ie, more suspicious for disease), a score of 0.5 if the test results are the same, and a score of 0.0 if the test result for the patient with disease is lower (ie, less suspicious for disease). Second, take the sum of these scores. If there are N nondiseased patients and M diseased patients in the sample, then there are $M \times N$ scores. Finally, divide the sum of these scores by $(M \times N)$. This gives the estimate of the area under the empirical ROC curve.

Figure A1 depicts a fictitious data set. The process described and illustrated in the figure can be written mathematically as follows (8): Let X_j denote the test score of the jth patient with disease and Y_k denote the test score of the kth patient without disease. Then,

$$A = \frac{1}{(M \times N)} \sum_{(j=1)}^{M} \sum_{(k=1)}^{N} \text{score}(X_j, Y_k),$$

where A is the estimate of the area under the empirical ROC curve and $\text{score}(X_j, Y_k)$ is the score assigned to the pair composed of the jth patient with disease and the kth patient without disease. The score equals 1 if X_j is greater than Y_k, equals $\frac{1}{2}$ if X_j is equal to Y_k, and equals 0 if X_j is less than Y_k. The symbol in the following formula

$$\sum_{(k=1)}^{N} c_k$$

is called a summation sign. It means that we take the sum of all of the c_k values, where k is from 1 to N. So, if N is equal to 12, then

$$\sum_{(k=1)}^{N} c_k = c_1 + c_2 + c_3 + \cdots + c_{12}.$$

References

1. Lusted LB. Signal detectability and medical decision-making. Science 1971; 171: 1217–1219.
2. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39:561–577.
3. Campbell G. General methodology I: advances in statistical methodology for the evaluation of diagnostic and laboratory tests. Stat Med 1994; 13:499–508.
4. Dorfman DD, Alf E Jr. Maximum likelihood estimation of parameters of signal detection theory: a direct solution. Psychometrika 1968; 33:117–124.
5. Dwyer AJ. In pursuit of a piece of the ROC. Radiology 1997; 202:621–625.
6. Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging 1989; 29:307–335.
7. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8:283–298.
8. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J Math Psychol 1975; 12:387–415.
9. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.
10. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837–844.
11. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986; 21:720–733.
12. Metz CE. Some practical issues of experimental design and data analysis in radiologic ROC studies. Invest Radiol 1989; 24: 234–245.
13. McClish DK. Analyzing a portion of the ROC curve. Med Decis Making 1989; 9:190–195.
14. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology 1996; 201:745–750.
15. Mushlin AI, Detsky AS, Phelps CE, et al. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. JAMA 1993; 269:3146–3151.
16. Iinuma G, Ushio K, Ishikawa T, Nawano S, Sekiguchi R, Satake M. Diagnosis of gastric cancers: comparison of conventional radiography and digital radiography with a 4 million-pixel charge-coupled device. Radiology 2000; 214:497–502.