Analyzing a Portion of the ROC Curve

Donna Katzman McClish

Med Decis Making 1989 9: 190
DOI: 10.1177/0272989X8900900307

Analyzing a Portion of the ROC Curve

DONNA KATZMAN McCLISH, PhD

The area under the ROC curve is a common index summarizing the information contained in the curve. When comparing two ROC curves, though, problems arise when interest does not lie in the entire range of false-positive rates (and hence the entire area). Numerical integration is suggested for evaluating the area under a portion of the ROC curve. Variance estimates are derived. The method is applicable for either continuous or rating scale binormal data, from independent or dependent samples. An example is presented which looks at rating scale data of computed tomographic scans of the head with and without concomitant use of clinical history. The areas under the two ROC curves over an a priori range of false-positive rates are examined, as well as the areas under the two curves at a specific point.
Key words: ROC curves; area under the curve; numerical integration; method of differentials; binormal data. **(Med Decis Making 1989;9:190–195)**

A very popular tool used to evaluate screening and diagnostic tests and algorithms is the receiver operating characteristic (ROC) curve.[1-8] The ROC curve provides a graphic representation of the tradeoff between false-positive rates and true-positive rates.

A number of indices have been used to summarize the information contained in the ROC curve. Most popular is the area under the curve, which has been called the best index of detectability.[4-9] Others include the index d, d′, d_e, and β.[4-9] A drawback to these indices is that they are global tests that evaluate the performance of the ROC curve over its entire range. But there are times when this may not be appropriate.

When comparing two ROC curves, the curves may be similar (different) over some range of clinical interest, while one may be superior to (the same as) the other over the remainder of the curve. This subtlety may not be reflected in a global test with such measures as the entire area under the curve. There may also be times when one is interested in only a restricted range of false-positive rates. For example, if we know a priori that a particular diagnostic test would not be useful if its false-positive rate were greater than 0.2, say, we would want to restrict our attention to that portion of the ROC curve.

One solution would be to compare ROC curves at individual points on the curve. McNeil and Hanley[10] have developed methods to compare binormal ROC curves at any fixed false-positive or true-positive rate. The problem remains when there are more than one or two false-positive rates of interest at which ROC curves might be compared. Even if the false-positive rates are chosen a priori, it is really not appropriate to do multiple statistical tests without some form of statistical adjustment to keep the overall Type I error small.

An alternate solution would be to compare areas under only a portion of the ROC curve. This paper will show how to determine the area under any portion of an ROC curve in the case of binormal data and how to compare two such areas. The methods we present are applicable for either continuous or rating scale data, and for ROC curves based on either independent or dependent samples.

Methods

AREA UNDER A PORTION OF THE ROC CURVE

Most ROC curves are generated under the assumption that the underlying data either are binormal or can be so transformed by a monotone transformation. In fact, Hanley[11] and Swets[5,6] have shown that most ROC curves are indistinguishable from the binormal model. In that case, without loss of generality, we will assume that data are realizations of independent normally distributed random variables, $X \sim N(\mu_x, \sigma_x)$ and $Y \sim N(\mu_y, \sigma_y)$, where X represents data from patients without disease and Y represents data from those with disease. We can define false-positive and false-negative rates as:

$$FP(c) = P(X > c) = 1 - \Phi[(c - \mu_x)/\sigma_x] = \Phi[(\mu_x - c)/\sigma_x] \quad (1)$$

and

$$TP(c) = P(Y > c) = \Phi[\mu_y - c)/\sigma_y] \quad (2)$$

Received February 16, 1988, from the Department of Biostatistics, Medical College of Virginia, Richmond, Virginia. Revision accepted for publication February 7, 1989.

Address correspondence and reprint requests to Dr. McClish: Department of Biostatistics, Medical College of Virginia, Box 32, MCV Station, Richmond, VA 23298.

where c is the cutoff decision point and Φ is the cumulative normal distribution function.

The ROC curve can be traced out by the functions[12]

$$[FP(c), TP(c)] = \{\Phi[(\mu_x - c)/\sigma_x], \Phi[(\mu_y - c)/\sigma_y]\} \quad -\infty < c < \infty$$

If we want to look at the area under the ROC curve from some point c_1 to the point c_2, this is by definition

$$A = \int_{c_1}^{c_2} TP(c)\, FP'(c)\, dc$$

$$= \int_{c_1}^{c_2} \Phi[(\mu_y - c)/\sigma_y]\, \phi[(\mu_x - c)/\sigma_x] \times \left(-\frac{1}{\sigma_x}\right) dc \qquad (3)$$

We can simplify the integral somewhat by a change of variables. Let $v = (\mu_x - c)/\sigma_x$ so that $dv = -dc/\sigma_x$ and $c = \mu_x - \sigma_x v$. Then A can be written

$$A = \int_{c_1'}^{c_2'} \Phi[(\mu_y - (\mu_x - \sigma_x v))/\sigma_y]\, \phi(v)\, dv$$

$$= \int_{c_1'}^{c_2'} \Phi[a + bv]\, \phi(v)\, dv \qquad (3')$$

with $c_i' = (\mu_x - c_i)/\sigma_x$, $a = (\mu_y - \mu_x)/\sigma_y$ and $b = \sigma_x/\sigma_y$.

Now generally we think in terms of the area under the curve between certain false-positive rates, as opposed to the cutoff or decision points that correspond to those false-positive rates. But the one-to-one correspondence between the cutoff point c_i and the false-positive rate $FP_i = FP(c_i)$, as expressed in equation 1, allows the substitution

$$c_i' = (\mu_x - c_i)/\sigma_x = \Phi^{-1}(FP_i) \qquad (4)$$

in equation 3′. Thus one need not explicitly consider the value of the cutoff point c_i. The limits of integration will be $c_i' = \Phi^{-1}(FP_i)$ where FP_i i = 1,2 are the two false-positive rates defining the interval of interest. It should be clear that if we take $c_1' = -\infty$ and $c_2' = +\infty$ then equation 3′ evaluates the area under the entire ROC curve, and if $c_1' = c_2'$ then this is essentially the degenerate case of one false-positive rate.

The integral equation 3′ can be evaluated through numerical integration.[13] If continuous data are available, then $\Delta = \mu_y - \mu_x$ can be estimated as $\hat{\Delta} = \bar{Y} - \bar{X}$ and $\hat{\sigma}_y^2 = s_y^2$ and $\hat{\sigma}_x^2 = s_x^2$. If rating data are used, then the Dorfman and Alf method as implemented by Metz[14,15] or Swets and Pickett[8] can be used to estimate a and b directly.

The area under a portion of the ROC curve is still bounded by 0.5 and 1. But tighter bounds can be derived. The partial area of the ROC curve is bounded above by the area of the rectangle that encloses it, i.e., the rectangle with corners $(FP_1,0)$, $(FP_1,1)$, $(FP_2,0)$, $(FP_2, 1)$. This rectangle has sides of length 1 and $FP_2 - FP_1$, giving an area of $FP_2 - FP_1$.

The lower bound for the partial area can be found by looking at the trapezoid with corners $(FP_1,0)$, (FP_1,FP_1), $(FP_2,0)$, (FP_2,FP_2). This is the trapezoid bounded above by the line $y = x$ and below by the horizontal axis. The area is $(FP_2 - FP_1)(FP_2 + FP_1)/2$. Thus we have that the partial area has the somewhat tighter bounds

$$\frac{1}{2}(FP_2 - FP_1)(FP_2 + FP_1) \le A \le FP_2 - FP_1 \qquad (5)$$

Notice that we could "standardize" the partial area by dividing by the maximum, providing a measure with maximum of 1:

$$\frac{1}{2}(FP_2 + FP_1) \le \frac{A}{FP_2 - FP_1} \le 1 \qquad (6)$$

Some may still find some difficulty in interpreting the partial area as a measure of discrimination. This is because, while the maximum values of two partial area estimates may be the same, the ranges of possible values may be quite different. For example, suppose that the area underneath an ROC curve between false-positive rates 0.7 and 0.9 is A = 0.18. The maximum value this area could attain is 0.2 while, the minimum is 0.16. An area of 0.18 found between false positives 0.3 and 0.5 would have the same maximum of 0.2, but a minimum value of 0.08. We would probably not want to value the two areas the same. One alternative is to use the transformation

$$\frac{1}{2}\left[1 + \frac{A - \min}{\max - \min}\right] \qquad (7)$$

for interpretation purposes. This transformation would make an entire area simply equal to the area (since max = 1, min = 0.5). In addition, it has a maximum value of 1 (when A = max) and a minimum value of 1/2 (when A = min). Thus, this transformation lets one view the partial area on the same scale as the total area.

For the particular example described above, the first area would be transformed to 0.75 while the second partial area would be transformed to 0.92. This might imply that the diagnostic test evaluated by the second area had superior discrimination in the portion of the curve under consideration.

VARIANCE OF THE PARTIAL AREA

To determine the variance of the area estimate, we use the method of differentials[16] to get an approximation. If the area A is a function of parameters $U_1, \ldots, U_k$ then the general formula for the variance is

$$\begin{aligned}\mathrm{Var}(A(U_1, U_2, \ldots U_k)) &= \sum_{i=1}^{k} (\partial A/\partial\theta_i)^2 \,\mathrm{Var}(U_i) \\ &\quad + 2\sum_{i<j} (\partial A/\partial\theta_i)(\partial A/\partial\theta_j)\,\mathrm{Cov}(U_i, U_j) \end{aligned} \tag{8}$$

where $\partial A/\partial\theta_i$ denotes the partial derivative of A with respect to U_i, evaluated at the mean value θ_i. This general formula applies regardless of whether the false-positive limits of the partial area are chosen a priori or are random.

When continuous data are used to generate the ROC curve and the partial area is evaluated between a priori chosen false-positive rates, equation 8 resolves to

$$\begin{aligned}\mathrm{Var}(A(\hat{\Delta}, s_x^2, s_y^2)) &= \left(\frac{\partial A}{\partial \Delta}\right)^2 \mathrm{Var}(\hat{\Delta}) + \left(\frac{\partial A}{\partial \sigma_x^2}\right)^2 \mathrm{Var}(s_x^2) \\ &\quad + \left(\frac{\partial A}{\partial \sigma_y^2}\right)^2 \mathrm{Var}(s_y^2) \end{aligned} \tag{9}$$

Notice that since X and Y are independent, the covariance terms will be zero.

To calculate the variance, then, we need to know the partial derivatives of the area with respect to Δ, σ_x^2 and σ_y^2 as well as the variances of the parameter estimates. The partial derivatives have been derived elsewhere, and the derivations are available from the author upon request. The three derivatives are:

$$\frac{\partial A}{\partial \Delta} = \frac{e^{-a^2/2(1+b^2)}}{\sqrt{2\pi(1+b^2)\sigma_y^2}} [\Phi(\tilde{c}_2) - \Phi(\tilde{c}_1)] \tag{10}$$

$$\begin{aligned}\frac{\partial A}{\partial \sigma_x^2} &= \frac{e^{-a^2/2(1+b^2)}}{4\pi(1+b^2)\sigma_x\sigma_y} [e^{-\tilde{\tilde{c}}_1} - e^{-\tilde{\tilde{c}}_2}] \\ &\quad - \frac{abe^{-a^2/2(1+b^2)}}{2\sigma_x\sigma_y\sqrt{2\pi}(1+b^2)^{3/2}} [\Phi(\tilde{c}_2) - \Phi(\tilde{c}_1)] \end{aligned} \tag{11}$$

and

$$\frac{\partial A}{\partial \sigma_y^2} = -\frac{a}{2\sigma_y}\frac{\partial A}{\partial \Delta} - b^2\frac{\partial A}{\partial \sigma_x^2} \tag{12}$$

where

$$\tilde{c}_i = [\Phi^{-1}(FP_i) + ab/(1 + b^2)]\sqrt{1 + b^2}$$

and

$$\tilde{\tilde{c}}_i = (\tilde{c}_i)^2/2$$

The variances of the parameter estimates are

$$\mathrm{Var}(\hat{\Delta}) = \sigma_y^2/n_y + \sigma_x^2/n_x \tag{13}$$

$$\mathrm{Var}(s_x^2) = 2\sigma_x^4/(n_x - 1) \tag{14}$$

$$\mathrm{Var}(s_y^2) = 2\sigma_y^4/(n_y - 1) \tag{15}$$

Calculation of the area, then, involves substituting the partial derivatives, equations 10–12, and variances, equations 13–15, into equation 9, using the usual estimates of Δ, σ_x^2, σ_y^2. We can verify that the variance is correct by considering the limiting results as FP_1 approaches 0.0 and FP_2 approaches 1.0. In the limit, $\tilde{c}_1 = -\infty$, $\tilde{c}_2 = +\infty$, and $\tilde{\tilde{c}}_i = +\infty$, $i = 1,2$ so that $\Phi(\tilde{c}_1) = 0.0$, $\Phi(\tilde{c}_2) = 1.0$, $\exp(-\tilde{\tilde{c}}_1) = \exp(-\tilde{\tilde{c}}_2) = 0.0$. The resulting partial derivatives are those that apply to the area under the entire curve.

The variance for the area under a portion of the curve when rating data have been used follows in a similar manner. We have

$$\begin{aligned}\mathrm{Var}(A(a,b)) &= \left(\frac{\partial A}{\partial a}\right)^2 \mathrm{Var}(\hat{a}) + \left(\frac{\partial A}{\partial b}\right)^2 \mathrm{Var}(\hat{b}) \\ &\quad + 2\left(\frac{\partial A}{\partial a}\right)\left(\frac{\partial A}{\partial b}\right) \mathrm{Cov}(\hat{a},\hat{b}) \end{aligned} \tag{16}$$

where there now is a correlation between a and b.

$$\frac{\partial A}{\partial a} = \frac{e^{-a^2/2(1+b^2)}}{\sqrt{2\pi(1+b^2)}} [\Phi(\tilde{c}_2) - \Phi(\tilde{c}_1)] \tag{17}$$

and

$$\begin{aligned}\frac{\partial A}{\partial b} &= \frac{e^{-a^2/2(1+b^2)}}{2\pi(1+b^2)} (e^{-\tilde{\tilde{c}}_1} - e^{-\tilde{\tilde{c}}_2}) \\ &\quad - \frac{abe^{-a^2/2(1+b^2)}}{\sqrt{2\pi}(1+b^2)^{3/2}} [\Phi(\tilde{c}_2) - \Phi(\tilde{c}_1)] \end{aligned} \tag{18}$$

where $\tilde{c}_i$ and $\tilde{\tilde{c}}_i$ are as above.

The variance and covariance terms are calculated as part of the Swets and Pickett[8] and Metz[14,15] programs, along with the estimates of a and b. Again, all of these estimates can be substituted into equation 16 to yield an estimate of the variance of the partial area.

COMPARING AREAS

Suppose we want to compare the area under a portion of two ROC curves, or test a hypothesis about an individual area. Then we must know something about the distribution of the area estimate. The area under the entire ROC curve has asymptotically a normal distribution[17] except when A is close to one. Thus, it seems reasonable that the partial area would also follow a normal distribution.

The test statistic for comparing the partial areas of two ROC curves is

$$Z = \frac{A_1 - A_2}{\sqrt{\text{Var}(A_1 - A_2)}} \tag{19}$$

If the two areas are estimated from independent samples then $V_I(A_1 - A_2) = \text{Var}(A_1) + \text{Var}(A_2)$. On the other hand, if the two areas are from dependent samples (say, derived from two sets of measurements on the same people) then the variance of $A_1 - A_2$ has a covariance term associated with it and

$$\text{Var}_D(A_1 - A_2) = \text{Var}(A_1) + \text{Var}(A_2) - 2\text{Cov}(A_1, A_2)$$

Just as with the variances, the method of differentials[15] can be used to approximate the covariance term. The general formula for the covariance is

$$\begin{aligned}\text{Cov}(A_1(U_1, U_2, \ldots, U_k), A_2(U_1, U_2, \ldots, U_k)) &= \sum_{i=1}^{k} (\partial A_1/\partial\theta_i)(\partial A_2/\partial\theta_i)\text{Var}(U_i) \\ &+ \sum_{i \neq j} (\partial A_1/\partial\theta_i)(\partial A_2/\partial\theta_j)\text{Cov}(U_i, U_j)\end{aligned} \tag{20}$$

For continuous data the formula is

$$\begin{aligned}\text{Cov}(A_1, A_2) = &\left(\frac{\partial A_1}{\partial \Delta_1}\right)\left(\frac{\partial A_2}{\partial \Delta_2}\right)\text{Cov}(\hat{\Delta}_1, \hat{\Delta}_2) \\ &+ \left(\frac{\partial A_1}{\partial \sigma^2_{x_1}}\right)\left(\frac{\partial A_2}{\partial \sigma^2_{x_2}}\right)\text{Cov}(s^2_{x_1}, s^2_{x_2}) \\ &+ \left(\frac{\partial A_1}{\partial \sigma^2_{y_1}}\right)\left(\frac{\partial A_2}{\partial \sigma^2_{y_2}}\right)\text{Cov}(s^2_{y_1}, s^2_{y_2})\end{aligned} \tag{21}$$

where

$$\begin{aligned}\text{Cov}(\hat{\Delta}_1, \hat{\Delta}_2) &= \rho_x \sigma_{x_1}\sigma_{x_2}/n_x + \rho_y \sigma_{y_1}\sigma_{y_2}/n_y \\ \text{Cov}(s^2_{x_1}, s^2_{x_2}) &= 2\rho_x \sigma^2_{x_1}\sigma^2_{x_2}/(n_x - 1) \\ \text{Cov}(s^2_{y_1}, s^2_{y_2}) &= 2\rho_y \sigma^2_{y_1}\sigma^2_{y_2}/(n_y - 1)\end{aligned} \tag{22}$$

and $\rho_v(\rho_x)$ is the correlation between the two sets of dependent measurements in the diseased (nondiseased) population. These can be estimated using sample variances and correlation estimates.

If the areas are derived from rating scale data then the covariance between two partial areas is

$$\begin{aligned}\text{Cov}(A_1, A_2) = &\left(\frac{\partial A_1}{\partial a_1}\right)\left(\frac{\partial A_2}{\partial a_2}\right)\text{Cov}(\hat{a}_1, \hat{a}_2) \\ &+ \left(\frac{\partial A_1}{\partial a_1}\right)\left(\frac{\partial A_2}{\partial b_2}\right)\text{Cov}(\hat{a}_1, \hat{b}_2) \\ &+ \left(\frac{\partial A_1}{\partial b_1}\right)\left(\frac{\partial A_2}{\partial a_2}\right)\text{Cov}(\hat{b}_1, \hat{a}_2) \\ &+ \left(\frac{\partial A_1}{\partial b_1}\right)\left(\frac{\partial A_2}{\partial b_2}\right)\text{Cov}(\hat{b}_1, \hat{b}_2)\end{aligned} \tag{23}$$

The partial derivatives for each area are computed from equations 17 and 18, while the Swets and Pickett[8] or Metz[14,15] implementation of the Dorfman and Alf method provides the four covariances needed.

A couple of small simulation studies were conducted to determine whether it was reasonable to use the critical values from the standard normal tables to compare two partial areas, to test hypotheses about individual areas, or to construct confidence intervals. The simulations used QuickBASIC Version 3.0[18] on a Zenith 386 personal computer. Uniform random numbers were generated using the random-number generator described by Wichman and Hill[19] and transformed to normal random deviates using the Box-Muller method.

For comparing two partial areas, performance was measured by the accuracy of the significance levels $\alpha = 0.10, 0.05, 0.01$ for the test statistics

$$Z = \frac{A_1 - A_2}{\sqrt{\text{Var}(A_1 - A_2)}} \tag{24}$$

For individual partial areas, we evaluated the performance of the test statistic

$$Z = \frac{A - A_0}{\sqrt{\text{Var}(A)}} \tag{25}$$

This is essentially equivalent to comparing the partial area A with a normal distribution with mean A_0 and variance Var(A).

For each simulation run, we generated two ROC curves from identical underlying binormal distributions. Data representing 50 subjects without disease were assumed to be distributed normally with mean 0 and variance 1, while 50 subjects with disease would have values that were normally distributed with mean Δ and variance 1. We chose two values of Δ, $\Delta = 1.0$ and 1.5 (corresponding to total areas of 0.760 and 0.856,

Table 1 • Simulation Results for Individual Areas*

	Δ = 1.0				Δ = 1.5			
	Lower Curve		Upper Curve		Lower Curve		Upper Curve	
	Z	Z*	Z	Z*	,Z	Z*	Z	Z*
α = 0.10	o.113	0.099	0.119	0.105	0.101	0.093	0.143	0.111
α = 0.05	0.063	0.051	0.075	0.055	0.057	0.043	0.099	0.061
α = 0.01	0.018	0.013	0.039	0.013	0.025	0.011	0.064	0.014

*Lower curve, $FP_1 = 0.05$, $FP_2 = 0.30$; upper curve, $FP_1 = 0.50$, $FP_2 = 0.80$.

Table 2 • Simulation Results for Comparing Two Curves*

	Δ = 1.0				Δ = 1.5			
	Lower Curve		Upper Curve		Lower Curve		Upper Curve	
	Z	Z*	Z	Z*	Z	Z*	Z	Z*
α = 0.10	0.104	0.103	0.099	0.112	0.095	0.087	0.063	0.110
α = 0.05	0.064	0.057	0.045	0.046	0.053	0.049	0.019	0.043
α = 0.01	0.018	0.010	0.006	0.006	0.008	0.006	0.000	0.006

*Lower curve, $FP_1 = 0.05$, $FP_2 = 0.30$; upper curve, $FP_1 = 0.50$, $FP_2 = 0.80$.

respectively). Two partial areas were considered—the area in the lower part of the curve between $FP_1 = 0.05$ and $FP_2 = 0.30$ and the area in the upper part of the curve between $FP_1 = 0.50$ and $FP_2 = 0.80$.

One thousand runs of the data were simulated to compare two partial areas. This also provided 2,000 (1,000 per curve) estimates of the individual partial areas. The proportion of times Z exceeded the appropriate critical value was calculated (tables 1 and 2). Performance was measured by the accuracy of the significance levels for $\alpha = 0.10$, 0.05, and 0.01 (that is, the proportions of times the test statistic Z = mean/standard deviation exceeded the 0.10, 0.05, and 0.01 levels of the standard normal distribution).

Tables 1 and 2 have the results of the simulation. As can be seen, all three significance levels for A were fairly well estimated when the lower portion of the curve was considered. On the other hand, when the false-positive limits were 0.5 and 0.8, the significance levels were not as accurate. For the individual areas, the critical value was liberal, allowing too many rejections, while for the comparison of two curves, the significance level was conservative.

Areas in the upper portion of the ROC curve are generally near their maximum values. This is analogous to the situation for the area under the entire ROC curve when it is close to 1, which is known to behave as a binomial proportion and exhibit a skewed distribution (for the size of sample generally used in medical studies). A solution is to transform the area. A common transformation for a parameter ψ bounded by 0 and 1 is

$$\theta = \ln\left(\frac{1 + \psi}{1 - \psi}\right)$$

In our situation we can use

$$\theta = \ln\left(\frac{m + A}{m - A}\right) = \ln\left(\frac{1 + A/m}{1 - A/m}\right) \quad (26)$$

where $m = FP_2 - FP_1$ is the maximum and A/m is as in equation 6. The variance of this transformation is

$$\mathrm{Var}\left\{\ln\left(\frac{m + A}{m - A}\right)\right\} = \frac{4m^2}{(m^2 - A^2)^2}\,\mathrm{Var}(A) \quad (27)$$

and the covariance is

$$\mathrm{Cov}(A_1, A_2) = \frac{4m^2}{(m^2 - A_1^2)(m^2 - A_2^2)}\,\mathrm{Cov}(A_1, A_2) \quad (28)$$

Tables 1 and 2 also evaluate the performances of statistical tests under this transformation, where Z* denotes the test statistic for the transformed parameter. Using the transformation θ of equation 26 in the test statistics, equations 24 and 25, Z* seems to bring the estimates closer to the actual significance levels, even with partial areas in the lower portion of the curve. Thus, it seems reasonable to use the normal distribution when evaluating the partial areas alone or examining the difference between the areas under corresponding portions of two ROC curves, with Z* the statistic of choice, particularly for the upper portion of the curve.

Example

McNeil and Hanley[10] published data on two (dependent) ROC curves comparing ratings of computed tomograms of the head with and without concomitant use of clinical history. Since the data were paired, they used the Metz program to provide estimates of the

parameters and the variance–covariance matrix. Values of the parameters for the ROC curve that used the clinical history (ch) were $\hat{a}_{ch} = 3.60$, $\hat{b}_{ch} = 1.29$, $Var(\hat{a}_{ch}) = 1.2288$, $Var(\hat{b}_{ch}) = 0.4043$, and $Cov(\hat{a}_{ch}, \hat{b}_{ch}) = 0.6495$. For the ROC curve that used no clinical history (nch) the parameter estimates were $\hat{a}_{nch} = 1.80$, $\hat{b}_{nch} = 0.59$, $Var(\hat{a}_{nch}) = 0.1552$, $Var(\hat{b}_{nch}) = 0.0533$, and $Cov(\hat{a}_{nch}, \hat{b}_{nch}) = 0.0681$. In addition, since the two ROC curves were estimated from data on the same individuals, covariances between parameters were $Cov(\hat{a}_{nch}, \hat{a}_{ch}) = 0.1712$, $Cov(\hat{a}_{ch}, \hat{b}_{nch}) = 0.0757$, $Cov(\hat{b}_{ch}, \hat{a}_{nch}) = 0.0542$, and $Cov(\hat{b}_{ch}, \hat{b}_{nch}) = 0.0378$.

In their report, NcNeil and Hanley compared the two curves at the single false-positive rate of 10%, and found the corresponding true positive rates of $TP_{ch} = 97.44\%$ and $TP_{nch} = 85.31\%$ to be significantly different ($Z = 2.25$). They concluded that at a false-positive rate of 10% the true-positive rate for computed tomograms with history was statistically higher than that obtained without using history.

If the interest were, instead, to compare the two curves at all false-positive rates less than or equal to 10% then it would be appropriate to look at the partial areas under the curves between $FP_1 = 0$ and $FP_2 = 0.10$. Since the two curves were derived from dependent samples, we can use equations 3′, 16–18, and 23 to compare them when the false-positive rates are less than 10%. If we do just that, the partial area under the ROC curve when clinical history is used is 0.0883537 with variance 0.0000669, while the partial area under the ROC curve from readings without the history is 0.0774144 with variance 0.0000718. The difference between these two partial areas is not significantly different from zero ($Z = 1.24$, $p = 0.215$). This result differs from that of McNeil and Hanley. While it may be that the two curves differ at the specific false-positive rate of 10%, they do not differ when the range of false-positive rates up to 10% is compared. Some may consider that looking at a range of false-positive rates will often be more appropriate than looking at a single point. In this instance the choice of analysis leads to different conclusions.

Discussion and Conclusion

We have suggested a solution to the problem of evaluating and comparing ROC curves over only a portion of all possible false-positive rates. We have shown that if a range of false-positive rates can be specified a priori, the partial area can be evaluated using numerical integration, explicit variances can be derived, and two partial areas can be compared statistically. The method requires a computer, but a straightforward BASIC program on a personal computer can evaluate the partial area in seconds.

The partial area also provides an intermediate solution to comparing diagnostic and screening tests—one that is intermediate between evaluating the entire area and a single point. Looking at the entire ROC curve gives a global picture of the diagnostic or screening test and requires few a priori assumptions about the desirability of any particular subset of false-positive or true-positive values. On the other hand, the total area gives the same weight to all portions of the curve. At the other extreme, choosing a single false-positive rate at which to compare true-positive rates may be difficult, if not impractical, as it may require precise knowledge about prevalence, costs, and other concerns. Considering the partial area is a compromise, implying a decision concerning suitability of some false-positive rates over others, yet allowing for an entire range of values to be evaluated.

References

1. Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. Arch Pathol Lab Med. 1986;110:13–20.
2. Erdreich LS, Lee ET. Use of relative operating characteristic analysis in epidemiology: a method for dealing with subjective judgment. Am J Epidemiol. 1981;114:649–61.
3. Hermann GA, Herrera N, Sugiura HT. Comparison of interlaboratory survey data in terms of receiver operating characteristic (ROC) indices. J Nucl Med. 1982;23:525–31.
4. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol. 1979;14:109–21.
5. Swets JA. Indices of discrimination or diagnostic accuracy. Their ROCs and implied models. Psychol Bull. 1986;6:100–17.
6. Swets JA. Forms of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. Psychol Bull. 1986;99:181–98.
7. Swets JA, Pickett, RM, Whitehead SF, et al. Assessment of diagnostic technologies. Science. 1979;205:753–9.
8. Swets JA, Pickett RM. Evaluation of diagnostic systems: methods from signal detection theory. New York: Academic Press, 1982.
9. Simpson AJ, Fitter MJ. What is the best index of detectability. Psych Bull. 1973;80:481–8.
10. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making. 1984;4:137–50.
11. Hanley JA. The robustness of the "binormal" assumption used in fitting ROC curves. Med Decis Making. 1988;8:197–203.
12. Kitchen JW. Calculus of one variable. Reading, MA: Addison-Wesley, 1968.
13. Gerald CF, Wheatley PO. Applied numerical analysis. Reading, MA: Addison-Wesley, 1984.
14. Metz CE. FORTRAN programs ROCFIT, CORROC and ROCPWR. Available from C. Metz, Dept of Radiology, University of Chicago, Chicago, Il.
15. Metz CE, Wang PL, Kronman HB. A new approach for testing the significance of differences between ROC curves from correlated data. In: Deconink F, ed. Information processing in medical imaging. The Hague: Nojhoff, 1984;432–45.
16. Elandt-Johnson RC, Johnson NL. Survival models and data analysis. New York: John Wiley and Sons, 1980.
17. Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. J Math Psychol. 1969;6:487–96.
18. Microsoft Corporation. Microsoft QuickBASIC Compiler. Redmond, WA: Microsoft Corporation, 1985.
19. Wichman B, Hill D. Building a random number generator. Byte. March 1987;127–8.