Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing

Rosman AS, Korsten MA*

Section of Gastroenterology and Medicine Program, James J. Peters VAMC, Bronx, NY and Mount Sinai School of Medicine (NYU), New York, NY, USA

Abstract

Summary receiver operating characteristics (sROC) analysis is a recently developed statistical technique that can be applied to meta-analysis of diagnostic tests. This technique can overcome some of the limitations associated with pooling the sensitivities and specificities of published studies. The sROC curve is initially constructed by plotting the sensitivity (true positivity) and false positivity (1 – specificity) of each study. After mathematical manipulation of the true and false positivities, linear regression is performed to calculate the slope and y-intercept. These coefficients are then entered into the sROC equation to generate the sROC curve. There are three commonly used methods to assess the accuracy of the test: the exact area under the curve (AUC) for the sROC function, the homogeneous AUC, and the index Q*. Statistical formulas can compare these values from different diagnostic tests. With the introduction of sROC software and better understanding of this method, the application of sROC analysis should continue to increase.

Key words: summary receiver operating characteristic (sROC) curve, meta-analysis, diagnostic tests.

Introduction

New diagnostic tests are continuously being introduced in medicine. Clinical trials usually attempt to demonstrate the superiority of a new test by comparing its sensitivity and specificity to a conventional test. In many cases, published studies comparing diagnostic tests may have inconclusive or conflicting results. A meta-analysis of the published studies can be useful in evaluating these comparisons. Most meta-analysis studies of diagnostic tests generally provide a pooled estimate of the sensitivity and specificity. Recently, meta-analysis studies have begun to use summary receiver operating characteristics (sROC) analysis. This method consists of constructing a receiver operating curve from published studies that have determined the sensitivity and specificity of a test. The sROC curve can then be evaluated by a variety of statistical techniques. As many clinicians are unfamiliar with sROC analysis, the purpose of this review is to summarize the principles involved in this technique. Although the authors are gastroenterologists and the hypothetical example relates to colonic polyps, the concepts apply to most tests in clinical medicine.

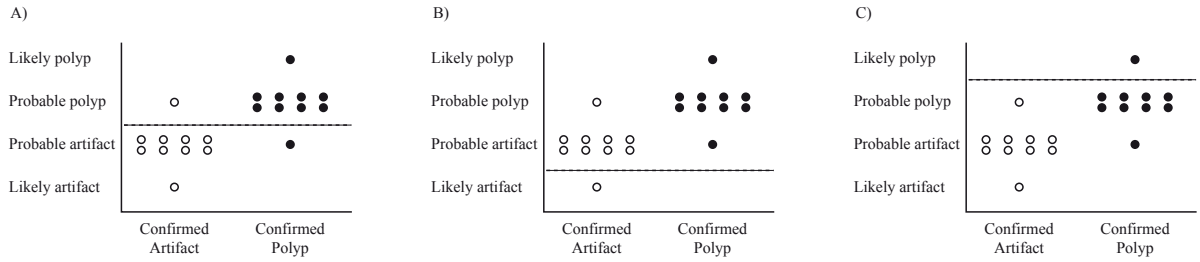
* CORRESPONDING AUTHOR:
Gastroenterology Practice (Suite F),
James J. Peters VA Medical Center
130 W. Kingsbridge Road, Bronx, NY 10468
Tel: 718 584-9000, ex. 3600
e-mail: Alan.Rosman@VA.Gov (Alan Rosman)

Received 02.07.2007 Accepted 10.07.2007

Limitations of sensitivity and specificity

The statistical technique for pooling sensitivities and specificities generally consists of weighing these rates by the inverse of their variance, summing the weighted rates, and dividing this sum by the sum of the inverses of their variance [1-3]. While pooled estimates of sensitivity and specificity are useful, they have a variety of limitations. First, sensitivity and specificity represent a trade-off as the threshold changes. By loosening the criteria (i.e., generally lowering the threshold), a test will become more sensitive but less specific. Raising the threshold will make a test more specific but less sensitive. Furthermore, sensitivity and specificity by themselves do not provide an overall evaluation of the accuracy of a test. For example, one may wish to compare two tests, one with a sensitivity and specificity both equal to 90% and the other with a sensitivity of 98% and a specificity of 80%. It is difficult to determine which one is more accurate from these characteristics.

Figure 1. **Three radiologists are evaluating the ability of CT colonography to discriminate polyps vs artifacts. In panel A, the first radiologists selects a threshold to optimize discrimination, resulting in a sensitivity and specificity both equal to 90%. In Panel B, the second radiologist selects a threshold to avoid missing polyps, resulting in a sensitivity of 100% and a specificity of 10%. In panel C, the third radiologist selects a threshold to minimize false positives, resulting in a sensitivity of 10% and a specificity of 100%**



In addition, pooling the sensitivities and specificities from multiple studies can occasionally result in a distorted estimate [4,5]. In a hypothetical example shown in *Fig. 1*, three radiologists are asked to determine whether filling defects seen on computed tomographic (CT) colonography represent polyps or artifacts. The radiologists classify the findings using the following scale: likely polyp, probable polyp, probable artifact, and likely artifact. All three of these radiologists have identical skills in visual perception and their samples of polyps are similar. When evaluating the ten endoscopically confirmed polyps, each one independently reports identical results as follows: likely polyp (n=1), probable polyp (n=8), and probable artifact (n=1). When evaluating the ten endoscopically confirmed artifacts, each one independently reports identical results as follows: likely artifact (n=1), probable artifact (n=8), and probable polyp (n=1). The radiologists are then asked to select a threshold and then report their results using a dichotomous scale: polyp or artifact. The first radiologist uses a threshold that maximizes the correct classification, resulting in a sensitivity of 90% and a specificity of 90% (*Fig. 1A*). The second radiologist, concerned about the legal consequences of a missed polyp, will diagnose a lesion as a polyp even it appears to be a probable artifact. The overly anxious radiologist will then report a sensitivity of 100% and a specificity of 10% (*Fig. 1B*). A third radiologist is more cavalier, believing that most polyps rarely progress to cancer. In order to minimize the number of colonoscopies being performed at his hospital, he will only diagnose a polyp on CT colonography if it has the appearance of a definite polyp. He then reports a sensitivity of 10% and a specificity of 100% (*Fig. 1C*). Using the statistical methods for pooling rates [1-3], the pooled sensitivity and specificity will both be equal to 67% (95% confidence intervals: 47-83%). These pooled values represent a distortion from the optimal sensitivity and specificity (both equal to 90%) that one would have obtained using an appropriate threshold. Analogous to Gresham's law of currency ("bad money drives good money out of circulation"), bad studies are able to distort good studies in a statistical pooling of sensitivity and specificity.

Advantages of sROC

In contrast, meta-analysis using sROC analysis will generate a composite statistic that reflects the discriminating ability of a diagnostic test. A traditional ROC curve plots the true positivity (sensitivity) as a function of false positivity (equal to 1-specificity) of a test at different thresholds [6]. In sROC analysis, one first plots the sensitivity and false positivity for each study [7]. A sROC curve is then constructed to fit these points. The area under the sROC curve is then determined to assess the discriminating ability. An area under the curve (AUC) close to 1.0 signifies that the test has almost perfect discrimination while an AUC close to 0.5 suggest poor discrimination [6]. An AUC significantly less than 0.5 would indicate that the criteria for "normal" and "abnormal" should be reversed. This scoring system is analogous to that of a true-false test. A student who knows all of the answers would score 100% while a student who knows none of the material should be able to score 50% by random guessing. A score significantly less than 50% would suggest an aberrant testing technique (e.g., confusing the symbols for true and false).

Many clinicians have not been well acquainted to sROC analysis for several reasons. This statistical technique is relatively new, having first been described in 1993 [7]. In addition, there have been relatively few published review articles which attempt to explain the mathematical techniques to most clinicians. Finally, many commonly used statistical software packages do not currently include sROC analysis. On the other hand, the techniques for performing sROC analysis are not complicated. The techniques include transformation using logarithms, linear regression, curve fitting, and understanding the relationship between integration and the area of the curve. These mathematical techniques are taught in most pre-medical school curriculums. Thus, sROC analysis should be accessible for most clinicians. In addition, sROC software is available to simplify the process.

Method of sROC

The sROC curve is a plot of the true positive rate (sensitivity) as a function of the false positive rate (1-specificity). The sROC equation is as follows:

$$TPR = \frac{e^{a/(1-b)} \times (FPR/(1-FPR))^{(1+b)/(1-b)}}{1+e^{a/(1-b)} \times (FPR/(1-FPR))^{(1+b)/(1-b)}} \quad (1)$$

where TPR is the true positive rate, FPR is the false positive rate, and a and b are coefficients which need to be deter-

Table 1. **Diagnostic characteristics of a test**

	Patients with disease	Patients without disease
Abnormal test	True positives	False positives
Normal test	False negatives	True negatives
Total	Number of patients with disease	Number of patients without disease

Table 2. **Transformation of the sensitivities and specificities in order to construct the sROC curve using linear regression**[1]

Study	Sensitivity (true positive rate)	Specificity	False positive rate (1-specificity)	Corrected sensitivity (true positive rate)[2]	Corrected false positive rate[2]	Logit of the true positive rate (TPR)[3]	Logit of the false positive rate (FPR)[4]	D (logit TPR - logit FPR)	S (logit TPR + logit FPR)
Optimal radiologist	9/10 (0.9)	9/10 (0.9)	1/10 (0.1)	9.5/11 (0.86)	1.5/11 (0.14)	1.85	-1.85	3.69	.00
Anxious radiologist	10/10 (1.0)	1/10 (0.1)	9/10 (0.1)	10.5/11 (0.95)	9.5/11 (.86)	3.04	1.85	1.20	4.89
Cavalier radiologist	1/10 (0.1)	10/10 (0.0)	10/10 (1.0)	1.5/11 (0.14)	0.5/11 (0.05)	-1.85	-3.04	1.20	-4.89

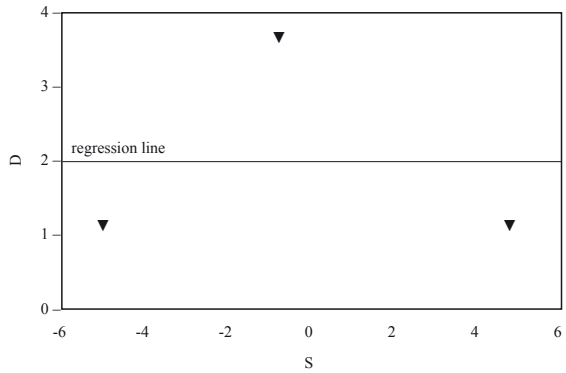
[1] – Number in parenthesis indicate rate; [2] – The rates were corrected by adding 0.5 to the numerator and 1.0 to the denominator; [3] – The logit of the true positive rate is the natural log of [true positive rate/(1-true positive rate)]; [4] – The logit of the false positive rate is the natural log of [false positive rate/(1-false positive rate)]

mined in order to fit the sROC curve [8]. At a first glance, the equation appears to be too daunting for a simple curve fitting technique. Nevertheless, Moses et al. [7] proposed a method that is relatively straightforward. The diagnostic characteristics of each study (i.e., true positivity and false positivity) can be transformed to other variables and then fitted using linear regression. As shown in *Tab. 1*, one first has to determine the number of true positives, false negatives, true negatives, and false positives for each study. One then has to calculate their sensitivity (true positives/number of patients with disease), specificity (true negatives/number of patients without disease), and false positive rate (1-specificity). If any study has a sensitivity or specificity that is either equal to 0 or 1.0, a continuity correction is required. Moses et al. [7] suggests adding 0.5 to all of the four cells of *Tab. 1* for all included studies.

One has to then transform the true positive rate (TPR) and false positive rates (FPR) into their corresponding logits. The logit of the true positive rate is the natural log of [TPR/(1-TPR)] while the logit of the false positive rate is the natural log of [FPR/(1-FPR)]. If the values for the TPR or FPR are either 0 or 1, the continuity correction will prevent obtaining an undefined value which would either occur by dividing by zero or taking the natural log of zero. One then calculates two parameters, D and S. D is defined as the difference of the logits (logit TPR - logit FPR) while S is defined as the sum of the logits (logit TPR + logit FPR). These transformations can be readily calculated using spreadsheet software by specifying the equations. Alternatively, many sophisticated statistical packages permit users to develop short programs for custom transformation of variables. *Tab. 2* illustrates the transformations of true and false positive rates using the example of the three radiologists.

These transformations will facilitate constructing the sROC curve using linear regression. A linear model using these transformed variables (D and S) can be constructed as follows:

Figure 2. **Linear regression of the transformed D and S variables from the hypothetical study of the three radiologists. D is defined as the difference of the logit of the true positive rate and the logit of the false positive rate. S is defined as the sum of the logit of the true positive rate and the logit of the false positive rate. The y-intercept, also known as coefficient *a*, is 2.03±1.18 while the slope, also known as coefficient *b* is 0±0.29. The sROC curve can be generated by plugging in the values for coefficients *a* and *b* in the sROC equation**



$$D = b \times S + a \qquad (2)$$

in which coefficients *a* and *b* are the y-intercept and slope, respectively. It can be mathematically shown that the coefficients a and b are identical to those in the sROC equation. These coefficients can be determined by performing linear regression using D as the dependent variable and S as the independent variable. Using the example of the three radiologists, coefficient *a* (the y intercept) is 2.03 while coefficient *b* (the slope) is 0 (shown in *Fig. 2*). The standard errors of coefficients *a* and *b* are 1.18 and 0.29, respectively.

One can construct the sROC curve by substituting the values for coefficients *a* and *b* into the sROC function (equation 1) and then plotting the true positive rate over the range

***Figure 3*. Comparison of two sROC curves. Panel A shows an sROC curve constructed assuming that coefficient *a* is 100 and coefficient *b* is 0. The curve approximates a right angle with an AUC near 1, signifying high discriminating ability. Panel B shows an sROC curve constructed assuming that coefficient *a* and *b* are zero. The curve is a 45 degree straight line with an AUC equal to 0.5, signifying poor discriminating ability**

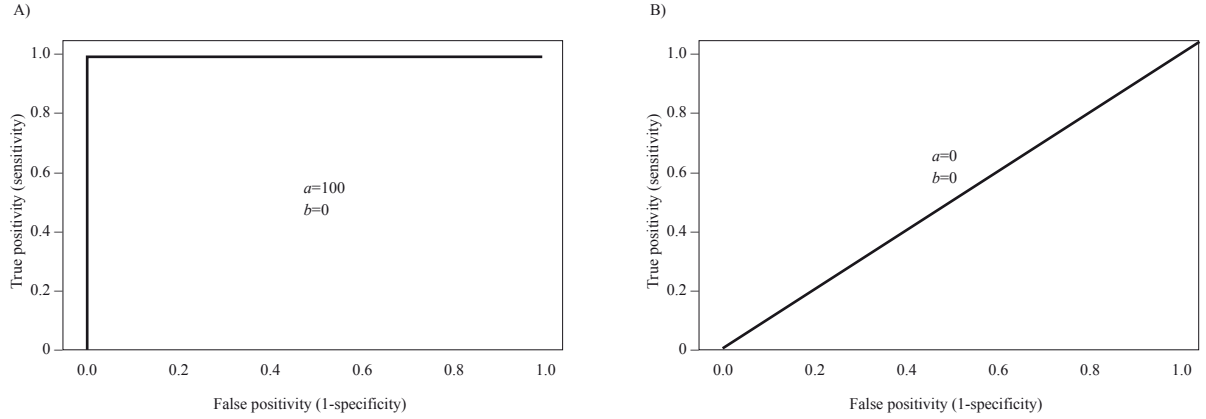
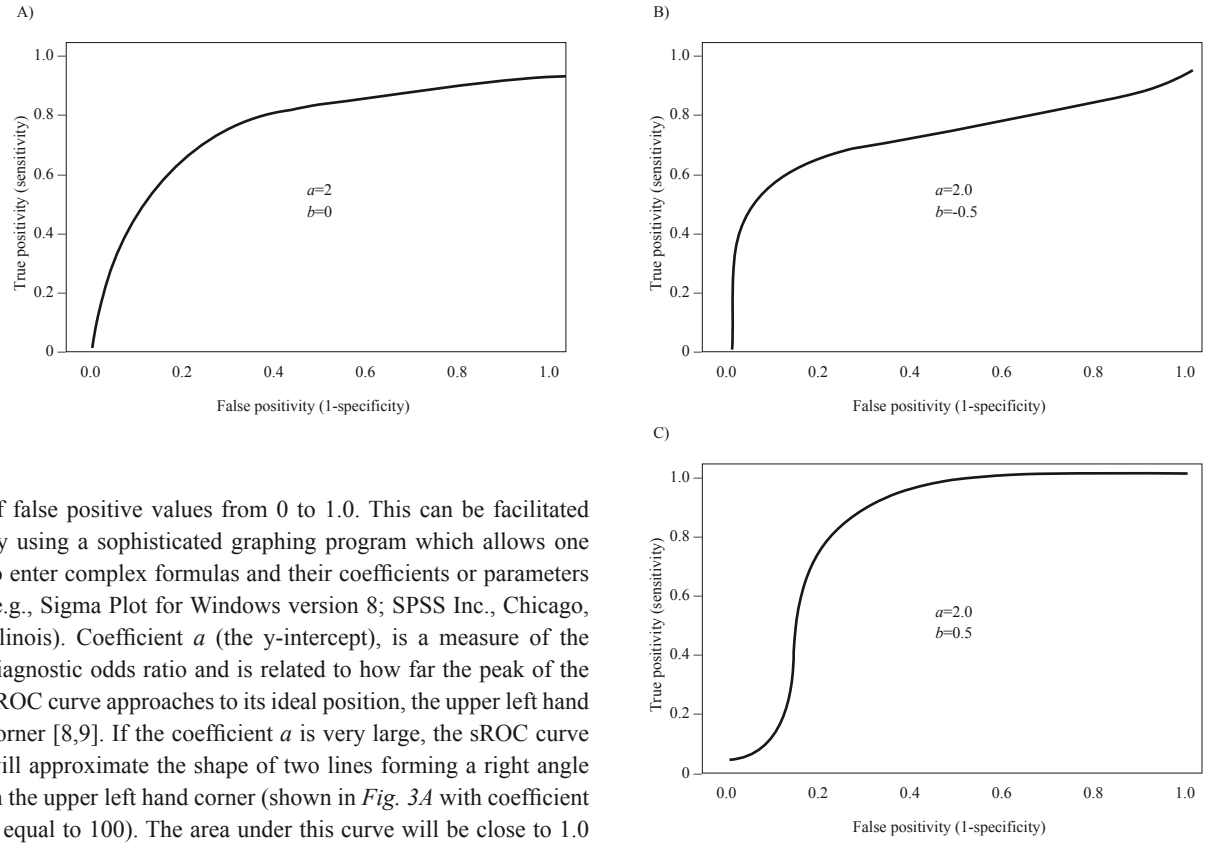


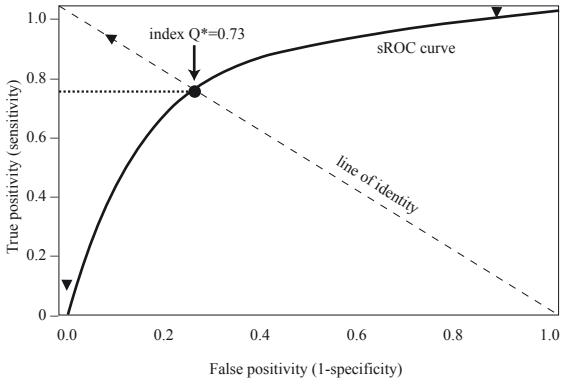
***Figure 4*. Comparison of three sROC curves. All three curves assume that coefficient *a* equals 2. Panel A shows an sROC curve constructed assuming that coefficients *b* equals 0, resulting in a symmetric shape. An sROC curve with coefficient *b* equals 0 is called the homogeneous case and has properties that make the AUC easier to calculate. Panels B and C assume the coefficient *b* is -0.5 and 0.5, respectively. These curves have an asymmetric shape and their exact AUCs are more complicated to calculate**



of false positive values from 0 to 1.0. This can be facilitated by using a sophisticated graphing program which allows one to enter complex formulas and their coefficients or parameters (e.g., Sigma Plot for Windows version 8; SPSS Inc., Chicago, Illinois). Coefficient *a* (the y-intercept), is a measure of the diagnostic odds ratio and is related to how far the peak of the sROC curve approaches to its ideal position, the upper left hand corner [8,9]. If the coefficient *a* is very large, the sROC curve will approximate the shape of two lines forming a right angle in the upper left hand corner (shown in *Fig. 3A* with coefficient *a* equal to 100). The area under this curve will be close to 1.0 and the test will have a high degree of discrimination. In contrast, if coefficient *a* is close to 0, the sROC curve will assume the shape of 45 degree line (shown in *Fig. 3B*) [4,8]. The area under this curve is 0.5 and the test will have poor discriminating ability. Coefficient *b* (i.e., the slope of the regression line) will affect the shape of the sROC curve. If coefficient *b*=0, the sROC curve will be symmetrical (shown in *Fig. 4A*). If coefficient *b* is significantly less than or greater than zero, the sROC curve will be markedly asymmetric (shown in *Fig. 4B* and *4C*) [8]. The sROC curve of the example of the three radiologists is shown in *Fig. 5*.

Once the sROC curve has been constructed, there are three methods to assess the discriminating ability of the test: the exact AUC, the homogeneous AUC, and the index Q* [8]. However, an exact calculation of the AUC for an sROC curve can be somewhat complicated. In many situations, the area under a curve can be calculated by integrating the function. However, the general sROC function (equation 1) cannot be integrated using calculus. Thus, the exact AUC needs to be determined by numerical integration [8] in which the sROC curve is bro-

Figure 5. **The sROC curve summarizing the diagnostic characteristics of the three radiologists. The true and false positivities of each radiologist are plotted using a triangle, ▼. The exact and homogenous AUC are both equal to 0.80±0.01. The "identity line" corresponds to the point in which sensitivity (true positivity) equals specificity (1 - false positivity). Index Q* can be visualized as the intersection of the sROC curve and the "line of identity". The point of intersection is shown as a closed circle, ●. Index Q* corresponds to the value of the true positivity at the point of intersection (shown as a horizontal line projecting from the point of intersection to the y-axis)**



ken up into very small rectangles or trapezoids (depending on the algorithm being used). Mathematical software programs such as Mathcad (MathSoft, Inc., Cambridge, MA), Mathematica (Wolfram Research, Inc., Champaign, IL), and Maple (MapleSoft, Waterloo, Ontario) can readily perform numerical integration of most functions. Walter [8] also provided the formulas for calculating the standard error of the exact AUC. This calculation also involves numerical integration and using the standard errors of the coefficients *a* and *b* and their covariance. These three terms are generally provided by by many statistical programs that perform linear regression. In the hypothetical example of the three radiologists, the exact AUC is 0.80 with a standard error of 0.01.

A simpler method for estimating the AUC was also provided by Walter [8]. Most sROC curves of clinical tests have a coefficient *b* close to zero. In mathematical models in which one assumes that one of the coefficients is zero, the example is then called the homogenous case. In the homogenous case of the sROC function, one assumes that coefficient *b*=0. The sROC equation can then be simplified to:

$$TPR = \frac{e^{a} \times (FPR/(1\text{-}FPR))}{1 + e^{a} \times (FPR/(1\text{-}FPR))} \qquad (3)$$

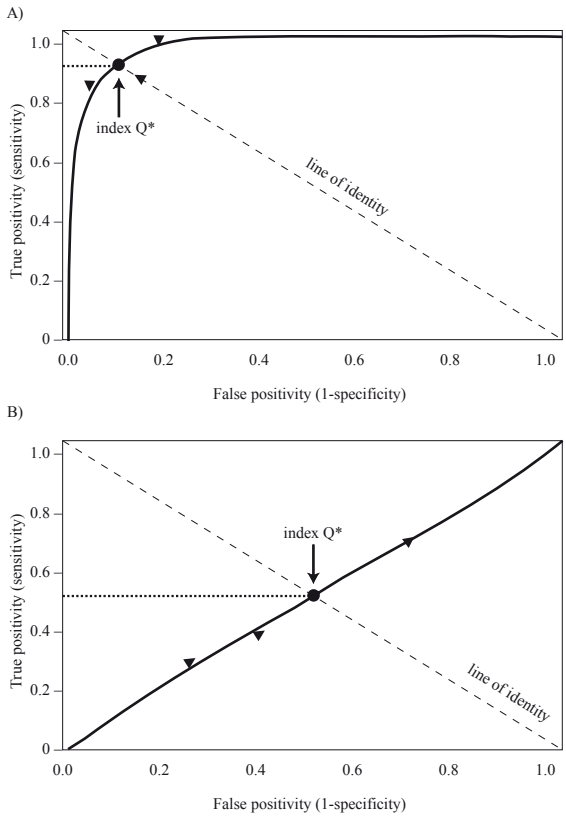
This function can be integrated with respect to FPR using standard calculus techniques, resulting in the following equation for the area under the homogeneous sROC curve:

$$\text{Homogeneus AUC} = \frac{e^{a} \times (e^{a}\text{-}1\text{-}a)}{(e^{a}\text{-}1)^{2}} \qquad (4)$$

The standard error of the homogeneous AUC [SE (homogenous AUC)] was shown by Walter [8] as follows:

$$SE(\text{Homogeneous}) = \frac{e^{a} \times [(e^{a}+1) \times a\text{-}2(e^{a}\text{-}1)] \times SE(a)}{(e^{a}\text{-}1)^{3}} \qquad (5)$$

Figure 6. **Comparison of a test with good discrimination (Panel A) with a test with poor discrimination (Panel B). Three hypothetical studies of the "good" test (shown in Panel A) have reported the following pairs of sensitivities and specificities: 90% and 88%, 98% and 85%, and 88% and 97%. The test's sROC curve and the points for the true and false positivities (using a triangle, ▼) are plotted in Panel A. The exact and homogeneous AUCs are both equal to 0.97±0.01. Index Q* corresponds to the point of intersection of the sROC curve and the "line of identity". The value for index Q*, 0.93±0.02, corresponds to the true positivity of the intersection (shown as a horizontal line projecting from the point of intersection to the y-axis). The exact AUC, homogeneous AUC, and index Q* are all close to 1, the value for a test with perfect discrimination. Three hypothetical studies of the "poor" test (shown in Panel B) have reported the following pairs of sensitivities and specificities: 40% and 60%, 30%, and 75%, and 70% and 30%. The test's sROC curve and the points for the true and false positivities (using a triangle, ▼) are plotted in Panel B. The exact and homogeneous AUCs are both equal to 0.51±0.01. The index Q*, the point of intersection of the sROC curve and the "line of identity" is also equal to 0.51±0.01. These values are close to 0.5, signifying that the test has poor discrimination**



where SE (a) is the standard error of coefficient *a* (i.e., the y-intercept). This value is generally provided by most statistical programs that perform linear regression. While these formulas are tedious to calculate by hand, they are much easier to compute than numerical integration. The homogenous AUC will be very close to the exact AUC in cases in which coefficient *b* is close to zero. When coefficient *a*=0, equation 3 will degenerate into an undefined value. Using an alternative formula, Walter [8] showed that the AUC of this case is equal 0.5.

Comparing two sROC curves

The AUC of two sROC curves can be statistically compared using a formula provided by Hanley and McNeil [10] for evaluating tradition ROC curves:

$$Z = \frac{AUC_1 - AUC_2}{\sqrt{SE(AUC_1)^2 + SE(AUC_2)^2}} \qquad (6)$$

where z is the z statistic, AUC_1 and AUC_2 are the area under the sROC curves of the two tests, and $SE(AUC_1)$ and $SE(AUC_2)$ are their standard errors. The p value for the z statistic can be determined by using a z table or an internet-based calculator.

Another method for assessing the accuracy of an sROC curve is the index Q*. This method is less intuitive to clinicians than AUC but is easier to calculate. The index Q* corresponds to the upper most point on the sROC curve in which true positivity (or sensitivity) equals specificity [7,8]. This can be shown graphically by drawing a "line of identity" in which true positivity = specificity on the sROC graph. As the x-axis on the sROC curve is the false positivity, one has to transform the variable specificity to 1-false positivity. Thus, the appropriate equation for this "line of identity" would be true positivity =1 – false positivity. This line would have a y-intercept of 1 and a slope of -1. Graphically, the index Q* corresponds to the value of the true positive rate at the point of the intersection of the sROC curve and the "line of identity" (shown in *Fig. 5*). A test close to ideal has an sROC curve that approximates a right angle; therefore, it would intersect the "line of identity" close to the upper left hand corner, resulting in an index Q* close to 1 (shown in *Fig. 6A*). In contrast, a test of poor discriminatory ability has an sROC curve that approximates a diagonal line. It would intersect the "line of identity" near their mid-points, resulting in an index Q* close to 0.5 (shown in *Fig. 6B*). The index Q* can also be calculated using the formula described by Moses et al. [7] and Walter [8]:

$$\text{index } Q^* = \frac{e^{a/2}}{1+e^{a/2}} \qquad (7)$$

The formula for the standard error of the index Q* is as follows:

$$SE(\text{index } Q^*) = \frac{e^{a/2}}{2(1+e^{a/2})^2} \times SE(a) \qquad (8)$$

where is the standard error of coefficient *a* (y-intercept).

The index Q* values of two tests can be statistically compared using a formula analogous to that of AUC as described by Moses et al. [7]:

$$z = \frac{\text{index } Q_1^* - \text{index } Q_2^*}{\sqrt{SE(\text{idex } Q_1^*)^2 + SE(\text{idex } Q_1^*)^2}} \qquad (9)$$

where z is the z statistic, index Q_1* and index Q_2* are the index Q* for test 1 and test 2, respectively, and $SE(\text{index } Q^*_1)$ and $SE(\text{index } Q^*_2)$ are their corresponding standard errors.

The following example will illustrate how sROC can be used to compare two hypothetical tests. The first test has good discrimination as three studies have reported the following pairs of sensitivities and specificities: 90% and 88%, 98% and 85%, and 88% and 97%. After applying linear regression, coefficients *a* and *b* are computed to be 5.08±0.65 and 0.06±0.43, respectively. *Fig. 6A* shows the sROC curve constructed using these coefficients. The exact and homogeneous AUCs are both equal to 0.97±0.01. These values are close to 1, suggesting that the test has nearly perfect discrimination. The point of intersection of the sROC curve and the "line of identity" has a true positivity (or sensitivity) equal to 0.93±0.02 which corresponds to the index Q*. The second test has poor discrimination as three studies have reported the following pairs of sensitivities and specificities: 40% and 60%, 30%, and 75%, and 70% and 30%. The coefficients *a* and *b* are 0.06 ±0.08 and -0.06±0.05, respectively. *Fig. 6B* shows the sROC curve constructed using these coefficients. The exact and homogenous AUCs are both equal to 0.51±0.01. Index Q* for this test (shown in *Fig. 6B*) is equal to 0.51±0.01. All of these values are close to 0.5, the value for a worthless test. Statistical comparisons of the exact and homogeneous AUCs result in a z-test of 25 ($p<0.0001$). Statistical comparisons of their index Q* result in a z-test of 17 ($p<0.001$).

Software has been developed in order to facilitate sROC analysis. Meta-test is a DOS based program that was developed by Dr. Joseph Lau at New-England Medical Center in Boston [11,12]. Meta-DiSc is a Windows based program that uses a graphic interface that was developed at the Clinical Biostatistics Unit at the Ramón y Cajal Hospital in Madrid [13].

Use of sROC

From 1995 to June 2007, approximately 180 published articles have discussed sROC analysis (SilverPlatter's MEDLINE, Ovid Technologies, New York). Furthermore, the number of papers has increased in the last 5 years. Published studies which used sROC analysis have evaluated CT colonography for polyp detection [14,15], diagnostic tests for hepatocellular carcinoma [16], diagnosing intravascular device-related bloodstream infection [17], ultrasonography for temporal arteritis [18], helical CT scans for diagnosing pulmonary embolism [19,20], and stress tests for risk stratification of coronary artery disease [21]. It is the hope of the authors that this paper will further familiarize clinicians with the principles of sROC analysis and further increase its application.

References

1. Midgette AS, Stukel TA, Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med Decis Making, 1993; 13: 253-7.

2. Yeung LT, King SM, Roberts EA. Mother-to-infant transmission of hepatitis C virus. Hepatology, 2001; 34: 223-9.

3. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess, 2005; 9: 1-113.

4. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med, 1994; 120: 667-76.

5. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol, 1995; 48: 119-30.

6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 1982; 143: 29-36.

7. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med, 1993; 12: 1293-316.

8. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med, 2002; 21: 1237-56.

9. Walter SD. The partial area under the summary ROC curve. Stat Med, 2005; 24: 2025-40.

10. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, 1983; 148: 839-43.

11. Engels EA, Terrin N, Barza M, Lau J. Meta-analysis of diagnostic tests for acute sinusitis. J Clin Epidemiol, 2000; 53: 852-62.

12. Gill CJ, Lau J, Gorbach SL, Hamer DH. Diagnostic accuracy of stool assays for inflammatory bacterial gastroenteritis in developed and resource-poor countries. Clin Infect Dis, 2003; 37: 365-75.

13. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol, 2006; 6: 31.

14. Rosman AS, Korsten MA. Meta-analysis comparing CT colonography, air contrast barium enema, and colonoscopy. Am J Med, 2007; 120: 203-10.

15. Purkayastha S, Athanasiou T, Tekkis PP, Constantinides V, Teare J, Darzi AW. Magnetic resonance colonography vs computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison. Colorectal Dis, 2007; 9: 100-11.

16. Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, Duca P. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. Am J Gastroenterol, 2006; 101: 513-23.

17. Safdar N, Fine JP, Maki DG. Meta-analysis: methods for diagnosing intravascular device-related bloodstream infection. Ann Intern Med, 2005; 142: 451-66.

18. Karassa FB, Matsagas MI, Schmidt WA, Ioannidis JP. Meta-analysis: test performance of ultrasonography for giant-cell arteritis. Ann Intern Med, 2005; 142: 359-69.

19. Harvey RT, Gefter WB, Hrung JM, Langlotz CP. Accuracy of CT angiography versus pulmonary angiography in the diagnosis of acute pulmonary embolism: evaluation of the literature with summary ROC curve analysis. Acad Radiol, 2000; 7: 786-97.

20. Hayashino Y, Goto M, Noguchi Y, Fukui T. Ventilation-perfusion scanning and helical CT in suspected pulmonary embolism: meta-analysis of diagnostic performance. Radiology, 2005; 234: 740-8.

21. Navare SM, Mather JF, Shaw LJ, Fowler MS, Heller GV. Comparison of risk stratification with pharmacologic and exercise stress myocardial perfusion imaging: a meta-analysis. J Nucl Cardiol, 2004; 11: 551-61.