Published in final edited form as:

J Comput Graph Stat. 2012 July 1; 21(3): 662–676. doi:10.1080/10618600.2012.663303.

A semiparametric separation curve approach for comparing correlated ROC data from multiple markers

Liansheng Larry Tang and

Department of Statistics, George Mason University, Fairfax, VA 22030

Xiao-Hua Zhou

Northwest HSR&D Center of Excellence, VA Puget Sound Health Care System, Seattle, WA 98101, USA; Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; arbin Medical University, Harbin, Heilongjiang, 150081, China

Liansheng Larry Tang: ltang1@gmu.edu; Xiao-Hua Zhou: azhou@u.washington.edu

Abstract

In this article we propose a separation curve method to identify the range of false positive rates for which two ROC curves differ or one ROC curve is superior to the other. Our method is based on a general multivariate ROC curve model, including interaction terms between discrete covariates and false positive rates. It is applicable with most existing ROC curve models. Furthermore, we introduce a semiparametric least squares ROC estimator and apply the estimator to the separation curve method. We derive a sandwich estimator for the covariance matrix of the semiparametric estimator. We illustrate the application of our separation curve method through two real life examples.

Keywords

Confidence band; Empirical distribution function; Least squares

1 Introduction

For biomarkers that generate continuous test results, receiver operating characteristic (ROC) curves are preferred for comparing the accuracy of biomarkers (Zhou et al., 2002), because an ROC curve combines all possible pairs of sensitivities (probabilities of identifying a diseased subject when the disease truly exists) and 1-specificities (probabilities of correctly ruling out a non-diseased subject when the disease is truly absent) from different decision thresholds, and thus describes the accuracy of biomarkers apart from decision thresholds. Many studies, especially in comparative biomarker studies, involve measuring multiple biomarkers on every subject. For example, in an Alzheimer disease (AD) biomarker study, cerebrospinal fluid samples were obtained via lumbar puncture on each of cognitively normal patients and AD patients (Clark et al., 2003). From each fluid sample, measurements of two most promising biomarkers reflecting the pathological AD processes, cerebrospinal fluid τ and $A\beta$, were taken. One of the main interests in this study was to investigate the accuracy of these two biomarkers to discriminate AD patients from cognitively normal patients. Another example is a popular pancreatic cancer biomarkers data set used in Wieand

Supplementary Materials: Online Appendix: Theorems 1 and 2 are proven. (appendix.pdf)

Online Supplement: The supplement contains the computer code and a dataset to illustrate how to use the proposed method. It can be obtained via a single download. (Supplement.zip)

The technical proofs, a dataset and the computer code are available in the online supplements.

et al. (1989). In this study, two biomarkers, CA 19-9 and CA 125, were measured on 90 pancreatic cancer patients and 51 pancreatitis patients. Main interest lies in comparing the diagnostic accuracy of these two biomarkers. In this article, we propose a new semiparametric separation curve method to compare ROC curves for this type of correlated biomarker data and derive large sample inference.

Let X be a continuous measurement of a biomarker on a subject from the diseased population with a survival function S_0 , and Y on a subject from the non-diseased population with a survival function S_1 . Without loss of generality, we assume that a larger biomarker measurement is more likely to indicate the disease status. At a decision threshold c , the corresponding true positive rate (TPR) and false positive rate (FPR) are written as follows:

$$S_0(c) = P(X > c), \text{ and } S_1(c) = P(Y > c),$$

respectively. An ROC curve is defined as a plot of TPRs vs FPRs when c varies in the range of $(-\infty, +\infty)$, and the ROC curve, hereafter denoted by $R(u)$, can be written as

$$R(u) = S_0(S_1^{-1}(u)), \text{ for } 0 \leq u \leq 1, \text{ where } u \text{ is some FPR.}$$

In this paper, we propose a separation curve method to determine over which range of FPRs the curves differ significantly. The separation curve method requires parameter estimators in the ROC model (1) to be introduced in the next section, and the estimator for the covariance matrix of these parameters. ROC methods for estimating these parameters and the covariance matrix have been proposed by various authors (Pepe, 1997; Cai, 2004; Zhang and Pepe, 2005; Tang and Zhou, 2009). Zou et al. (1997) and Lloyd (1998) introduced kernel smoothing ROC methods for estimating a single ROC curve by combining the kernel smoothers of S_0 and S_1 . These kernel methods can be extended to compare biomarkers. That is, the kernel smoothers of K biomarkers for diseased and non-diseased subjects are first obtained. The pair of smoothers for each biomarker are used to construct the estimated ROC curve. The ROC summary measures are then estimated using the ROC estimates. Inference can be performed by bootstrapping diseased and non-diseased subjects separately. However, the kernel method does not allow testing equality among biomarkers. An intuitive parametric least squares (PLS) ROC method proposed by Zhang and Pepe (2005) requires no iteration and thus takes much less computation time than the aforementioned methods. In this article, we will introduce a semiparametric least squares method to nonparametrically estimate the baseline function in the model (1). The proposed method allows modeling ROC curve differences parametrically so that the resulting separation curves are regular. The other advantages of the proposed method are that it is simple to implement, and can be used to test equality directly. In addition, it has an explicit form for the covariance matrix.

The article is organized as follows. In Section 2, we will propose a semiparametric separation curve method to find the region of FPRs where ROC curves are significantly different. The semiparametric least squares method is introduced in Section 3. We did simulation studies to evaluate and compare the robustness and the efficiency of the semiparametric least squares (SLS) method with existing methods. In Section 4, we illustrate the application of the newly proposed separation curve method using SLS estimators through the aforementioned real-life examples. Some discussions are presented in Section 5.

2 Determining a range of FPRs for which ROC curves differ

Suppose we have $K + 1$ biomarkers. The ROC curve of the k th biomarker is given by

$R_k(u) = S_{0,k}(S_{1,k}^{-1}(u))$, $k=0, 1, \dots, K$, where $S_{0,k}$ and $S_{1,k}$ are respective survival functions of the diseased and nondiseased for the k th biomarker. Its empirical estimator is

$\hat{R}_k(u) = \hat{S}_{0,k}(\hat{S}_{1,k}^{-1}(u))$. We propose a model for the $K + 1$ ROC curves

$$G^{-1}(R_k(u)) = H_0(u) + \alpha_k + \beta_k G_k^{-1}(u), \quad (1)$$

for known distribution functions G and G_k . For the model (1) to be identifiable, we let $\alpha_0 = \beta_0 = 0$, so that $R_0(u) = G(H_0(u))$. The baseline function H_0 is a transform of the ROC curve for the chosen baseline group and can be modeled parametrically or nonparametrically. The differences between ROC curves will be modeled parametrically. The parameters to be estimated are the other (α_k, β_k) and the baseline function $H_0(u)$. The advantage of the model is that the ROC curves, instead of underlying distributions of the biomarkers, are modeled. This is particularly useful when the underlying distributions of the biomarkers are on different scales.

The parameters α_k measure superiority of marker k over the baseline marker. The parameters β_k measure strength of interaction or difference in the shape of biomarkers. The β_k allows marker k to have a differently shaped ROC curve. The baseline function H_0 can be modeled parametrically or nonparametrically. When H_0 is given by a parametric form as

$H_0 = \alpha_0 + \beta_0 G_0^{-1}$ with unknown parameters α_0, β_0 and a known basis function G_0 , the model (1) becomes Zhang and Pepe (2005) model. When G_k are chosen to have median 0, then $G_k^{-1}(u) = 0$ relates to $u = 0.5$, and the α_k measures biomarker effectiveness at 50% FPR. If one is interested in the entire ROC curve, u can be chosen in the range of $(0, 1)$; if the partial ROC curve on $u \in [u_0, u_1]$ ($0 < u_0 < u_1 < 1$) is of primary interest, u can be chosen to be within the region of interest. It is important to note that interaction terms, $\beta_k G_k^{-1}(u)$, in the model (1) allow ROC curves to intersect.

ROC curves may be similar for some region of FPRs while one may be superior to another over other regions. Sometimes, we may only be interested in comparing ROC curves over a particular region of FPRs, say the region of FPRs of less than 0.1 in early cancer screening. If there is a prior region of interested FPRs, we can compare ROC curves over this prior region. However, in many applications, we do not have a prior set of FPRs to consider. To find a range of FPRs for the k th and reference ROC curves differ, we need to find the region of u that satisfies that $\hat{R}_k(u) - \hat{R}_0(u)$ in the range of $[u_0, u_1]$. We call

$$\Delta_k(u) = G^{-1}(R_k(u)) - G^{-1}(R_0(u)), k=1, \dots, K, \quad (2)$$

a separation curve between the k th and the reference ROC curves. Since the link function G is a monotone function, such the region is given by $\{u: \Delta_k(u) > 0\}$. Similar results follow when we are interested in finding the region of FPRs, where $\hat{R}_k(u) > \hat{R}_0(u)$, by identifying $\{u: \Delta_k(u) > 0\}$, assuming that G is an increasing function without loss of generality. A similar separation curve between the i th ROC curve and the j th ROC curve, $i, j = 1, \dots, K$ and $i \neq j$, defined by

$$\Delta_{i,j}(u) = G^{-1}(R_i(u)) - G^{-1}(R_j(u)), \quad (3)$$

models the difference between two transformed ROC functions.

The separation curve is estimated by substituting respective parameter estimators. To visualize variability associated with the estimated separation curve, we will construct its confidence band. Denote $\theta_k = (\alpha_k, \beta_k)^T$, and its estimator $\hat{\theta}_k$. Denote $\theta_{i,j} = (\alpha_i, \beta_i, \alpha_j, \beta_j)^T$, and its estimator $\hat{\theta}_{i,j}$. Let Σ_k be the covariance matrix of $\hat{\theta}_k$, and let $\Sigma_{i,j}$ be the covariance matrix of $\hat{\theta}_{i,j}$. The explicit expression of the confidence band is given in Theorem 1 below.

Theorem 1

Denote $H_k = (1, G_k^{-1}(u))$ and $\tilde{H}_{i,j} = (H_i, -H_j)$. Under mild regularity conditions, the $(1 - \alpha)100\%$ simultaneous confidence band of the separation curve between the k th ROC curve and the reference ROC curve in $[u_0, u_1]$ is constructed as follows:

$$H_k \hat{\theta}_k \pm \sqrt{\chi_{2,\alpha}^2 H_k \sum_k H_k^T},$$

where $\chi_{2,\alpha}^2$ is the α th quantile of the chi-square distribution with d.f. = 2; the $(1 - \alpha)100\%$ simultaneous confidence band of the separation curve between the i th ROC curve and the j th ROC curve in $[u_0, u_1]$ is constructed as follows:

$$\tilde{H}_{i,j}(\hat{\theta}_i, \hat{\theta}_j)^T \pm \sqrt{\chi_{4,\alpha}^2 \tilde{H}_{i,j} \sum_{i,j} \tilde{H}_{i,j}^T},$$

where $\chi_{4,\alpha}^2$ is the α th quantile of the chi-square distribution with d.f. = 4;

Proof. See Appendix A in the online appendix.

Theorem 1 provides a useful way to construct the approximate confidence band of a separation curve to indicate the region of FPRs for which there is significant difference between two ROC curves. This new separation curve approach is transformation invariant and the confidence band has a nice theoretical property. Such confidence band provides us a global measure of sampling variability of estimated separation curves. We can also visualize the difference between each pair of $K + 1$ estimated ROC curves by comparing the confidence band of corresponding separation curve with a horizontal line at $\Delta(u) = 0$. This will be illustrated in Section 4.

The separation curve provides a way to conduct hypothesis tests on the equality of ROC curves in a restricted range, $u_0 \leq u \leq u_1$, when we choose jump points within this range. Our model implies that the reference ROC curve and the k th ROC curve differ by θ_k . Consequently, testing the equality of these two ROC curves in $[u_0, u_1]$ is equivalent to testing $H_0: \theta_k = 0$. If $\hat{\theta}_k$ follows multivariate normality, the test statistic κ_k has an asymptotic chi-square distribution with two degrees of freedom under H_0 :

$$\kappa_k = (\hat{\theta}_k - \theta_k)^T \sum_k^{-1} (\hat{\theta}_k - \theta_k).$$

Testing the equality of two ROC curves of the i th and j th biomarkers in $[u_0, u_1]$ is also reduced to a χ^2 test. The null hypothesis becomes $H_0 : (\alpha_i - \alpha_j, \beta_i - \beta_j) = 0$. The resulting χ^2_4 statistic has the following expression:

$$\kappa_{i,j} = \begin{pmatrix} (\widehat{\alpha}_i - \widehat{\alpha}_j) - (\alpha_i - \alpha_j) \\ (\widehat{\beta}_i - \widehat{\beta}_j) - (\beta_i - \beta_j) \end{pmatrix}^T \left(A \sum_{i,j} A^T \right)^{-1} \begin{pmatrix} (\widehat{\alpha}_i - \widehat{\alpha}_j) - (\alpha_i - \alpha_j) \\ (\widehat{\beta}_i - \widehat{\beta}_j) - (\beta_i - \beta_j) \end{pmatrix},$$

where $A = (I_2, -I_2)$ with a 2×2 identity matrix I_2 .

3 A semiparametric least squares estimation procedure

In this section, we propose a semiparametric least squares procedure to estimate the unknown baseline function H_0 and the ROC parameter vector θ . The procedure is inspired by a simple transformation of equations in Model (1). The $K+1$ empirical ROC curves can be transformed by G^{-1} and the resulting model is given by

$$\begin{aligned} G^{-1} \left\{ \widehat{R}_0(u) \right\} &= H_0(u) + \varepsilon_0, \\ G^{-1} \left\{ \widehat{R}_1(u) \right\} &= H_0(u) + \alpha_1 + \beta_1 G^{-1}(u) + \varepsilon_1, \\ &\vdots \\ G^{-1} \left\{ \widehat{R}_K(u) \right\} &= H_0(u) + \alpha_K + \beta_K G^{-1}(u) + \varepsilon_K, \end{aligned} \quad (4)$$

where $\varepsilon_0, \dots, \varepsilon_K$ are zero mean random errors. The baseline function H_0 and the ROC parameter vector θ are then simultaneously estimated by solving for least squares estimators.

In practice, we use discretized values of $u = (u_1, \dots, u_P)^T$ which is a vector of FPR values within $[u_0, u_1]$ ($0 < u_0 < u_1 < 1$) for estimating ROC curves in the range of $[u_0, u_1]$, where P can be a fixed number. Also, for simplicity, we let $S(u) = (S(u_1), S(u_2), \dots, S(u_P))^T$ be any process or function S . Let $H_{0,p} = H_0(u_p), p = 1, \dots, P$. Denote the P points on the baseline ROC curve as $H_0 = (H_{0,1}, H_{0,2}, \dots, H_{0,P})^T$. The following three steps outline our estimation procedure:

- *Step 1:* Obtain K transformed empirical ROC curves by

$$\widetilde{Y}_0 = G^{-1} \left\{ \widehat{R}_0(u) \right\} \text{ and } \widetilde{Y}_k = G^{-1} \left\{ \widehat{R}_k(u) \right\}.$$

- *Step 2:* Perform linear regression on the PK dimensional vector

$$\widetilde{Y} = \left(\widetilde{Y}_0^T, \widetilde{Y}_1^T, \dots, \widetilde{Y}_K^T \right)^T;$$

$$\widetilde{Y} = L \left(H_0^T, \theta^T \right)^T + \varepsilon.$$

Here the $PK \times (P + 2K)$ matrix L is given by:

$$\mathbf{L} = \begin{pmatrix} \mathbf{I}_P & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_P & \mathbf{L}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_P & \mathbf{0} & \mathbf{L}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{I}_P & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L}_K \end{pmatrix}$$

with $P \times P$ identity matrices, $\mathbf{I}_P$, zero submatrices, $\mathbf{0}$, and $P \times 2$ submatrices

$$\mathbf{L}_k = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ G_k^{-1}(u_1) & G_k^{-1}(u_2) & \cdots & G_k^{-1}(u_P) \end{pmatrix}^T,$$

and $\boldsymbol{\varepsilon}$ is an error vector.

- *Step 3:* Obtain the least squares estimator $(\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T$ of $(\mathbf{H}_0^T, \boldsymbol{\theta}^T)^T$ as follows:

$$(\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \tilde{\mathbf{Y}}.$$

The resulting $\widehat{\mathbf{H}}_0$ and $\widehat{\boldsymbol{\theta}}$ can then be plugged in the model (4) to get semiparametric ROC curve estimators. When there is only one biomarker, i.e., $K = 0$, the estimated ROC curve for this biomarker becomes its empirical ROC curve. As pointed by a referee, the method does not change if the baseline biomarker is changed to the k th biomarker because essentially the original baseline function H_0 is transformed by a linear transformation: $H_0(u) - \alpha_k - \beta_k G_k(u)$. In addition, the model (4) is identifiable if $P \geq 3 + 2(K - 1)$. So $P = 5$ will always be enough for $K \leq 2$.

3.1 Asymptotic variance of $\widehat{\boldsymbol{\theta}}$

In order to apply the SLS estimators to construct newly proposed separation curves, we will need to establish the large sample properties of $\widehat{\boldsymbol{\theta}}$. Let $(\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T$ be the solutions to (4).

The following theorem then gives the asymptotic normality result of $(\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T$ with an explicit formula for its asymptotic covariance matrix.

Theorem 2—Under mild regularity conditions, assuming Equation (4) holds, when $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, $(\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T$ satisfies the following asymptotic result:

$$\sqrt{m} \left((\widehat{\mathbf{H}}_0^T, \widehat{\boldsymbol{\theta}}^T)^T - (\mathbf{H}_0^T, \boldsymbol{\theta}^T)^T \right) \xrightarrow{D} N_{K+1} \left(0, \sum_{h, \theta} \mathbf{J} \boldsymbol{\Sigma}^* \mathbf{J}^T \right),$$

where explicit forms for $\mathbf{J}$, $\boldsymbol{\Sigma}^*$ are given in the Appendix B.

Proof: see Appendix B in the online appendix.

The asymptotic covariance matrix $\boldsymbol{\Sigma}$ of $\widehat{\boldsymbol{\theta}}$ is a sandwich estimator, consisting the last $2K \times 2K$ rows and columns of $\boldsymbol{\Sigma}_{h, \theta}$. For the purpose of estimating $\boldsymbol{\theta}$, we apply the procedures outlined in Section 3. The covariance estimator $\widehat{\boldsymbol{\Sigma}}^*$ of $\boldsymbol{\Sigma}^*$ requires kernel smoothing

techniques to estimate the derivative function H'_0 because of unknown H_0 . In our simulation studies and the examples, we used the Epanechnikov kernel function

$E(x) = \frac{3}{4}(1-x^2)I(|x| \leq 1)$ with the bandwidth of $4/\max(\min(m, n)^{4/5}, 50)$. Selecting the optimal kernel bandwidth for estimating $\hat{\Sigma}^*$ is not trivial.

3.2 Simulation studies

In simulation studies, we evaluated the finite sample performance and robustness of the newly proposed semiparametric least squares method (SLS) and the parametric least squares (PLS) method, the placement value method (PV) and a kernel smoothing (KS) method mentioned in the introduction. In Section 3.2.1 we simulated biomarker results without any interaction term between biomarker types and FPRs, and compared the robustness of SLS, PLS and the PV method. In Section 3.2.2 we compared finite sample performance of PLS and SLS on hypothesis testing. We also compared finite sample performance of SLS, PLS and KS when there was an interaction term between biomarker types and FPRs in Section 3.2.3.

3.2.1 Robustness without an interaction between FPRs and covariates—

Bivariate normal data were simulated. Assume bivariate normal models had the forms of $(X_1, X_2) \sim \mathcal{N}(1, 3), \tilde{\Sigma}_1)$ and $(Y_1, Y_2) \sim \mathcal{N}(0, 1), \tilde{\Sigma}_1)$, where $\tilde{\Sigma}_1$ had variances 1 and 2 with a correlation parameter ρ . We simulated 1000 data sets under sample sizes $m = n = 200$ with $\rho = 0.5$. The true baseline function was given by $H_0 = 1/\sqrt{2} + 1/\sqrt{2}\Phi^{-1}$, $\alpha_1 = \sqrt{2} - 1$, and $\beta_1 = 0$. In the PLS approach, we purposely misspecified the baseline function to be $H_0 = \beta_0\Phi^{-1}$. In the SLS method, we allowed the baseline functions H_0 to be estimated nonparametrically. We compared these methods regarding their estimation bias and square root of MSE (RMSE). The results are presented in Table 1. As we can see from this table, PLS leads to large bias for estimating α_1 . From the simulation results, estimated α_1 is biased upwards 70% on the average, and thus estimated ROC curves departed much more than the truth. Although the PV method is unable to estimate the slope parameter β_1 , it still has good performance in estimating α_1 because the interaction term, $\beta_1\Phi^{-1}(u)$, does not exist in the simulated data. Our SLS method has virtually no bias in estimating both α_1 and β_1 .

3.2.2 Finite sample performance of hypothesis testing on equal ROC curves—

The bivariate normal data were again simulated. Assume that the bivariate normal models had the forms of $(X_1, X_2) \sim \mathcal{N}(1, 1), \tilde{\Sigma}_1)$ and $(Y_1, Y_2) \sim \mathcal{N}(0, 0), \tilde{\Sigma}_1)$. We applied the PLS method with correctly specified baseline function $H_0 = \alpha_1 + \beta_1\Phi^{-1}$ and the SLS method. The null hypothesis of equal ROC curves, $H_0 : (\alpha_1, \beta_1) = (0, 0)$, can be tested by the χ^2 test statistic, $\kappa = (\hat{\alpha}_1, \hat{\beta}_1)\hat{\Sigma}^{-1}(\hat{\alpha}_1, \hat{\beta}_1)^T$, with two degrees of freedom, where $\hat{\Sigma}$ is the estimator of Σ , and was calculated using asymptotic results of PLS in Tang and Zhou (2009) and SLS derived in Theorem 2. We simulated 1000 data sets under the null hypothesis with $m = n = (50, 100, 200)$ and $\rho = 0.5$. The nominal rejection rate was set to be 0.05. In the simulated data sets, we obtained the covariance matrices of the PLS estimators from bootstrapping and the covariances of SLS estimators by the asymptotic results. Since the PV method is unable to estimate β_1 , it is not applicable in this hypothesis test setting. We thus exclude the PV method from Table 3.2.4. The PLS method has good small sample performance, while the SLS method yields rejection rates closer to the nominal level as sample sizes get larger.

In addition, we investigated the finite sample performance of the proposed simultaneous confidence band. We simulated 1000 bivariate binormal data sets using the same model in the previous paragraph with $m = n = (100, 200, 400)$ and $\rho = 0.5$. For each data set, we applied the SLS method to get the parameter estimates and the resulting 95% confidence band using Theorem 1. We counted the number of times that the confidence band covered

the true difference curve out of 1000 simulated data sets. The resulting coverage percentages are close to the nominal level, as indicated in Table 3.

3.2.3 Robustness and finite sample performance with an interaction between FPRs and covariates—

Assume the bivariate normal models had the forms of $(X_1, X_2) \sim \mathcal{N}(1, 3), \tilde{\Sigma}_2)$ and $(Y_1, Y_2) \sim \mathcal{N}(0, 1), \tilde{\Sigma}_3)$, where $\tilde{\Sigma}_2 = \begin{pmatrix} 2 & \sqrt{10}\rho \\ \sqrt{10}\rho & 5 \end{pmatrix}$ and $\tilde{\Sigma}_3 = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix}$. Therefore, we got $H_0 = 1/\sqrt{2} + 1/\sqrt{2}\Phi^{-1}$ and $\alpha_1 = \beta_1 = 2/\sqrt{5} - 1/\sqrt{2}$. We simulated 1000 data sets under sample sizes $m = n = (50, 100, 200)$ when $\rho = 0.5$. For each simulated data set, we applied the PLS method with the correctly specified baseline function, our SLS method and the PV method to estimate parameters α_1 and β_1 . Biases and RMSEs for each of these methods are shown in Table 4. Both the PLS and SLS methods estimate the parameters well with small biases and RMSEs. Since the interaction between the FPRs and covariates was nonzero in this simulation, the PV method has substantially large bias and RMSE when estimating α_1 . This implies that estimated ROC curves from PV are likely to have much larger gaps than the truth.

Under the same simulation setting, we compared PLS, SLS and KS. When we implemented KS, we chose the kernel bandwidth recommended by Lloyd (1998). Again, we simulated 1000 data sets under sample sizes $m = n = (50, 100, 200)$ when $\rho = 0.5$. For each simulated data, we applied the three methods and estimated the separation curve at $u = (0.25, 0.50, 0.75)$. Table 5 shows biases and RMSEs for these methods. PLS has the smallest biases and RMSEs among the three methods. The KS method, which does not require parametric assumptions in PLS and SLS, has the largest biases and RMSEs.

3.2.4 Improvement on the coverage of confidence intervals by using bootstrap

—As suggested by a referee, we carried out a simulation study to evaluate the coverage of confidence intervals obtained by bootstrapping the approximate pivotal $SE(\hat{\beta}_1)^{-1}(\hat{\beta}_1 - \beta_1)$. We simulated bivariate binormal data sets from the same model as Section 3.2.3 with $m = n = (50, 100, 200)$ and $\rho = 0.5$. For each data set, we applied the SLS method, and obtained $\hat{\beta}_1$ and its standard error $\hat{SE}(\hat{\beta}_1)$. The limits of a standard $(1 - \alpha)\%$ confidence interval were obtained by $\hat{\beta}_1 \pm \hat{SE}(\hat{\beta}_1)z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1 - \alpha)$ th normal percentile. The coverage is given in Table 6. We then carried out the bootstrap procedure to investigate whether the coverage may be improved. We bootstrapped the original simulated data 199 times and applied the SLS method to each of bootstrap sample. For the b th bootstrap sample, $b = 1, \dots, 199$, denote $\hat{\beta}_{1,b}$ as the parameter estimate for β_1 , $\hat{SE}(\hat{\beta}_{1,b})$ as the estimated standard error of $\hat{\beta}_{1,b}$. The pivotal statistic was then $z^b = \hat{SE}(\hat{\beta}_{1,b})^{-1/2}(\hat{\beta}_{1,b} - \beta_1)$. The resulting $(1 - \alpha)\%$ confidence interval was given by

$$\widehat{\beta}_1 - \widehat{SE}(\widehat{\beta}_1)^{1/2} z_{(200 \times (1-\alpha))}^*, \widehat{\beta}_1 + \widehat{SE}(\widehat{\beta}_1)^{1/2} z_{(200 \times \alpha)}^*,$$

where $z_{(200 \times \alpha)}^*$ is the α th percentile of z^b . Based on the coverage given in Table 6, the bootstrap method based on the pivotal statistic offers a nice improvement over the standard confidence interval, especially for smaller sample sizes.

4 Examples

We plotted the empirical ROC curves from aforementioned two data sets in Figure 1. It is clear from Figure 1(a) that two empirical ROC curves cross each other in the region of large

FPRs. Figure 1(b) shows the empirical ROC curves of AD biomarkers adjusting for the age covariate. To construct the empirical ROC curves, we first dichotomized the age of subjects using a cutoff value 74, and then plotted four empirical ROC curves generated from all combinations of the dichotomized age covariate and biomarker types. Clearly, these ROC curves have intersections.

We constructed separation curves and their confidence bands for the examples by substituting the SLS estimators in the explicit expressions given by Theorem 1. The SLS estimator, θ and its covariance matrix estimator $\hat{\Sigma}$ were obtained with $G = G_1 = \Phi$. In the first example, we used the updated data set from the original AD biomarker study (Clark et al., 2003) described in Section 1. In the data set, 105 subjects were confirmed as AD patients by autopsy, and 63 subjects were confirmed to be absent of AD condition by autopsy. We applied PLS and our newly proposed SLS method to this data set and evaluated the accuracy of biomarkers τ and $A\beta$ in discriminating AD patients from non-AD patients. τ was used as the reference marker. The χ^2 statistic for testing the equality of the two ROC curves is 0.72 with the p -value 0.69, which does not show a significant difference between two AD biomarkers. The estimated semiparametric separation curve and its 95% simultaneous confidence band are plotted in Figure 2. It is clear that the separation curve with its 95% confidence band does not show a significant difference between two AD biomarkers at any region of FPRs. We also evaluated the effect of different kernel bandwidths and different P . When we increased the bandwidth from 4/100 to 4/10, we obtained the same χ^2 statistic with the same p -value. Although the χ^2 statistic reduced slightly to 0.66 with the p -value of 0.72 when we increased the value of P from 60 to 80, the conclusion remains unchanged.

We were also interested in whether the covariate such as patient's age at the time of lumbar puncture affects the diagnostic accuracy of these two biomarkers. We dichotomized the ages of all subjects at the time of lumbar puncture into two groups using 74 as the cutoff point. We let $G_1 = G_2 = G_3 = \Phi$ and used the ROC curve corresponding to the combination of τ and Age > 74 as the reference curve. By substituting the SLS estimator, we obtained the χ^2 statistics for comparing two AD biomarkers within each of the age groups. For the group of subjects who were older than 74 at the time of lumbar puncture, the χ^2 statistic for comparing τ and $A\beta$ is 6.58 with the p -value 0.04; for the other age group who were at age 74 or younger, the χ^2 statistic for comparing two biomarkers is 1.22 with the p -value 0.87. Therefore, τ and $A\beta$ have significantly different diagnostic accuracy adjusting for the age covariate based on our ROC analysis. Figure 3 shows the estimated separation curves. The confidence band of the separation curves at Age > 74 shows difference between the accuracy of two biomarkers with the specificity $u = 0.098$.

The second example we considered here is the aforementioned cancer biomarker data. CA 19-9 was used as the reference marker. In this example we first evaluated the proposed model with different choices for the link function including Probit link, exponential link, and logistic link. We obtained the link transformed empirical ROC curves for two biomarkers, and plotted the difference between two transformed curves in Figure 4. The suitable candidate link function should have the difference curve that matches well with $\alpha_1 + \beta_1 \Phi^{-1}(u)$. Figure 4 indicates that the difference curve from the Probit link is the closest to $-0.45 + 0.6\Phi^{-1}(u)$. This indicates that the proposed model with the link function Φ^{-1} will potentially be a nice fit for the biomarker data.

We constructed the separation curves and their confidence bands by substituting the SLS estimators. Figure 5 shows the separation curves for two cancer biomarkers. To test the equivalence of two ROC curves, we computed χ^2 statistic, which has a value of 23.65, with the p -value around 7.29×10^{-6} . Thus, it is reasonable to conclude that two cancer biomarkers have different diagnostic accuracy based on their ROC curves. Figure 5(c)

shows the 95% confidence band of the separation curve. From this plot, we see that CA 19-9 has significantly better accuracy than CA 12-5 in small values of FPRs. Calculation based on the upper confidence band gives the region of FPRs, $u = 0.42$, where CA 19-9 dominates CA 12-5. This result is different from the one obtained in Tang and Zhou (2009) which overlaid one simultaneous ROC band over the other and inspected the value of FPR where the two bands differ. The reason is that using the overlay of two bands does not consider the correlation of these bands. This leads to increased uncertainty of the difference. Here we obtained the simultaneous confidence band of the difference curve which takes into account the correlation. Thus, the band derived here actually shows larger difference between two ROC curves than Tang and Zhou (2009)'s method. We also evaluated the effect of different kernel bandwidths and different P . The χ^2 statistic did not change with different bandwidths. It increased slightly to 26.81 with the p -value around 1.5×10^{-6} when we reduced the value of P from 51 to 30, the conclusion remains unchanged.

5 Discussion

We proposed a separation curve between ROC curves and derived its asymptotic confidence band to identify the regions of FPRs where ROC curves differ. The confidence band provides a visualization tool for the difference among ROC curves. We also introduced a semiparametric LS method to estimate ROC parameters. Since the advantage of using the semiparametric LS estimator is clear over both existing parametric LS method and the semiparametric PV method in our simulation studies, we applied SLS estimators in constructing separation curves and associated confidence bands for two real examples.

This paper focuses on the comparison among biomarkers. This is particularly useful when some biomarkers are much more expensive to obtain, but do not perform as well as others. If all biomarkers are relatively inexpensive, it is feasible to combine the biomarkers using the logistic regression in (Lloyd, 1998).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank two anonymous referees, the associate editor and the editor for their constructive comments and useful suggestions. The project described here was supported by Award Number R15CA150698 from the National Cancer Institute under the American Recovery and Reinvestment Act of 2009, by Award Number H98230-11-1-0196 from the National Security Agency, and by Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, project grant #XVA 61-036. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, the National Institutes of Health, or the Department of Veterans Affairs.

References

- Cai T. Semiparametric ROC Regression Analysis. *Biostatistics*. 2004; 5:45–56. [PubMed: 14744827]
- Clark CM, Xie S, Chittams J, Ewbank D, Peskind E, Galasko D, Morris JC, Daniel W, McKeel J, Farlow M, Weitlauf SL, Quinn J, Kaye J, Knopman D, Arai H, Doody RS, DeCarli C, Leight S, Lee VMY, Trojanowski JQ. Cerebrospinal fluid tau and Amyloid: How well do these biomarkers reflect autopsy-confirmed dementia diagnoses. *Archives of Neurology*. 2003;60, 1696–1702.
- Lloyd CJ. Using Smoothed Receiver Operating Characteristic Curves to Summarize and Compare Diagnostic Systems. *Journal of The American Statistical Association*. 1998;93, 1356–1364.
- Pepe MS. A regression modeling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*. 1997;84, 595–608.

- Tang L, Zhou XH. Semiparametric Inferential Procedures for Comparing Multivariate ROC Curves with Interaction Terms. *Statistica Sinica*. 2009;19, 1203–1221.
- Wieand S, Gail MH, James BR, James KL. A Family of Nonparametric Statistics for Comparing Diagnostic Markers with Paired or Unpaired Data. *Biometrika*. 1989;76, 585–592.
- Zhang Z, Pepe MS. A Linear Regression Framework for Receiver Operating Characteristic(ROC) Curve Analysis. Tech rep, UW Biostatistics Working Paper Series, Working Paper. 2005; 253 <http://www.bepress.com/uwbiostat/paper253>.
- Zhou, XH.; McClish, DK.; Obuchowski, N. *Statistical Methods in Diagnostic Medicine*. New York: Wiley; 2002.
- Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*. 1997;16, 2143–2156.

Figure 1. Empirical ROC curves from two examples. (a) shows empirical ROC curves for biomarkers 19-9 and 125 in the cancer biomarker data; (b) shows empirical ROC curves for τ and $A\beta$ biomarkers with the dichotomized age covariate.

Figure 2. Estimated separation curve using the semiparametric least squares procedure in AD biomarker data without adjusting for the age covariate: solid line, separation curve; shaded region, 95% confidence band; dashed line, horizontal line at $\Delta(u) = 0$.

Figure 3.

Estimated separation curves using the semiparametric least squares estimators. (a) shows the separation curve for comparing τu and $A\beta$ for patients older than 74, and (b) shows the separation curve for comparing τu and $A\beta$ for patients younger than 74. Solid line, separation curve; shaded region, 95% confidence band; dashed line, horizontal line at $\Delta(u) = 0$.

Figure 4.

The difference curves from various link functions. Dashed line, the curve generated by the Probit link; dotted line, exponential link; dash-dot line, logistic link; solid line, $-0.45 + 0.6\Phi^{-1}(u)$.

Figure 5. Estimated separation curve using the semiparametric least squares estimators. Solid line, separation curve; shaded region, 95% confidence band; dashed line, horizontal line at $\Delta(u) = 0$.

Table 1
Bias and RMSE of parameter estimators from semiparametric models and a misspecified parametric model. Results are based on 1000 realizations of bivariate normal model

	PLS		SLS		PV	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
α_1	70.43%	0.71	0.72%	0.13	-0.67%	0.12
β_1	-0.19%	0.09	-0.66%	0.10	NA	NA

Table 2
Rejection rates from correctly specified models at nominal level 0.05. Results are based on 1000 realizations of normal model. The 95% prediction interval of the rejection rate is $(5.0\% \pm 1.4\%)$

m	n	PLS	SLS
50	50	4.70%	7.90%
100	100	4.50%	6.50%
200	200	5.60%	6.20%

Table 3
Coverage percentages of the 95% simultaneous confidence band

m	n	SLS
100	100	96.7%
200	200	97.1%
400	400	96.3%

Table 4
Bias and RMSE of parameter estimators from correctly specified models with interaction term. Results are based on 1000 realizations of bivariate normal model

m=n		PLS		SLS		PV	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α_1	-0.25%	0.2272	-0.60%	0.2285	-69.75%	0.7162
	β_1	-0.56%	0.1893	-1.24%	0.1897	NA	NA
100	α_1	-0.15%	0.1540	-0.32%	0.1547	-68.85%	0.7289
	β_1	-0.17%	0.1318	-0.54%	0.1323	NA	NA
200	α_1	-0.24%	0.1038	-0.29%	0.1040	-70.13%	0.7095
	β_1	-0.02%	0.0905	-0.15%	0.0905	NA	NA

Table 5
Bias and RMSE of three points on the separation curve. Results are based on 1000 realizations of bivariate normal model

m=n		SLS			PLS			KS		
		u=.25	0.5	0.75	u=.25	0.5	0.75	u=.25	0.5	0.75
50	Bias	-1.28%	-1.53%	-1.77%	-1.16%	-1.82%	-2.48%	3.15%	0.79%	1.88%
	RMSE	0.2233	0.2349	0.3052	0.2232	0.2342	0.3035	0.2622	0.2872	0.5321
100	Bias	-0.30%	-0.22%	-0.15%	-0.30%	-0.59%	-0.87%	3.74%	1.07%	-0.38%
	RMSE	0.1432	0.1538	0.2076	0.1430	0.1553	0.2108	0.1773	0.1921	0.2579
200	Bias	-0.25%	-0.33%	-0.41%	-0.10%	0.04%	0.18%	4.25%	0.71%	-1.91%
	RMSE	0.1045	0.1051	0.1383	0.0986	0.1029	0.1359	0.1245	0.1255	0.1699

Table 6
Coverage percentages: bootstrap versus standard confidence interval

m	n	Bootstrap	Standard
50	50	4.70%	12.90%
100	100	4.50%	8.20%
200	200	5.60%	5.60%