



UW Biostatistics Working Paper Series

4-29-2008

Semiparametric Inferential Procedures for Comparing Multivariate ROC Curves with Interaction Terms

Liansheng Tang

Xiao-Hua Zhou

Suggested Citation

Tang, Liansheng and Zhou, Xiao-Hua, "Semiparametric Inferential Procedures for Comparing Multivariate ROC Curves with Interaction Terms" (April 2008). *UW Biostatistics Working Paper Series*. Working Paper 326.
<http://biostats.bepress.com/uwbiostat/paper326>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Semiparametric Inferential Procedures for Comparing Multivariate ROC Curves with Interaction Terms

Liansheng Tang¹, Xiao-Hua Zhou^{2,3,*}

¹George Mason University, ²University of Washington, ³ VA Puget Sound Health Care System

Abstract: Multivariate ROC curve models that include an interaction term between biomarker type and false positive rate is important in comparative biomarker studies, because such interaction allows ROC curves of different biomarkers to cross each other. However, there has been limited work in drawing inference for comparing multivariate ROC curves, especially when the interaction terms are present. In this article we derive the asymptotic covariance of three estimators for multivariate ROC models. These covariance estimates have not been readily available in the literature, and bootstrap methods have to be used to obtain covariance estimates. With the readily available variance estimates, we can easily perform hypothesis testing among ROC curves while bootstrap tests are not so easily performed. The asymptotic results are applied to compare ROC curves and their areas under ROC curves. Moreover, we derive simultaneous confidence bands for multivariate ROC curves. We evaluate and compare the finite sample performance of our asymptotic covariance estimators. We also discuss the advantage of using our asymptotic results over bootstrap procedures. Finally, we illustrate our approach through a well-known pancreatic cancer study.

Key words and phrases: Diagnostic Accuracy, Simultaneous Confidence Band, Brownian Bridge Process.

1. Introduction

Research in early cancer detection involves developing diagnostic tools, such as a biomarker, to distinguish diseased patients from non-diseased patients. Biomarkers often yield continuous measurements. A popular tool to evaluate and compare the accuracy of biomarkers is a receiver operating characteristic (ROC) curve (Zhou et al., 2002), which is a plot of true positive rates vs false positive rates across all thresholds. Estimating a single binormal ROC curve of a continuous-scale biomarker has been well studied in the literature (Metz,

Herman and Shen, 1998; Cai and Moskowitz, 2004). However, inferential procedures for comparing multivariate ROC curve models with interaction terms between biomarker type and false positive rates (FPRs) have not been well studied, mainly because it is sometimes difficult to draw inferences in the presence of these interaction terms. However, such interactions are important because they allow ROC curves to cross each other. For example, Wieand et al. (1989) studied pancreatic cancer biomarkers, CA 19-9 and CA 125, which were measured on 51 pancreatitis patients and 90 pancreatic cancer patients. The empirical ROC curves were generated from this data set and their plots are shown in Figure 1. It is clear that two ROC curves cross each other when FPR gets close to 1. This fact shows the existence of interaction terms between biomarker type and FPRs.

Insert Figure 1 here.

Metz, et al. (1984) proposed a maximum likelihood estimator (MLE) for estimating bivariate binormal models from ordinal data. But their method requires estimating correlation parameters, besides the location and scale parameters in the marginal normal distributions. It would be more difficult to further extending their MLE method to more than two ROC curves when many more correlation parameters are to be estimated. As the number of biomarkers gets large, the MLE method becomes nonapplicable. For multiple independent ROC curves, Zhang (2004) and Zhang and Pepe (2005) proposed an intuitive least squares (LS) method, and Pepe (2000) presented an elegant generalized linear model (GLM) approach. Since it is complicated to derive the asymptotic results, the authors did not consider the large sample inference for the LS and GLM estimators with clustered data. Cai and Pepe (2002) considered an interesting semiparametric generalized estimating equation (GEE) method to allow an unknown baseline function when estimating ROC curves from correlated biomarker data. They derived asymptotic results for their estimators, but they did not consider how to compare multiple ROC curves from clustered data with the presence of interactions between biomarker type and FPRs.

In this article we adapt the LS method to clustered ROC curve data with the presence of interactions between biomarker type and FPRs. We derive explicit covariance structures between empirical ROC curves and use our results to derive

the asymptotic covariances of the LS estimator. We also adapt the GLM and GEE methods to this type of biomarker data and derive their asymptotic sandwich covariance estimators. These covariances have not been readily available in the literature, and bootstrap methods have to be used to obtain covariance estimates. With the readily available variance estimates, we can easily perform hypothesis testing among ROC curves while bootstrap tests are not so easily performed. For example, if we want to test whether two ROC curves vary by a certain amount δ at a specified FPR u_0 , i.e., $H_0 : ROC_1(u_0) - ROC_2(u_0) = \delta$, it is not clear how to bootstrap from the null distribution, while it is straightforward to perform such hypothesis tests using our asymptotic results and the delta method, which will be introduced in Section 4. We derive inferential procedures for comparing multivariate ROC curves that include the interaction terms, which have multivariate binormal ROC curves as a special case. In particular, we develop methods for comparing ROC curves and the areas under these ROC curves. We derive asymptotic simultaneous confidence bands for ROC curves. Such asymptotic results of simultaneous confidence bands of ROC curves are rarely studied. Instead, computer intensive methods are often employed to construct confidence bands (Cai and Pepe, 2002). Our confidence bands provide an easy-to-use tool to illustrate the sampling variability of the ROC curve estimates. In addition, we develop a method for comparing multiple ROC curves at some specified FPR.

This article is organized as follows. In Section 2 we adapt the LS methods to estimate multivariate ROC curves, and we also adapt GLM and a simplified GEE method. In Section 3 we derive the asymptotic results for the LS method when estimating multivariate ROC models. The asymptotic results are also derived for GLM and GEE methods. In Section 4 we apply the results to draw inference on comparing ROC curves and the areas under the ROC curves. In addition, asymptotic simultaneous confidence bands are derived for multivariate ROC curves. We carry out large scale simulation studies to evaluate and compare the finite sample performance of our covariance estimators of the LS, GLM and GEE methods. We also carry out simulation studies to evaluate the advantages of using asymptotic results over using bootstrap procedures. The simulation results are summarized in Section 5. A comparative pancreatic cancer diagnostic trial

serves as an illustrative example in Section 6, and some discussions are presented in Section 7.

2. Three estimators of multivariate ROC curves

In this section, we adapt the LS and GLM methods to clustered ROC data. We also give a simplified version of the GEE method for estimating multivariate ROC curves. Let $X_\ell = (X_{\ell,1}, X_{\ell,2}, \dots, X_{\ell,m})$ denote measurements of the ℓ th biomarker on m diseased subjects, and $Y_{\tilde{\ell}} = (Y_{\tilde{\ell},1}, Y_{\tilde{\ell},2}, \dots, Y_{\tilde{\ell},n})$ denote measurements of the $\tilde{\ell}$ th biomarker on n healthy subjects, where $\ell, \tilde{\ell} = 1, \dots, K$. For the ℓ th and $\tilde{\ell}$ th different biomarkers measured on the i th diseased subject, $i = 1, \dots, m$, the measurements $X_{\ell,i}$ and $X_{\tilde{\ell},i}$ follow the bivariate survival function $F_{\ell,\tilde{\ell}}$ with the marginal distributions F_ℓ and $F_{\tilde{\ell}}$, respectively, and the measurements of the j th healthy subject, $Y_{\ell,j}$ and $Y_{\tilde{\ell},j}$ with $j = 1, \dots, n$, follow the bivariate survival function $G_{\ell,\tilde{\ell}}$ with the marginal distributions G_ℓ and $G_{\tilde{\ell}}$, respectively. The ROC curve of the ℓ th biomarker is then given by $Q_\ell(u) = F_\ell(G_\ell^{-1}(u))$, and its empirical form is $\tilde{Q}_\ell(u) = \hat{F}_\ell(\hat{G}_\ell^{-1}(u))$, where \hat{F}_ℓ and \hat{G}_ℓ are empirical functions of F_ℓ and G_ℓ , respectively.

Let Z_k be a dummy variable for the k th biomarker, $k = 2, \dots, K$. The multivariate ROC curves are then given by the following expressions:

$$\begin{aligned} Q_1(u) &= g\{\theta_{10} + \theta_{11}h(u)\}, \\ \text{and } Q_k(u) &= g[\theta_{10} + \theta_{11}h(u) + \sum_{i=2}^K \{\theta_{i0}Z_i + \theta_{i1}Z_i h(u)\}], \end{aligned} \quad (2.1)$$

for $0 < a \leq u \leq b < 1$, where g is some specified link function and h is some specified baseline function. In the regression ROC modeling, $\theta_{10} + \theta_{11}h(u)$ is the baseline function, usually denoted as $h_0(u)$. In this article we let this baseline function be a known function up to two unknown parameters, θ_{10} and θ_{11} , as in Pepe (2000), Zhang (2004) and Zhang and Pepe (2005). The important components of the model (2.1) are $\theta_{k1}Z_k h(u)$, which are the interaction terms between dummy variables indicating biomarker types and FPRs. Such interaction terms play an important role in estimating ROC curves especially when estimating the multivariate binormal ROC curves. With these terms, the model (2.1) includes the multivariate binormal ROC model as a special case, which is commonly used in the literature (Metz, et al., 1984). Specifically, let g be Φ , let h be Φ^{-1} and

let Z_k be an indicator variable for the k th biomarker, the model (2.1) has the following expressions:

$$\begin{aligned} Q_1(u) &= \Phi\{\theta_{10} + \theta_{11}\Phi^{-1}(u)\} \\ \text{and } Q_k(u) &= \Phi\{\theta_{10} + \theta_{11}\Phi^{-1}(u) + \theta_{k0} + \theta_{k1}\Phi^{-1}(u)\}, \end{aligned} \quad (2.2)$$

for $k = 2, \dots, K$. When there is only one biomarker, the model (2.2) reduces to the commonly used binormal model (Zhou, et al., 2002).

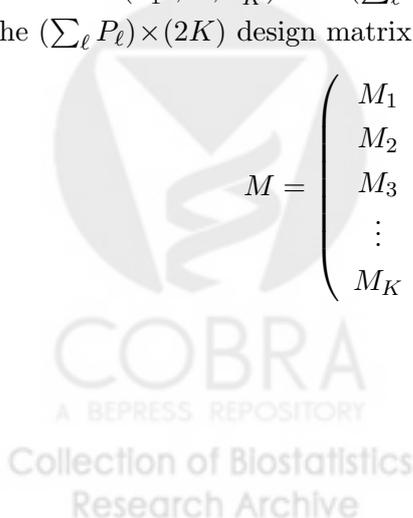
Let $u_\ell = (u_{\ell,1}, \dots, u_{\ell,P_\ell})^T$ be some fixed partition points in the range of $[a, b]$ on the ℓ th empirical ROC curve. Here P_ℓ is arbitrarily chosen for the ℓ th ROC curve where $0 < a = u_{\ell,1} < u_{\ell,2} < \dots < u_{\ell,P_\ell} = b < 1$. For example, if we choose 50 jump points for the 1st ROC curve, we can choose $u_1 = (1/51, 2/51, \dots, 50/51)$. Also, for simplicity, we denote $L(u_\ell) = (L(u_{\ell,1}), L(u_{\ell,2}), \dots, L(u_{\ell,P_\ell}))^T$, where $\ell = 1, \dots, K$, for any process or function L .

Zhang (2004) and Zhang and Pepe (2005) proposed a least squares (LS) approach to estimate multiple ROC curves. They let Z_k be the indicator variables for the k th biomarker. In the LS estimating procedure, the ROC curve corresponding to a reference biomarker is chosen as the reference ROC curve. For the ℓ th empirical ROC curve, the partition points, $u_\ell = (u_{\ell,1}, \dots, u_{\ell,P_\ell})^T$, are chosen within interval boundaries $[a, b]$. When we are interested in the entire ROC curve, a and b can be chosen to be close to 0 and 1, respectively. If a partial ROC curve is of interest, a and b can be chosen accordingly. By plugging in the empirical functions \hat{F}_ℓ and \hat{G}_ℓ^{-1} , the ℓ th empirical ROC curve is computed by $\tilde{Q}_\ell(u_\ell) = \hat{F}_\ell(\hat{G}_\ell^{-1}(u_\ell))$. Denote $\tilde{Y}_\ell = g^{-1}(\tilde{Q}_\ell(u_\ell))$. We combine $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_K$ and get a linear regression equation as follows:

$$\tilde{Y} = M\theta + \epsilon, \quad (2.3)$$

where $\tilde{Y} = (\tilde{Y}_1^T, \dots, \tilde{Y}_K^T)^T$ is a $(\sum_\ell P_\ell) \times 1$ vector with its element $\tilde{Y}_\ell = g^{-1}(\tilde{Q}_\ell(u_\ell))$. The $(\sum_\ell P_\ell) \times (2K)$ design matrix M is given by:

$$M = \begin{pmatrix} M_1 & 0 & 0 & \cdots & 0 \\ M_2 & M_2^* & 0 & \cdots & 0 \\ M_3 & 0 & M_3^* & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_K & 0 & 0 & \cdots & M_K^* \end{pmatrix},$$



with its $P_\ell \times 2$ submatrices

$$M_\ell = \begin{pmatrix} 1 & \cdots & 1 \\ h(u_{\ell,1}) & \cdots & h(u_{\ell,P_\ell}) \end{pmatrix}^T, \quad \text{and} \quad M_k^* = \begin{pmatrix} Z_k & \cdots & Z_k \\ Z_k h(u_{\ell,1}) & \cdots & Z_k h(u_{\ell,P_\ell}) \end{pmatrix}^T.$$

Also, the error term ϵ has a multivariate normal distribution given by $\epsilon \sim N(0, \Sigma_\epsilon)$. The detailed proof is given in the on-line version of the paper at <http://www.stat.sinica.edu.tw/statistica>. Based on the regression equations in (2.3), the LS estimator $\hat{\theta}^{LS}$ of θ is given by $\hat{\theta}^{LS} = (M^T M)^{-1} M^T \tilde{Y}$.

There are several other estimating equation methods for estimating ROC parameters. Pepe (2000) observed that the expected value of indicator variables $I\{X_{\ell,i} \geq \hat{G}_\ell^{-1}(u_{\ell,p})\}$ converges to the true ROC curve of the ℓ th biomarker and proposed a GLM method to estimate the ROC curve model (2.1). If partial ROC curves on $[a, b]$ are of interest, $u_{\ell,p}$ are chosen within this range. The GLM approach estimates parameters in the model (2.1) by the following estimating equations:

$$\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \sum_{i=1}^m \frac{g'(\tilde{M}_{\ell,p}^T \theta)}{g(\tilde{M}_{\ell,p}^T \theta)(1 - g(\tilde{M}_{\ell,p}^T \theta))} \tilde{M}_{\ell,p} \left\{ I\{X_{\ell,i} \geq \hat{G}_\ell^{-1}(u_{\ell,p})\} - g(\tilde{M}_{\ell,p}^T \theta) \right\} = 0,$$

or

$$\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \tilde{M}_{\ell,p} \{ \tilde{Q}_\ell(u_{\ell,p}) - g(\tilde{M}_{\ell,p}^T \theta) \} = 0,$$

for $\ell = 1, \dots, K$ and $p = 1, \dots, P_\ell$, with $\theta = (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \dots, \theta_{K0}, \theta_{K1})^T$ and a weight function $w_\ell(u_{\ell,p}) = \{g'(\tilde{M}_{\ell,p}^T \theta)\} / [g(\tilde{M}_{\ell,p}^T \theta)\{1 - g(\tilde{M}_{\ell,p}^T \theta)\}]$. Here $\tilde{M}_{1,p}$ is a $1 \times 2K$ vector with the first two elements being 1 and $h(u_{1,p})$, respectively, and the rest elements being zeros. $\tilde{M}_{k,p}$ have the first two elements being 1 and $h(u_{k,p})$, respectively, the $(2k + 1)$ th and $(2k + 2)$ th elements being Z_k and $Z_k h(u_{k,p})$, respectively, and the rest elements being zeros. The asymptotic results of the parameter estimator not provided in Pepe (2000) will be given in Section 3.

The GEE method by Cai and Pepe (2002) also used the indicator variables $I\{X_{\ell,i} \geq \hat{G}_\ell^{-1}(u_{\ell,p})\}$ and relied on the fact that points on ROC curves can be interpreted as conditional expectations of these indicator variables. Their method is flexible to allow an unknown baseline function h_0 in the model (2.1). When the baseline function has the form of $h_0(u) = \theta_{10} + \theta_{11}h(u)$, the GEE method

is similar to the GLM method. Specifically, the GEE method is to solve the following estimating equations:

$$\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \widetilde{M}_{\ell,p} \{ \widetilde{Q}_\ell(u_{\ell,p}) - g(\widetilde{M}_{\ell,p}^T \theta) \} = 0.$$

It is observed that if the baseline function has a known form, the GLM method differs from the GEE method by including the weight function w_ℓ .

3. Large sample theory of ROC estimators

We derive the asymptotic covariance structure of multiple empirical ROC curves for clustered data. The result is then applied to derive asymptotic covariances of the LS estimator. We also derive asymptotic sandwich covariances of GLM and GEE estimators. Asymptotic results of the LS and GLM estimators for clustered data have not been provided in the literature (Pepe, 2000; Zhang, 2004; Zhang and Pepe, 2005). Although Cai and Pepe (2002) have studied the large sample theory of the GEE estimator, their covariance estimator has a complicated form. For the multivariate ROC models we considered here, the covariance estimator of $\hat{\theta}^{GEE}$ has a simplified covariance estimator that is easy to apply. These covariance results are essential for drawing inference on comparing ROC curves and constructing simultaneous confidence ROC bands as discussed in Section 4. We derive the covariance structure between empirical ROC curves in Appendices 1 and 2, and summarize the results in Theorem 1. Theorem 1 builds a basis for deriving asymptotic covariances of the LS estimator, which is discussed in this section.

Theorem 1. Under mild regularity conditions, i.e., 1) \bar{F}_ℓ and \bar{G}_ℓ have continuous densities \bar{F}'_ℓ and \bar{G}'_ℓ , respectively, 2) the first derivative Q'_ℓ of Q_ℓ is bounded in (a, b) , when $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, $\text{cov}[\sqrt{m}\{\widetilde{Q}_\ell(s) - Q_\ell(s)\}, \sqrt{m}\{\widetilde{Q}_\ell(t) - Q_\ell(t)\}]$ converges in distribution to

$$[F_{\ell,\bar{\ell}}\{G_\ell^{-1}(s), G_{\bar{\ell}}^{-1}(t)\} - Q_\ell(s)Q_{\bar{\ell}}(t)] + \lambda[Q'_\ell(s)Q'_{\bar{\ell}}(t)\{G_{\ell,\bar{\ell}}\{G_\ell^{-1}(s), G_{\bar{\ell}}^{-1}(t)\} - st\}],$$

for s, t in $[a, b]$.

3.1 Asymptotic covariance of LS estimator

Denote $\tilde{\theta}_{\ell 1} = \theta_{11}$ when $\ell = 1$; and $\tilde{\theta}_{\ell 1} = \theta_{11} + \theta_{\ell 1}$ when $\ell \geq 2$. Let us define

the following $2K \times 2K$ square matrix:

$$J = \begin{pmatrix} KD & Z_2D & \cdots & Z_KD \\ Z_2D & Z_2^2D & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ Z_KD & O & \cdots & Z_K^2D \end{pmatrix}^{-1} \begin{pmatrix} I_2 & I_2 & \cdots & I_2 \\ O & I_2 & \cdots & O \\ & & \ddots & \\ O & O & \cdots & I_2 \end{pmatrix},$$

where

$$D = \begin{pmatrix} b - a & \int_a^b h(u)du \\ \int_a^b h(u)du & \int_a^b h^2(u)du \end{pmatrix},$$

for $0 < a < b < 1$, and I_2 is a 2×2 identity matrix. Let us define the following quantity:

$$V_\ell(s, t) = \frac{Q_\ell(s \wedge t) - Q_\ell(s)Q_\ell(t)}{g'(g^{-1}(Q_\ell(s)))g'(g^{-1}(Q_\ell(t)))} + \lambda \tilde{\theta}_{\ell 1}^2 h'(s)h'(t)(s \wedge t - st),$$

where $\ell = 1, \dots, K$, and

$$\begin{aligned} \tilde{V}_{\ell, \tilde{\ell}}(s, t) &= \frac{F_{\ell, \tilde{\ell}}(G_\ell^{-1}(s), G_{\tilde{\ell}}^{-1}(t)) - Q_\ell(s)Q_{\tilde{\ell}}(t)}{g'[g^{-1}\{Q_\ell(s)\}]g'(g^{-1}(Q_{\tilde{\ell}}(t)))} \\ &+ \lambda \tilde{\theta}_{\ell 1} \tilde{\theta}_{\tilde{\ell} 1} h'(s)h'(t)\{G_{\ell, \tilde{\ell}}(G_\ell^{-1}(s), G_{\tilde{\ell}}^{-1}(t)) - st\}, \end{aligned}$$

for $\ell, \tilde{\ell} = 1, 2, \dots, K$, and $\ell \neq \tilde{\ell}$, where $\lambda = \lim_{m, n \rightarrow \infty} m/n$. The following Theorem 2 gives the results of the LS estimator. The detailed proof of Theorem 2 is given in the on-line version of the paper at <http://www.stat.sinica.edu.tw/statistica>.

Theorem 2. *Under mild regularity conditions stated in Theorem 1, when $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, and $P_\ell \rightarrow \infty$, the regression parameter estimator $\hat{\theta}^{LS}$ has the following asymptotic multivariate normal distribution:*

$$\sqrt{m}(\hat{\theta}^{LS} - \theta) \xrightarrow{D} N(0, \Sigma^{LS} = J\Sigma^y J^T).$$

Here Σ^y is a $2K \times 2K$ matrix and

$$\Sigma^y = \begin{pmatrix} \Sigma_{11}^y & \Sigma_{12}^y & \cdots & \Sigma_{1K}^y \\ \Sigma_{21}^y & \Sigma_{22}^y & \cdots & \Sigma_{2K}^y \\ & & \ddots & \\ \Sigma_{K1}^y & \Sigma_{K2}^y & \cdots & \Sigma_{KK}^y \end{pmatrix},$$

with 2×2 diagonal symmetric submatrices, $\Sigma_{\ell\ell}^y$, whose elements are

$$\begin{aligned}\sigma_{\ell\ell}^{(1,1)} &= \int_a^b \int_a^b V_\ell(s, t) ds dt, & \sigma_{\ell\ell}^{(2,2)} &= \int_a^b \int_a^b h(s)h(t)V_\ell(s, t) ds dt, \\ \sigma_{\ell\ell}^{(1,2)} &= \sigma_{\ell\ell}^{(2,1)} = \int_a^b \int_a^b h(s)V_\ell(s, t) ds dt,\end{aligned}$$

and 2×2 off-diagonal symmetric submatrices, $\Sigma_{\ell, \tilde{\ell}}^y$, whose elements are

$$\begin{aligned}\sigma_{\ell, \tilde{\ell}}^{(1,1)} &= \int_a^b \int_a^b \tilde{V}_{\ell, \tilde{\ell}}(s, t) ds dt, & \sigma_{\ell, \tilde{\ell}}^{(2,2)} &= \int_a^b \int_a^b h(s)h(t)\tilde{V}_{\ell, \tilde{\ell}}(s, t) ds dt, \\ \sigma_{\ell, \tilde{\ell}}^{(1,2)} &= \sigma_{\ell, \tilde{\ell}}^{(2,1)} = \int_a^b \int_a^b h(s)\tilde{V}_{\ell, \tilde{\ell}}(s, t) ds dt.\end{aligned}$$

In practice, $F_{\ell, \tilde{\ell}}$, $G_{\ell, \tilde{\ell}}$, F_ℓ and G_ℓ are unknown and are estimated by their respective empirical functions. If ROC data are unclustered, the off-diagonal submatrices, $\Sigma_{\ell, \tilde{\ell}}^y$'s, become zero matrices. If we further let $h = g^{-1}$ and Z_k be indicator variables in the model (1), we get the same asymptotic result as in Zhang (2004).

3.2 Asymptotic Covariances of the GLM and GEE estimators

The GLM estimator is estimated by solving the following estimating equations:

$$U(\theta) = \sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \tilde{M}_{\ell,p} \{ \tilde{Q}_\ell(u_{\ell,p}) - g(\tilde{M}_{\ell,p}^T \theta) \} = 0,$$

where $\tilde{M}_{\ell,p}$ is defined in Section 2.1. To solve these estimating equations, we will need the Newton-Raphson method to obtain the GLM estimator $\hat{\theta}^{GLM}$. From the Newton-Raphson algorithm, it follows that

$$\begin{aligned}cov(\hat{\theta}^{GLM}) &= \left(-\frac{\partial U(\theta)}{\partial \theta} \right)^{-1} var \left(\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \tilde{M}_{\ell,p} \{ \tilde{Q}_\ell(u_{\ell,p}) - g(\tilde{M}_{\ell,p}^T \theta) \} \right) \\ &\quad \left(-\frac{\partial U(\theta)}{\partial \theta} \right)^{-1},\end{aligned}$$

where $\partial U(\theta)/\partial \theta$ is the partial derivative of $U(\theta)$ with regard to θ . Let

$$\begin{aligned}U_{1i} &= \sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \tilde{M}_{\ell,p} \left\{ I(X_{\ell i} \geq G_\ell^{-1}(u_{\ell,p})) - g(\tilde{M}_{\ell,p}^T \theta) \right\}, \\ \text{and } U_{2j} &= \sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \tilde{M}_{\ell,p} Q'_\ell(u_{\ell,p}) \left\{ I(Y_{\ell j} \geq G_\ell^{-1}(u_{\ell,p})) - u_{\ell,p} \right\}.\end{aligned}$$

Therefore, our result in Theorem 1 gives that under mild regularity conditions, when $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, the GLM estimator $\hat{\theta}^{GLM}$ satisfies the following asymptotic normality:

$$\sqrt{m}(\hat{\theta}^{GLM} - \theta) \xrightarrow{D} N(0, \Sigma^{GLM}),$$

where

$$\Sigma^{GLM} = \left(\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \widetilde{M}_{\ell,p} g'(\widetilde{M}_{\ell,p}^T \theta) \right)^{-1} \lim_{m,n \rightarrow \infty} \left\{ \sum_{i=1}^m U_{1i} U_{1i}^T + \lambda \sum_{v=1}^n U_{2j} U_{2j}^T \right\} \left(\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} w_\ell(u_{\ell,p}) \widetilde{M}_{\ell,p} g'(\widetilde{M}_{\ell,p}^T \theta) \right)^{-1}.$$

The asymptotic property of the modified GEE estimator can be similarly derived. We denote

$$\begin{aligned} U_{1i}^* &= \sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \widetilde{M}_{\ell,p} \left\{ I(Y_{\ell i} \geq G_\ell^{-1}(u_{\ell,p})) - g(\widetilde{M}_{\ell,p}^T \theta) \right\}, \\ \text{and } U_{2j}^* &= \sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \widetilde{M}_{\ell,p} Q'_\ell(u_{\ell,p}) \left\{ I(Y_{\ell j} \geq G_\ell^{-1}(u_{\ell,p})) - u_{\ell,p} \right\}. \end{aligned}$$

When $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, the GEE estimator $\hat{\theta}^{GEE}$ satisfies the following asymptotic normality:

$$\sqrt{m}(\hat{\theta}^{GEE} - \theta) \xrightarrow{D} N(0, \Sigma^{GEE}),$$

where

$$\Sigma^{GEE} = \left(\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \widetilde{M}_{\ell,p} g'(\widetilde{M}_{\ell,p}^T \theta) \right)^{-1} \lim_{m,n \rightarrow \infty} \left\{ \sum_{r=1}^m U_{1i}^* U_{1i}^{*T} + \lambda \sum_{v=1}^n U_{2j}^* U_{2j}^{*T} \right\} \left(\sum_{\ell=1}^K \sum_{p=1}^{P_\ell} \widetilde{M}_{\ell,p} g'(\widetilde{M}_{\ell,p}^T \theta) \right)^{-1}.$$

These covariance estimators are sandwich estimators. We will evaluate their finite sample performance in our simulation studies.

4. Multivariate ROC Analysis

Estimated ROC curves, denoted by \widehat{Q}_ℓ , are estimated by replacing θ with the estimator $\hat{\theta}$ in the model (2.1). $\hat{\theta}$ is estimated via either of three aforementioned methods. Here $\hat{\theta}$ is a general notation, which can also be estimated via

other available methods. But since asymptotic results are derived for the three methods, it is convenient to utilize these results. Our asymptotic results give the corresponding covariance matrix estimator $\hat{\Sigma}$ of Σ for $\hat{\theta}$. For further notational convenience, let $\Sigma_{\ell, \tilde{\ell}}$ be the 2×2 submatrices of Σ , i.e.,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1K} \\ & \vdots & \\ \Sigma_{K1} & \cdots & \Sigma_{KK} \end{pmatrix}_{2K \times 2K}.$$

The estimator $\hat{\Sigma}_{\ell, \tilde{\ell}}$ of $\Sigma_{\ell, \tilde{\ell}}$ is the corresponding submatrix of $\hat{\Sigma}$. In this section we derive methods for pairwise comparison of ROC curves. We also develop methods for comparing more than two areas under ROC curves under multivariate binormal assumptions. In addition, we derive inferential procedures for simultaneous confidence ROC bands and for comparing multiple ROC curves at some specified FPR.

4.1 Pairwise comparison of ROC curves

It is often of interest to compare ROC curves to investigate the accuracy of biomarkers. The reference ROC curve $Q_1(u)$ and the k th ROC curve $Q_k(u)$ in $[a, b]$ only differ by a parameter vector $\theta_k = (\theta_{k0}, \theta_{k1})^T$. Consequently, testing the equality of these two ROC curves is equivalent to testing $H_0 : \theta_k = (0, 0)^T$. It follows from the multivariate normality of $\hat{\theta}$ in Section 3 that the asymptotic distribution of the test statistic, $\kappa_k = (\hat{\theta}_k - \theta_k)^T \Sigma_{kk}^{-1} (\hat{\theta}_k - \theta_k)$ is χ_2^2 . Similarly, testing the equality of two ROC curves of the ω th and ν th different biomarkers in $[a, b]$, $\omega, \nu = 2, \dots, K$, is also reduced to a χ^2 test. The null hypothesis becomes $H_0 : (\theta_{\omega 0} - \theta_{\nu 0}, \theta_{\omega 1} - \theta_{\nu 1}) = 0$. Denote that $\theta_{\omega, \nu} = (\theta_{\omega 0}, \theta_{\omega 1}, \theta_{\nu 0}, \theta_{\nu 1})^T$. The resulting chi-square statistic is given by the following expression:

$$\kappa_{\omega, \nu} = \begin{pmatrix} (\hat{\theta}_{\omega 0} - \hat{\theta}_{\nu 0}) - (\theta_{\omega 0} - \theta_{\nu 0}) \\ (\hat{\theta}_{\omega 1} - \hat{\theta}_{\nu 1}) - (\theta_{\omega 1} - \theta_{\nu 1}) \end{pmatrix}^T (A \Sigma_{\omega, \nu} A^T)^{-1} \begin{pmatrix} (\hat{\theta}_{\omega 0} - \hat{\theta}_{\nu 0}) - (\theta_{\omega 0} - \theta_{\nu 0}) \\ (\hat{\theta}_{\omega 1} - \hat{\theta}_{\nu 1}) - (\theta_{\omega 1} - \theta_{\nu 1}) \end{pmatrix},$$

where $A = (I_2, -I_2)$ with a 2×2 identity matrix I_2 , and $\Sigma_{\omega, \nu}$ is the covariance matrix of $\theta_{\omega, \nu}$. $\Sigma_{\omega, \nu}$ is a 4×4 principal submatrix of Σ , and can be subtracted by elements in $\hat{\Sigma}$.

4.2 Comparing the areas under multivariate binormal ROC curves

Our asymptotic results in Section 3 are also applicable for comparing areas under multivariate ROC curves, especially for multivariate binormal ROC curves. Let $A = (A_1, \dots, A_K)$ be a vector, where A_ℓ is the area under the ℓ th ROC curve. It is well known that under the binormal assumption, A_ℓ is a simple function of θ given by the following:

$$A_1(\theta_{10}, \theta_{11}) = \Phi \left\{ \frac{\theta_{10}}{\sqrt{1 + \theta_{11}^2}} \right\},$$

and

$$A_k(\theta_{10}, \theta_{11}, \theta_{k0}, \theta_{k1}) = \Phi \left\{ \frac{\theta_{10} + \theta_{k0}}{\sqrt{1 + (\theta_{11} + \theta_{k1})^2}} \right\}. \quad (4.1)$$

Let q be a second-order differentiable and real valued function of A . It follows that when $m/n \rightarrow \lambda$ as $m, n \rightarrow \infty$, $\sqrt{m}\{q(\hat{A}) - q(A)\}$ asymptotically has a normal distribution given by

$$\sqrt{m}\{q(\hat{A}) - q(A)\} \xrightarrow{D} N(0, \sigma_q^2),$$

where the variance σ_q^2 is given by :

$$\begin{aligned} \sigma_q^2 = & \lim_{m, n \rightarrow \infty} m \left\{ \frac{\partial q}{\partial A_1} \frac{\partial q}{\partial A_1} \text{var}(A_1) + 2 \sum_{k=2}^K \frac{\partial q}{\partial A_1} \frac{\partial q}{\partial A_k} \text{cov}(A_1, A_k) \right. \\ & \left. + \sum_{k=2}^K \sum_{\bar{k}=2}^K \frac{\partial q}{\partial A_k} \frac{\partial q}{\partial A_{\bar{k}}} \text{cov}(A_k, A_{\bar{k}}) \right\}. \end{aligned}$$

By Taylor expansions on A_ℓ and our asymptotic results, we get that

$$\begin{aligned} \text{var}(A_1) &= B_1^T \Sigma_{11} B_1, \\ \text{cov}(A_1, A_k) &= B_1^T (\Sigma_{11} + \Sigma_{k1}) B_k, \\ \text{and} \quad \text{cov}(A_k, A_{\bar{k}}) &= B_k^T (\Sigma_{11} + \Sigma_{k\bar{k}}) B_{\bar{k}}, \end{aligned}$$

where

$$B_1 = \left\{ \phi \left(\frac{\theta_{10}}{\sqrt{1 + \theta_{11}^2}} \right) \frac{1}{\sqrt{1 + \theta_{11}^2}}, -\phi \left(\frac{\theta_{10}}{\sqrt{1 + \theta_{11}^2}} \right) \frac{\theta_{11}}{(\sqrt{1 + \theta_{11}^2})^{\frac{3}{2}}} \right\}^T,$$

$$B_k = \left[\phi \left\{ \frac{\theta_{10} + \theta_{k0}}{\sqrt{1 + (\theta_{11} + \theta_{k1})^2}} \right\} \frac{1}{\sqrt{1 + (\theta_{11} + \theta_{k1})^2}}, -\phi \left\{ \frac{\theta_{10} + \theta_{k0}}{\sqrt{1 + (\theta_{11} + \theta_{k1})^2}} \right\} \frac{\theta_{11} + \theta_{k1}}{\{\sqrt{1 + (\theta_{11} + \theta_{k1})^2}\}^{\frac{3}{2}}} \right]^T.$$

and $\Sigma_{\ell\tilde{\ell}}$, for $\ell, \tilde{\ell} = 1, \dots, K$, is estimated from asymptotic results. Delong, et al. (1988) gave a similar formula for comparing the areas under nonparametric ROC curves. Although their approach is robust, semiparametric approaches may be more appealing to derive smooth ROC curves for continuous biomarker data. If q is some linear function, the theoretical result is simplified to a similar formula in Delong, et al. (1988). A simple example is to let E be a vector with the ℓ th element being 1, the $\tilde{\ell}$ th element being -1 and other elements being zero. Then we have $q(\hat{A}) = E\hat{A}$ corresponds to $\hat{A}_\ell - \hat{A}_{\tilde{\ell}}$, whose variance estimator follows from the variance result. Inference is then easily drawn for testing $H_0 : A_\ell = A_{\tilde{\ell}}$, and for constructing a confidence interval.

4.3 Simultaneous confidence ROC bands

The variance of estimated ROC curves at each FPR can be derived from the parameter estimator $\hat{\theta}$ and its covariance matrix estimator $\hat{\Sigma}$. Denote $H = (1, h(u))$, $\tilde{H} = (H, H)$. We get the following corollary about the variance of estimated ROC curves, $\hat{Q}_\ell(u)$:

Corollary 2. *The variance of estimated ROC curves at u are given by:*

$$\sigma_1^2(u) = g'[g^{-1}\{Q_1(u)\}]^2 H \Sigma_{11} H^T,$$

$$\text{and } \sigma_k^2(u) = g'[g^{-1}\{Q_k(u)\}]^2 \tilde{H} \begin{pmatrix} \Sigma_{11} & \Sigma_{1k} \\ \Sigma_{k1} & \Sigma_{kk} \end{pmatrix} \tilde{H}^T,$$

respectively, for $k = 2, \dots, K$.

Therefore, the $(1 - \alpha)100\%$ pointwise confidence interval of $Q_\ell(u)$ is given by

$$\hat{Q}_\ell(u) \pm z_{\alpha/2} \hat{\sigma}_\ell(u), \quad 0 \leq u \leq 1.$$

In Theorem 3 below, we give explicit expressions of simultaneous bands for multivariate ROC curves. The detailed proof of Theorem 3 is given in the on-line version of the paper at <http://www.stat.sinica.edu.tw/statistica>.

Theorem 3. *Under mild conditions, the $(1 - \alpha)100\%$ simultaneous confidence bands for multivariate ROC curves in $[a, b]$ are constructed as follows:*

$$g \left\{ H \hat{\theta}_1 \pm \sqrt{\chi_{2,\alpha}^2 H \Sigma_{11} H^T} \right\},$$

and

$$g \left\{ \tilde{H} (\hat{\theta}_1^T, \hat{\theta}_k^T)^T \pm \sqrt{\chi_{4,\alpha}^2 \tilde{H} \begin{pmatrix} \Sigma_{11} & \Sigma_{1k} \\ \Sigma_{k1} & \Sigma_{kk} \end{pmatrix} \tilde{H}^T} \right\},$$

respectively, for $k = 2, \dots, K$.

Note that the reason why simultaneous confidence bands for ROC curves have such simplified expressions is that we assume that g and h are known. The estimated ROC curves are fully determined by two parameters for the reference biomarker and four parameters for other biomarkers. Therefore, χ_2^2 and χ_4^2 distributions arise, and the derivation of simultaneous bands is naturally simplified.

4.4 Comparing multiple ROC curves at some specified FPR

Denote $\widehat{Q}(u) = (\widehat{Q}_1(u), \dots, \widehat{Q}_K(u))^T$. Similarly as in Section 4.2, suppose that q is a real-valued and second-order derivable function on the vector \widehat{Q} . We have that $q(\widehat{Q}(u_0))$ at a specified FPR u_0 converges to a normal distribution with mean zero and the variance given by

$$\begin{aligned} \tilde{\sigma}_q^2(u_0) = & \lim_{m, n \rightarrow \infty} m \left[\frac{\partial q}{\partial Q_1} \frac{\partial q}{\partial Q_1} \text{var}\{Q_1(u_0)\} + 2 \sum_{k=2}^K \frac{\partial q}{\partial Q_1} \frac{\partial q}{\partial Q_k} \text{cov}\{Q_1(u_0), Q_k(u_0)\} \right. \\ & \left. + \sum_{k=2}^K \sum_{\tilde{k}=2}^K \frac{\partial q}{\partial Q_k} \frac{\partial q}{\partial Q_{\tilde{k}}} \text{cov}\{Q_k(u_0), Q_{\tilde{k}}(u_0)\} \right]. \end{aligned}$$

Here we have

$$\begin{aligned} \text{var}\{Q_1(u_0)\} &= C_1(u_0)^T \Sigma_{11} C_1(u_0), \\ \text{cov}\{Q_1(u_0), Q_k(u_0)\} &= C_1(u_0)^T (\Sigma_{11} + \Sigma_{k1}) C_k(u_0), \end{aligned}$$

and

$$\text{cov}\{Q_k(u_0), Q_{\tilde{k}}(u_0)\} = C_k(u_0)^T (\Sigma_{11} + \Sigma_{k\tilde{k}}) C_{\tilde{k}}(u_0),$$

where

$$C_1(u) = (g'\{\theta_{10} + \theta_{11}h(u)\}, \theta_{11}g'\{\theta_{10} + \theta_{11}h(u)\})^T,$$

and

$$C_k(u) = [g'\{\theta_{10} + \theta_{11}h(u) + \theta_{k0} + \theta_{k1}h(u)\}, (\theta_{11} + \theta_{k1})g'\{\theta_{10} + \theta_{11}h(u) + \theta_{k0} + \theta_{k1}h(u)\}]^T.$$

Again, if q is a linear function, the result can be greatly simplified.

5. Simulation studies

5.1 Finite sample performance of hypothesis testing

We ran a large set of simulation studies to evaluate and compare the finite sample performance of our asymptotic covariance estimators for the LS, GLM

and GEE methods. The bivariate normal data were simulated from $N((1, 1), \Sigma_0)$ for the diseased and $N((0, 0), \Sigma_0)$ for the healthy, where Σ_0 has the variances 1 and 2 with a correlation parameter ρ . True ROC curves of tests 1 and 2 have the same form given by $Q_1(u) = Q_2(u) = \Phi\{1/\sqrt{2} + 1/\sqrt{2}\Phi^{-1}(u)g\}$, for $0 \leq u \leq 1$. Thus, the true value of the parameter vector is $\theta = (1/\sqrt{2}, 1/\sqrt{2}, 0, 0)$. We fitted a bivariate binormal model to the data; that is, we let $K = 2$ in the model (2.2). The null hypothesis of equal ROC curves, $H_0 : (\theta_{20}, \theta_{21}) = (0, 0)$, can be tested by the χ^2 test statistic $\kappa = (\hat{\theta}_{20}, \hat{\theta}_{21})\hat{\Sigma}_{22}(\hat{\theta}_{20}, \hat{\theta}_{21})^T$ with 2 degrees of freedom, where $(\hat{\theta}_{20}, \hat{\theta}_{21})$'s covariance matrix estimate, $\hat{\Sigma}_{22}$, is calculated using asymptotic results in Theorem 2. We simulated 1000 data sets under the null hypothesis with various combinations of $m = (50, 100, 200)$ and $n = (50, 100, 200)$ under $\rho = (0, 0.25, 0.5, 0.75)$. The nominal rejection rate was set to be 5%. In the simulation, the variances of LS estimators were estimated using asymptotic results developed in Theorem 2. The variances of GLM and GEE's estimators were obtained using asymptotic results in Section 3. For a small sample size such as 50, GEE method sometimes did not converge and we had to run more than 1000 simulations in order to obtain 1000 valid estimates. Since the LS method does not require iterations, the computation time is greatly reduced compared to that of GLM or GEE. Under the same circumstances, the computing time of LS is less than half of that of GLM or GEE. In particular, we conducted our simulation study on the same Unix machine. It took 53 seconds for the LS procedure to estimate the parameters from 1000 simulated data sets when $m = n = 200$, while the computational times of GLM and GEE were 870 seconds and 348 seconds, respectively. When $m = n = 50$, the computation time of LS was reduced to 24 seconds, while the times of GLM and GEE were reduced to 210 seconds and 98 seconds, respectively.

Table 1 presents the rejection rates from these three approaches. As shown in Table 1, asymptotic results of LS work well for all combinations of sample sizes as the rejection rates are close to the nominal level, 5%. Even for a sample size as small as 50 for both diseased and healthy groups, the rejection rates do not have much departure from the nominal level. Moreover, the rejection rates of LS are not affected by values of the correlation parameter, ρ . From Table 1, the GEE and GLM approaches behave similarly to each other. Both approaches

have over-rejection rates when sample sizes are as small as 50. This is mainly due to the much variability of sandwich variance estimators for the GEE and GLM method. Readers are referred to Kauermann and Carroll (2001) for more details. It is also noticeable in Table 1 that as sample sizes for the healthy get larger, rejection rates of GEE and GLM get closer to the nominal level even when sample sizes for the diseased are small. However, rejection rates for the diseased and small sample sizes for the healthy are not improved with large sample sizes. We tried bootstrap methods in these situations. The bootstrap method performed similarly as the LS method on the rejection rates regardless of sample sizes. The bootstrap performed better than GLM and GEE when sample sizes were small, and similarly as GLM and GEE when sample sizes were large.

Insert Table 2 here.

5.2 Finite sample performance of point and interval estimates

We used the same setting as in the previous section to evaluate and compare estimation precision of three methods in this simulation study. We again simulated 1000 data sets under sample sizes $m = n = (50, 200, 400)$ with $\rho = 0.5$. The nominal coverage probability of confidence intervals was 95%. We applied LS, GEE and GLM to simulated data sets to get estimates of the ROC parameter vector $(\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21})$. Confidence intervals for the parameters were calculated based on asymptotic results. We then compared these methods based on bias, square root of MSE (RMSE) and the coverage probabilities of confidence intervals. The results are shown in Table 2. All three methods have good accuracy for estimating the parameters. The coverage probabilities differ among these approaches. Our simulation results show that the LS approach has nice finite sample property as the coverage probabilities of all parameters are close to the nominal level for small sample sizes. When sample sizes are small, confidence intervals computed from sandwich covariance estimators for GLM and GEE approaches cover the intercept parameters properly, but these confidence intervals over-cover slope parameters. As sample sizes approach 400, the coverages of GLM and GEE estimators get closer to the nominal level for slope parameters.

Insert Table 2 here.

5.3 The advantage of our asymptotic results over bootstrap procedures

Many authors applied bootstrap methods to estimate covariance matrices for the LS and GLM estimators when the asymptotic results were not derivable (Pepe, 2000; Zhang and Pepe, 2005). However, it can sometimes take much more computation time to bootstrapping than using asymptotic results. In this simulation study, we compared coverage percentages of bootstrap covariance estimates with those of asymptotic covariance estimates for the LS approach. We used the same setting in Section 5.1 with $m = n = (50, 200, 400)$. Under each combination of sample sizes, we simulated 1000 data sets with $\rho = 0.5$. For each data set we applied bootstrap procedures to get covariance estimates of the LS approach. The number of bootstrap was set to be 1000. We then used covariance estimates to get confidence intervals and their coverage percentages. We showed coverage percentages of bootstrap methods in Table 2. As can be seen from Table 2, our asymptotic results were as good as bootstrap results because their coverage percentages were very close. More importantly, asymptotic covariance estimates were computed much faster than bootstrapped covariance estimates. For example, when using the LS method as $m = n = 400$ it took 30 seconds to obtain a bootstrap covariance estimate for one data set on a PC, while it took only 5 seconds to obtain an asymptotic covariance estimate for the same data set on the same PC.

6. Application to Pancreatic cancer biomarkers

Main interest in the aforementioned biomarker example is to determine whether ROC curves generated by two biomarkers are equal. If not, it would be interesting to tell which biomarker can better distinguish the diseased from the healthy. We applied the LS estimation procedure to this data set. With the probit link, $g = \Phi$, and $h = \Phi^{-1}$, ROC curves of these biomarkers have the same structure as those in (2.2) when $K = 2$. We got the estimate of the parameter vector as $(\hat{\theta}_{10}^{LS}, \hat{\theta}_{11}^{LS}, \hat{\theta}_{20}^{LS}, \hat{\theta}_{21}^{LS}) = (1.18, 0.47, -0.49, 0.55)$. Its covariance matrix estimate was calculated based on the asymptotic result in Theorem 2. The GLM and GEE parameter estimates were very close to the LS estimate, and thus were not listed. The χ^2 was then calculated to be 18.16 with the p-value 0.0001, which indicates significant difference in diagnostic accuracy between two biomarkers. We also calculated the difference between two areas under estimated ROC curves using the procedure in Section 4.2 and found significant difference

between two areas with the p-value 0.02. To visualize the sampling variability of estimated ROC curves, simultaneous ROC bands were constructed using the result in Theorem 3. Figure 2 shows the estimated ROC curves and their simultaneous bands. These ROC curves fit very close to the empirical curves. Although our results on ROC curves and their areas show that two biomarkers are different, it is clear in Figure 3 that two ROC curves intersect. The simultaneous bands can help us determine the region of FPRs where two ROC curves are different. Figure 3 shows overlapped confidence bands. It is obvious that the ROC curve for CA 19-9 is significantly better than that for CA 125 when the false positive rate is less than around 0.17, and two ROC curves do not have much difference elsewhere.

Insert Figures 2-3 here.

7. Discussion

The interaction terms in our multivariate ROC model are completely different from the interactions of the K diagnostic tests on the measurement level. In fact, the interactions we refer to are between test types and FPRs. Multivariate ROC models such as multivariate binormal models play an important role in ROC analysis of clustered biomarkers. This article derived asymptotic covariances of the LS, GLM and GEE estimators for multivariate ROC models with the presence of interaction terms between biomarker type and FPRs. We developed three theorems in this paper. Theorems 1-2 were developed mainly for the LS procedure. We applied some results in empirical process theory to show the asymptotic properties in Theorems 1 and 2. We then derived the asymptotic properties of correlated ROC curves with interaction terms in the model. To our knowledge, such asymptotic properties with correlated data have not been addressed in empirical process theory. Theorem 3 for confidence bands can be used with all three aforementioned estimators with their respective variance estimates. The confidence band method gave an intuitive way to visualize the variability of ROC curves and the difference between ROC curves.

Due to the nature of these aforementioned ROC methods, no matter which baseline biomarker is selected, the estimated ROC curves remain the same for each of the methods. That is, all these methods respect exchangeability of baseline markers. For example, in the setting of two ROC curves, $\hat{\theta} = (\hat{\theta}_{10}, \hat{\theta}_{11}, \hat{\theta}_{20}, \hat{\theta}_{21})$

is obtained using one of the methods when biomarker 1 is chosen as the baseline marker. The resulting estimated ROC curves are then $\tilde{Q}_1(u) = g\{\hat{\theta}_{10} + \hat{\theta}_{11}h(u)\}$ for biomarker 1, and $\tilde{Q}_2(u) = g[\hat{\theta}_{10} + \hat{\theta}_{20} + \{\hat{\theta}_{11} + \hat{\theta}_{21}\}h(u)]$ for biomarker 2. Suppose now we instead choose biomarker 2 as the baseline and obtain parameter estimate $\hat{\theta}^* = (\hat{\theta}_{10}^*, \hat{\theta}_{11}^*, \hat{\theta}_{20}^*, \hat{\theta}_{21}^*)$. The resulting estimated ROC curves are $\tilde{Q}_1^*(u) = g\{\hat{\theta}_{10}^* + \hat{\theta}_{11}^*h(u)\}$ for biomarker 2 and $\tilde{Q}_2^*(u) = g[\hat{\theta}_{10}^* + \hat{\theta}_{20}^* + \{\hat{\theta}_{11}^* + \hat{\theta}_{21}^*\}h(u)]$ for biomarker 1. All the three parametric methods give $\hat{\theta}_{10} = \hat{\theta}_{10}^* + \hat{\theta}_{20}^*$, $\hat{\theta}_{11} = \hat{\theta}_{11}^* + \hat{\theta}_{21}^*$, $\hat{\theta}_{10}^* = \hat{\theta}_{10} - \hat{\theta}_{20}^*$ and $\hat{\theta}_{11}^* = \hat{\theta}_{11} - \hat{\theta}_{21}^*$. Thus the estimated ROC curves remain unchanged.

Our new contributions also include procedures to compare AUCs and ROC curves, especially when the interaction terms between FPR's and biomarker type are present. We drew inference for pairwise comparison between ROC curves, multiple comparisons of areas under ROC curves under binormal assumptions. Our inferential procedures in Section 4 for comparing ROC curves are very general for multivariate ROC models. Besides the three estimators we discussed, if other estimators and their covariances were available, they can also be used in these procedures.

Acknowledgment

The authors would like to thank an associate editor and two referees for their constructive comments and suggestions. This work is supported in part by NIH grant R01EB005829. This report presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

References

- Cai, T. and Moskowitz, C. (2004). Semiparametric estimation of the binormal ROC curve. *Biostatistics* **5**, 573-586.
- Cai, T. and Pepe, M.S. (2002). Semi-parametric ROC analysis to evaluate biomarkers for disease. *Journal of the American Statistical Association* **97**, 1099-1107.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837-845.
- Kauermann, G. and Carroll, R. J. (2001). A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association*

- tion **96**, 1387-1396.
- Metz, C.E., Wang, P., and Kronman, H. B. (1984). A new approach for testing the significance of differences between ROC curves measured from correlated data. In *Information Processing in Medical Imaging*, edited by F. Deconinck, 432-445.
- Metz, C.E., Herman B.A. and Shen J-H. (1998). Maximum-likelihood estimation of ROC curves from continuously-distributed data. *Statistics in Medicine* **17**, 1033-1053.
- Pepe, M.S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352-359.
- Wieand, S., Gail, M.H., James, B.R. and James, K.L. (1989). A family of non-parametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76**, 585-592.
- Zhang, Z. and Pepe, M.S. (2005). A linear regression framework for receiver operating characteristic(ROC) curve analysis. *UW Biostatistics Working Paper Series*. Working Paper 253. <http://www.bepress.com/uwbiostat/paper253>.
- Zhang, Z. (2004). Semiparametric least squares analysis of the receiver operating characteristic curve. *University of Washington Doctoral Dissertation*.
- Zhou, X.H., McClish, D.K. and Obuchowski, N.A. (2002). *Statistical Methods in Diagnostic Medicine*. New York: Wiley.



Table 1: Rejection rates (in %) with the nominal level $\alpha = 0.05$ from asymptotic results

m	n	LS				GEE				GLM			
		$\rho=0$	0.25	0.5	0.75	$\rho=0$	0.25	0.5	0.75	$\rho=0$	0.25	0.5	0.75
50	50	3.7	6.0	3.6	6.9	13.1	11.8	14.2	11.8	15.2	16.2	14.9	10.3
50	100	4.8	4.5	3.6	7.2	6.9	7.8	7.7	6.9	9.3	8.8	10.0	8.3
50	200	4.1	6.5	4.9	5.8	5.6	5.1	5.6	4.5	5.4	5.9	4.5	5.8
100	50	5.3	5.3	5.0	6.7	16.0	14.7	12.9	13.6	15.3	14.1	11.9	11.3
100	100	5.7	5.2	6.4	6.8	8.2	8.3	9.1	8.8	11.2	9.8	10.1	9.5
100	200	4.9	5.5	4.6	5.2	4.4	4.3	4.2	5.0	4.7	5.0	5.6	5.6
200	50	5.1	6.2	4.7	4.7	16.8	14.5	16.3	13.6	18.6	15.7	14.4	13.2
200	100	4.7	5.2	4.9	5.4	9.2	9.4	8.1	10.5	11.0	11.9	9.8	11.4
200	200	4.7	4.7	5.3	5.6	5.5	4.9	4.5	4.6	5.5	4.8	5.1	6.0

The rejection rate with 1000 realizations of normal model. The 95% prediction interval of the rejection rate is $(5.0\% \pm 1.4\%)$.

Table 2: Bias, RMSE, coverage probability (CP) of parameter estimators

m(n)		LS				GEE			GLM		
		Bias	RMSE	CP	CP(BT)	Bias	RMSE	CP	Bias	RMSE	CP
50	θ_{10}	-2.36%	0.19	94.80%	95.20%	0.66%	0.19	97.20%	0.60%	0.19	98.70%
	θ_{11}	0.80%	0.12	95.20%	95.00%	-7.88%	0.15	99.90%	-8.71%	0.15	100.00%
	θ_{20}	-0.55%	0.20	93.80%	95.70%	-0.92%	0.20	99.40%	-0.69%	0.21	99.10%
	θ_{21}	0.28%	0.16	96.20%	95.80%	0.49%	0.16	100.00%	0.24%	0.15	99.90%
200	θ_{10}	0.07%	0.09	95.20%	94.20%	0.02%	0.10	94.70%	0.81%	0.10	94.40%
	θ_{11}	0.52%	0.06	94.60%	94.50%	-1.76%	0.07	99.60%	-1.63%	0.06	99.70%
	θ_{20}	-0.28%	0.10	93.90%	93.90%	-0.09%	0.10	94.30%	-0.21%	0.10	95.00%
	θ_{21}	0.01%	0.08	94.70%	95.20%	0.19%	0.08	99.00%	-0.08%	0.08	99.60%
400	θ_{10}	-0.03%	0.07	95.40%	94.40%	-0.07%	0.07	94.70%	0.02%	0.06	96.50%
	θ_{11}	0.37%	0.04	94.80%	95.00%	-0.72%	0.04	96.40%	-0.86%	0.04	97.10%
	θ_{20}	-0.42%	0.07	95.90%	94.90%	0.02%	0.07	95.00%	0.06%	0.07	96.00%
	θ_{21}	-0.08%	0.05	95.30%	95.90%	-0.08%	0.05	97.70%	0.01%	0.05	97.60%

CP is the coverage percentage for 95% confidence intervals using asymptotical standard errors with a normal quantile. CP(BT) is the coverage percentage for 95% confidence intervals using bootstrap. Results are based on 1000 realizations of bivariate normal model.

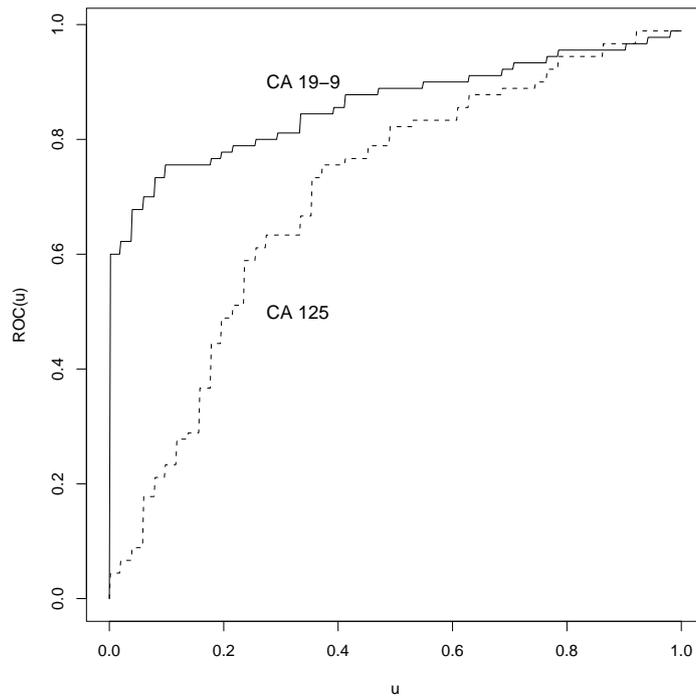


Figure 1: Empirical ROC curves for CA 19-9 and CA 125: solid line, CA 19-9; dashed line, CA 125.



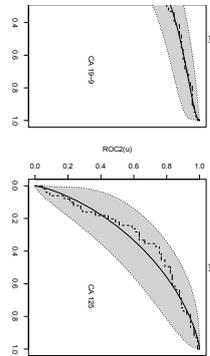


Figure 2: Estimated ROC curves and their 95% confidence bands for CA 19-9 and CA 125: dashed lines, empirical ROC; solid lines, estimated ROC; shaded regions, 95% confidence band; dotted lines, confidence band boundaries.

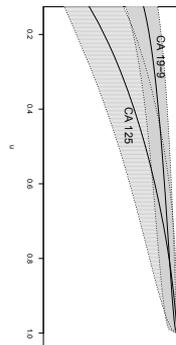


Figure 3: Overlapped 95% confidence bands for CA 19-9 and CA 125: solid lines, estimated ROC; shaded regions, 95% confidence band; dotted lines, confidence band boundaries.