Unified Least Squares Methods for the Evaluation of Diagnostic Tests With the Gold Standard

Cancer Informatics
1–12
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1176935116686063

Liansheng Larry Tang[1,2], Ao Yuan[2,3], John Collins[1,2], Xuan Che[2] and Leighton Chan[2]

[1]Department of Statistics, George Mason University, Fairfax, VA, USA. [2]Rehabilitation Medicine Department, NIH Clinical Center, Bethesda, MD, USA. [3]Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC, USA.

ABSTRACT: The article proposes a unified least squares method to estimate the receiver operating characteristic (ROC) parameters for continuous and ordinal diagnostic tests, such as cancer biomarkers. The method is based on a linear model framework using the empirically estimated sensitivities and specificities as input "data." It gives consistent estimates for regression and accuracy parameters when the underlying continuous test results are normally distributed after some monotonic transformation. The key difference between the proposed method and the method of Tang and Zhou lies in the response variable. The response variable in the latter is transformed empirical ROC curves at different thresholds. It takes on many values for continuous test results, but few values for ordinal test results. The limited number of values for the response variable makes it impractical for ordinal data. However, the response variable in the proposed method takes on many more distinct values so that the method yields valid estimates for ordinal data. Extensive simulation studies are conducted to investigate and compare the finite sample performance of the proposed method with an existing method, and the method is then used to analyze 2 real cancer diagnostic example as an illustration.

KEYWORDS: ROC curve, least squares, sensitivity, specificity

RECEIVED: September 1, 2016. **ACCEPTED:** November 24, 2016.

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 687 words, excluding any confidential comments to the academic editor.

TYPE: Review

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported in part by the Intramural Research Program of the National Institutes of Health and the US Social Security Administration.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Liansheng Larry Tang, Rehabilitation Medicine Department, NIH Clinical Center, Bethesda, MD 20892, USA. Email: ltang1@gmu.edu

Introduction

In diagnostic test development, one is concerned about whether a newly developed test is more accurate than traditional ones to correctly discriminate a subject with a certain condition ("the case") from a subject without the condition ("the control").[1] Early diagnosis of serious diseases plays an important role because late detection can have serious consequences. For example, a patient with lung cancer might have a higher chance of surviving if detected early and the lesion is surgically removed. But the person will die if the diagnosis is incorrect and necessary surgery is not performed.[2]

For diagnostic tests that generate binary results, their accuracy can be summarized in terms of the sensitivity (ie, probability of identifying a case when the subject truly has the condition) and specificity (ie, probability of correctly identifying a control when the subject does not have the condition). The sensitivity is also called as the true-positive rate (TPR), and the false-positive rate (FPR) is 1 − specificity. For tests that generate continuous or ordinal results, the receiver operating characteristic (ROC) curve is a standard statistical tool to describe and compare the accuracy of diagnostic tests.[1] The ROC curve, commonly used in medical diagnostic studies, is a plot of TPR versus FPR at different possible thresholds. It is widely used in radiology, psychophysical, and medical imaging research for detection performance, military monitoring, and industrial quality control. It is used to examine the trade-off between the TPR and FPR under different thresholds and overcomes the limitation of having to dichotomize the test results to use isolated measurements of TPR and FPR. The ROC curve is plotted by connecting all the points generated by possible thresholds.[1] A test with 100% TPR and 0% FPR is a perfect predictor, ie, all the case patients have positive test results and all the control patients have negative test results.

Most ROC curves are concave and above the chance diagonal which is the line segment between $(0,0)$ and $(1,1)$. However, some of them are below the chance diagonal and are called improper curves.[3] The closer the curve is to the upper left corner, the larger the area under ROC curve (AUC) is and the better distinguishing ability the diagnostic test has. The perfect test has an AUC of 1. Figure 1 provides the illustration of the ROC curves for 3 biomarkers with different diagnostic accuracies. The ROC curve for biomarker 1 is uniformly above the other 2 ROC curves. This means that biomarker 1 has the best performance in detecting the case and control among the 3 biomarkers.

The ROC analysis of continuous data from a single test has been extensively investigated since the seminal work by Dorfman and Alf.[4] Diagnostic test studies generate correlated results when the same subject undergoes 2 or more different tests.[5] An important area in ROC research with multiple markers is the comparison of tests' accuracy. Parametric and semiparametric methods have been proposed to estimate ROC curves from this type of correlated data in the literature. Parametric methods assume distributions for measurements,[1] but these methods may not perform well if the parametric

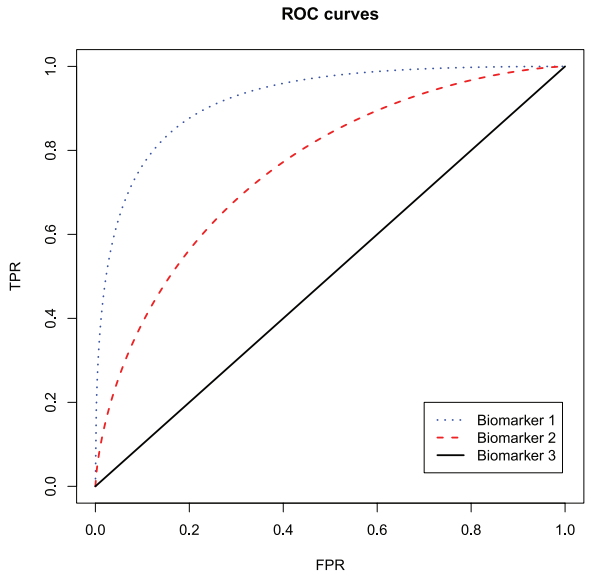

Figure 1. ROC curves for 3 biomarkers: dotted curve—biomarker 1 (AUC = 0.9), dashed curve—biomarker 2 (AUC = 0.7), and solid curve—biomarker 3 (AUC = 0.5). AUC indicates area under ROC curve; FPR, false-positive rate; ROC, receiver operating characteristic; TPR, true-positive rate.

assumptions are invalid. An intuitive parametric least squares (LS) ROC method proposed by Zhang and Pepe[6] requires no iteration and thus takes much less computation time than the ROC methods using iterations. The asymptotic covariance of their LS estimator is derived by Tang and Zhou.[7] An essential assumption of the procedure by Zhang and Pepe[6] is that the basis function of the ROC curve is known. A recent paper by Tang and Zhou[8] relaxes this assumption and estimates the basis function nonparametrically.

Besides continuous test data, ordinal data occur frequently in radiology when radiologists or computer algorithms are used to read subjects' medical images and provide ordinal ratings regarding their belief in the severity of subjects' disease status. Several methods for estimating a single ROC curve from ordinal data have been proposed by various authors.[4,9] Morris et al[10] provide a detailed summary of these methods for ordinal data. The maximum likelihood estimation (MLE) method by Dorfman and Alf[4] is the most widely used procedure for ordinal data. Metz and colleagues[11,12] consider 2 modalities. Hsieh and Turnbull[9] develop a generalized LS approach. As the number of markers becomes larger than 2, the MLE method by Metz et al[11] becomes inapplicable. It is also not trivial to extend the single ROC method by Hsieh and Turnbull[9] to multiple binormal ROC curves because the correlation structure among empirical ROC curves is unknown.

In this article, we propose a unified linear regression method to estimate the ROC curve from pairs of consistent sensitivity and specificity estimates. The proposed method estimates a pair of sensitivity and specificity for a given cutoff point. For a set of chosen cutoff points on the continuous data, a number of pairs can be obtained, and the estimates in the pairs can be values for the response variable and covariate in the linear regression setting. The method provides valid ROC parameter estimates for both continuous data and ordinal data.

Notations and Methods

Suppose that multiple tests are applied to a case sample with m subjects and a control sample with n subjects. For test ℓ, the test result for the ith case subject, $X_{\ell i}$, and test result for the jth control subject, $Y_{\ell j}$, are available, where $i = 1,\ldots,m$ and $j = 1,\ldots,n$ for $\ell = 1,\ldots,L$. At some given thresholds $c_{\ell r}$ $(r = 1,\ldots,R)$, let $Se_{\ell r}$ and $Sp_{\ell r}$ be the sensitivity and specificity of the ℓth test, respectively. The observed results $X_{\ell i}$ and $Y_{\ell j}$ may be continuous or ordinal. In the latter case, they are derived from some underlying variables $\tilde{Y}_{\ell i}$ and $\tilde{X}_{\ell j}$. At a threshold c, the TPR or the sensitivity is given by $Se_\ell(c) = P(\tilde{Y}_{\ell i} > c \mid D = 1)$ and the specificity or (1 – FPR) is given by $Sp_\ell(c) = P(\tilde{X}_{\ell j} \leqslant c \mid D = 0)$, where D is the indicator for disease status with 1 being a case and 0 being a control. For the continuous diagnostic tests, the observed test results are identical to the underlying results.

For the ordinal diagnostic test, the observed ordinal ratings $X_{\ell i}$ and $Y_{\ell j}$ are considered to be obtained by applying decision thresholds to the latent variables. For the ℓth ordinal test, $X_{\ell i}$ and $Y_{\ell j}$ take on ordinal ratings, $1,\ldots,R_\ell$. These ratings are considered to be obtained by applying $R_\ell - 1$ decision thresholds, $-\infty = c_0 < c_1 < \cdots < c_{R_\ell - 1} < c_{R_\ell} = +\infty$, to the latent variables. Specifically, the rating $X_{\ell i} = r_\ell$ $(Y_{\ell j} = r_\ell)$ is given to a case subject (or a control subject) if $\tilde{Y}_{\ell i}$ (or $\tilde{X}_{\ell j}$) falls between c_{r-1} and c_r $(r = 1,\ldots,R)$.

The Hsieh method for one binormal ROC curve

We first consider one test with ordinal test results X_{1i} and Y_{1j} for cancer and control subjects, respectively. Because the ROC curve is invariant to any monotonic transformation of the underlying test results, $\tilde{Y}_{1i}$ and $\tilde{X}_{1j}$ can be considered to have already been transformed by some unknown monotone function so that $\tilde{Y}_{1i} \sim N(\alpha_1, \beta_1^2)$ and $\tilde{X}_{1j} \sim N(0,1)$. Let P_{r1} be the probability of having the rating $X_{1i} = r$ for the ith case subject, and let P_{r0} be the probability of having the rating $Y_{1j} = r$ for the jth control subject. The sensitivity and specificity at a threshold c are $Se(c) = 1 - \Phi((c - \alpha_1)/\beta_1)$ and $Sp(c) = \Phi(c)$, respectively. From this, we may write the probabilities P_{r1} and P_{r0} as $P_{r1} = Se_{r-1} - Se_r$ and $P_{r0} = Sp_r - Sp_{r-1}$, where $Se_r = Se(c_r)$ and $Sp_r = Sp(c_r)$. The log-likelihood function is given as follows:

$$\sum_{r=1}^{R} f_{r1} \log P_{r1} + \sum_{r=1}^{R} f_{r0} \log P_{r0}$$

where f_{r1} and f_{r0} are the observed numbers of responses in the rth category from the cancer and control subjects, respectively. Dorfman and Alf[4] solve the score equation of the log-likelihood function and obtain the MLE estimators of α_1, β_1, and $c_1,\ldots,c_{R-1}$.

Hsieh and Turnbull[9] developed a generalized LS approach. They estimated the empirical ROC curve at a fixed number of FPRs and applied the generalized LS method to the transformed empirical ROC curve to obtain the parameter estimators. The regression method by Hsieh and Turnbull[9] for estimating 1 ROC curve is similar to the method of Dorfman and Alf[4]. The essential difference between them is that the former only requires the estimated sensitivities and specificities, whereas the latter requires the actual observations. For the result r, we have $Se_r = 1 - \sum_{k=1}^{r} P_{k1}$ and $Sp_{c_r} = \sum_{k=1}^{r} P_{k0}$ for $r = 1,\ldots,R_1$. Hsieh and Turnbull[9] observed that

$$Se_r = 1 - \Phi\left(\frac{c_r - \alpha_1}{\beta_1}\right), \quad Sp_r = \Phi\left(c_r\right)$$

for $r = 1,\ldots,R$. The equations above can be written as follows:

$$\Phi^{-1}(Sp_r) = \alpha_1 + \beta_1 \Phi^{-1}(1 - Se_r)$$

for $r = 1,\ldots,R$. Thus, by assuming a perfect gold standard, the authors use $\sum_{k=1}^{r} f_{k0}/m$ and $\sum_{k=1}^{r} f_{k1}/n$ to estimate Sp_r and $1 - Se_r$, respectively, and obtained the following linear regression model with error terms:

$$\begin{aligned} \Phi^{-1}\left(\frac{1}{m}\sum_{k=1}^{r} f_{k0}\right) &= \alpha_1 + \beta_1 \Phi^{-1}(1 - Se_r) + \varepsilon_{0r} \\ \Phi^{-1}\left(\frac{1}{n}\sum_{k=1}^{r} f_{k1}\right) &= \Phi^{-1}(Sp_{1r}) + \varepsilon_{1r} \end{aligned} \tag{1}$$

for $r = 1,\ldots,R$, where $(\varepsilon_{01},\ldots,\varepsilon_{0R})^T$ and $(\varepsilon_{11},\ldots,\varepsilon_{1R})^T$ are mean 0 random error vectors. These random vectors are independent, but the error terms within each vector are correlated. Based on the regression model, Hsieh and Turnbull[9] propose to obtain a generalized LS estimator for α_1 and β_1.

The proposed method for multiple binormal ROC curves

The least squares method of Hsieh and Turnbull[9] only deals with 1 diagnostic test. It is possible to extend it to allow multiple diagnostic tests. Our extension still builds on the intrinsic property of the ROC curve that the ROC curve is invariant to any monotonic transformation of the test results. We assume that after some unknown transformation, the latent test results follow normal distributions for the case and control subjects. Suppose that for test 1, after some monotone transformation, $\tilde{Y}_{1j} \sim N(\mu_{1,1}, \sigma_{1,1}^2)$ in the cancer group and $\tilde{X}_{1i} \sim N(\mu_{\ell,0}, \sigma_{1,0}^2)$ in the control group:

$$\begin{aligned} P\left(\tilde{Y}_{1j} \leq c_{1,1}\right) &= \Phi\left(\frac{c_{1,1} - \mu_{1,1}}{\sigma_{1,1}}\right), \\ P\left(\tilde{X}_{1i} \leq c_{1,0}\right) &= \Phi\left(\frac{c_{1,0} - \mu_{1,0}}{\sigma_{1,0}}\right) \end{aligned} \tag{2}$$

The equations above lead to $\Phi^{-1}(1 - Se_{1,1}) = (\mu_{1,0} - \mu_{1,1})/\sigma_{1,1} + \sigma_{1,0}/\sigma_{1,1}\Phi^{-1}(Sp_{1,1})$. Let $\alpha_1 = -(\mu_{1,0} - \mu_{1,1})/\sigma_{1,1}$ and $\beta_1 = \sigma_{1,0}/\sigma_{1,1}$. Denote $Se_{1,1} = Se(c_{1,1})$ and $Sp_{1,1} = Sp(c_{1,1})$. We have the following equation:

$$\Phi^{-1}\left(1 - Se_{1,1}\right) = -\alpha_1 + \beta_1 \Phi^{-1}\left(Sp_{1,1}\right) \tag{3}$$

The resulting ROC curve for test 1 can be written as $ROC_1(u) = \Phi(\alpha_1 + \beta_1 \Phi^{-1}(u))$. Consider the test ℓ, for $\ell = 2,\ldots,L$. Suppose that after some monotone transformation, $\tilde{X}_{\ell j} \sim N(\mu_{\ell,1}, \sigma_{\ell,1}^2)$ and $\tilde{Y}_{\ell i} \sim N(\mu_{\ell,0}, \sigma_{\ell,0}^2)$. Note that the transformation may vary among modalities, but it should be the same for the cancer and control subjects for the same modality. For test ℓ, the following equations give the relationship between the rating categories and normal distribution parameters for the more general setting in which the nondiseased population can take on any normal distribution:

$$\begin{aligned} P\left(\tilde{X}_{1j} \leq c_{1,1}\right) &= \Phi\left(\frac{c_{1,1} - \mu_{1,1}}{\sigma_{1,1}}\right), \\ P\left(\tilde{Y}_{1i} \leq c_{1,0}\right) &= \Phi\left(\frac{c_{1,0} - \mu_{1,0}}{\sigma_{1,0}}\right) \\ P\left(\tilde{X}_{\ell j} \leq c_{\ell,1}\right) &= \Phi\left(\frac{c_{\ell,1} - \mu_{\ell,1}}{\sigma_{\ell,1}}\right), \\ P\left(\tilde{Y}_{\ell i} \leq c_{\ell,0}\right) &= \Phi\left(\frac{c_{r,0} - \mu_{\ell,0}}{\sigma_{\ell,0}}\right) \end{aligned} \tag{4}$$

Let $\alpha_\ell = -(\mu_{\ell,0} - \mu_{\ell,1})/\sigma_{\ell,1} - (\mu_{1,0} + \mu_{1,1})/\sigma_{1,1}$ and $\beta_\ell = \sigma_{\ell,0}/\sigma_{\ell,1} - \sigma_{1,0}/\sigma_{1,1}$. The relationship between the sensitivities and specificities at varying cutoff points for multiple tests can be expressed as follows:

$$\begin{aligned} \Phi^{-1}\left(1 - Se_{1r_1}\right) &= -\alpha_1 + \beta_1 \Phi^{-1}\left(Sp_{1r_1}\right) \\ \Phi^{-1}\left(1 - Se_{\ell r_\ell}\right) &= -\left(\alpha_1 + \alpha_\ell\right) + \\ &\left(\beta_1 + \beta_\ell\right)\Phi^{-1}\left(Sp_{\ell r_\ell}\right), \quad \ell = 2,\ldots,L \end{aligned} \tag{5}$$

Here, the ℓth ROC curve is given by $ROC_\ell(u) = \Phi((\alpha_1 + \alpha_\ell) + (\beta_1 + \beta_\ell)\Phi^{-1}(u))$.

Let u be the FPR, which is 1 – specificity. The empirical functions of $Se_{\ell r_\ell}$ and $Sp_{\ell r_\ell}$ are defined by DeLong et al[5] and Tang and Zhou[7] as follows:

$$\hat{Se}_{\ell r_\ell} = \frac{1}{m}\sum_{i=1}^{m} I\left(X_{\ell i} > c_{\ell, r_\ell}\right) \text{ and } \hat{Sp}_{\ell r_\ell} = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_{\ell j} \leq c_{\ell, r_\ell}\right) \tag{6}$$

We will substitute the estimated proportions in the above model. The regression equations with error terms can be written as follows:

$$\begin{aligned} &\Phi^{-1}\left(1 - \hat{Se}_{1r_1}\right) = -\alpha_1 + \beta_1 \Phi^{-1}\left(Sp_{1r_1}\right) + \varepsilon_{1,r_1,0} \\ &\Phi^{-1}\left(\hat{Sp}_{1r_1}\right) = \Phi^{-1}\left(Sp_{1r_1}\right) + \varepsilon_{1,r_1,1}, \text{ for } r_1 = 1,\ldots,R_1 \\ &\Phi^{-1}\left(1 - \hat{Se}_{\ell r_\ell}\right) = -\left(\alpha_1 + \alpha_\ell\right) + \left(\beta_1 + \beta_\ell\right)\Phi^{-1}\left(Sp_{\ell r_\ell}\right) + \varepsilon_{\ell,r_\ell,0} \\ &\Phi^{-1}\left(\hat{Sp}_{\ell r_\ell}\right) = \Phi^{-1}\left(Sp_{\ell r_\ell}\right) + \varepsilon_{\ell,r_\ell,1}, \text{ for } r_\ell = 1,\ldots,R_\ell, \\ &\text{and } \ell = 2,\ldots,L \end{aligned} \tag{7}$$

Our parametric procedure is based on the model (equation (7)). We outline our parametric LS procedure as follows:

- *Step 1*. Obtain the empirical functions $\hat{S}e_{\ell r_\ell}$ and $\hat{S}p_{\ell r_\ell}$ at the threshold c_{ℓ,r_ℓ}, for $r_\ell = 1,\ldots,R_\ell$ and $\ell = 1,\ldots,L$.
- *Step 2*. Transform the sample proportions, $1-\hat{S}e_{\ell r_\ell}$, by Φ, and define the following vectors:

$$\boldsymbol{Y}_\ell = \left(\Phi^{-1}\left(1-Se_{\ell 1}\right),\Phi^{-1}\left(1-Se_{\ell 2}\right),\ldots,\Phi^{-1}\left(1-Se_{\ell R_\ell}\right)\right)^T$$
$$\hat{\boldsymbol{Y}}_\ell = \left(\Phi^{-1}\left(1-\hat{S}e_{\ell 1}\right),\Phi^{-1}\left(1-\hat{S}e_{\ell 2}\right),\ldots,\Phi^{-1}\left(1-\hat{S}e_{\ell R_\ell}\right)\right)^T$$

- *Step 3*. Combine $\boldsymbol{Y}_1,\boldsymbol{Y}_2,\ldots,\boldsymbol{Y}_L$ to get the following linear regression equation:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}_0 \tag{8}$$

where $\boldsymbol{\theta} = (\alpha_1,\ldots,\alpha_L,\beta_1,\ldots,\beta_L)^T$, $\boldsymbol{Y} = ((\boldsymbol{Y}_1)^T,\ldots,(\boldsymbol{Y}_L)^T)^T$ is a $(\sum_\ell R_\ell)\times 1$ vector, and the $(\sum_\ell R_\ell)\times(2L)$ design matrix $\boldsymbol{X}$ is as follows:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 & 0 & 0 & \cdots & 0 \\ \boldsymbol{X}_2 & \boldsymbol{X}_2 & 0 & \cdots & 0 \\ \boldsymbol{X}_3 & 0 & \boldsymbol{X}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_L & 0 & 0 & \cdots & \boldsymbol{X}_L \end{pmatrix}$$

with its $R_\ell \times 2$ submatrices

$$\boldsymbol{X}_\ell = \begin{pmatrix} -1 & \cdots & -1 \\ \Phi^{-1}\left(Sp_{\ell 1}\right) & \cdots & \Phi^{-1}\left(Sp_{\ell R_\ell}\right) \end{pmatrix}^T$$

Also, the $(\sum_\ell R_\ell)\times 1$ error vector e_0 is given by $\boldsymbol{\varepsilon}_0 = (\varepsilon_{1,1,0},\ldots,\varepsilon_{1,R_1,0},\ldots,\varepsilon_{L,1,0},\ldots,\varepsilon_{L,R_L,0})^T$.

- *Step 4*. Replace $Sp_{\ell R_\ell}$ using the $\hat{S}p_{\ell R_\ell}$ in $\boldsymbol{X}$ to obtain $\hat{\boldsymbol{X}}$.
- *Step 5*: Obtain the final ordinal LS estimator

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}^T\hat{\boldsymbol{Y}} \tag{9}$$

The method by Tang and Zhou[7] also first creates the response variable based on the test results and constructs a linear model so that the parameters are estimated using the LS method. However, the key difference between the proposed method and the LS method in a study by Tang and Zhou[7] lies in the response variable. The response variable in the latter is transformed empirical ROC curves at different thresholds. It takes on many values for continuous test results, but few values for ordinal test results. The limited number of values for the response variable makes it impractical to use the method by Tang and Zhou[7] for ordinal data. However, the response variable in the proposed method takes on values for the specificities and sensitivities. The number of values is at least 2 times the number of distinct ordinal test results.

Asymptotic properties of the parameter estimates

We study the asymptotic properties of the LS parameter estimates in the context of 2 tests for simplicity. Denote $\mathbf{S}_p = (Sp_{\ell,r_\ell} : r_\ell = 1,\ldots,R_\ell,\ \ell = 1,\ldots,L)^T$, $\mathbf{S}_e = (Se_{\ell,r_\ell} : r_\ell = 1,\ldots,R_\ell,\ \ell = 1,\ldots,L)^T$, and $\mathbf{S} = (\mathbf{S}_p^T,\mathbf{S}_e^T)^T$ with components given by expression (6). Denote $\mathbf{S}_0$ the true value of $\mathbf{S}$ generating the observed data, and $\hat{\mathbf{S}} = (\hat{\mathbf{S}}_{\ell,r_\ell} : r_\ell = 1,\ldots,R_\ell,\ \ell = 1,\ldots,L)$, with each $\hat{\mathbf{S}}_{\ell,r_\ell}$ being the empirical estimators of $\mathbf{S}_0(c_{\ell,r_\ell})$ at the threshold c_{ℓ,r_ℓ}. The formula is shown in equation (6). Let

$$\mathbf{P} = \left(p_{ij}\right)_{1\leqslant i,j\leqslant L}, \text{ with } p_{ij} = P\left(X_{i1} > c_{i,r_i}, X_{j1} > c_{j,r_j}\right)$$
$$\mathbf{Q} = \left(q_{ij}\right)_{1\leqslant i,j\leqslant L}, \text{ with } q_{ij} = P\left(Y_{i1} > c_{i,r_i}, Y_{j1} > c_{j,r_j}\right)$$

Denote $E\Phi^{-1}(Sp_\ell) = \lim R_\ell^{-1}\sum_{r=1}^{R_\ell}\Phi^{-1}(Sp_{\ell,r})$, $E(\Phi^{-1}(Sp_l))^2 = \lim R_\ell^{-1}\sum_{r=1}^{R_\ell}(\Phi^{-1}(Sp_{\ell,r}))^2$, $E\Phi^{-1}(1-Se_\ell) = \lim R_l^{-1}\sum_{r=1}^{R_\ell}\Phi^{-1}(1-Se_{\ell,r})$, and $E[\Phi^{-1}(Sp_\ell)\Phi(1-Se_\ell)] = \lim R_\ell^{-1}\sum_{r=1}^{R_\ell}\Phi^{-1}(Sp_{\ell,r_\ell})\Phi^{-1}(1-Se_{\ell,r_\ell})$:

$$\Omega_\ell = \begin{pmatrix} 1 & E\Phi^{-1}(Sp_\ell) \\ E\Phi^{-1}(Sp_\ell) & E(\Phi^{-1}(Sp_\ell))^2 \end{pmatrix}, \quad (\ell = 1,\ldots,L)$$

and with $\lambda_\ell = \lim R_\ell/R$. Let Ω be the $2L\times 2L$ matrix:

$$\Omega = \begin{pmatrix} \sum_{\ell=1}^{L}\lambda_\ell\Omega_\ell & \lambda_2\Omega_2 & \lambda_3\Omega_3 & \cdots & \lambda_L\Omega_L \\ \lambda_2\Omega_2 & \lambda_2\Omega_2 & 0 & \cdots & 0 \\ \lambda_3\Omega_3 & 0 & \lambda_3\Omega_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_L\Omega_L & 0 & 0 & \cdots & \lambda_L\Omega_L \end{pmatrix}$$

Recall $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$.

Denote $\xrightarrow{a.s.}$ for convergence almost surely and $\xrightarrow{D}$ for convergence in distribution. The following condition will be used:

(C1). Ω is positive definite.

Theorem 1

1. For fixed $R_\ell(\ell = 1,\ldots,L)$, let $\eta = \lim n/(m+n)$. Then, as $m,n\to\infty$,

$$\hat{\mathbf{S}}\xrightarrow{a.s.}\mathbf{S}_0, \text{ and } \sqrt{\frac{mn}{m+n}}\left(\hat{\mathbf{S}}-\mathbf{S}_0\right)\xrightarrow{D}N(\mathbf{0},\Lambda),$$
$$\Lambda = \begin{pmatrix} \eta\mathbf{P} & \mathbf{0} \\ \mathbf{0} & (1-\eta)\mathbf{Q} \end{pmatrix}$$

2. Assume (C1) and $R_\ell/R\to\lambda_\ell > 0$, for all $1\leqslant\ell\leqslant L$, with L fixed. Let $m,n\to\infty$ first and then $R_\ell\to\infty(\ell = 1,\ldots,L)$, then,

$$\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0 \text{ and } \sqrt{R}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{D} N\left(\mathbf{0}, \sigma^2 \Omega^{-1}\right)$$

Simulation Studies

Estimates from normal test results

We conduct simulation studies to investigate the finite sample performance of the proposed method. Two simulation settings are used to simulate data sets. The first setting is under the normal distributions for the cancer and control populations, and the second is under the lognormal distributions. Given the same parameter values, the true ROC curve should be the same for normal or lognormal data sets due to the monotonic invariant property of the ROC curve. Because the proposed method only deals with the sensitivities and specificities, the estimated ROC curve should be valid for both distributions given the correctly specified link and baseline functions.

Under the normal setting, we simulate normal observations for 2 diagnostic tests. The bivariate normal data were simulated as outcomes from paired tests. Assume that the bivariate normal models had the forms $(X_1, X_2) \sim N((1,1), \Sigma_1)$ and $(Y_1, Y_2) \sim N((0,0), \Sigma_1)$, where

$$\Sigma_1 = \begin{pmatrix} 1 & \sqrt{2}\rho \\ \sqrt{2}\rho & 1 \end{pmatrix}$$

with ρ denoting the correlation in bivariate outcomes.

We simulate 1000 replications with all combinations of $m = (40, 100, 200)$ and $n = (50, 150, 300)$ under ρ = (0.1, 0.2, 0.4, 0.5). For each replication, the threshold points are chosen to be normal quantiles of 100 equally spaced points ranging from 0.001 to 0.999. The threshold points are used to dichotomize the continuous observations. The dichotomized data are used to obtain empirical sensitivities and specificities which are the proportions of the test results greater than the threshold point for the cancer group and the proportions of the test results less than the threshold points for the noncancer group, respectively. Model (7) is then fit to the estimated sensitivities and specificities to obtain the estimates for α_1, β_1, α_2, and β_2. The LS method by Tang and Zhou[7] is also fit to the simulated data sets for comparison with the proposed method. The difference in the Tang and Zhou (TZ) LS method is that it estimates the empirical ROC curve at 100 equally spaced points ranging from 0.001 to 0.999. The transformed empirical ROC curves at these points are the observations for the response variable in the linear model. Table 1 presents the biases and root-mean-square errors (RMSEs) of these ROC parameter estimates by the proposed method and the TZ method. The biases are generally small for the proposed method. As the sample sizes become larger, the biases do not change much. The RMSEs tend to become smaller when both sample sizes for the cancer and noncancer groups become larger. We can also see that the correlation between the 2 tests does not affect the biases and RMSEs. Table 2 presents the biases and square RMSEs of these ROC parameter estimates by the TZ method. The biases and RMSEs are close to those by the proposed method.

Estimates from lognormal test results

We use the same setting as in the previous section to simulate the bivariate normal results first. We then take the exponential of the normal results to generate bivariate lognormal results. We again apply the proposed method and TZ method to the simulated data sets. Table 3 shows the biases and RMSEs of the ROC parameter estimates for the simulated lognormal data under all combinations of the sample sizes and correlation values for the proposed method. The simulation results show that the proposed approach has nice finite sample property as the biases and RMSEs are small even for small sample sizes. The results are similar as those for normal test results. This indicates that the proposed method is robust to monotonic transformation of the test results. As the sample sizes for both cancer and control groups become larger, the RMSEs tend to decrease. Table 4 shows the biases and RMSEs of the ROC parameter estimates for the simulated lognormal data under all the combinations of the sample sizes and correlation values for the TZ method. The biases and RMSEs are close to those by the proposed method.

Applications to Cancer Diagnostic Biomarkers

We apply our method to 2 real data sets. The first example investigates the diagnostic accuracy of serum biomarkers on pancreatic cancer, and the second example investigates the accuracy of gene expression biomarkers on ovarian cancer.

Pancreatic cancer tests

We use the cancer diagnostic example in Wieand et al[13] to illustrate the proposed method. The example is popular for the illustration of methodologies on estimating ROC curves from correlated data. Two pancreatic cancer tests, CA 19-9 and CA 125, were measured on 51 patients with pancreatitis and 90 patients with pancreatic cancer. It is of interest to estimate the ROC curves for these 2 tests. The test results approximately follow normal distributions after some monotonical transformation. We can assume a bivariate binormal ROC model for these tests. The estimation procedure follows the steps in section "Simulation Studies." We first define $R_1 = R_2 = 100$ cutpoints for 2 tests. For each test, we take the minimum and maximum of the combined results from both cancer and control groups as the lower and upper bounds and then obtain 100 equally spaced points within the bounds. The sets of cutpoints, c_{ℓ, r_ℓ}, are different for 2 tests. The sensitivity at a cutpoint for a test, $\hat{S}e_{\ell r_1}$, is calculated as the proportions that the test results are greater than the cutpoint, and the specificity, $\hat{S}p_{\ell r_1}$, is calculated as the proportion that the test results are less than or equal to the cutpoint. The pairs of sensitivities and specificities are then obtained for each test. The response vector in the linear model is given as follows:

Table 1. Biases (in %) and RMSEs for normal data—proposed method.

ρ		$M = 40$						$M = 100$						$M = 200$					
		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
0.10	α_1	−0.78	0.22	−1.67	0.22	−1.61	0.21	−2.20	0.15	−1.33	0.13	−0.98	0.13	−0.57	0.12	−0.95	0.10	−1.52	0.10
	β_1	−1.17	0.11	0.03	0.10	−0.70	0.09	−1.77	0.09	−1.20	0.07	−0.58	0.06	−1.48	0.07	−1.00	0.06	−1.11	0.05
	α_2	−2.59	0.33	−0.59	0.32	−2.90	0.31	−0.50	0.22	−0.50	0.19	−1.61	0.20	−1.27	0.16	−0.65	0.15	−0.27	0.14
	β_2	0.29	0.16	−0.22	0.13	0.58	0.13	0.23	0.12	0.43	0.09	−0.55	0.09	−0.24	0.10	−0.45	0.08	−0.12	0.07
0.20	α_1	−1.95	0.22	−0.94	0.20	−1.38	0.21	−0.77	0.15	−0.59	0.13	−1.39	0.13	−0.82	0.12	−0.82	0.10	−1.04	0.10
	β_1	−1.51	0.12	−0.65	0.10	−0.59	0.09	−1.78	0.08	−1.41	0.07	−0.88	0.06	−1.82	0.08	−1.42	0.06	−0.89	0.05
	α_2	−2.55	0.34	−2.45	0.33	−1.36	0.32	−3.04	0.23	−1.82	0.19	−0.74	0.19	−1.26	0.16	−0.56	0.14	−0.64	0.13
	β_2	0.71	0.17	0.03	0.14	0.67	0.13	−0.32	0.11	0.14	0.09	0.11	0.08	0.09	0.10	0.04	0.07	−0.41	0.06
0.40	α_1	−1.41	0.21	−1.05	0.22	−0.16	0.20	−1.44	0.15	0.25	0.14	−0.56	0.13	−1.26	0.12	−0.32	0.10	−1.06	0.10
	β_1	−0.36	0.11	−0.42	0.09	0.29	0.09	−1.60	0.08	−0.56	0.07	−0.54	0.06	−1.78	0.07	−1.32	0.06	−1.09	0.05
	α_2	−2.02	0.30	−2.02	0.28	−2.44	0.28	−1.17	0.19	−1.85	0.17	−1.64	0.17	−0.75	0.14	−1.25	0.13	−0.36	0.13
	β_2	−0.55	0.15	0.11	0.13	−0.71	0.12	−0.07	0.11	−0.48	0.09	−0.51	0.08	0.54	0.09	0.17	0.07	−0.04	0.07
0.50	α_1	−1.03	0.23	−1.09	0.20	−0.74	0.20	−1.93	0.15	−0.48	0.13	−0.76	0.13	−0.40	0.12	−0.77	0.10	−0.63	0.10
	β_1	−0.92	0.12	−0.32	0.09	−0.16	0.09	−1.99	0.09	−0.93	0.07	−0.78	0.06	−1.68	0.08	−1.49	0.06	−0.95	0.05
	α_2	−2.82	0.28	−3.50	0.27	−2.29	0.27	−0.47	0.18	−1.80	0.17	−0.79	0.16	0.12	0.13	−1.79	0.12	−1.14	0.12
	β_2	0.14	0.16	−0.72	0.13	0.25	0.13	0.33	0.11	−0.05	0.09	−0.11	0.08	0.43	0.09	0.10	0.07	−0.07	0.06

Abbreviation: RMSEs, root-mean-square errors.
Results are based on 1000 realizations of bivariate normal model.

Table 2. Biases (in %) and RMSEs for normal data—TZ method.

ρ		$M = 40$						$M = 100$						$M = 200$					
		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
0.10	α_1	−0.95	0.23	−1.73	0.22	−1.26	0.21	0.74	0.15	−0.58	0.14	−0.30	0.13	0.60	0.12	0.70	0.10	0.28	0.10
	β_1	−1.74	0.13	−0.86	0.11	−1.04	0.10	−0.19	0.09	−0.96	0.08	−0.87	0.07	−0.64	0.07	−0.51	0.06	−0.46	0.05
	α_1	−2.88	0.36	−1.93	0.34	−3.36	0.33	−1.73	0.22	−0.72	0.21	−1.00	0.21	−0.54	0.16	−1.62	0.15	−1.12	0.15
	β_1	1.10	0.19	−0.07	0.16	0.21	0.16	−0.99	0.12	−0.43	0.11	−0.14	0.10	0.00	0.10	−0.34	0.08	−0.44	0.08
0.20	α_1	−0.17	0.21	−1.57	0.22	−2.62	0.22	0.65	0.15	−0.92	0.13	−1.15	0.13	1.94	0.12	0.22	0.10	−0.30	0.10
	β_1	−1.41	0.12	−1.19	0.11	−1.17	0.11	−1.29	0.09	−1.45	0.08	−1.16	0.07	−0.37	0.07	−0.52	0.06	−0.81	0.05
	α_1	−3.71	0.34	−0.92	0.34	−2.01	0.32	−1.39	0.20	−0.76	0.20	−1.13	0.20	−1.30	0.16	−0.76	0.15	−0.61	0.14
	β_1	0.50	0.18	0.81	0.16	0.79	0.16	0.24	0.12	0.35	0.11	−0.15	0.10	−0.19	0.10	−0.10	0.08	0.08	0.08
0.40	α_1	0.21	0.24	−3.07	0.23	−1.23	0.21	1.22	0.15	0.48	0.14	−0.86	0.13	1.88	0.12	0.25	0.10	0.00	0.09
	β_1	−1.16	0.13	−1.43	0.11	−0.86	0.11	−0.84	0.09	−1.08	0.08	−1.06	0.07	−0.42	0.08	−0.43	0.06	−0.94	0.06
	α_1	−2.90	0.31	−1.42	0.32	−3.72	0.30	−2.20	0.19	−1.81	0.19	−1.15	0.19	−1.57	0.15	−0.82	0.13	−0.19	0.13
	β_1	0.72	0.18	1.08	0.16	−0.60	0.16	−0.11	0.12	0.10	0.11	−0.50	0.10	0.06	0.10	−0.23	0.08	0.35	0.07
0.50	α_1	−1.12	0.24	−0.83	0.22	−1.91	0.22	1.21	0.15	−0.15	0.14	−1.05	0.14	1.15	0.12	−0.21	0.10	−1.00	0.10
	β_1	−0.61	0.13	−0.81	0.12	−1.28	0.11	−0.87	0.09	−1.23	0.08	−1.28	0.07	−0.29	0.07	−0.71	0.06	−0.62	0.05
	α_1	−0.51	0.29	−2.26	0.30	−3.07	0.28	−1.70	0.19	−2.08	0.18	−1.05	0.17	−0.65	0.14	−0.50	0.12	−0.70	0.13
	β_1	0.53	0.17	−0.31	0.17	0.25	0.15	0.30	0.12	−0.10	0.10	0.34	0.10	−0.27	0.09	0.01	0.08	−0.07	0.07

Abbreviations: RMSEs, root-mean-square errors; TZ, Tang and Zhou.
Results are based on 1000 realizations of bivariate normal model.

Table 3. Biases (in %) and RMSEs for lognormal data—proposed method.

ρ		$M = 40$						$M = 100$						$M = 200$					
		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
0.10	α_1	−1.78	0.23	−1.48	0.20	−1.25	0.20	−1.21	0.14	−0.90	0.14	−0.92	0.13	−1.26	0.12	−1.20	0.10	−0.83	0.10
	β_1	−0.86	0.11	0.01	0.10	−0.24	0.09	−1.59	0.09	−0.90	0.07	−0.62	0.06	−1.62	0.08	−1.46	0.06	−1.04	0.05
	α_2	−2.01	0.36	−1.92	0.32	−2.52	0.31	−1.69	0.22	−1.29	0.21	−2.01	0.20	−0.57	0.17	−0.59	0.15	−0.91	0.14
	β_2	−0.57	0.16	−0.57	0.14	−0.08	0.13	−0.10	0.11	−0.18	0.09	−0.68	0.08	−0.06	0.10	0.46	0.08	−0.32	0.07
0.20	α_1	−2.38	0.21	−1.21	0.21	−1.30	0.21	−0.53	0.15	−1.21	0.13	−0.77	0.13	−0.38	0.12	−1.24	0.10	−0.90	0.10
	β_1	−0.76	0.11	−0.47	0.10	−0.14	0.09	−1.42	0.08	−1.28	0.07	−0.79	0.06	−1.44	0.07	−1.49	0.06	−1.06	0.05
	α_2	−1.99	0.31	−3.06	0.31	−3.79	0.32	−1.80	0.20	−1.44	0.19	−1.19	0.19	−0.50	0.15	−0.99	0.14	−0.89	0.13
	β_2	0.28	0.16	−0.35	0.14	−0.23	0.13	−0.32	0.11	0.12	0.09	−0.15	0.08	0.50	0.09	0.25	0.07	−0.40	0.07
0.40	α_1	−1.85	0.23	−1.11	0.21	−1.55	0.20	−1.07	0.15	−1.40	0.13	−1.14	0.13	−1.52	0.12	−0.75	0.10	−0.26	0.10
	β_1	−1.48	0.12	−0.44	0.10	0.22	0.09	−1.19	0.08	−1.19	0.07	−0.69	0.06	−1.59	0.07	−1.24	0.06	−1.06	0.05
	α_2	−1.85	0.30	−1.70	0.29	−0.61	0.28	−1.30	0.19	−1.50	0.18	−0.21	0.17	0.53	0.14	−0.62	0.13	−0.76	0.13
	β_2	1.03	0.16	0.38	0.13	0.16	0.13	−0.12	0.11	−0.32	0.09	−0.18	0.08	0.13	0.09	−0.12	0.07	0.15	0.06
0.50	α_1	−2.52	0.21	−0.35	0.19	−0.41	0.21	−1.91	0.15	−0.94	0.14	−0.53	0.13	−1.42	0.12	−1.19	0.10	−1.00	0.10
	β_1	−0.76	0.11	−0.20	0.09	0.22	0.09	−1.81	0.09	−1.21	0.07	−0.23	0.06	−1.83	0.07	−1.27	0.06	−0.90	0.05
	α_2	−2.40	0.30	−2.56	0.27	−3.71	0.28	−0.64	0.18	−0.71	0.17	−1.20	0.16	−0.32	0.14	−0.39	0.12	−0.83	0.12
	β_2	−0.10	0.16	−0.24	0.13	−0.95	0.13	0.41	0.11	0.16	0.09	−0.64	0.08	−0.02	0.09	−0.09	0.07	−0.19	0.066

Abbreviation: RMSEs, root-mean-square errors.
Results are based on 1000 realizations of bivariate lognormal model.

Table 4. Biases (in %) and RMSEs for lognormal data—TZ method.

ρ		$M = 40$						$M = 100$						$M = 200$					
		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300		N = 50		N = 150		N = 300	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
0.10	α_1	−2.08	0.23	−2.34	0.22	−1.18	0.22	0.20	0.15	−0.69	0.14	−0.31	0.13	1.37	0.12	−0.36	0.10	−0.77	0.10
	β_1	−1.10	0.13	−0.96	0.11	−0.84	0.11	−0.73	0.09	−1.03	0.08	−1.04	0.07	−0.80	0.07	−0.85	0.06	−1.08	0.05
	α_1	−2.97	0.37	−0.88	0.34	−3.00	0.33	−1.11	0.22	−1.48	0.20	−2.27	0.21	−1.62	0.17	0.41	0.15	−0.26	0.14
	β_1	−0.49	0.19	0.42	0.16	0.50	0.15	−0.43	0.13	−0.68	0.11	−0.02	0.11	0.11	0.10	0.47	0.08	0.39	0.07
0.20	α_1	−1.05	0.23	−1.12	0.21	−1.34	0.21	0.21	0.16	−0.53	0.14	−1.32	0.13	1.47	0.12	−0.02	0.10	−0.72	0.10
	β_1	−0.51	0.12	−0.53	0.11	−1.01	0.11	−1.22	0.10	−0.94	0.07	−1.13	0.07	−0.36	0.07	−0.74	0.06	−0.58	0.05
	α_1	−1.43	0.34	−4.06	0.32	−2.64	0.33	−1.41	0.22	−1.90	0.20	−0.90	0.19	−1.68	0.16	−0.82	0.14	−0.52	0.14
	β_1	−0.07	0.17	−1.24	0.16	0.25	0.16	0.23	0.13	−0.36	0.11	−0.13	0.10	−0.38	0.10	0.09	0.08	−0.38	0.08
0.40	α_1	0.44	0.22	−2.02	0.22	−2.80	0.22	0.40	0.15	−0.77	0.14	−1.45	0.14	1.25	0.12	0.75	0.10	−0.86	0.09
	β_1	−1.17	0.13	−0.96	0.11	−1.72	0.11	−0.86	0.09	−0.72	0.08	−1.42	0.07	−0.48	0.07	−0.33	0.06	−0.58	0.06
	α_1	−3.74	0.32	−3.45	0.32	−2.72	0.32	−0.91	0.19	−1.52	0.19	−0.87	0.18	−0.78	0.14	−1.16	0.13	−0.30	0.13
	β_1	−0.32	0.18	−0.12	0.17	0.35	0.16	−0.19	0.12	−0.46	0.11	0.23	0.10	0.00	0.10	−0.08	0.08	−0.09	0.08
0.50	α_1	−1.78	0.23	−2.79	0.22	−3.26	0.22	0.31	0.15	−1.46	0.14	−0.59	0.14	2.04	0.12	−0.18	0.10	−0.08	0.10
	β_1	−1.44	0.13	−1.40	0.11	−1.07	0.11	−0.98	0.09	−1.39	0.08	−1.00	0.07	−0.40	0.07	−0.54	0.06	−0.70	0.05
	α_1	−2.84	0.30	−3.24	0.27	−0.74	0.30	−1.84	0.19	−0.72	0.17	−1.22	0.17	−1.54	0.13	−0.61	0.12	−1.19	0.13
	β_1	−0.41	0.17	−0.12	0.15	0.42	0.16	0.12	0.11	0.05	0.10	0.02	0.10	0.13	0.09	−0.45	0.08	0.00	0.08

Abbreviations: RMSEs, root-mean-square errors; TZ, Tang and Zhou.
Results are based on 1000 realizations of bivariate lognormal model.

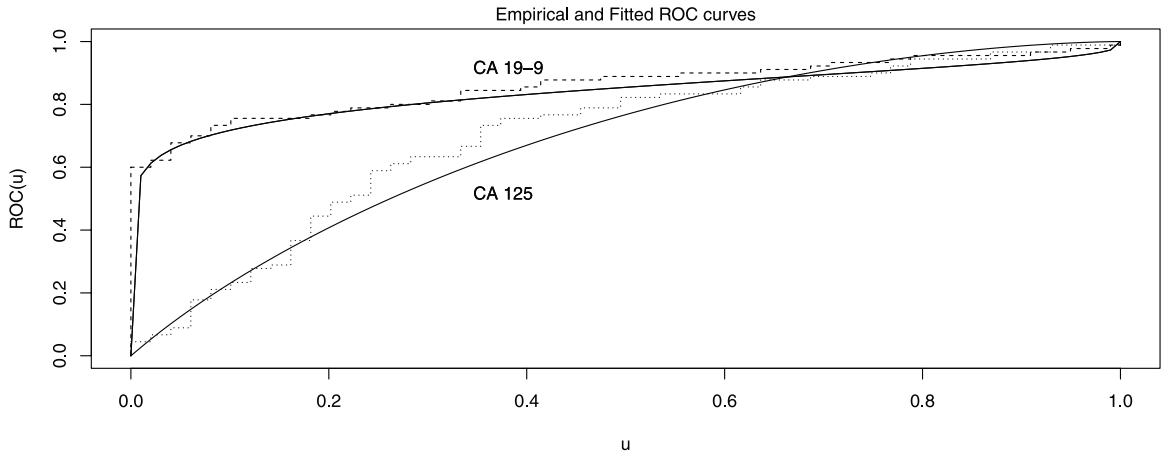

Figure 2. ROC curves for CA 19-9 and CA 125: solid lines, the proposed method; dashed and dotted lines, empirical ROC curves. ROC indicates receiver operating characteristic.

$$\boldsymbol{Y}_\ell = \left(\Phi^{-1}\left(1-\hat{S}e_{1,1}\right),\Phi^{-1}\left(1-\hat{S}e_{1,2}\right),\ldots,\Phi^{-1}\left(1-\hat{S}e_{1,100}\right)\right.$$
$$\left.\Phi^{-1}\left(1-\hat{S}e_{2,1}\right),\Phi^{-1}\left(1-\hat{S}e_{2,2}\right),\ldots,\Phi^{-1}\left(1-\hat{S}e_{2,100}\right)\right)^T \quad (10)$$

and the design matrix is as follows:

$$\widehat{X}=\begin{pmatrix} \begin{pmatrix} -1 & \cdots & -1 \\ \Phi^{-1}\left(\hat{S}p_{\ell 1}\right) & \cdots & \Phi^{-1}\left(\hat{S}p_{\ell,100}\right) \end{pmatrix}^T & 0 \\ \begin{pmatrix} -1 & \cdots & -1 \\ \Phi^{-1}\left(\hat{S}p_{\ell 1}\right) & \cdots & \Phi^{-1}\left(\hat{S}p_{\ell,100}\right) \end{pmatrix}^T & \begin{pmatrix} -1 & \cdots & -1 \\ \Phi^{-1}\left(\hat{S}p_{\ell 1}\right) & \cdots & \Phi^{-1}\left(\hat{S}p_{\ell,100}\right) \end{pmatrix}^T \end{pmatrix}$$

The final LS estimates are obtain through the following equation:

$$\hat{\boldsymbol{\theta}} = ((\widehat{X})^T\widehat{X})^{-1}(\widehat{X})^T\boldsymbol{Y} \quad (11)$$

The parameter estimates are $(\hat{\alpha}_1,\hat{\beta}_1,\hat{\alpha}_2,\hat{\beta}_2) = (1.0550, 0.3748, -0.3252, 0.7710)$. Based on these estimates, the estimated ROC curves are given by $ROC_1(u)$ = $\Phi(1.0550 + 0.3748\Phi^{-1}(u))$ for CA19-9 and $ROC_2(u)$ = $\Phi(0.7298 + 1.1457\Phi^{-1}(u))$ for CA 125. Figure 2 shows the fitted ROC curves for 2 tests. Both fitted curves are close to the empirical ROC curves. Unlike the rough empirical curves, the fitted curves are much smoother.

Ovarian cancer tests

We also illustrate our method using a gene expression data set previously analyzed using a suite of traditional ROC methods in a study by Pepe et al.[14] Briefly, these data report messenger RNA (mRNA) expression levels for 1536 gene clones in 30 subjects with ovarian cancer and 23 without cancer. We will focus our example on 2 gene clones, *SPINT2* and *TACSTD1*, among the top 10 ranking clones identified in this prior work. *SPINT2* is associated with ovarian cancer[15] and fallopian tube carcinomas specifically.[16] Elevated expression levels of *TACSTD1*, also called EPCAM, have been associated with local and metastatic prostate cancer[17] and colorectal cancer[18] while potentially being protective against ovarian cancer.[19]

We follow the same approach from the first example to define $R_1 = R_2 = 100$ cutpoints for 2 biomarkers and obtain 100 equally spaced points within the bounds. The parameter estimates based on the proposed method are $(\hat{\alpha}_1,\hat{\beta}_1,\hat{\alpha}_2,\hat{\beta}_2) = (2.0940, 0.6561, -0.3417, 1.0591)$. Based on these estimates, the estimated ROC curves are given by $ROC_1(u) = \Phi(2.0940 + 0.6561\Phi^{-1}(u))$ for *SPINT2* and $ROC_2(u) = \Phi(1.7523 + 1.7152\Phi^{-1}(u))$ for *TACSTD1*. The TZ method[7] is also applied to the data set to show the difference in the fitted ROC curves. The parameter estimates based on the proposed method are $(\hat{\alpha}_1^*,\hat{\beta}_1^*,\hat{\alpha}_2^*,\hat{\beta}_2^*) = (1.7078, 0.3242, -0.2759, 0.7907)$. Based on these estimates, the estimated ROC curves are given by $ROC_1^*(u) = \Phi(1.7078 + 0.3242\Phi^{-1}(u))$, for *SPINT2* and $ROC_2^*(u) = \Phi(1.4319 + 1.1149\Phi^{-1}(u))$ for *TACSTD1*. Figure 3 shows the fitted ROC curves from the

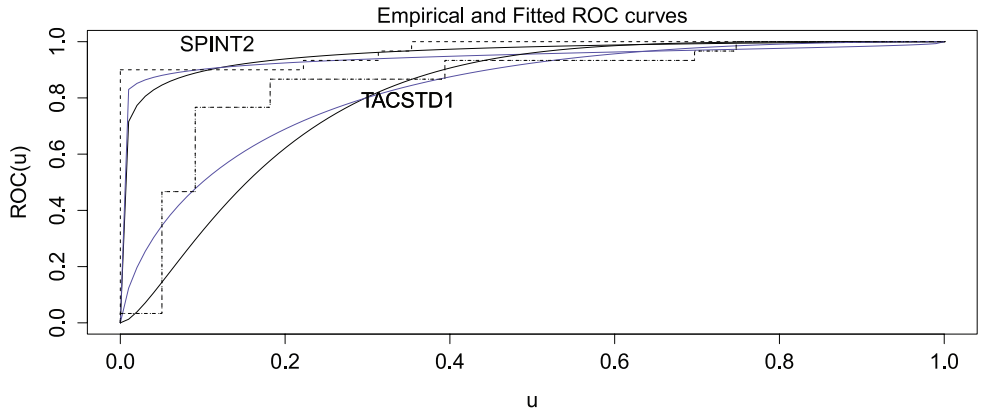

Figure 3. ROC curves for *SPINT2* and *TACSTD1*: solid lines, the fitted ROC curves with the proposed method in black and TZ method in blue; dashed lines, empirical ROC curves. ROC indicates receiver operating characteristic; TZ, Tang and Zhou.

proposed method and the TZ method for 2 tests. The fitted curves are close to the empirical ROC curves. Unlike the rough empirical curves, the fitted curves are much smoother.

Discussion

This article proposes an LS method to estimate the ROC parameters. The method builds on the estimated sensitivities and specificities. This method differs from that of Tang and Zhou[7] by handling the case of continuous response data and ordinal response data. The key difference between the 2 methods lies in the response variable. The response variable in the latter is transformed empirical ROC curves at different thresholds. It takes on many values for continuous test results, but few values for ordinal test results. The limited number of values for the response variable makes it impractical for ordinal data. However, the response variable in the proposed method takes on many more distinct values so that the method yields valid estimates for ordinal data. The simulation studies show that the proposed method has good finite sample performance for both simulated normal and lognormal data. The method also shows satisfactory results in cancer diagnostic examples.

As demonstrated by Hanley,[20] the binormal ROC curve tends to fit data to other distributions reasonably well. However, the assumption of the binormal ROC curve may seem quite strong because the data need to be normal after some unknown transformation. As a future research topic, more simulation studies need to be conducted for other distributions and for ordinal data to investigate the finite sample performance of the proposed method.

The method proposed here assumes that the gold standard is known. Future topics include the estimation of the ROC curves without the presence of the gold standard or when the gold standard is imperfect. It is challenging to do so because our method requires valid estimates for sensitivities and specificities. The method of Hui and Walter[21] may be applied, but 2 or more populations are required for the estimation. If 2 binary tests are to be evaluated from the samples in 1 population, the sensitivity and specificity cannot be estimated with the absence of a gold standard. Under this situation, the 5 parameters to be estimated involve the prevalence, sensitivities, and specificities for 2 tests. However, only 3 degrees of freedom are allowed with testing within 1 population and are not sufficient for estimating 5 parameters. Testing on the samples from 2 populations increases the degree of freedom to 6. Hui and Walter[21] consider the setting in which multiple tests are applied to several populations and discuss the approaches to estimate the sensitivities and specificities. The estimated sensitivities and specificities can potentially be used as response variables in our method to generate valid ROC curve estimators for continuous data.

Author Contributions

Conceived and designed the experiments: LT. Analyzed the data: LT, XC. Wrote the first draft of the manuscript: LT, AY. Made critical revisions and approved final version: JC, LC. All authors reviewed and approved of the final manuscript.

REFERENCES

1. Zhou XH, McClish DK, Obuchowski N. *Statistical Methods in Diagnostic Medicine*. New York, NY: Wiley; 2002.
2. Silvestri GA, Tanoue LT, Margolis ML, Barker J, Detterbeck F. Noninvasive staging of non-small cell lung cancer. *Chest*. 2003;123:137S–146S.
3. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
4. Dorfman DD, Alf E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: rating-method data. *J Math Psychol*. 1969;6:487–496.
5. DeLong ER, DeLong D, Clarke-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
6. Zhang Z, Pepe MS. A linear regression framework for receiver operating characteristic (ROC) curve analysis. http://www.bepress.com/uwbiostat/paper253. UW Biostatistics working paper series, Working paper 253. Published May 2005.
7. Tang LL, Zhou XH. Semiparametric inferential procedures for comparing multivariate ROC curves with interaction terms. *Stat Sinica*. 2009;19:1137–1145.
8. Tang LL, Zhou XH. A semiparametric separation curve approach for comparing correlated ROC data from multiple markers. *J Comput Graph Stat*. 2012;21:662–676.
9. Hsieh F, Turnbull BW. Non- and semi- parametric estimation of the receiver operating characteristics (ROC) curve. *Ann Stat*. 1996;24:25–40.
10. Morris DE, Pepe MS, Barlow WE. Contrasting two frameworks for ROC analysis of ordinal ratings. *Med Decis Making*. 2010;30:484–498.
11. Metz CE, Wang P-L, Kronman H. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Proceedings of the 8th Conference Information Processing in Medical Imaging (ed F Deconinck); August 29-September 2, 1983; Brussels, Belgium. The Hague, The Netherlands: Martinus Nijhoff.

12. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Med Decis Making*. 1998;18:110–121.
13. 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.
14. Pepe MS, Longton G, Anderson GL, Schummer M. Selecting differentially expressed genes from microarray experiments. *Biometrics*. 2003;59:133–142.
15. Nakamura K, Abarzua F, Kodama J, et al. Expression of hepatocyte growth factor activator inhibitors (HAI-1 and HAI-2) in ovarian cancer. *Int J Oncol*. 2009;34:345–353.
16. Nowee M, Snijders A, Rockx D, et al. DNA profiling of primary serous ovarian and fallopian tube carcinomas with array comparative genomic hybridization and multiplex ligation-dependent probe amplification. *J Pathol*. 2007;213:46–55.
17. Massoner P, Thomm T, Mack B, et al. EpCAM is overexpressed in local and metastatic prostate cancer, suppressed by chemotherapy and modulated by MET-associated miRNA-200c/205. *Br J Cancer*. 2014;111:955–964.
18. Chai X, Song R, Xu F. Expression changes in epithelial cell adhesion molecule during colorectal cancer tumorigenesis. *Genet Mol Res*. 2015;14:7624–7629.
19. Battista MJ, Cotarelo C, Jakobi S, et al. Overexpression of epithelial cell adhesion molecule protein is associated with favorable prognosis in an unselected cohort of ovarian cancer patients. *J Cancer Res Clin*. 2014;140:1097–1102.
20. Hanley JA. The robustness of the "binormal" assumptions used in fitting ROC curves. *Med Decis Making*. 1988;8:197–203.
21. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980;36:167–171.

Appendix 1

Proof of Theorem 1

1. Because each component of $\hat{\mathbf{S}}$ is an empirical version of the corresponding component in $\mathbf{S}_0$, $\hat{\mathbf{S}} \to \mathbf{S}_0$ (a.s.) is a direct consequence of the strong law of large numbers.

Note that $\hat{\mathbf{S}}_p$ and $\hat{\mathbf{S}}_e$ are independent; by central limit theorem, $\sqrt{m}(\hat{\mathbf{S}}_e - \mathbf{S}_{e,0}) \xrightarrow{D} N(\mathbf{0}, P)$ and $\sqrt{n}(\hat{\mathbf{S}}_p - \mathbf{S}_{p,0}) \xrightarrow{D} N(\mathbf{0}, Q)$. This gives the desired result.

2. For fixed L and R_ℓ $(\ell = 1,\ldots,L)$, as $n \to \infty$, $(\hat{Se}_{\ell,r_\ell}, \hat{Sp}_{\ell,r_\ell}) \to (Se_{\ell,r_\ell}, Sp_{\ell,r_\ell})$ $(r_\ell = 1,\ldots,R_\ell;\ \ell=1,...,L)$. So, we can replace $\hat{Sp}_{\ell,r_\ell}$ in $\mathbf{X}$ by Sp_{ℓ,r_ℓ} and replace $\hat{Se}_{\ell,r_\ell}$ in $\mathbf{Y}$ by Se_{ℓ,r_ℓ}. Then, as $R_\ell \to \infty$:

$$\frac{1}{R_\ell}\left(\mathbf{X}_\ell\right)' \mathbf{X}_\ell \xrightarrow{a.s.} \Omega_\ell := \begin{pmatrix} 1 & E\Phi^{-1}\left(Sp_\ell\right) \\ E\Phi^{-1}(Sp_\ell) & E\left(\Phi^{-1}\left(Sp_\ell\right)\right)^2 \end{pmatrix}, \quad (\ell = 1,\ldots,L)$$

Recall $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}$, we have the following equation:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{X}\boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}\right)$$
$$= \boldsymbol{\theta}_0 + \left(R^{-1}\mathbf{X}^T\mathbf{X}\right)^{-1}\frac{1}{R}\boldsymbol{\varepsilon} \xrightarrow{a.s.} \boldsymbol{\theta}_0$$

Note that (Ω is a $2L \times 2L$ matrix)

$$\frac{1}{R}\mathbf{X}^T\mathbf{X} = \frac{1}{R}\begin{pmatrix} \sum_{\ell=1}^{L}\mathbf{X}_\ell^T\mathbf{X}_\ell & \mathbf{X}_2^T\mathbf{X}_2 & \mathbf{X}_3^T\mathbf{X}_3 & \cdots & \mathbf{X}_L^T\mathbf{X}_L \\ \mathbf{X}_2^T\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{X}_2 & 0 & \cdots & 0 \\ \mathbf{X}_3^T\mathbf{X}_3 & 0 & \mathbf{X}_3^T\mathbf{X}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_L^T\mathbf{X}_L & 0 & 0 & \cdots & \mathbf{X}_L^T\mathbf{X}_L \end{pmatrix} \to \Omega$$

Also, as ε_r is iid with $E(\varepsilon_r) = 0$ and $Var(\varepsilon_r) = \sigma^2$, by central limit theorem and the Slutsky theorem, $\sqrt{R}(1/R)\mathbf{X}^T\boldsymbol{\varepsilon} \xrightarrow{D} N(\mathbf{0}, \Omega)$, and so,

$$\sqrt{R}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(R^{-1}\mathbf{X}^T\mathbf{X}\right)^{-1}\sqrt{R}\frac{1}{R}\mathbf{X}^T\boldsymbol{\varepsilon} \xrightarrow{D} N\left(\mathbf{0}, \sigma^2\Omega^{-1}\right)$$