Published in final edited form as:

Stat Med. 2017 November 20; 36(26): 4141–4152. doi:10.1002/sim.7415.

A semiparametric method for comparing the discriminatory ability of biomarkers subject to limit of detection

Lixuan Yin¹, Guoqing Diao¹, and Aiyi Liu²

¹Department of Statistics, George Mason University, Fairfax, 22030, VA, USA

²Biostatistics and Bioinformatics Branch, DIPHR, NICHD, NIH, Rockville, 20852, MD, USA

Abstract

Receiver operating characteristic curves and the area under the curves (AUC) are often used to compare the discriminatory ability of potentially correlated biomarkers. Many biomarkers are subject to limit of detection due to the instrumental limitation in measurements and may not be normally distributed. Standard parametric methods assuming normality can lead to biased results when the normality assumption is violated. We propose new estimation and inference procedures for the AUCs of biomarkers subject to limit of detection by using the semiparametric transformation model allowing for heteroscedasticity. We obtain the nonparametric maximum likelihood estimators by maximizing the likelihood for the observed data with limit of detection. The proposed estimators are shown to be consistent, asymptotically normal, and asymptotically efficient. Additionally, we propose a Wald type test statistic to compare the AUCs of 2 potentially correlated biomarkers with limit of detection. Extensive simulation studies demonstrate that the proposed method is robust to non-normality while performing as well as its parametric counterpart when the normality assumption is true. An application to an autism study is provided.

Keywords

AUC; limit of detection; nonparametric maximum likelihood estimator; ROC curve; semiparametric transformation model

1 | INTRODUCTION

Receiver operating characteristic (ROC) curve is a commonly used tool for evaluating the performance of biomarkers.^{1,2} The ROC curve of a biomarker is a plot of its true positive rate versus its false positive rate as its discrimination threshold varies. The true positive rate is known as sensitivity whereas the false positive rate as 1—specificity. As a summary statistic, the area under the ROC curve (AUC) is often used to quantify the diagnostic performance of a biomarker. Let U and V denote scores of a biomarker for the diseased group and the control group, respectively. Bamber³ showed that $AUC = P(U > V)$ that can be interpreted as the probability that the biomarker value is higher for a diseased subject in a

Correspondence: Department of Statistics, George Mason University, Fairfax, VA 22030, USA. gdiao@gmu.edu.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

randomly selected pair of healthy and diseased individuals. The closer the AUC is to 1, the better performance of the biomarker.³ An AUC value of 0.5 indicates that the biomarker has no discriminatory ability.

Both parametric and nonparametric estimation of the AUC has been well developed in the areas related to diagnostic medicine. One parametric approach is to assume U and V are independent with $U \sim N(\mu_U, \sigma_U^2)$ and $V \sim N(\mu_V, \sigma_V^2)$. Thus, AUC can be expressed as

$$AUC = \Phi \left(\frac{\mu_U - \mu_V}{\sqrt{\sigma_U^2 + \sigma_V^2}} \right),$$

where Φ is the cumulative standard normal distribution function. The AUC can then be estimated by replacing the unknown parameters with their estimators. When the normality assumption is violated, parametric transformations such as the log-transformation are often used.^{4,5} Alternatively, one can consider a nonparametric estimator of the AUC, which is equivalent to computing the well known Mann-Whitney statistic, after estimating the ROC curve as a step-function based on empirical cumulative distribution functions for the diseased group and the control group, respectively.^{2,3,6} One may also fit the ROC curve using kernel smoothing and then the AUC can be estimated through integration.^{7,8} For detailed comparison among different approaches, readers are referred to Zhou et al² and Faraggi and Reiser.⁹

Most of the aforementioned methods assume that the biomarker is fully observed. Many biomarkers, however, are subject to limit of detection (LOD), which is usually caused by the instrumental limitation in measurements being undetectable below or beyond some level, eg, previous studies.^{10–14} Missing data due to the LOD can give rise to biased parameter estimation and inefficient testing using standard methods. Various parametric methods have been developed in literature. Assuming normality, Mumford et al¹⁵ used a likelihood-based approach to estimating μ_U , μ_V , σ_U^2 , and σ_V^2 . Moreover, Perkins, Schisterman, and Vexler¹⁶ examined different correction methods and demonstrated that maximum likelihood estimation would lead to asymptotically unbiased estimators of AUC for both normally and gamma distributed biomarker scores. Vexler, Liu, and Schisterman¹² treated the biomarker data subject to LOD as censored data.

Under many circumstances, it is often of interest whether one biomarker has the same discriminatory ability as the other, and this is achieved by comparing their AUCs.¹⁷ Furthermore, it is interesting to estimate the confidence interval of AUCs and the difference between AUCs.¹⁸ However, when the measures of the biomarkers are subject to LOD, traditional methods may inflate type I error rate. Vexler et al¹² addressed these issues and proposed a general likelihood-based method to estimate the AUCs subject to LOD. Moreover, Vexler et al¹⁴ proposed a maximum likelihood ratio test (MLRT), using the likelihood function proposed by Lyles, Williams, and Chuachoowong¹⁰ to compare AUCs of 2 biomarkers. Additionally, they assumed the measurements of the 2 biomarkers follow a bivariate normal distribution, possibly after a known transformation. However, it is difficult

to identify the true transformation, and incorrect transformations can result in the inflation of type I error.

In this paper, we propose a semiparametric approach to estimating AUCs and comparing the AUCs of potentially correlated biomarkers that are subject to the LOD. Specifically, we propose to fit the data using the semiparametric transformation model (previous works,^{19–22} among others) allowing for heteroscedasticity, under which the data are normally distributed after an unspecified transformation. Recently, Kim, Zeng, and Zhou²³ considered a similar transformation model for multiple biomarkers in ROC analysis. Under the transformation model in Kim et al,²³ there is a constraint such that the transformation function equals 0 at 0 and an intercept term is required. The authors then proposed to estimate the positive and negative parts of the transformation function as well as the finite-dimensional parameters by using the nonparametric maximum likelihood approach. However, the resulting estimators may be different if signs of the biomarkers values are changed under an increasing transformation. In this paper, we propose a new estimation procedure that does not require the constraint on the transformation function or the intercept term in the model and the proposed estimators are invariant to any increasing transformations on the biomarkers.

The rest of the article is organized as follows. In Section 2, we introduce the semiparametric transformation model allowing for heteroscedasticity and derive the nonparametric maximum likelihood estimation. The nonparametric maximum likelihood estimator (NPMLE) of the AUC is shown to be consistent, asymptotically normal, and asymptotically efficient. Additionally, we propose a Wald test for comparing the AUCs of 2 biomarkers subject to LOD that appropriately accounts for their potential correlation. We examine the finite sample performance of the proposed NPMLEs and Wald test through extensive simulation studies in Section 3. In Section 4, an application to an autism study is provided. We conclude with a brief discussion in Section 5.

2 | METHODS

2.1 | Estimation of AUC

Let X denotes the score of a biomarker of interest. Suppose that there are $n = n_1 + n_2$ subjects, with the first n_1 subjects from the diseased group and the rest from the control group. The underlying scores are $(X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n)$. The AUC is equal to $P(X_i > X_j)$, for any $i = 1, \dots, n_1$ and $j = n_1 + 1, \dots, n$, with larger AUC values indicating better diagnostic accuracy. In the presence of LOD, instead of observing X_j completely for $i = 1, \dots, n$, we observe

$$\tilde{X}_i = \begin{cases} X_i, & \text{if } X_i \geq d_x; \\ \text{NA (not available)}, & X_i < d_x. \end{cases}$$

where d_x is the value of the LOD and is known or can be easily estimated by $\min\{\tilde{X}, i = 1, \dots, n\}$.¹² Therefore, the observed data consist of $\{(\tilde{X}, \Delta_i = I(X_i \geq d_x), Z_i), i = 1, \dots, n\}$, [Correction added on 15 September 2017, after first

online publication: $\leq$ in the preceding equations have been changed to $\geq$] where Z_i takes value 1 for subjects in the diseased group and 0 otherwise.

For data subject to the LOD, they can be treated as left-censored data; therefore, by modeling $-X$ instead of X , we obtain right-censored data that can fit well in the framework of survival analysis.

We propose to use the following semiparametric transformation model,^{19–22,24}

$$H(-X_i) = -\beta Z_i + \sigma(Z_i)\epsilon_i, i = 1, \dots, n, \quad (1)$$

where H is a strictly increasing otherwise unspecified function, β is a regression coefficient, $\sigma(0) = 1$, $\sigma(1) = \sigma$, and $\epsilon_i \sim N(0, 1)$. Model (1) implies that data in each group are normally distributed after an unknown transformation. Specifically, $H(-X_i) \sim N(-\beta, \sigma^2)$ for $i = 1, \dots, n_1$, and $H(-X_i) \sim N(0, 1)$ for $i = n_1 + 1, \dots, n$. We include $\sigma(Z_i)$ in model (1) to allow for heteroscedasticity. Note that we set $\sigma(0) = 1$ and fix the mean and variance of ϵ_i to ensure model identifiability.

Under model (1), we can show that the AUC for biomarker X is

$$\begin{aligned} AUC_x &= P(X_1 > X_{n_1+1}) \\ &= P(-X_1 < -X_{n_1+1}) \\ &= P\left[H(-X_1) < H(-X_{n_1+1})\right] \\ &= \Phi\left(\frac{\beta}{\sqrt{1+\sigma^2}}\right). \end{aligned}$$

Write $\Lambda(x) = e^{H(x)}$. Let θ denotes the unknown parameters β , σ , and Λ . The likelihood function for θ based on the observed data is

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2(Z_i)}} \exp\left\{-\frac{(H(-X_i) + \beta Z_i)^2}{2\sigma^2(Z_i)}\right\} \frac{\Lambda'(-X_i)}{\Lambda(-X_i)} \right]^{\Delta_i} \times \\ &\quad \left[1 - \Phi\left\{\frac{H(-X_i) + \beta Z_i}{\sigma(Z_i)}\right\} \right]^{1-\Delta_i}, \end{aligned} \quad (2)$$

where $\Lambda'(x)$ is the first derivative of $\Lambda(x)$.

Naturally, we would like to maximize the above likelihood (2) to obtain the maximum likelihood estimators of θ . However, the maximum of the likelihood does not exist since one can fix $\Lambda(-X_i)$ and let $\Lambda'(-X_i) = \infty$ for a subject with $\Delta_i = 1$. Therefore, we allow $\Lambda(\cdot)$ to be right continuous and replace $\Lambda'(t)$ with the jump size of $\Lambda(\cdot)$ at t . The same technique was

used in the estimation of the baseline cumulative hazard function under the Cox model. For detailed discussion, readers are referred to Zeng and Lin,²² Lin,²⁵ and Kalbfleisch and Prentice.²⁶ The resultant nonparametric likelihood, still denoted as $L_n(\boldsymbol{\theta})$ for simplicity, takes the form

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2(Z_i)}} \exp \left\{ -\frac{(H(-X_i) + \beta Z_i)^2}{2\sigma^2(Z_i)} \right\} \frac{\Lambda\{-X_i\}}{\Lambda(-X_i)} \right]^{\Delta_i} \times \left[1 - \Phi \left\{ \frac{H(-X_i) + \beta Z_i}{\sigma(Z_i)} \right\} \right]^{1 - \Delta_i}. \quad (3)$$

We then maximize the nonparametric likelihood with respect to the parameters β , σ , and jump sizes $\Lambda\{-X_{(1)}\}$, $\Lambda\{-X_{(2)}\}$, $\dots$, $\Lambda\{-X_{(m)}\}$, where m is the total number of observed distinct values. The resulting nonparametric maximum likelihood estimators (NPMLEs) are denoted by $\hat{\boldsymbol{\theta}}_n \equiv (\hat{\beta}_n, \hat{\sigma}_n, \hat{\Lambda}_n)$.²⁷ The maximization procedure can be performed using quasi-Newton algorithm.²⁸ In our experience, the optimization algorithm works very well even when sample size is in the thousands. We have checked the gradients at the NPMLEs, and they are all very close to 0 (eg, $<10^{-6}$). If sample size is extremely large, we can also consider the iterative procedure described in Yin and Zeng.²⁹

Following the arguments in Zeng and Lin,²² under some regularity conditions, we can show that $\hat{\boldsymbol{\theta}}_n$ are consistent and asymptotically normal. Furthermore, the asymptotic covariances of $(\hat{\beta}_n, \hat{\sigma}_n)$ attain the semiparametric efficiency bound as defined in Bickel et al.²⁷ To estimate the variance-covariance matrix of $\hat{\boldsymbol{\theta}}_n$, we view (3) as a parametric log-likelihood with β , σ^2 , and the jump sizes of Λ at the observed data points as unknown parameters. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}_n$ can then be consistently estimated by inverting the observed information matrix according to the parametric likelihood theory.³⁰

Recall that AUC_x involves parameters β and σ . Therefore, we can replace the unknown parameters (β, σ) with their NPMLEs $(\hat{\beta}_n, \hat{\sigma}_n)$ to obtain the NPMLE of AUC_x , that is,

$$\hat{AUC}_{n,x} = \Phi \left(\frac{\hat{\beta}_n}{\sqrt{1 + \hat{\sigma}_n^2}} \right).$$

The following theorem establishes the consistency, asymptotic normality, and asymptotic efficiency of $\hat{AUC}_{n,x}$.

We first impose the following regularity conditions:

- (C1). The true values β_0 and σ_0 for parameters β and σ belong to the interior of a known compact set $\mathcal{B}_0$ in $\mathbb{R} \times \mathbb{R}^+$.

(C2). The true function Λ_0 belongs to the following class

$$\mathcal{A}_0 = \left\{ \Lambda : \Lambda \text{ is a strictly increasing function in } (-\infty, -d_x] \text{ and is continuously differentiable with } \Lambda(-\infty) = 0 \text{ and } \Lambda(-d_x) < \infty \right\}.$$

Theorem 1. Under the above regularity conditions (C1) and (C2),

- a. The NPMLEs $(\hat{\beta}_n, \hat{\sigma}_n, \hat{\Lambda}_n)$ exist. Furthermore, $\hat{\beta}_n \rightarrow \beta_0$, $\hat{\sigma}_n \rightarrow \sigma_0$, and $\sup_{(-\infty, -d_x)} |\hat{\Lambda}_n - \Lambda_0| \rightarrow 0$ almost surely.
- b. $n^{1/2}(\hat{\beta}_n - \beta_0, \hat{\sigma}_n - \sigma_0, \hat{\Lambda}_n - \Lambda_0)$ converges weakly to a zero-mean Gaussian process in the metric space $l^\infty(\mathcal{A})$, where

$$\mathcal{A} = \left\{ (h_1, h_2) : h_1 \in \mathbf{R}^2, h_2 \text{ is a function on } (-\infty, -d_x); \|h_1\| \leq 1, \|h_2\|_V \leq 1, \text{ where } \|\cdot\| \text{ is the Euclidean norm, } \|h_2\|_V \text{ is the total variation of } h_2 \right\}.$$

- c. With probability one

$$\hat{AUC}_{n,x} \rightarrow AUC_{0,x},$$

where $AUC_{0,x}$ is the true value for AUC_x . Furthermore, $n^{1/2}(\hat{AUC}_{n,x} - AUC_{0,x})$ converges weakly to a zero-mean normal distribution.

Proof. The consistency and asymptotic normality of the NPMLEs $(\hat{\beta}_n, \hat{\sigma}_n, \hat{\Lambda}_n)$ have been proved by Zeng and Lin²² and Zeng et al²⁴ using arguments from empirical process theory and semiparametric efficiency theory. Based on this result and continuous mapping theorem, it is straightforward that $\hat{AUC}_{n,x}$ is a strongly consistent estimator for AUC_x .

THEOREM 1(b) shows that the random element $n^{1/2}\{\hat{\Lambda}_n(\cdot) - \Lambda_0(\cdot), \hat{\beta}_n - \beta_0, \hat{\sigma}_n - \sigma_0\}$ converges weakly to a zero-mean Gaussian process and its asymptotic covariance matrix can be estimated by the inverse of the observed information matrix.

To prove the second part of THEOREM 1(c), we use the functional delta method. We first show $AUC = R(\beta, \sigma^2)$ is Hadamard differentiable,

$$\left(\frac{\partial AUC}{\partial \beta}, \frac{\partial AUC}{\partial \sigma^2} \right) = \left(\frac{\phi\left(\frac{\beta}{\sqrt{1+\sigma^2}}\right)}{\sqrt{1+\sigma^2}}, -\frac{\beta\phi\left(\frac{\beta}{\sqrt{1+\sigma^2}}\right)}{2\sqrt{(1+\sigma^2)^3}} \right) \equiv (\eta_1, \eta_2),$$

where $\phi(\cdot)$ is the density function of standard normal distribution.

It follows that $n^{1/2}(\widehat{AUC}_{n,x} - AUC_{0,x})$ is asymptotically equivalent to

$$n^{1/2}(\hat{\beta}_n - \beta)\eta_1 + n^{1/2}(\hat{\sigma}_n^2 - \sigma^2)\eta_2.$$

From the proof of asymptotic properties in Zeng and Lin,³¹ $n^{1/2}(\hat{\beta}_n - \beta)\eta_1 + n^{1/2}(\hat{\sigma}_n^2 - \sigma^2)\eta_2$ is asymptotically equivalent to

$$n^{1/2}(\mathcal{P}_n - \mathcal{P})\left(\mathcal{S}_{\beta, \sigma^2, \Lambda}\right)\mathcal{I}_{\beta, \sigma^2, \Lambda}^{-1}(\eta_1, \eta_2, 0),$$

where $(\mathcal{S}_{\beta, \sigma^2, \Lambda})$ is the score operator for $(\beta, \sigma^2, \Lambda)$, $\mathcal{I}_{\beta, \sigma^2, \Lambda}$ denotes information operator for those parameters, and $\mathcal{P}_n$ and $\mathcal{P}$ denote the empirical measure and the expectation under the true model, respectively. The score and information operators are calculated by differentiating over the submodels $\epsilon \mapsto (\beta, \sigma^2, \Lambda) + \epsilon\{h_1, h_2, \int_{-\infty}^{\cdot} h_3(s)d\Lambda(s)\}$, where $(h_1, h_2) \in \mathbb{R}^2$ and $h_3 \in D(-\infty, -d_x]$ of bounded variation. Therefore, $n^{1/2}(\widehat{AUC}_{n,x} - AUC_{0,x})$ converges weakly to a zero-mean Gaussian **distribution with variance** $E(\xi^2)$, **where**

$$\xi = (\eta_1, \eta_2, 0)\mathcal{I}_{\beta, \sigma^2, \Lambda}^{-1}(\mathcal{S}_{\beta, \sigma^2, \Lambda}).$$

Under the regularity conditions (C1) and (C2), the score operator is continuously invertible and onto; therefore, the variance $E(\xi^2)$ exists and can be consistently estimated by

$$\frac{1}{n} \sum_{i=1}^n \hat{\xi}_i \hat{\xi}_i^T, \text{ where } \hat{\xi}_i \text{ is the sample version of } \xi \text{ for the } i\text{th subject evaluated at } (\hat{\beta}_n, \hat{\sigma}_n^2, \hat{\Lambda}_n).$$

We obtain $\hat{\xi}_i$ as follows. By treating β, σ^2 , and the jump sizes of $\Lambda(\cdot)$ as parameters, we calculate the score vector of those parameters for the i th subject, denoted by $\widehat{U}_i$, and the observed information matrix for those parameters, denoted by $\widehat{\mathcal{I}}_{m+2}$. Then we have $\hat{\xi}_i$ as

$$\{\hat{\eta}_1, \hat{\eta}_2, 0_{1 \times m}\} \widehat{\mathcal{I}}_{m+2}^{-1} \widehat{U}_i$$

where

$$\hat{\eta}_1 = \frac{\partial AUC}{\partial \beta} \Big|_{\hat{\beta}_n, \hat{\sigma}_n^2} = \frac{\phi(\frac{\hat{\beta}_n}{\sqrt{1 + \hat{\sigma}_n^2}})}{\sqrt{1 + \hat{\sigma}_n^2}},$$

$$\hat{\eta}_2 = \frac{\partial AUC}{\partial \sigma^2} \Big|_{\hat{\beta}_n, \hat{\sigma}_n^2} = - \frac{\hat{\beta}_n \phi\left(\frac{\hat{\beta}_n}{\sqrt{1 + \hat{\sigma}_n^2}}\right)}{2\sqrt{(1 + \hat{\sigma}_n^2)^3}}.$$

The justification of calculating the above score vector and information matrix by treating the jump sizes of $\Lambda(\cdot)$ as parameters was provided by Parner et al.³⁰ $\square$

2.2 | Comparison of 2 AUCs

Let X and Y denote the scores of 2 biomarkers. The underlying scores are $(X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n)$ and $(Y_1, \dots, Y_{n_1}, Y_{n_1+1}, \dots, Y_n)$. The AUCs of the biomarkers are $AUC_x = P(X_i > X_j)$, $AUC_y = P(Y_i > Y_j)$, for any $i = 1, \dots, n_1$ and $j = n_1 + 1, \dots, n$. In the presence of LOD, instead of observing X_i, Y_i completely for $i = 1, \dots, n$, we observe

$$\tilde{X}_i = \begin{cases} X_i & \text{if } X_i \geq d_x, \\ \text{NA(available),} & X_i < d_x, \end{cases}$$

$$\tilde{Y}_i = \begin{cases} Y_i & \text{if } Y_i \geq d_y, \\ \text{NA (not available),} & Y_i < d_y, \end{cases}$$

where d_x and d_y are the corresponding LOD values for biomarkers X and Y . Therefore, the observed data consist of $\{(\tilde{X}_i, \Delta_{i,x} = I(X_i \geq d_x), \tilde{Y}_i, \Delta_{i,y} = I(Y_i \geq d_y), Z_i), i = 1, \dots, n\}$.

[Correction added on 15 September 2017, after first online publication: $\leq$ in the preceding equations have been changed to $\geq$]

We assume the semiparametric transformation model for the distributions of X_i and Y_i

$$H_x(-X_i) = -\beta_x Z_i + \sigma_x(Z_i) \epsilon_{i,x}, i = 1, \dots, n,$$

$$H_y(-Y_i) = -\beta_y Z_i + \sigma_y(Z_i) \epsilon_{i,y}, i = 1, \dots, n,$$

where H_x and H_y are strictly increasing otherwise unspecified functions, β_x, β_y are regression coefficients, $\sigma_x(0) = 1, \sigma_x(1) = \sigma_x, \sigma_y(0) = 1, \sigma_y(1) = \sigma_y$, and both $\epsilon_{i,x}$ and $\epsilon_{i,y}$ are $N(0, 1)$. Note that $\epsilon_{i,x}$ and $\epsilon_{i,y}$ may be correlated and we do not have to assume that $(\epsilon_{i,x}, \epsilon_{i,y})$ is bivariate normal.

For testing $H_0 : AUC_x = AUC_y$ versus $H_1 : AUC_x \neq AUC_y$, we propose the following Wald type test statistic

$$W_n = \frac{(\hat{AUC}_{n,x} - \hat{AUC}_{n,y})^2}{\hat{Var}(\hat{AUC}_{n,x} - \hat{AUC}_{n,y})},$$

where $\hat{AUC}_{n,x}$ and $\hat{AUC}_{n,y}$ are the NPMLEs of $AUC_{0,x}$ and $AUC_{0,y}$, respectively. The denominator term is provided in the following proof. We have the following theoretical results on the proposed test statistic W_n .

Theorem 2. *Under the regularity conditions (C1) and (C2) for both biomarkers and the null hypothesis that $AUC_x = AUC_y$, the Wald test statistic W_n converges weakly to a chi-square distribution with 1 degree of freedom.*

Proof. We have proved that

$$n^{1/2}(\hat{AUC}_{n,x} - AUC_{0,x}) = n^{1/2}\mathcal{P}_n\xi_x + o_p(1),$$

and

$$n^{1/2}(\hat{AUC}_{n,y} - AUC_{0,y}) = n^{1/2}\mathcal{P}_n\xi_y + o_p(1),$$

where ξ_x and ξ_y are defined same as ξ but for biomarker X and biomarker Y, respectively.

Therefore, under the null hypothesis,

$$n^{1/2}(\hat{AUC}_{n,x} - \hat{AUC}_{n,y}) = n^{1/2}\mathcal{P}_n(\xi_x - \xi_y) + o_p(1).$$

It follows that, under the null hypothesis, $n^{1/2}(\hat{AUC}_{n,x} - \hat{AUC}_{n,y})$ converges in distribution to a zero-mean normal distribution and $Var\{n^{1/2}(\hat{AUC}_{n,x} - \hat{AUC}_{n,y})\}$ can be consistently estimated by $\frac{1}{n} \sum_{i=1}^n (\hat{\xi}_{i,x} - \hat{\xi}_{i,y})^2$, where $\hat{\xi}_{i,x}$ and $\hat{\xi}_{i,y}$ are the sample version of ξ_x and ξ_y , respectively, for the i th subject evaluated at $(\hat{\beta}_n, \hat{\sigma}_n^2, \hat{\Lambda}_n)$. Therefore, THEOREM 2 is proved. $\square$

3 | SIMULATION STUDIES

We conduct extensive simulation studies to examine the finite-sample performance of the proposed methods based on the semiparametric transformation model and compare it with that of the parametric methods assuming normality as described in Vexler et al.¹⁴ In the first set of simulations, we evaluate the performance of the proposed NPMLE of AUC under various scenarios. We generate data from the following model

$$X_i = \beta_x Z_i + \sigma_x(Z_i)\epsilon_i, i = 1, \dots, n,$$

where Z_i takes value 1 for subjects in the diseased group and 0 otherwise and $\sigma_x(0) = 1$, $\sigma_x(1) = \sigma_x$. Under the above model, $H(x) = x$.

We use similar simulation settings as stated in Vexler et al,¹⁴ specifically, by setting $\beta_x = 2$ and $\sigma_x = 1.2$. Therefore, the true value of AUC is 0.9. The values of the LOD are set at -0.87 , -0.32 , and 0.91 corresponding to approximate 10%, 20%, and 50% censoring rates, respectively. For each scenario, we consider the sample sizes of 100 and 300 with equal allocations and generate 10 000 samples. The variance-covariance matrix of the NPMLEs is estimated by the inverse of the observed information matrix.

Table 1 presents the summary statistics of the NPMLEs of the unknown parameters β_x , σ_x , and AUC with different LOD values. We observe that the biases of the NPMLEs are small under all simulation scenarios. The standard error estimates reflect the true variation. The coverage probability of the 95% confidence interval estimates is close to the nominal level. Additionally, as sample size increases from 100 to 300, the coverage probability improves as expected. Also, the standard deviation decreases approximately by a factor of $\sqrt{3}$ that suggests root n convergence rate of the estimators. We also consider additional simulation studies for a larger sample size and smaller true values of AUCs. We obtain similar results that are summarized in the Web Appendix in the ‘‘Supplementary Materials’’ section.

For comparison, we also consider the parametric approach proposed by Vexler et al.¹⁴ We follow the same simulation settings as above except that the samples generated are not normally distributed. To sample data from the nonnormal distribution, we use the transformation function used by Diao and Lin²¹ as well as the Gaussian CDF transformation described in Liu et al³² as follows:

$$H^{-1}(y) = \{25\log(y - d + 1)\}^{2/3} \text{sign}\{\log(y - d + 1)\}, \quad (4)$$

$$H^{-1}(y) = \frac{g_0(y) - \int g_0(t) \phi\left(\frac{t-\mu}{\sigma}\right) dt}{\sqrt{\int (g_0(y) - \phi\left(\frac{t-\mu}{\sigma}\right) dt)^2 \phi\left(\frac{y-\mu}{\sigma}\right) dy}}, \quad (5)$$

Where $g_0(t) = \Phi\left(\frac{t-0.05}{0.4}\right)$. As stated in §1, Vexler et al¹² assumed the measurements of the biomarkers follow a normal or gamma distribution after a known transformation; therefore, we examine the robustness of our methods by applying the parametric approach under mis-specified transformations including the identity, square root, and logarithm transformations and compare the mean squared errors.

From Table 2, we can see that under almost all scenarios, the proposed estimator is substantially more efficient than the estimators based on the parametric approach with respect to mis-specified transformation functions. The proposed estimator appears to be reasonably robust to a variety of model mis-specifications. Comparing to the estimator under

correct model specification, the proposed estimator has limited loss of efficiency. Note that the NPMLEs of (β, σ) are invariant to any increasing transformation H .

We also compare the proposed estimator with the nonparametric estimator of AUC ignoring LOD. As is shown in the online supplemental materials, the nonparametric estimator of AUC tends to be biased especially when the censoring rate is high. Additionally, the proposed estimator is substantially more efficient than the nonparametric estimator.

The next set of simulation studies is concerned with examining the type I error rates and powers of our proposed test for comparing the performance of 2 biomarkers. We adopt a similar simulation setting used in Vexler et al.¹⁴ We generate values of $X_i, i = 1, \dots, n_1$ for the diseased group from the normal distribution with mean $\mu_{1,x}$ and variance 1. We then generate $Y_i = aX_i + e_i, i = 1, \dots, n_1$. For the control group, we generate $\{X_i, i = n_1 + 1, \dots, n\}$ from a normal distribution with mean 1 and variance 0.25 and then generate $Y_i = bX_i + e_i, i = n_1 + 1, \dots, n$. The random variables $e_i, i = 1, \dots, n$ are i.i.d. standard normal. We set $\mu_{1,x} = 1.3$ yielding $AUC_x = 0.6$. Additionally, by setting $a = 0.5$ with each value of $b = 0.246, -0.1, -0.6, -0.8, \text{ and } -1.5$, we have $AUC_y = 0.6, 0.7, 0.8, 0.85, \text{ and } 0.9$, respectively. The limit of detection is chosen as $d = 0, -1, -3$ for both biomarkers. The correlations between the 2 biomarkers for the diseased and the control group can then be calculated as $\rho_D = a(1 + a^2)^{1/2}$ and $\rho_{\bar{D}} = 0.5b/(1 + 0.25b^2)^{1/2}$, respectively. For each setting, we compare the tests based on the true model, the proposed model, and the parametric model with mis-specified transformations.

Figure 1 presents type I error rates and the powers based on 10,000 replicates. The proposed test provides accurate control of type I error rates and has little loss of power compared to the parametric method with the true transformation. For all the scenarios under both transformation functions (a) and (b), the parametric tests with the 3 mis-specified transformation functions all end up with inflated type I errors compared to the semiparametric model and the true model. Similar observations are found in literature (eg, Diao and Lin¹⁹). In some cases, the parametric methods with mis-specified transformation can lead to type I error rates in the ranges of 40% to 95%. Table 3 presents results of the Type I error rates for all scenarios under both transformation functions, and similar results were observed.

4 | APPLICATION TO AN AUTISM STUDY

We apply the proposed methods to data from a study of biomarkers of autism spectrum disorder and autism from the National Institute of Child Health and Human Development, National Institutes of Health. In this case-control study, measurements of biomarkers dehydroepiandrosterone (DHEA) levels (in ng/dL) and DHEA sulphate (DHEAS) levels (in $\mu\text{g/dL}$) are recorded from $n = 71$ boys with autism spectrum disorder (ASD) or autism and $n = 59$ aged-matched controls. The sensitivities of the assays or detection limits are 10 ng/dL for DHEA and 10 $\mu\text{g/dL}$ for DHEAS. Consequently, DHEA levels are detectable in 70 of 71 subjects with ASD/autism and 55 of 59 controls while DHEAS levels are detectable in only 29 of 71 subjects with ASD/autism and 14 of 59 controls.³³ Among the individuals with observed data in the case group, DHEA level ranges from 10 to 591 with mean 67.7 and

standard deviation 82.2 and DHEAS level ranges from 11 to 117 with mean 28.5 and standard deviation 23.3. And in the control group, DHEA level ranges from 11 to 238 with mean 65 and standard deviation 54.9 and DHEAS level ranges from 14 to 56 with mean 29.8 and standard deviation 15.0. Among all the observed individuals, Wilcoxon rank sum test reveals that there is no significant difference in the distribution of DHEA between the case and control groups (P value = .72) and also no significant difference in the distribution of DHEAS between the case and control groups (P value = .20). By fitting a parametric Tobit model, DHEA level has a normal distribution truncated at 10 with mean 58.3 for the control group and 65.6 for the case group and DHEAS level has a normal distribution truncated at 10 with mean -12.9 for the control group and -0.1 for the case group.

To evaluate the performance of both biomarkers accounting for the missing data, we use the proposed test compared with likelihood ratio test proposed by Vexler et al.¹⁴ Since exploratory analyses (see the histograms and normal Q-Q plots displayed in Figures 2 and 3, respectively) suggest that the distributions for DHEA and DHEAS are not normal, comparisons are made under different transformed values. In particular, we fit the models under the following transformation assumptions: no transformation, log transformation, and square root transformation.

Under different transformations, we obtain the estimates of AUCs. From Table 4, we can see that the proposed method is robust to nonnormality, whereas the estimates of AUCs from parametric approach are sensitive to the transformation assumptions. In particular, the parametric methods appear to yield more significant results. However, as is evident in the simulation studies, the parametric tests may lead to inflated type I error rates, and therefore, one should interpret the results from the parametric tests with caution in the presence of nonnormality.

5 | DISCUSSION

We have proposed new semiparametric methods for estimating the AUC of a biomarker subject to LOD and comparing the AUCs of 2 biomarkers. We establish the large sample properties of the proposed estimators and test statistic. Compared to its parametric counterpart, the proposed semiparametric method gains much robustness with very little loss of efficiency or power when the parametric assumption is true. We have implemented the proposed methods in a computationally efficient program in C language. An R interface to the program is freely available at <http://mason.gmu.edu/~gdiao/software/sAUC>.

For the comparison of the AUCs of 2 biomarkers, we proposed to use 2 separate semiparametric transformation models for these 2 biomarkers that do not require bivariate normality of the residual errors. We then propose a Wald test statistic that appropriately accounts for the potential correlation between the 2 biomarkers. And it is shown that our approach has comparable efficiency to the parametric model under the correctly specified transformation. It is possible to improve our approach by considering a bivariate semiparametric transformation model; however, the gain of efficiency would be limited. Additionally, it is computationally more expensive to calculate the likelihood under the joint

model assuming bivariate normal after unknown transformation since it involves analytically intractable 2-dimensional integrals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Research of A. Liu was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

Funding information

Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health; Human Development

REFERENCES

1. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;12(14):1293–1316. [PubMed: 8210827]
2. Zhou XH, McClish DK, Obuchowski NA. *Statistical Methods in Diagnostic Medicine*, Vol. 569: New York: John Wiley & Sons; 2009.
3. Bamber D The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol*. 1975;12 (4):387–415.
4. Goddard MJ, Hinberg I. Receiver operator characteristic (ROC) curves and non-normal data: an empirical study. *Stat Med*. 1990;9(3):325–337. [PubMed: 2343225]
5. Reiser B, Faraggi D. Confidence intervals for the generalized ROC criterion. *Biometrics*. 1997;53(2):644–652. [PubMed: 9235120]
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29–36. [PubMed: 7063747]
7. Zou KH, Tompary CM, Fielding JR, Silverman SG. Original smooth receiver operating characteristic curve estimation from continuous data: statistical methods for analyzing the predictive value of spiral CT of ureteral stones. *Acad Radiol*. 1998;5(10):680–687. [PubMed: 9787838]
8. Lloyd CJ. Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *J Am Stat Assoc*. 1998;93(444):1356–1364.
9. Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Stat Med*. 2002;21(20):3093–3106. [PubMed: 12369084]
10. Lyles RH, Williams JK, Chuachoowong R. Correlating two viral load assays with known detection limits. *Biometrics*. 2001;57(4):1238–1244. [PubMed: 11764265]
11. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect*. 2004;112(17):1691–1696.
12. Vexler A, Liu A, Schisterman EF. Efficient design and analysis of biospecimens with measurements subject to detection limit. *Biom J*. 2006;48(5):780–791. [PubMed: 17094343]
13. Schisterman EF, Vexler A, Whitcomb BW, Liu A. The limitations due to exposure detection limits for regression models. *Am J Epidemiol*. 2006;163(4):374–383. [PubMed: 16394206]
14. Vexler A, Liu A, Eliseeva E, Schisterman EF. Maximum likelihood ratio tests for comparing the discriminatory ability of biomarkers subject to limit of detection. *Biometrics*. 2008;64(3):895–903. [PubMed: 18047527]
15. Mumford SL, Schisterman EF, Vexler A, Liu A. Pooling biospecimens and limits of detection: effects on ROC curve analysis. *Biostatistics*. 2006;7(4):585–598. [PubMed: 16531470]

16. Perkins NJ, Schisterman EF, Vexler A. Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error. *Stat Med*. 2009;28(13):1841–1860. [PubMed: 19340817]
17. Molodianovitch K, Faraggi D, Reiser B. Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biom J*. 2006;48(5):745–757. [PubMed: 17094340]
18. Tian L, Vexler A, Yan L, Schisterman F. Confidence interval estimation of the difference between paired AUCs based on combined biomarkers. *J Stat Plan Inference*. 2009;139(10):3725–3732. [PubMed: 19946609]
19. Diao G, Lin DY. A powerful and robust method for mapping quantitative trait loci in general pedigrees. *Am J Human Genetics*. 2005;77(1):97–111. [PubMed: 15918154]
20. Diao G, Lin DY. Improving the power of association tests for quantitative traits in family studies. *Genet Epidemiol*. 2006;30(4):301–313. [PubMed: 16607624]
21. Diao G, Lin DY. Semiparametric variance-component models for linkage and association analyses of censored trait data. *Genet Epidemiol*. 2006;30(7):570–581. [PubMed: 16858699]
22. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *J R Stat Soc B (Statistical Methodology)*. 2007;69(4):507–564.
23. Kim E, Zeng D, Zhou XH. Semiparametric transformation models for multiple continuous biomarkers in ROC analysis. *Biom J*. 2015;57(5):808–833. [PubMed: 26138227]
24. Zeng D, Lin D, Lin X. Semiparametric transformation models with random effects for clustered failure time data. *Stat Sin*. 2008;18(1):355–377. [PubMed: 19809573]
25. Lin D. On the breslow estimator. *Lifetime Data Anal*. 2007;13(4):471–480. [PubMed: 17768681]
26. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*, Vol. 360: Hoboken, New Jersey: John Wiley & Sons; 2011.
27. Bickel PJ, Klaassen CA, Bickel PJ, Ritov Y, et al. *Efficient and Adaptive Estimation for Semiparametric Models*: Baltimore: Johns Hopkins University Press Baltimore; 1993.
28. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition Cambridge: Cambridge University Press; 1992.
29. Yin G, Zeng D. Efficient algorithm for computing maximum likelihood estimates in linear transformation models. *J Comput Graph Stat*. 2006;15(1):228–245.
30. Parner E et al. Asymptotic theory for the correlated gamma-frailty model. *Annals Stat*. 1998;26(1):183–214.
31. Zeng D, Lin D. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006;93(3):627–640.
32. Liu H, Han F, Yuan M, Lafferty J, Wasserman L, et al. High-dimensional semiparametric Gaussian copula graphical models. *Annals Stat*. 2012;40(4):2293–2326.
33. Mills JL, Hediger ML, Molloy CA, et al. Elevated levels of growth-related hormones in autism and autism spectrum disorder. *Clin Endocrinol*. 2007;67(2):230–237.

FIGURE 1.
Power curves based on 10,000 replicates

FIGURE 2.
Histograms of dehydroepiandrosterone (DHEA) and DHEA sulphate (DHEAS) levels in case and control groups

FIGURE 3.

Q-Q plots of dehydroepiandrosterone (DHEA) and DHEA sulphate (DHEAS) levels in case and control groups

TABLE 1

Summary statistics of the nonparametric maximum likelihood estimators

$n_1 = n_2$	CR, %	Parameters	Bias	SE	SEE	CP, %
50	10	β_x	-0.039	0.178	0.161	94.29
		σ_x	-0.057	0.125	0.121	87.38
		<i>AUC</i>	0.000	0.029	0.030	91.92
	20	β_x	-0.038	0.180	0.162	94.41
		σ_x	-0.056	0.131	0.124	87.66
		<i>AUC</i>	0.000	0.030	0.030	92.04
	50	β_x	-0.055	0.186	0.148	94.28
		σ_x	-0.107	0.188	0.160	84.76
		<i>AUC</i>	0.007	0.035	0.034	88.95
150	10	β_x	-0.012	0.103	0.101	94.54
		σ_x	-0.018	0.075	0.074	92.16
		<i>AUC</i>	0.000	0.017	0.018	93.96
	20	β_x	-0.012	0.104	0.102	94.52
		σ_x	-0.020	0.079	0.077	92.24
		<i>AUC</i>	0.000	0.017	0.018	94.15
	50	β_x	-0.016	0.109	0.104	94.85
		σ_x	-0.033	0.117	0.112	91.29
		<i>AUC</i>	0.002	0.021	0.021	92.73

Abbreviations: Bias, sampling average of the biases of the estimates; CR, censoring rate of sample (proportion of objects that are not observed numerically); SE, average of the standard estimates; SEE, empirical standard deviations of the estimates; *CP*, empirical coverage probability of 95 % confidence interval estimates. Results are based on 10 000 replicates.

TABLE 2

Comparison of the AUC estimators under different transformations

$n_1 = n_2$	CR, %	New	Identity		Log		Square root		True	
		MSE($\times 10^{-3}$)	Bias	R.E.	Bias	R.E.	Bias	R.E.	Bias	R.E.
True transformation: $H^{-1}(y) = \{25\log(y - d + 1)\}^{2/3} \text{sign}\{\log(y - d + 1)\}(\text{a})$										
50	10	0.89	-0.024	1.798	-0.063	5.476	-0.126	18.537	0.001	0.968
50	20	0.90	-0.018	1.379	-0.049	3.710	-0.091	10.187	0.001	0.977
50	50	1.22	0.015	1.034	0.023	1.163	0.021	1.044	0.000	1.065
150	10	0.31	-0.025	3.171	-0.066	16.664	-0.149	77.846	0.000	0.972
150	20	0.31	-0.018	2.085	-0.050	8.973	-0.090	27.219	0.000	0.975
150	50	0.45	0.017	1.389	0.024	1.924	0.021	1.629	-0.001	0.976
True transformation: $H^{-1}(y) = \frac{g_0(y) - \int g_0(t) \phi(\frac{1-\mu}{\sigma}) dt}{\sqrt{\int (g_0(y) - \phi(\frac{t-\mu}{\sigma}) dt)^2 \phi(\frac{y-\mu}{\sigma}) dy}}(\text{b})$										
50	10	0.89	-0.140	24.306	-0.155	27.831	-0.165	32.090	0.001	0.968
50	20	0.90	-0.092	10.917	-0.096	11.979	-0.104	14.446	0.001	0.977
50	50	1.22	0.016	1.542	0.016	0.952	0.016	0.948	0.000	1.065
150	10	0.31	-0.143	79.257	-0.154	78.365	-0.165	93.332	0.000	0.972
150	20	0.31	-0.091	29.235	-0.097	35.325	-0.105	42.949	0.000	0.975
150	50	0.45	0.017	2.092	0.017	1.314	0.017	1.613	-0.001	0.976

Abbreviations: MSE, mean-squared error; R.E., relative efficiency calculated as the ratio of MSE of other methods to the MSE of the new one. Results are based on 10 000 replicates.

TABLE 3

Type I error rates under transformation functions

	$n_1 = n_2$	LOD (d)	New	Identity	Log	Square root	True
Transformation (a)	50	-3	0.061	0.068	0.091	0.114	0.055
		-1	0.060	0.121	0.172	0.201	0.055
		0	0.059	0.206	0.391	0.503	0.053
	150	-3	0.052	0.088	0.164	0.229	0.051
		-1	0.052	0.234	0.419	0.502	0.052
		0	0.053	0.459	0.805	0.903	0.051
	50	-3	0.051	0.627	0.651	0.649	0.055
		-1	0.059	0.628	0.647	0.645	0.055
		0	0.058	0.628	0.646	0.608	0.053
Transformation (b)	150	-3	0.052	0.977	0.985	0.987	0.051
		-1	0.052	0.977	0.984	0.985	0.052
		0	0.053	0.969	0.976	0.957	0.051

Abbreviation: LOD, limit of detection.

TABLE 4

Data analysis of autism data

	New	Identity	Log	Square root
Estimation				
AUC_x	0.521	0.528	0.525	0.500
AUC_y	0.682	0.680	0.666	0.674
Hypothesis Testing				
Test statistic	6.98	7.14	9.15	8.60
<i>P</i> value	.008	.008	.002	.003