Statistica Sinica **18**(2008), 817-836

REGRESSION ANALYSIS FOR THE PARTIAL AREA UNDER THE ROC CURVE

Tianxi Cai and Lori E. Dodd

Harvard University and National Cancer Institute

Abstract: Performance evaluation of any classification method is fundamental to its acceptance in practice. Evaluation should consider the dependence of a classifier's accuracy on relevant covariates in addition to its overall accuracy. When developing a classifier with a continuous output that allocates units into one of two groups, receiver operating characteristic (ROC) curve analysis is appropriate. The partial area under the ROC curve (pAUC) is a summary measure of the ROC curve used to make statistical inference when only a region of the ROC space is of interest. We propose a new pAUC regression method to evaluate covariate effects on the diagnostic accuracy. We provide asymptotic distribution theory and inference procedures that allow for correlated observations. Graphical methods and goodness-of-fit statistics for model checking are also developed. Simulation studies demonstrate that the large-sample theory provides reasonable inference in small samples and the new estimator is considerably more efficient than the estimator proposed by Dodd and Pepe (2003a). Application to an analysis of prostate-specific antigen (PSA), a biomarker for early detection of prostate cancer, demonstrates the utility of the method in practice.

Key words and phrases: Diagnostic accuracy, generalized linear model, model checking.

1. Introduction

Binary classification is a relevant undertaking in a wide variety of statistical fields. Algorithms such as support vector machines and neural networks have been applied, for example, to detect automobile insurance claim fraud (Viaene et al. (2002)) or to predict peptide binding (Brusic et al. (1998)). In the medical field, a multitude of medical tests, such as biomarkers and imaging modalities have been developed to screen and diagnose disease, as well to predict outcome and monitor response to therapy. Rigorous evaluation of any classification method is a prerequisite to its wide-spread use. A method must be shown to be accurate and factors influencing the accuracy of a method must be adequately understood.

Accuracy may be summarized by the percent of correct classifications. However, a more refined analysis of accuracy considers the false positive error and the

false negative error separately, as each has a unique associated cost. For a continuous outcome variable, Y, let $Y \geq c$ denote a positive classification. Throughout, the two states are referred to as "diseased" and "disease-free", however more general terminology could be used. Additionally, the term "test" refers generally to the continuous output of a classifier, such as a biomarker or a neural network result. The true positive rate (TPR) is defined as $S_{\mathrm{D}}(c) \equiv P(Y \geq c \mid \text{diseased})$, while the false positive rate (FPR) is defined as $S_{\bar{\mathrm{D}}}(c) \equiv P(Y \geq c \mid \text{disease-free})$. The receiver operating characteristic (ROC) curve plots $\{(S_{\bar{\mathrm{D}}}(c),\ S_{\mathrm{D}}(c)),\ c \in (-\infty, \infty)\}$ or, equivalently, $\{(u, \mathrm{ROC}(u)), u \in (0,1)\}$. The curve describes the inherent capacity of the test in discriminating the two states, without linking the test to any specific positivity criterion.

A single summary index is useful as a descriptive of overall test performance and for hypothesis testing. The most common summary index of the ROC curve is the area under the curve (AUC) (Bamber (1975) and DeLong, DeLong and Clarke-Pearson (1988)). The AUC can be interpreted as the probability that a randomly selected case with disease will be regarded with greater suspicion than a randomly selected disease-free case. Often, interest does not lie in the entire range of FPRs and, consequently, only part of the area under the ROC curve is relevant. For example, very low false positive rates such as FPR ≤ 0.05 have been advocated in settings such as cancer screening (Baker and Pinsky (2001)) and then analysis should be restricted to that portion of the curve. Alternatively, a restricted region of TPRs may be of interest (Jiang, Metz and Nishikawa (1996)). Noting that a definition with respect to TPRs is straightforward, we consider the partial AUC (pAUC) for a range of FPRs, without loss of generality, say FPR $\in (0, u]$ for some $u \leq 1$. The pAUC is given as $\mathrm{pAUC}(u) = \int_0^u \mathrm{ROC}(u)du$ (McClish (1989) and Thompson and Zucchini (1989)), which has a value of u when a test is perfect, and of $u^2/2$ when a test is uninformative. Another reason to analyze the pAUC, rather than the entire AUC, is that a summary of the entire ROC curve fails to consider the plot as a composite of different segments with different diagnostic implications (Dwyer (1996)). This is particularly important if prominent differences between ROC plots in specific regions are muted or reversed when the total area is considered.

Methods for estimating and comparing pAUCs are available (McClish (1989), Wieand et al. (1989), Zhang et al. (2002), Pepe (2003) and Dodd and Pepe (2003a)). Generalizations of these methods to regression modeling assists with further characterization of a classifier. As an example, consider PSA, a biomarker for prostate cancer. Since a biomarker that detects cancer prior to the onset of clinical symptoms is of clinical interest, a model of PSA accuracy with a covariate representing the time prior to clinical diagnosis is of interest. This will provide information about by how much PSA advances diagnosis. In addition, if there is a relationship between PSA accuracy and age, a model that includes age as a covariate might identify ages for targeting PSA screening programs.

Two approaches to the pAUC regression analysis have been proposed (Thompson and Zucchini (1989) and Dodd and Pepe (2003a)). The method proposed by Thompson and Zucchini (1989) does not accommodate continuous covariates and is not applicable to many types of data. Dodd and Pepe (2003a) present a more flexible pAUC regression method; however, they do not provide theoretical justification for their estimator and rely on the bootstrap to estimate the variance. Furthermore, when the covariates are continuous, their methods require making unnecessary assumptions. Specifically, similar to Dodd and Pepe (2003b), they model the pAUC by comparing test results of diseased subjects, Y_{D}, with covariate $\mathbf{Z}_1$ to test results of disease-free, $Y_{\bar{\text{D}}}$, with covariate $\mathbf{Z}_0$ as

$$\text{pAUC}_{\mathbf{Z}_1,\mathbf{Z}_0}(u) = \eta\{\beta_0 + \boldsymbol{\beta}_1^{\text{T}}\mathbf{Z}_1 + \boldsymbol{\beta}_2^{\text{T}}(\mathbf{Z}_1 - \mathbf{Z}_0)\} \tag{1}$$

for a given link function $\eta : (-\infty, \infty) \to [0, u]$. This formulation requires modeling the effect of $\mathbf{Z}_1 - \mathbf{Z}_0$, the difference between the covariate levels in the two populations in addition to the quantity of interest, $\boldsymbol{\beta}_1^{\text{T}}\mathbf{Z}_1$. However, when assessing the test accuracy adjusting for covariates, the interest only lies in comparing the distribution of Y_{D} and $Y_{\bar{\text{D}}}$ among subjects when $\mathbf{Z}_0 = \mathbf{Z}_1$. Thus $\boldsymbol{\beta}_2$ is not of scientific interest and (1) imposes unnecessary modelling. Although this assumption be may relaxed by only including comparisons between Y_{D} and $Y_{\bar{\text{D}}}$ if the covariate level $\mathbf{Z}_1$ is close to the covariate level $\mathbf{Z}_0$, one may improve the robustness of the model by making such comparisons only when $\mathbf{Z}_0 = \mathbf{Z}_1$.

In this article, we propose to model the covariate specific pAUC assuming (1) only when $\mathbf{Z}_0 = \mathbf{Z}_1$. Our estimation approach is based on the concept of *placement values* (Hanely and Hajian-Tilaki (1997) and Pepe and Cai (2004)), defined as particular standardizations of the raw measurements relative to the reference populations. In Section 2, we introduce placement values and illustrate how they can be used to estimate the pAUC when there is no covariate. In Section 3, we propose a marginal regression model for the pAUC and derive inference procedures for the regression parameters allowing for clustered data. Simulation studies in Section 4 suggest that the new approach performs well. Furthermore the new estimator, while being more robust, is considerably more efficient that the Dodd and Pepe (2003a) estimator. To examine whether the specified regression model is appropriate for the data, in Section 5 we present both graphical procedures and goodness of fit testing statistics for model checking. Section 6 gives results from the application of the proposed method to a PSA dataset. Some discussion is provided in Section 7.

2. Placement Values and pAUC Estimation

As in Pepe and Cai (2004), we choose the disease-free population as the reference population and define the placement value for Y_{D} as $U_{\text{D}} \equiv S_{\bar{\text{D}}}(Y_{\text{D}})$. Then U_{D} quantifies the degree of separation between the two populations. Moreover,

$$P(U_{\text{D}} \le u) = P\{S_{\bar{\text{D}}}(Y_{\text{D}}) \le u\} = P\{Y_{\text{D}} \ge S_{\bar{\text{D}}}^{-1}(u)\} = \text{ROC}(u),$$

and

$$E(U_{\text{D}}) = \int_0^1 u \; d \; \text{ROC}(u) = 1 - \int_0^1 \text{ROC}(u) du = 1 - \text{AUC}.$$

DeLong, DeLong and Clarke-Pearson (1988) and Hanely and Hajian-Tilaki (1997) interpreted the nonparametric estimate of the AUC as one minus the sample mean of the empirically estimated placement values. Placement values have been used recently to make inference about ROC regression models (Pepe and Cai (2004) and Cai (2004)). Here, we propose to make inference about the pAUC based on *truncated* placement values.

We first illustrate our proposal by constructing a non-parametric estimator for the pAUC in the absence of covariates. Suppose we have $N_{\text{D}} = \sum_{i=1}^{n_{\text{D}}} K_i$ data records for n_{D} diseased subjects, $\{Y_{\text{D}ik},\ k = 1, \ldots, K_i,\ i = 1, \ldots, n_{\text{D}}\}$, and $N_{\bar{\text{D}}} = \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} K_j$ data records for $n_{\bar{\text{D}}}$ disease-free subjects, $\{Y_{\bar{\text{D}}jl},\ l = 1, \ldots, K_j,\ j = n_{\text{D}}+1, \ldots, n_{\text{D}}+n_{\bar{\text{D}}}\}$. We assume that K_i and K_j are relatively small with respect to n_{D} and $n_{\bar{\text{D}}}$. Observations from the same subject may be correlated but are independent between subjects, with $S_{\bar{\text{D}}}(y) \equiv P(Y_{\bar{\text{D}}jl} \geq y) = P(Y_{\bar{\text{D}}j'l'} \geq y)$ and $S_{\text{D}}(y) \equiv P(Y_{\text{D}ik} \geq y) = P(Y_{\text{D}i'k'} \geq y)$.

Let $U_{\text{D}ik}^{(u)} \equiv \min(U_{\text{D}ik}, u)$ denote the truncated placement value and $\widehat{U}_{\text{D}ik}^{(u)} \equiv \min(\widehat{U}_{\text{D}ik}, u)$ be the empirical estimator of $U_{\text{D}ik}^{(u)}$, where $U_{\text{D}ik} \equiv S_{\bar{\text{D}}}(Y_{\text{D}ik})$, $\widehat{U}_{\text{D}ik} = \widehat{S}_{\bar{\text{D}}}(Y_{\text{D}ik})$ and

$$\widehat{S}_{\bar{\text{D}}}(y) = N_{\bar{\text{D}}}^{-1} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \sum_{l=1}^{K_j} I(Y_{\bar{\text{D}}jl} \geq y).$$

Using integration by parts, we find that the marginal mean of the truncated placement values relates to the pAUC through

$$E(U_{\text{D}ik}^{(u)}) = \int_0^u \{1 - \text{ROC}(v)\} dv = u - \text{pAUC}(u).$$

This motivates us to estimate the pAUC(u) with

$$\widehat{\text{pAUC}}(u) = u - \frac{1}{N_{\text{D}}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} \widehat{U}_{\text{D}ik}^{(u)} = \frac{1}{N_{\text{D}}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} \widehat{V}_{\text{D}ik}^{(u)},$$

where $\widehat{V}_{\text{D}ik}^{(u)} = u - \widehat{U}_{\text{D}ik}^{(u)}$. When $K_i = K_j = 1$, this estimator is equivalent to the non-parametric estimate proposed by Dodd and Pepe (2003a). Since Dodd and Pepe (2003a) did not provide large sample theory for $\widehat{\text{pAUC}}(u)$, we show in appendix A the consistency of $\widehat{\text{pAUC}}(u)$, and that the distribution of $n_{\text{D}}^{1/2}\{\widehat{\text{pAUC}}(u) - \text{pAUC}(u)\}$ is approximately $N(0, \widehat{\sigma}^2)$ accounting for within cluster correlation, where $\widehat{\sigma}^2 = n_{\text{D}}^{-1} \sum_{i=1}^{n_{\text{D}}} \widehat{\mathcal{P}}_{\text{D}i}^2 + n_{\bar{\text{D}}}^{-1} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \widehat{\mathcal{P}}_{\bar{\text{D}}j}^2$, $\widehat{\mathcal{P}}_{\text{D}i} = n_{\text{D}} N_{\text{D}}^{-1} \sum_k$

$\widehat{V}^{(u)}_{\text{D}ik} - \widehat{\text{pAUC}}(u)$, and $\widehat{\mathcal{P}}_{\bar{\text{D}}j} = (n_{\text{D}}/n_{\bar{\text{D}}})^{1/2} N_{\text{D}}^{-1} \sum_l \sum_{i,k} I(\widehat{U}_{\text{D}ik} \le u)\{\widehat{U}_{\text{D}ik} - I(Y_{\bar{\text{D}}jl} \ge Y_{\text{D}ik})\}$.

3. Partial AUC Regression

Next, we use truncated placement values to develop estimating equations for pAUC regression models. Let $\mathbf{X}_{ik} = (\mathbf{Z}_{\text{D}ik}, \mathbf{Z}_{ik})$ denote the covariates associated with $Y_{\text{D}ik}$, and $\mathbf{Z}_{jl}$ be the covariates associated with $Y_{\bar{\text{D}}jl}$. Covariates denoted by $\mathbf{Z}$ are relevant to both diseased and disease-free subjects. Examples include the subject's age or the type of biomarker represented by Y (Pepe (2003, Chap. 6)). Covariates denoted by $\mathbf{Z}_{\text{D}}$ are specific to diseased subjects, but not applicable to disease-free subjects. Examples include severity of disease and timing of biomarker measurement prior to disease onset. In the presence of $\mathbf{Z}_{\text{D}}$, one would be interested in comparing the distribution of Y among those diseased subjects with covariates $\mathbf{X} = (\mathbf{Z}, \mathbf{Z}_{\text{D}})$ to the distribution of Y among those disease-free subjects with covariates $\mathbf{Z}$.

We assume a marginal model for the covariate specific pAUC:

$$\int_0^u \text{ROC}_{\mathbf{X}_{ik}}(v)dv \equiv \text{pAUC}_{\mathbf{X}_{ik}}(u) = \eta(\boldsymbol{\beta}_0^{\text{T}} \vec{\mathbf{X}}_{ik}), \tag{2}$$

where $\text{ROC}_{\mathbf{X}}(v) = P\{Y_{\text{D}} \ge S^{-1}_{\bar{\text{D}},\mathbf{Z}}(v) \mid \mathbf{X} = (\mathbf{Z}_{\text{D}}, \mathbf{Z})\}$, $S_{\bar{\text{D}},\mathbf{Z}}(y) = P(Y_{\bar{\text{D}}jl} \ge y \mid \mathbf{Z}_{jl} = \mathbf{Z})$, and $\vec{\mathbf{X}}_{ik} = (1, \mathbf{X}_{ik})$. To estimate $\boldsymbol{\beta}_0$, we define the placement value for the test result $Y_{\text{D}ik}$ with covariate $\mathbf{X}_{ik}$ as $U_{\text{D}ik} \equiv S_{\bar{\text{D}},\mathbf{Z}_{ik}}(Y_{\text{D}ik})$. It is straightforward to show that $E\{V^{(u)}_{\text{D}ik} \mid \mathbf{X}_{ik}\} = \int_0^u \{\text{ROC}_{\mathbf{X}_{ik}}(v)\}dv = \text{pAUC}_{\mathbf{X}_{ik}}(u)$, where $V^{(u)}_{\text{D}ik} = u - \min(U_{\text{D}ik}, u)$. If $S_{\bar{\text{D}},\mathbf{Z}}(\cdot)$ is known, then one can easily estimate the effect of $\mathbf{X} = (\mathbf{Z}, \mathbf{Z}_{\text{D}})$ on pAUC by solving

$$\frac{1}{N_{\text{D}}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} w(\vec{\mathbf{X}}_{ik}) \vec{\mathbf{X}}_{ik} \left\{ V^{(u)}_{\text{D}ik} - \eta(\boldsymbol{\beta}^{\text{T}} \vec{\mathbf{X}}_{ik}) \right\} = 0,$$

where $w(\cdot)$ is a given positive weight function. However, $S_{\bar{\text{D}},\mathbf{Z}}(\cdot)$ is unknown in general and thus $V^{(u)}_{\text{D}ik}$ needs to be estimated in order to make inference about $\boldsymbol{\beta}_0$. If the covariates $\mathbf{Z}$ are discrete, $S_{\bar{\text{D}},\mathbf{Z}}(y)$ can be estimated non-parametrically within covariate specific subsets. When continuous covariates are included, we recommend semi-parametric regression models for $S_{\bar{\text{D}},\mathbf{Z}}(y)$. For example, one could assume a flexible semi-parametric location-scale model (Pepe (1998) and Heagerty and Pepe (1999)). Other types of semi-parametric models, such as linear transformation models (Han (1987) and Cai, Wei and Wilcox (2000)), could also be considered. We do not assume any specific model for $S_{\bar{\text{D}},\mathbf{Z}}(y)$, but require that the resulting estimator of $S_{\bar{\text{D}},\mathbf{Z}}(y)$ be $n_{\bar{\text{D}}}^{-1/2}$-consistent. We note that

the Dodd and Pepe estimator also requires semi-parametric regression models for the conditional quantile of $Y_{\bar{\text{D}}}$ (Dodd and Pepe (2003a, p.620)). With $S_{\bar{\text{D}},\mathbf{Z}}(\cdot)$ estimated by $\widehat{S}_{\bar{\text{D}},\mathbf{Z}}(\cdot)$, we propose to estimate $\boldsymbol{\beta}_0$ by solving

$$\frac{1}{N_{\text{D}}}\sum_{i=1}^{n_{\text{D}}}\sum_{k=1}^{K_i} w(\vec{\mathbf{X}}_{ik})\vec{\mathbf{X}}_{ik}\left\{\widehat{V}^{(u)}_{\text{D}ik} - \eta(\boldsymbol{\beta}^{\mathsf{T}}\vec{\mathbf{X}}_{ik})\right\} = 0, \tag{3}$$

where $\widehat{V}^{(u)}_{\text{D}ik} = u - \min(\widehat{U}_{\text{D}ik}, u)$, $\widehat{U}_{\text{D}ik} = \widehat{S}_{\bar{\text{D}},\mathbf{Z}_{ik}}(Y_{\text{D}ik})$.

Let $\widehat{\boldsymbol{\beta}}$ denote the solution to (3). We show in appendix B that, $\widehat{\boldsymbol{\beta}}$ is unique and consistent. To obtain interval estimates of specific components of $\boldsymbol{\beta}_0$, we also show in appendix B that, accounting for the correlation within each subject, $n_{\text{D}}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically equivalent to a sum of independent terms indexed by subjects:

$$n_{\text{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx \mathbb{A}^{-1}\left\{ n_{\text{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\text{D}}} \mathfrak{B}_{\text{D}i} + n_{\bar{\text{D}}}^{-\frac{1}{2}}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathfrak{B}_{\bar{\text{D}}j}\right\},$$

where $\mathbb{A} = E\{\dot{\eta}(\boldsymbol{\beta}_0^{\mathsf{T}}\vec{\mathbf{X}}_{ik})\vec{\mathbf{X}}_{ik}^{\otimes 2}\}$, $\dot{\eta}(x) = d\eta(x)/dx$, $\mathfrak{B}_{\text{D}i} = K_{\text{D}}^{-1}\sum_{k=1}^{K_i} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}$ $\{V^{(u)}_{\text{D}ik} - \eta(\boldsymbol{\beta}_0^{\mathsf{T}}\vec{\mathbf{X}}_{ik})\}$, $\mathfrak{B}_{\bar{\text{D}}j}$ is the limit of $(p_{10}^{1/2}/N_{\text{D}})\sum_{i=1}^{n_{\text{D}}}\sum_{k=1}^{K_i} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}\int_0^u I_{\bar{\text{D}}j}$ $(v;\mathbf{Z}_{ik})d\text{ROC}_{\mathbf{X}_{ik}}(v)$, and $I_{\bar{\text{D}}j}(v,\mathbf{Z})$ is defined in appendix B. It follows from the Multivariate Central Limit Theorem that the distribution of $n_{\text{D}}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be approximated by $N(0,\boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}$ can be consistently estimated by

$$\widehat{\mathbb{A}}^{-1}\left\{ n_{\text{D}}^{-1}\sum_{i=1}^{n_{\text{D}}} \widehat{\mathfrak{B}}_{\text{D}i}\widehat{\mathfrak{B}}_{\text{D}i}^{\mathsf{T}} + n_{\bar{\text{D}}}^{-1}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \widehat{\mathfrak{B}}_{\bar{\text{D}}j}\widehat{\mathfrak{B}}_{\bar{\text{D}}j}^{\mathsf{T}}\right\}\widehat{\mathbb{A}}^{-1},$$

where $\widehat{\mathbb{A}} = \frac{1}{N_{\text{D}}}\sum_{i=1}^{n_{\text{D}}}\sum_{k=1}^{K_i}\dot{\eta}(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\vec{\mathbf{X}}_{ik})\vec{\mathbf{X}}_{ik}\vec{\mathbf{X}}_{ik}^{\mathsf{T}}$, $\widehat{\mathfrak{B}}_{\text{D}i} = K_{\text{D}}^{-1}\sum_{k=1}^{K_i} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}\{V^{(u)}_{\text{D}ik} -$ $\eta(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\vec{\mathbf{X}}_{ik})\}$, $\widehat{\mathfrak{B}}_{\bar{\text{D}}j} = (p_{10}^{1/2}/N_{\text{D}})\sum_{i=1}^{n_{\text{D}}}\sum_{k=1}^{K_i} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}I(\widehat{U}_{\text{D}ik} \le u)\widehat{I}_{\bar{\text{D}}j}(\widehat{U}_{\text{D}ik};\mathbf{Z}_{ik})$, and $\widehat{I}_{\bar{\text{D}}j}(v,\mathbf{Z})$ is obtained by replacing all the theoretical quantities in $I_{\bar{\text{D}}j}(v,\mathbf{Z})$ by their empirical counterparts.

4. Model Checking Procedures

The proposed inference procedures require the specification of a link function $\eta(\cdot)$. Here, we present a graphical method, as well as statistical tests, to assess whether model (2) with a given link function $\eta(\cdot)$ is appropriate for the data. Noting that $\text{pAUC}_{\mathbf{X}_{ik}}(u)$ is the conditional mean of $V^{(u)}_{\text{D}ik}$, we define the residuals for fitting (2) as $\widehat{e}_{ik} = \widehat{V}^{(u)}_{\text{D}ik} - \eta(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\vec{\mathbf{X}}_{ik})$. To examine the appropriateness of (2), we first check the functional form for each component of the covariate $\mathbf{X}$. For

$q = 1, \ldots, p$, we consider the following moving sum of the $\widehat{e}_{ik}$'s over the $X_{ik}^{(q)}$:

$$\bar{W}_q(x;b) = \frac{1}{N_{\text{D}}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} I(x - b < X_{ik}^{(q)} \leq x) \widehat{e}_{ik} \tag{4}$$

for a pre-specified positive block size b, where $X_{ik}^{(q)}$ is the qth element of $\mathbf{X}_{ik}$. Moving sums of residuals were proposed by Lin, Wei and Ying (2002) to test the goodness of fit for generalized linear models. When $b = \infty$, (4) corresponds to the partial residual process considered by Su and Wei (1991).

Under H_0 that model (2) holds, $\bar{W}_q(x;b)$ is expected to fluctuate around 0. To obtain its large sample distribution, let $\mathcal{L} = (\mathcal{L}_1, \ldots, \mathcal{L}_{n_{\text{D}}+n_{\bar{\text{D}}}})$ be a random sample from the standard normal distribution, independent of the data. Define

$$\begin{aligned}
n_{\text{D}}^{\frac{1}{2}} \widehat{W}_q(x;b) &= n_{\text{D}}^{\frac{1}{2}} \sum_{i=1}^{n_{\text{D}}} \widehat{W}_{\text{D}qi}(x) \mathcal{L}_i + n_{\bar{\text{D}}}^{\frac{1}{2}} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \widehat{W}_{\bar{\text{D}}qj}(x) \mathcal{L}_j \,, \\
\widehat{W}_{q_{\text{D}i}}(x;b) &= K_{\text{D}}^{-1} \sum_{k=1}^{K_i} I(x - b < X_{ik}^{(q)} \leq x) \left\{ \widehat{V}_{\text{D}ik}^{(u)} - \eta(\widehat{\boldsymbol{\beta}}^{\mathsf{T}} \vec{\mathbf{X}}_{ik}) \right\} \\
&\quad + \widehat{\mathbf{R}}_q(x;b)^{\mathsf{T}} \widehat{\mathbb{A}}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\mathfrak{B}}_{\text{D}i} \,, \\
\widehat{W}_{q_{\bar{\text{D}}j}}(x;b) &= \frac{p_{10}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} I(x - b < X_{ik}^{(q)} \leq x) I(\widehat{U}_{\text{D}ik} \leq u) \widehat{I}_{\bar{\text{D}}j}(\widehat{U}_{\text{D}ik}, \mathbf{Z}_{ik}) \\
&\quad + \widehat{\mathbf{R}}_q(x;b)^{\mathsf{T}} \widehat{\mathbb{A}}^{-1} \widehat{\mathfrak{B}}_{\bar{\text{D}}j} \,,
\end{aligned}$$

and $\widehat{\mathbf{R}}_q(x;b) = (1/N_D) \sum_{i,k} I(x - b < X_{ik}^{(q)} \leq x) \dot{\eta}(\widehat{\boldsymbol{\beta}}_0^{\mathsf{T}} \vec{\mathbf{X}}_{ik}) \vec{\mathbf{X}}_{ik}$. In Appendix C we show that, under H_0, the conditional distribution of $n_{\text{D}}^{1/2} \widehat{W}_q(x;b)$ given the data is the same in the limit as the unconditional distribution of $n_{\text{D}}^{1/2} \bar{W}_q(x;b)$. To approximate the null distribution of $W_q(x;b)$, we simulate a number of realizations from $n_{\text{D}}^{1/2} \widehat{W}_q(x;b)$ by repeatedly generating the normal samples of $\mathcal{L}$ while fixing the data at their observed values. To assess how unusual the observed process $\bar{W}_q(x;b)$ is under H_0, one may plot $\bar{W}_q(x;b)$ along with a few realizations from $\widehat{W}_q(x;b)$, and supplement the graphical display with an estimated p-value from a supremum-type test statistic $S_q = \sup_x |\bar{W}_q(x;b)|$. An unusually large observed value s_q would suggest improper specification of the functional form of X_q. In practice, the p-value, $P(S_q \geq s_q)$, can be approximated by $P(\widehat{S}_q \geq s_q)$, where $\widehat{S}_q = \sup_x |\widehat{W}_q(x;b)|$. We estimate $P(\widehat{S}_q \geq s_q)$ by generating a large number $\mathcal{J}$, say $\mathcal{J} = 5,000$, of realizations from $\widehat{W}_q(\cdot;b)$.

To assess the linearity of the model given in (2), and more generally the link

function $\eta(\cdot)$, we consider the moving sum of residuals over the fitted values:

$$\bar{W}_\eta(x;b) = n_{\bar{\text{D}}}^{\frac{1}{2}} \sum_{i=1}^{n_{\text{D}}} \sum_{k=1}^{K_i} I(x-b < \widehat{\boldsymbol{\beta}}^{\mathsf{T}} \vec{\mathbf{X}}_{ik} < x) \widehat{e}_{ik}.$$

The null distribution of $\bar{W}_\eta(x;b)$ can be approximated by the conditional distribution of $\widehat{W}_\eta(x;b)$, which is obtained from $\widehat{W}_q(x;b)$ by replacing $I(x-b < X_{ik}^{(q)} \le x)$ with $I(x-b < \widehat{\boldsymbol{\beta}}^{\mathsf{T}} \vec{\mathbf{X}}_{ik} \le x)$. As noted in Lin, Wei and Ying (2002), although S_η is referred to as the link function test, anomalies in $\bar{W}_\eta$ may reflect misspecification of the link function, of the functional form of the response variable, or of the linear predictor.

5. Simulation Studies

5.1. Asymptotic inference in finite samples

To evaluate the finite sample performance of the method, we first examine the variance estimator for $\widehat{\text{pAUC}}(u)$ when there is no covariate. We simulate Y_{D} from $N(10, 1.5^2)$ and $Y_{\bar{\text{D}}}$ from $N(9,1)$. The induced ROC curve has a partial area of 0.0726 for FPR ≤ 0.2. The results, summarized in Table 1, show that the standard error estimates based on large sample approximation are close to the true sampling standard errors. In addition, for confidence intervals the empirical coverage probabilities are close to their nominal counterparts.

Next, we examine the validity of the large sample approximations in the regression setting for making inference in finite sample sizes. We simulate data from the following models:

$$Y_{\text{D}i} = 10 + 1.3Z_i - \epsilon_{\text{D}i}, \qquad \text{for} \quad i = 1, \ldots, n_{\text{D}}, \tag{5}$$

$$Y_{\bar{\text{D}}j} = \ 9 + 0.5Z_j - \epsilon_{\bar{\text{D}}j}, \qquad \text{for} \quad j = 1, \ldots, n_{\bar{\text{D}}}, \tag{6}$$

where Z is generated from Uniform(0, C). We first set $C = 1$ and generate $\epsilon_{\text{D}i} \sim N(0, 1.5^2)$ and $\epsilon_{\bar{\text{D}}j} \sim N(0,1)$. The induced pAUC model is

$$\text{pAUC}_z(u) = \eta_u(1 + 0.8z)\,, \qquad \text{where} \quad \eta_u = \int_0^u \Phi\Big\{\frac{x + \Phi^{-1}(v)}{1.5}\Big\} dv\,.$$

Table 1. The Bias, sampling standard error (SSE), sample average of the estimated standard errors (ESE), and empirical coverage probability (CovP) of the 95% confidence interval for $\widehat{\text{pAUC}}$. Results are based on 1,000 simulated datasets.

	$n_{\bar{\text{D}}} = 100$				$n_{\bar{\text{D}}} = 200$			
	Bias	SSE	ESE	CovP	Bias	SSE	ESE	CovP
$n_{\text{D}} = 100$	0.0005	0.011	0.011	0.946	0.0005	0.010	0.010	0.945
$n_{\text{D}} = 200$	0.0007	0.010	0.010	0.940	0.0002	0.008	0.008	0.944

Table 2. The Bias, sampling standard error (SSE), average of the estimated standard error estimator (ESE), and the coverage probability (CovP) of the 95% confidence interval. Each entry is based on 1,000 simulation samples.

(a) $N(0,1)$ versus $N(0,1.5^2)$

$(n_{\bar{D}}, n_D)$	β_0				β_1			
	Bias	SSE	ESE	CovP	Bias	SSE	ESE	CovP
(100, 100)	0.008	0.430	0.431	0.950	0.044	0.714	0.724	0.960
(100, 200)	0.023	0.332	0.347	0.960	0.000	0.551	0.576	0.957
(200, 100)	0.021	0.376	0.389	0.963	-0.009	0.643	0.663	0.959
(400, 100)	-0.002	0.369	0.367	0.949	0.020	0.635	0.632	0.955

(b) Extreme Value versus Extreme Value

$(n_{\bar{D}}, n_D)$	β_0				β_1			
	Bias	SSE	ESE	CovP	Bias	SSE	ESE	CovP
(100, 100)	0.031	0.453	0.472	0.950	0.007	0.323	0.349	0.967
(100, 200)	0.020	0.408	0.419	0.947	0.011	0.290	0.304	0.963
(200, 100)	0.002	0.385	0.391	0.955	0.008	0.277	0.294	0.966
(400, 100)	-0.010	0.334	0.343	0.951	0.017	0.257	0.263	0.954

We refer to this as the normal-normal model. We choose $u = 0.2$ and fit the data with $\text{pAUC}_z(u) = \eta_u(\beta_0 + \beta_1 z)$. To estimate the FPR conditional on covariates, we use a semi-parametric location model (Heagerty and Pepe (1999)): $S_{D,Z}(y) = S_0(y - \gamma Z)$, where γ and S_0 are unspecified. In Table 2(a), we present the bias, the sampling standard error, average of the standard error estimates, and the coverage probability of the 95% confidence intervals for β_0 and β_1. The standard error estimates are close to the true sampling standard errors. In addition, the empirical coverage probabilities are close to their nominal counterparts.

In another study, we also use models (5) and (6), but simulate ϵ_{Di} and $\epsilon_{\bar{D}j}$ from extreme value distributions and Z from Uniform$(0, 2)$. The corresponding link function η_u is then

$$\eta_u(x) = u - \frac{1 - \exp\{-(1-u)^{1+\exp(x)}\}}{1 + \exp(x)}.$$

The results for $u = 0.2$, summarized in Table 2(b), also show that the asymptotic approximations behave reasonably in finite samples.

5.2. Comparison with existing method

To compare the proposed method to the Dodd and Pepe (2003a) approach, we simulate data from models (5) and (6) with ϵ_{Di} and $\epsilon_{\bar{D}j}$ generated from zero-mean normal distributions and extreme value distributions. For each simulated

Table 3. Estimates of β_0 and β_1 compared with their respective actual values $\beta_0 = 1$ and $\beta_1 = 0.8$, based on the Dodd and Pepe approach (D&P) and on the new approach (New). Results are based on 1,000 simulated datasets.

(a) N$(0, 1)$ versus N$(0, 1.5^2)$

$(n_{\bar{\text{D}}}, n_{\text{D}})$	Bias				Mean Squared Error			
	β_0^{New}	$\beta_0^{\text{D\&P}}$	β_1^{New}	$\beta_1^{\text{D\&P}}$	β_0^{New}	$\beta_0^{\text{D\&P}}$	β_1^{New}	$\beta_1^{\text{D\&P}}$
$(100, 100)$	0.008	−0.004	0.044	0.028	0.185	0.623	0.512	2.182
$(100, 200)$	0.023	0.015	0.000	−0.017	0.111	0.559	0.304	1.994
$(200, 100)$	9.021	0.004	−0.009	−0.010	0.148	0.327	0.414	1.103
$(400, 100)$	−0.002	−0.002	0.020	0.013	0.136	0.237	0.404	0.787

(b) Extreme Value versus Extreme Value

$(n_{\bar{\text{D}}}, n_{\text{D}})$	Bias				Mean Squared Error			
	β_0^{New}	$\beta_0^{\text{D\&P}}$	β_1^{New}	$\beta_1^{\text{D\&P}}$	β_0^{New}	$\beta_0^{\text{D\&P}}$	β_1^{New}	$\beta_1^{\text{D\&P}}$
$(100, 100)$	0.031	−0.045	0.007	0.064	0.206	0.678	0.104	0.542
$(100, 200)$	0.020	−0.010	0.011	0.023	0.167	0.540	0.084	0.435
$(200, 100)$	0.002	−0.002	0.008	0.010	0.148	0.342	0.077	0.251
$(400, 100)$	−0.010	−0.032	0.017	0.029	0.111	0.217	0.066	0.153

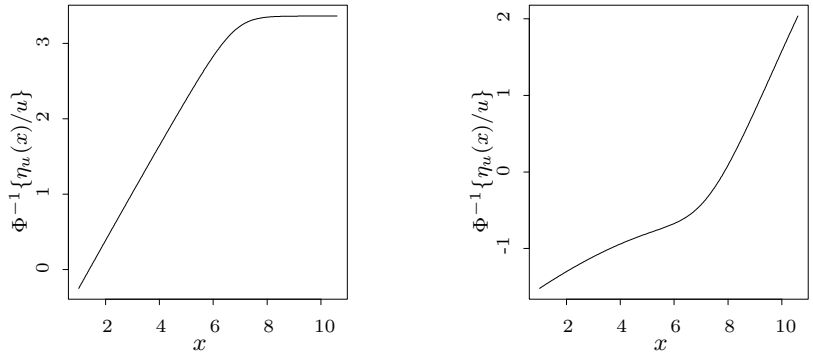
data, we obtain point estimates of β_0 and β_1 with the proposed approach by fitting $\text{pAUC}_z(u) = \eta_u(\beta_0 + \beta_1 z)$, and with Dodd and Pepe (2003a) by fitting $\text{pAUC}_{z_{\text{D}}, z_{\bar{\text{D}}}}(u) = \eta_u\{\beta_0 + \beta_1 z_{\text{D}} + \beta_2(z_{\text{D}} - z_{\bar{\text{D}}})\}$. The results in Table 3 show that even though the new approach uses a more robust model, the new estimator is more efficient than the Dodd and Pepe (2003a) estimator. At sample sizes of $n_{\bar{\text{D}}} = 400$ and $n_{\text{D}} = 100$, the empirical efficiency of the Dodd and Pepe (2003a) method relative to the new method is 57% for β_0 and 51% for β_1 when $\epsilon_{\bar{\text{D}}} \sim N(0, 1)$ and $\epsilon_{\text{D}} \sim N(0, 1.5^2)$. When ϵ_{D} and $\epsilon_{\bar{\text{D}}}$ are generated from the extreme value distribution, the relative efficiency is 51% for β_0 and 43% for β_1.

5.3. Mis-specified link function

To examine the properties of the estimator under a mis-specified link function, we simulate data from models (5) and (6) with $Z \sim \text{Uniform}(0, 12)$, and fit the data to the model

$$\text{pAUC}_z(u) = u\Phi(\beta_0 + \beta_1 z)\,. \tag{7}$$

We generate $\epsilon_{\bar{\text{D}}}$ from a standard normal. For ϵ_{D} we consider two scenarios: (1) $N(0, 1.5^2)$, and (2), a mixture of $N(2, 3^2)$ with probability 0.3 and $N(7, 1)$ with probability 0.7. To explore how far from (7) the true underlying link functions



(a) Normal (b) Mixture

Figure 1. Plot of function $\Phi^{-1}\{\eta_u(x)/u\}$ for $u = 0.2$.

Table 4. Bias and mean squared error (MSE) of the predicted pAUC. For each dataset, we fit with two models: $\text{pAUC}_z(0.2) = 0.2\Phi(\beta_0+\beta_1 z)$ (Linear) and $\text{pAUC}_z(0.2) = 0.2\Phi\{\beta_0 + \boldsymbol{\beta}_1^{\intercal}\mathcal{R}(z)\}$ (Spline). The results are based on 1,000 simulated datasets with sample size $n_{\text{D}} = n_{\bar{\text{D}}} = 200$.

		$\epsilon_{\text{D}} \sim N(0, 1.5^2)$				$\epsilon_{\text{D}} \sim$ Normal Mixture			
	True	Bias		MSE		Bias		MSE	
z	$\text{pAUC}_z(0.2)$	Linear	Spline	Linear	Spline	Linear	Spline	Linear	Spline
2	2.4E-2	5.1E-4	−8.8E-4	1.1E-4	2.4E-4	−9.8E-3	1.5E-4	1.3E-4	1.3E-4
4	3.6E-2	9.1E-5	3.5E-4	1.6E-5	3.5E-5	−3.0E-3	−3.1E-4	7.0E-5	1.3E-4
6	4.8E-2	−1.1E-4	−8.2E-4	6.8E-7	1.3E-5	1.7E-2	−6.0E-4	3.6E-4	1.3E-4
8	8.0E-2	4.2E-5	8.1E-5	1.3E-8	1.0E-8	2.6E-2	1.9E-3	7.5E-4	2.3E-4
10	1.6E-1	7.3E-5	8.0E-5	5.7E-7	6.1E-5	−1.3E-2	1.6E-4	2.4E-4	2.0E-4

are, we examine the linearity of $\Phi^{-1}\{\eta_u(x)/u\}$ in x, where η_u is the true link function. In Figure 1, we can see that (7) is a fair approximation for the first setting, especially for $x \leq 8$, but not so for the second setting.

As shown in Table 4, the predicted pAUC based on the linear model in (7) has little bias in the first setting, but the bias is substantial in the second setting. To improve the approximation, we instead fit a quadratic spline model for the covariate effect:

$$\text{pAUC}_z(u) = u\Phi\Big\{\beta_0 + \boldsymbol{\beta}_1 z + \boldsymbol{\beta}_2 z^2 + \sum_{k=1}^{K} b_k (z - \kappa_k)_+^2\Big\}, \tag{8}$$

where $x_+ \equiv \max(0, x)$ and $\kappa_1, \ldots, \kappa_K$ are the pre-specified knots. In this study, we use three knots at 3, 6, and 9. The results, also presented in in Table 4, suggest that the spline model (8) is rather robust with respect to the mis-specification of the link function.

6. Example: Early detection of prostate cancer with PSA

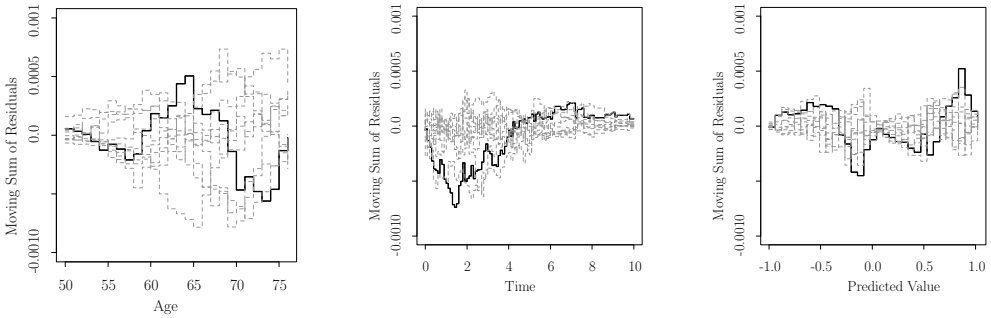
PSA levels in serum are used to screen men for prostate cancer. However, considerable controversy exists as to its value. A longitudinal case-control study of PSA as a screening marker for prostate cancer was nested within the Beta-Carotene and Retinol Efficacy Trial, in an effort to evaluate the accuracy of PSA, prior to onset of clinical symptoms, in diagnosing prostate cancer (Thornquist et al. (1993) and Etzioni et al. (1999)). As part of the protocol, serum was drawn periodically from study participants, and stored. 88 subjects developed prostate cancer during the study and their serum samples were analyzed for PSA levels. An age-matched set of 88 control subjects also had their stored serum samples analyzed for PSA levels. The median number of PSA measurements per subject is 4 and the median time interval between two consecutive measurements is 1 year.

Among subjects that develop cancer it is likely that PSA measured closer to the time of onset of clinical symptoms is more predictive of disease than measures taken earlier in time. Additionally, increasing age is associated with increasing serum PSA level and could affect the discriminatory capacity of PSA. To understand the time and age effect on PSA accuracy, we consider a pAUC model with a covariate T, defined as the time (in years) between the onset of symptoms and the time at which the serum sample was drawn, and an additional covariate z = age at the time of measurement (in years). We choose the upper bound of FPR as 0.02, considered in Baker (2000) for PSA screening, and fit the model

$$\text{pAUC}_{z,T}(0.02) = 0.02\Phi(\beta_0 + \beta_z z + \beta_t T), \tag{9}$$

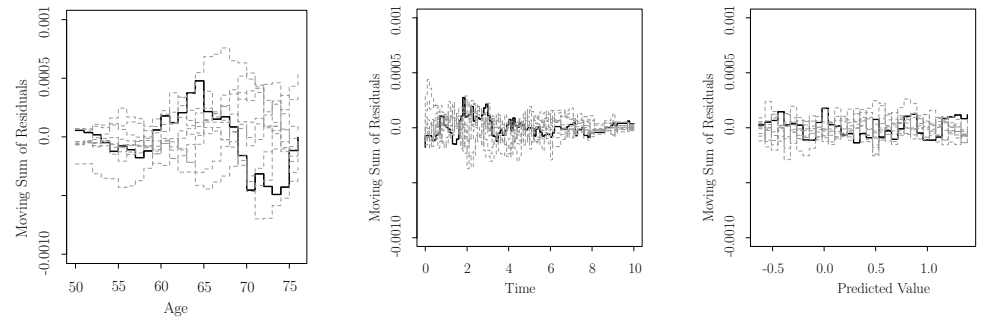
to the data. Using our approach, the estimate of β_t is -0.091 (s.e.= 0.053) per year and the coefficient for age, β_z, is estimated as 0.0053 (s.e.= 0.020) per year of age . The negative coefficient for T implies that discrimination improves as T decreases, i.e., when PSA is measured closer to diagnosis. The coefficient for age is almost 0 (p-value= 0.79) suggesting that discrimination is about the same in younger men as in older men. We also fit the model using the Dodd and Pepe method where the comparison between a diseased subject and a non-diseased subject is only included if the age difference is no greater than 2 years. The estimated coefficients are -0.11 (s.e.= 0.14) for the time lag and 0.075 (s.e.= 0.18) for age.

To examine whether model (9) is appropriate for the data, we consider $\bar{W}_z$ and $\bar{W}_T$ for checking the linearity in specific covariate effects, and $\bar{W}_\eta$ for checking the link function. Figures 2(a)$-$(c) display the observed processes $\bar{W}_z$, $\bar{W}_T$, $\bar{W}_\eta$ along with realizations of $\widehat{W}_z$, $\widehat{W}_T$ and $\widehat{W}_\eta$. The p-values based on the sup-statistics with $\mathcal{J} = 5,000$ are 0.38 for the linearity in age, 0.0085 for the linearity



(a) p-value= 0.38 (b) p-value= 0.0085 (c) p-value= 0.026

(a)−(c) : Linear covariate effect model



(d) p-value= 0.42 (e) p-value= 0.52 (f) p-value= 0.50

(d)−(f) : Cubic time effect and linear age effect model

Figure 2. Plot of moving sums of residuals: (a) and (d) for testing linear age effect with $b = 10$ (interquartile range of age); (b) and (e) for testing linear time effect with $b = 3$; (c) and (f) for testing the linearity of the model with $b = 1$; the observed pattern is shown by the thick solid curve, and 10 simulated realizations under the null are shown by the dotted curve.

in T, and 0.026 for the link function. Thus, the linearity assumption in the time effect is problematic. This motivates us to consider the model

$$\text{pAUC}_{z,T}(0.02) = 0.02\Phi(\beta_0 + \beta_z z + \beta_T T + \beta_{T^2} T^2 + \beta_{T^3} T^3) \tag{10}$$

to allow for a non-linear time effect. The resulting estimate of the age effect is $\widehat{\beta}_z = 0.0047$ (s.e. $= 0.021$). The estimated time effects in model (10) are $\widehat{\beta}_T = -0.59$ (s.e. $= 0.14$), $\widehat{\beta}_{T^2} = -0.10$ (s.e.$= 0.032$) and $\widehat{\beta}_{T^3} = -0.0053$ (s.e.$= 0.0022$). We apply the model checking procedure again for model (10). The residual plots, shown in Figure 2(d)−(f), along with the p-values (0.42 for S_z, 0.52 for S_T and 0.50 for S_η), indicate that the revised model is reasonable.

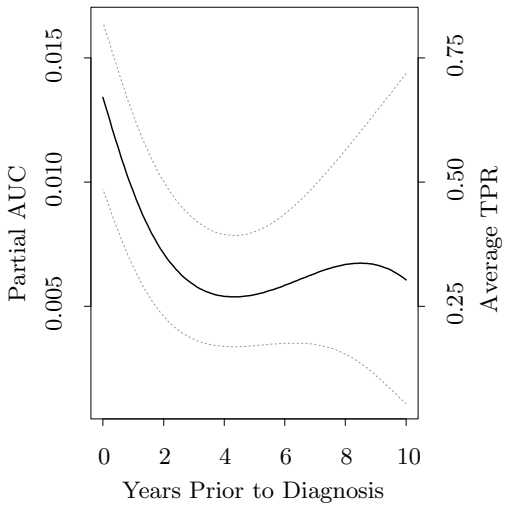


Figure 3. Predicted pAUC for PSA as a biomarker of prostate cancer in 60 year $X_{5,5} = x$, old men. Shown also are their 95% confidence intervals.

Figure 3 displays the estimated pAUCs and their 95% confidence bands for patients who are 60 years old at different times before clinical diagnosis. For example, when $T = 2$ years, the estimated pAUC(0.02) is 0.0071 (s.e. = 0.0014). Therefore, if we define the restricted reference population to be all 60-year old disease-free men with PSA value exceeding its corresponding 98th percentile, there is a $0.0071/0.02 = 36\%$ chance that a randomly selected 60-year old man with cancer whose PSA is measured at 2 years prior to diagnosis is higher than that of a man randomly selected from the restricted reference population. This probability can also be viewed as the average TPR over the range of FPR ≤ 0.02. Thus the average TPR fluctuates around 30% when $T \geq 3$ years, then improves quickly to 57% when T decreases to 6 months. This indicates that PSA may not be accurate for early detection of prostate cancer. To fully understand the predictive accuracy of PSA, one needs to further evaluate the positive and negative predictive values of PSA assessed through prospective studies.

7. Remarks

This paper provides an alternative pAUC regression method to Dodd and Pepe (2003a). Advantages of the proposed method include large-sample theory, improved efficiency, and model checking procedures. When $u = 1$, the proposed estimator provides an alternative to the AUC regression approach developed by Dodd and Pepe (2003b). The proposed inference procedure also accounts for possible within-cluster correlation. It is important to note that both the

proposed method and the Dodd and Pepe (2003a) method require modelling of the conditional distribution of $Y_{\bar{\text{D}}}$ when there are continuous covariates. Existing model checking procedures (e.g., Lin, Wei and Ying (2002) and Cai and Zheng (2007)) may be used to examine the adequacy of the proposed model for $Y_{\bar{\text{D}}}$. The model checking procedures for the proposed pAUC model are based on a simulation technique that has a minimal computational burden relative to other re-sampling methods such as the bootstrap. This offers a formal goodness of fit method that is not available with existing AUC and pAUC regression methods. Additional simulation studies indicated that the proposed tests have proper sizes at least when $\min(n_{\bar{\text{D}}}, n_{\text{D}}) \geq 200$. The power of the tests would depend on the degree of the model mis-specification and this remains to be investigated. When applied to the PSA example, the procedure indicated an important non-linearity, which resulted in a revised and better-fitting model. The validity of the proposed inference procedure also requires the correct specification of the FPR model. Goodness of fit for typical FPR models, such as the semi-parametric location scale model, may be examined based on existing methods such as those proposed in Lin, Wei and Ying (2002) and Cai and Zheng (2007).

Although the focus here is on a general regression model, the method is easily adapted to compare the accuracies of two tests, as considered by Wieand et al. (1989). It is straightforward to extend our procedures to make inference about the difference of two pAUCs for both paired and unpaired data. With a single covariate indicating test type, one can create a model based on (2) to examine the difference in the accuracy of two tests. The resulting estimator is equivalent to the estimator proposed by Wieand et al. (1989) when $K_{\bar{\text{D}}j} = K_{\text{D}i} = 1$.

Appendix

A. Large Sample Properties of $\widehat{\text{pAUC}}(u)$

For technical reasons, we assume that potentially every diseased subject has $\mathcal{K} = \max(K_1, \ldots, K_{n_{\text{D}}})$ records, and that the n_{D} sets of random vectors $\{\bar{\mathbf{Y}}_{\text{D}i}\}$, or $\{(\mathbf{Y}_{\text{D}i}, \bar{\mathbf{X}}_i)\}$ with covariates, are independent and identically distributed, where $\bar{\mathbf{Y}}_{\text{D}i} = (Y_{\text{D}i1}, \ldots, Y_{\text{D}iK_{\text{D}}})$ and $\bar{\mathbf{X}}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iK_{\text{D}}})$. Although not every subject with disease has $\mathcal{K}$ records, the presence or absence of individual records in a cluster does not depend on the observations. Similar assumptions are made for many marginal method based estimators, for example in Lee, Wei and Amato (1992) and Cai, Wei and Wilcox (2000). Corresponding assumptions are made for observations from disease-free subjects.

We assume that $\text{ROC}_{\mathbf{X}}(\cdot)$ is continuously differentiable. The uniform consistency of $\widehat{S}_{\bar{\text{D}}}(\cdot)$ and the Uniform Law of Large Numbers (Pollard (1990)) ensure the consistency of $\widehat{\text{pAUC}}(u)$. It remains to determine the large sample distribution of

$\widehat{\text{pAUC}}(u)$. To this end, let $\widehat{\mathcal{I}}_{\bar{\text{D}}}(u) = S_{\bar{\text{D}}}(\widehat{S}_{\bar{\text{D}}}^{-1}(u))$ and $\widetilde{\text{pAUC}}(u) = 1/N_{\text{D}} \sum_{i,k} V_{\text{D}ik}^{(u)}$. We note that

$$\begin{aligned} & n_{\text{D}}^{\frac{1}{2}} \Big\{ \widehat{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \quad = n_{\text{D}}^{\frac{1}{2}} \Big\{ \widetilde{\text{pAUC}}(\widehat{\mathcal{I}}_{\bar{\text{D}}}(u)) - \widetilde{\text{pAUC}}(u) \Big\} + n_{\text{D}}^{\frac{1}{2}} \Big\{ \widetilde{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \qquad + \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} I(U_{\text{D}ik} \le \widehat{\mathcal{I}}_{\bar{\text{D}}}(u)) \Big\{ u - \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - \widehat{U}_{\text{D}ik} + U_{\text{D}ik} \Big\}. \end{aligned}$$

It has been shown that $\sup_u |\widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - u| \to 0$ and $n_{\bar{\text{D}}}^{1/2} \{ \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - u \}$ is asymptotically equivalent to $n_{\bar{\text{D}}}^{-1/2} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} I_{\bar{\text{D}}j}(u)$, where $I_{\bar{\text{D}}j}(u) = \sum_{l=1}^{K_j} \{ u - I(Y_{\bar{\text{D}}jl} \ge S_{\bar{\text{D}}}^{-1}(u)) \}$ (Cai and Pepe (2002)). This, coupled with the equicontinuity of the process $n_{\text{D}}^{1/2} \{ \widetilde{\text{pAUC}}(u) - \text{pAUC}(u) \}$, ensures that

$$\begin{aligned} & n_{\text{D}}^{\frac{1}{2}} \Big\{ \widehat{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \quad \approx n_{\text{D}}^{\frac{1}{2}} \Big\{ \text{pAUC}(\widehat{\mathcal{I}}_{\bar{\text{D}}}(u)) - \text{pAUC}(u) \Big\} + n_{\text{D}}^{\frac{1}{2}} \Big\{ \widetilde{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \qquad + \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} I(U_{\text{D}ik} \le \widehat{\mathcal{I}}_{\bar{\text{D}}}(u)) \Big\{ u - \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - \widehat{U}_{\text{D}ik} + U_{\text{D}ik} \Big\}. \end{aligned}$$

It follows from a Taylor series expansion that $1/N_{\text{D}} \sum_{i,k} I(U_{\text{D}ik} \le u) \to \text{ROC}(u)$, and from the equicontinuity of the process $n_{\text{D}}^{1/2} \{ \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - u \}$ that

$$\begin{aligned} & n_{\text{D}}^{\frac{1}{2}} \Big\{ \widehat{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \quad \approx n_{\text{D}}^{\frac{1}{2}} \text{ROC}(u) \Big\{ \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - u \Big\} + n_{\text{D}}^{\frac{1}{2}} \Big\{ \widetilde{\text{pAUC}}(u) - \text{pAUC}(u) \Big\} \\ & \qquad - p_{10}^{\frac{1}{2}} \int_0^u n_{\bar{\text{D}}}^{\frac{1}{2}} \Big\{ \widehat{\mathcal{I}}_{\bar{\text{D}}}^{-1}(v) - v \Big\} d\text{ROC}(v) - n_{\text{D}}^{\frac{1}{2}} \text{ROC}(u) \Big\{ \widehat{\mathcal{I}}_{\bar{\text{D}}}(u) - u \Big\} \\ & \quad \approx n_{\text{D}}^{-\frac{1}{2}} \sum_{i=1}^{n_{\text{D}}} \mathcal{P}_{\text{D}i} + n_{\bar{\text{D}}}^{-\frac{1}{2}} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathcal{P}_{\bar{\text{D}}j} \,, \end{aligned}$$

where $\mathcal{P}_{\text{D}i} = K_{\text{D}}^{-1} \sum_{k=1}^{K_i} V_{\text{D}ik}^{(u)} - \text{pAUC}(u)$, $\mathcal{P}_{\bar{\text{D}}j} = p_{10}^{1/2} \int_0^u I_{\bar{\text{D}}j}(v) d\text{ROC}(v)$ and K_{D} is the limit of $N_{\text{D}}/n_{\text{D}}$. It follows from the Central Limit Theorem that $n_{\text{D}}^{1/2} \{ \widehat{\text{pAUC}}(u) - \text{pAUC}(u) \}$ converges in distribution to a zero-mean normal with variance σ^2, where σ^2 is the limit of $n_{\text{D}}^{-1} \sum_{i=1}^{n_{\text{D}}} \mathcal{P}_{\text{D}i}^2 + n_{\bar{\text{D}}}^{-1} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathcal{P}_{\bar{\text{D}}j}^2$. A consistent estimate of σ^2 is $\widehat{\sigma}^2$ which is obtained by replacing all the theoretical quantities in $n_{\text{D}}^{-1} \sum_{i=1}^{n_{\text{D}}} \mathcal{P}_{\text{D}i}^2 + n_{\bar{\text{D}}}^{-1} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathcal{P}_{\bar{\text{D}}j}^2$ by their empirical counterparts.

B. Large Sample Properties of $\widehat{\boldsymbol{\beta}}$

To show the existence and uniqueness of $\widehat{\boldsymbol{\beta}}$, we assume that the covariates $\mathbf{X} = (\mathbf{Z}, \mathbf{Z}_{\text{D}})$ are bounded, the estimators of $S_{\bar{\text{D}},\mathbf{Z}}(y)$ are uniformly consistent, and $n_{\bar{\text{D}}}^{1/2}\{\widehat{S}_{\bar{\text{D}},\mathbf{Z}}(y) - S_{\bar{\text{D}},\mathbf{Z}}(y)\}$ converges weakly to a Gaussian process uniformly in y and $\mathbf{Z}$. Without loss of generality, we also assume that $n_{\bar{\text{D}}}^{1/2}\{\widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}}(u) - u\}$ can be approximated by a sum of independent terms:

$$\sup_{u,\mathbf{Z}} \left| n_{\bar{\text{D}}}^{\frac{1}{2}} \left\{ \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}}(u) - u \right\} - n_{\bar{\text{D}}}^{-\frac{1}{2}} \sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} I_{\bar{\text{D}}j}(u, \mathbf{Z}) \right| \to 0 \tag{11}$$

in probability, where $\widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}}(u) = S_{\bar{\text{D}}}(\widehat{S}^{-1}_{\bar{\text{D}},\mathbf{Z}}(u))$. Let $\bar{\mathbf{V}}(\boldsymbol{\beta})$ denote the left hand side of (3). It is easy to see that $(\partial\bar{\mathbf{V}}(\boldsymbol{\beta}))/(\partial\boldsymbol{\beta}) = \widehat{\mathbb{A}}(\boldsymbol{\beta})$, where $\widehat{\mathbb{A}}(\boldsymbol{\beta}) = 1/N_{\text{D}} \sum_{i,k}$ $\dot{\eta}(\boldsymbol{\beta}^{\text{T}}\vec{\mathbf{X}}_{ik})\vec{\mathbf{X}}_{ik}^{\otimes 2}$ is nonnegative definite. Furthermore, $\widehat{\mathbb{A}}(\boldsymbol{\beta}_0) \to \mathbb{A}$. When $\vec{\mathbf{X}}_{ik}$ is non-degenerate, $\mathbb{A}$ is positive definite. Now, since $\bar{\mathbf{V}}(\boldsymbol{\beta}_0) \to 0$, by the Inverse Function Theorem, there exists a unique solution $\widehat{\boldsymbol{\beta}}$ to the equation $\bar{\mathbf{V}}(\boldsymbol{\beta})$ in a neighborhood of $\boldsymbol{\beta}_0$. This, coupled with the nonnegativity of $\widehat{\mathbb{A}}(\boldsymbol{\beta})$, ensures the uniqueness of the root of $\bar{\mathbf{V}}(\boldsymbol{\beta}) = 0$ in the entire domain of $\boldsymbol{\beta}$, asymptotically. The above proof also implies that $\widehat{\boldsymbol{\beta}}$ is strongly consistent.

By the consistency of $\widehat{\boldsymbol{\beta}}$ and a Taylor series expansion of $\bar{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$, we obtain

$$n_{\text{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx \mathbb{A}^{-1} n_{\text{D}}^{\frac{1}{2}} \bar{\mathbf{V}}(\boldsymbol{\beta}_0)\,. \tag{12}$$

Define $V_{\text{D}ik}^{(u)} = u - \min(u, U_{\text{D}ik})$, $e_{ik} = V_{\text{D}ik}^{(u)} - \text{pAUC}_{\mathbf{X}_{ik}}(u)$, $\widetilde{\mathbf{V}}(u) = 1/N_{\text{D}} \sum_{i,k}$ $w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}e_{ik}$ and $\bar{\mathbf{V}}_1 = 1/N_{\text{D}} \sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}(\widehat{V}_{\text{D}ik}^{(u)} - V_{\text{D}ik}^{(u)})$. Then $\bar{\mathbf{V}}(\boldsymbol{\beta}_0) =$ $\widetilde{\mathbf{V}}(u) + \bar{\mathbf{V}}_1$. We first show the large sample approximation for $n_{\text{D}}^{1/2}\bar{\mathbf{V}}_1$. Note that

$$\begin{aligned} n_{\text{D}}^{\frac{1}{2}}\bar{\mathbf{V}}_1 = {} & \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik} \left[I(U_{\text{D}ik} \le \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u)) \left\{ \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u) - U_{\text{D}ik} \right\} - V_{\text{D}ik}^{(u)} \right] \\ & + \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik} I(U_{\text{D}ik} \le \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u)) \left\{ u - \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u) - \widehat{U}_{\text{D}ik} + U_{\text{D}ik} \right\}. \end{aligned}$$

It follows from the equicontinuity of $n_{\text{D}}^{1/2}\widetilde{\mathbf{V}}(\cdot)$ and the uniform consistency of $\widehat{S}_{\bar{\text{D}},\mathbf{Z}}(\cdot)$ that

$$\begin{aligned} n_{\text{D}}^{\frac{1}{2}}\bar{\mathbf{V}}_1 \approx {} & \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik} \left\{ \text{ROC}_{\mathbf{X}_{ik}}(u) - I(U_{\text{D}ik} \le u) \right\} \left\{ \widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u) - u \right\} \\ & - \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}} \sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik} I(U_{\text{D}ik} \le u)(\widehat{U}_{\text{D}ik} - U_{\text{D}ik}). \end{aligned} \tag{13}$$

Since $n_{\bar{\text{D}}}^{1/2}\{\widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}}(u)-u\}$ converges weakly to a Gaussian process, using the Strong Law of Large Numbers and the Strong Representation Theorem (Pollard (1990)), one can show that (13) $\to 0$ in probability. Therefore,

$$n_{\text{D}}^{\frac{1}{2}}\bar{\mathbf{V}}_1 \approx \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}}\sum_{i,k} w(\mathbf{X}_{ik})\vec{\mathbf{X}}_{ik}\int_0^u \Big\{\widehat{\mathcal{I}}_{\bar{\text{D}},\mathbf{Z}_{ik}}(u)-u\Big\}d\text{ROC}_{\mathbf{X}_{ik}}(u)\,.$$

This, coupled with (11) and (12), implies that $n_{\text{D}}^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \approx \mathbb{A}^{-1}\{n_{\text{D}}^{-1/2}\sum_{i=1}^{n_{\text{D}}} \mathfrak{B}_{\text{D}i} + n_{\bar{\text{D}}}^{-1/2}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathfrak{B}_{\bar{\text{D}}j}\}$.

C. Large Sample Distribution of $\bar{W}(\mathbf{x})$ Under Model (2)

Let $I_{ik}^{(q)}(x;b)$ denote $I(x-b<X_{ik}^{(q)}\le x)$. By the consistency of $\widehat{\boldsymbol{\beta}}$ and the Taylor series expansion, uniformly in x, we have

$$\begin{aligned} n_{\text{D}}^{\frac{1}{2}}\bar{W}_q(x;b) \approx{}& \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}}\sum_{i,k} I_{ik}^{(q)}(x;b)(\widehat{V}_{\text{D}ik}^{(u)}-V_{\text{D}ik}^{(u)}) + \frac{n_{\text{D}}^{\frac{1}{2}}}{N_{\text{D}}}\sum_{i,k} I_{ik}^{(q)}(x;b)e_{ik} \\ & -\widehat{\mathbf{R}}_q(x;b)^{\intercal}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0). \end{aligned}$$

Furthermore, the Uniform Law of Large Numbers (Pollard (1990)) implies that $\widehat{\mathbf{R}}_q(x)$ converges, uniformly in x, to a non-random function, $\mathbf{R}_q(x)$. Using arguments similar to those given in Appendix B and the large sample properties of $n_{\text{D}}^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$, one can show that

$$\begin{aligned} n_{\text{D}}^{\frac{1}{2}}\bar{W}_q(x;b) \approx{}& n_{\text{D}}^{\frac{1}{2}}\sum_{i=1}^{n_{\text{D}}}\Big\{W_{q_{\text{D}_1 i}}(x;b)+W_{q_{\text{D}_2 i}}(x;b)\Big\} \\ & +n_{\bar{\text{D}}}^{-\frac{1}{2}}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}}\Big\{W_{q_{\bar{\text{D}}_1 j}}(x;b)+W_{q_{\bar{\text{D}}_2 j}}(x;b)\Big\}\,, \end{aligned}$$

where $W_{q_{\text{D}_1 i}}(x;b) = K_{\text{D}}^{-1}\sum_{k=1}^{K_i} I_{ik}^{(q)}(x;b)e_{ik}$, $W_{q_{\text{D}_2 i}}(x;b) = R_q(x;b)^{\intercal}\mathbb{A}^{-1}\mathfrak{B}_{\text{D}i}$, $W_{q_{\bar{\text{D}}_1 j}}(x;b)$ is the limit of $(p_{10}^{1/2})/N_{\text{D}}\sum_{i,k} I_{ik}^{(q)}(x;b)\int_0^u I_{\bar{\text{D}}j}(v,\mathbf{Z}_{ik})d\text{ROC}_{\mathbf{X}_{ik}}(v)$, and $W_{q_{\bar{\text{D}}_2 j}}(x;b) = R_q(x;b)^{\intercal}\mathbb{A}^{-1}\mathfrak{B}_{\bar{\text{D}}j}$.

For fixed x, $n_{\text{D}}^{-1/2}\sum_{i=1}^{n_{\text{D}}}\{W_{q_{\text{D}i}}(x;b)+W_{q_{\text{D}_2}}(x;b)\}$ and $n_{\bar{\text{D}}}^{-1/2}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}}\{W_{q_{\bar{\text{D}}_1 j}}(x;b) + W_{q_{\bar{\text{D}}_2 j}}(x;b)\}$ are essentially sums of independent and identically distributed zero-mean random variables. It follows from the Multivariate Central Limit Theorem that $W_q(x;b)$ converges in finite dimensional distributions to a zero-mean Gaussian process. Since $R_q(x;b)^{\intercal}\mathbb{A}^{-1}$ is non-random and $n_{\text{D}}^{-1/2}\sum_{i=1}^{n_{\text{D}}} \mathfrak{B}_{\text{D}i} + n_{\bar{\text{D}}}^{-1/2}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}} \mathfrak{B}_{\bar{\text{D}}j}$ does not involve x, $n_{\text{D}}^{-1/2}\sum_{i=1}^{n_{\text{D}}} W_{q_{\text{D}_2 i}}(x;b) + n_{\bar{\text{D}}}^{-1/2}$

$\sum_{j=n_{\bar{\mathrm{D}}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} W_{q_{\bar{\mathrm{D}}_2 j}}(x;b)$ is tight. Now, both $W_{q_{\mathrm{D}_1 i}}(\mathbf{x})$ and $W_{q_{\bar{\mathrm{D}}_1 j}}(x;b)$ are uniformly bounded monotone functions, which are clearly manageable (Pollard (1990, p.38)). It follows from the Functional Central Limit Theorem (Pollard (1990, p.53)) that $n_{\mathrm{D}}^{-1/2}\sum_{i=1}^{n_{\mathrm{D}}} W_{q_{\mathrm{D}_1 i}}(x;b)+n_{\bar{\mathrm{D}}}^{-1/2}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} W_{q_{\bar{\mathrm{D}}_1 j}}(x;b)$ is tight. Hence, $W_q(x;b)$ converges weakly to a zero-mean Gaussian process. Appealing to arguments similar to those given in Su and Wei (1991), we have that, conditional on the data, the process $n_{\mathrm{D}}^{1/2}\widehat{W}_q(x;b)$ converges weakly to the same limiting Gaussian process as that of $n_{\mathrm{D}}^{1/2}\bar{W}_q(x;b)$.

References

Baker, S. G. (2000). Identifying Combinations of Cancer Markers for Further Study As Triggers of Early Intervention. *Biometrics* **56**, 1082-1087.

Baker, S. G. and Pinsky, P. F. (2001). A proposed design and analysis for comparing digital and analog mammography: Special receiver operating characteristic methods for cancer screening. *J. Amer. Statist. Assoc.* **96**, 421-428.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psych.* **12**, 387-415.

Brusic, V., Rudy, G., Honeyman, M., Hammer, J. and Harrison, L. (1998). Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* **14**, 121-130.

Cai, T. (2004). Semiparametric ROC Regression Analysis with placement values. *Biostatistics* **5**, 45-6.

Cai, T. and Pepe, M. S. (2002). Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. *J. Amer. Statist. Assoc.* **97**, 1099-1107.

Cai, T., Wei, L. J. and Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika*, **87**, 867-878.

Cai, T. and Zheng, Y. (2007). Model checking for ROC regression analysis. *Biometrics*, **63**, 152-163.

DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L. (1988). Comparing the Areas Under Two Or More Correlated Receiver Operating Characteristic Curves: A nonparametric approach. *Biometrics*, **44**, 837-845.

Dodd, L. and Pepe, M. S. (2003a). Partial AUC estimation and regression, *Biometrics* **59**, 614-623.

Dodd, L. and Pepe, M. S. (2003b). Semi-parametric Regression for the Area under the Receiver Operating Characteristic Curve, *J. Amer. Statist. Assoc.* **98**, 409-417.

Dwyer (1996). In Pursuit of a Piece of the ROC. *Radiology* **201**, 621-625.

Etzioni, R., Pepe, M., Longton, G., Hu, C. and Goodman, G. (1999). Incorporating the Time Dimension in Receiver Operating Characteristic Curves: A case study of prostate cancer. *Medical Decision Making* **19**, 242-251.

Han, A. K. (1987). A non-parametric analysis of transformations. *J. Econometrics* **35**, 191-209.

Hanely, J. A. and Hajian-Tilaki, K. O. (1997). Sampling variability of nonparametric estimate of the areas under receiver operating characteristic curves: An update. *Academic Radiology* **4**, 49-58.

Heagerty, P. J. and Pepe, M. S. (1999). Semiparametric Estimation of Regression Quantiles With Application to Standardizing Weight for Height and Age in US Children. *Appl. Statist.* **48**, 533-551.

Jiang, Y. L., Metz, C. E. and Nishikawa, R. M. (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology* **201**, 745-750.

Lee, E. W., Wei, L. J. and Amato, D. A. (1992). Cox-type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations (Disc: P247). In *Survival Analysis: State of the Art*, (Edited by J. P. Klein and P. K. Goel), 237-247. Kluwer Academic Publishers Group.

Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-checking Techniques Based on Cumulative Residuals. *Biometrics* **58**, 1-12.

McClish, R. J. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making* **9**, 190-195.

Pepe, M. S. (1998). Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results. *Biometrics* **54**, 124-135.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, United Kingdom.

Pepe, M. S. and Cai, T. (2004). The Analysis of Placement Values for Evaluating Discriminatory Measures. *Biometrics* **60**, 528-535.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, Hayward, CA.

Su, J. Q. and Wei, L. J. (1991). A Lack-of-fit Test for the Mean Function in a Generalized Linear Model. *J. Amer. Statist. Assoc.* **86**, 420-426.

Thompson, M. L. and Zucchini, W. (1989). On the Statistical Analysis of ROC Curves. *Statist. Medicine* **8**, 1277-1290.

Thornquist, M. D., Omenn, G. S. and Goodman, G. E., et al. (1993). Statistical Design and Monitoring of the Carotene and Retinol Efficacy trial. *Controlled Clinical Trials* **14**, 308-324.

Viaene, S., Derrig, R. A., Baesens, B. and Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *J. Risk and Insurance* **69**, 372-421.

Wieand, S., Gail, M. H., James, B. R. and James, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76**, 585-592.

Zhang, D. D., Zhou, X.-H., Freeman, D. H., Jr. and Freeman, J. L. (2002). A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Statist. Medicine* **21**, 701-715.

Department of Biostatistics, Harvard University, Boston, MA 02115, USA.

E-mail: tcai@hsph.harvard.edu

Division of Cancer Treatment and Diagnosis, National Cancer Institute 6130 Executive Blvd, MSC 7434 Bethesda, MD 20892, USA.

E-mail: doddl@mail.nih.gov

(Received November 2005; accepted October 2006)