UW Biostatistics Working Paper Series

1-27-2003

Semiparametric Receiver Operating Characteristic Analysis to Evaluate Biomarkers for Disease

Tianxi Cai
Harvard University, tcai@hsph.harvard.edu

Margaret S. Pepe
University of Washington, mspepe@u.washington.edu

Semiparametric Receiver Operating Characteristic Analysis to Evaluate Biomarkers for Disease

Tianxi CAI and Margaret Sullivan PEPE

The receiver operating characteristic (ROC) curve is a popular method for characterizing the accuracy of diagnostic tests when test results are not binary. Various methodologies for estimating and comparing ROC curves have been developed. One approach, due to Pepe, uses a parametric regression model $\mathrm{ROC}_{\mathbf{x}}(u) = g(h_0(u) + \boldsymbol{\theta}_0'\mathbf{x})$ with the baseline function $h_0(u)$ specified up to a finite-dimensional parameter. In this article we extend the regression models by allowing arbitrary nonparametric baseline functions. We also provide asymptotic distribution theory and procedures for making statistical inference. We illustrate our approach with dataset from a prostate cancer biomarker study. Simulation studies suggest that the extra flexibility inherent in the semiparametric method is gained with little loss in statistical efficiency.

KEY WORDS: Diagnostic tests; Disease screening; Estimating equation; Generalized linear model; Sensitivity.

1. INTRODUCTION

New technologies, such as gene expression microarrays and protein mass spectrometry, promise to yield a multitude of biomarkers for disease. Once a biomarker is identified, it can be used as the basis for diagnosing disease (Srivastava and Kramer 2000) or for monitoring patients during and after treatment. Biomarkers such as CA-125 and prostate-specific antigen (PSA) are used for these purposes in the management of ovarian cancer and prostate cancer, respectively. Another application for biomarker research is in the development of treatment strategies, where disease-specific biomarkers can suggest new targets for therapeutic drugs (Elmer-Dewitt et al. 2001). Statistical methods are needed to evaluate a biomarker's capacity for distinguishing subjects with disease from those without and for comparing biomarkers (Pepe et al. 2001). One potentially useful statistical tool in this setting is the receiver operating characteristic (ROC) curve, which has long been popular in medical diagnostic research, particularly in radiology (Hanley 1989). In this article we consider new statistical methodology for making inference about ROC curves and apply the methods to data from a prostate cancer biomarker study.

Let D be a binary variable taking the value 1 for diseased subjects and 0 for nondiseased subjects. Let the variable Y denote the biomarker, and use the convention that higher values of Y are considered more indicative of disease. The ROC curve is motivated as follows: If a threshold value c is used to classify subjects as diseased or not on the basis of Y, then the true-positive and false-positive rates can be written as

$$S_{\mathrm{D}}(c) = P(Y \geq c | D = 1)$$

and

$$S_{\bar{\mathrm{D}}}(c) = P(Y \geq c | D = 0).$$

A perfect biomarker is one for which at some threshold c^*, $S_{\mathrm{D}}(c^*) = 1$ and $S_{\bar{\mathrm{D}}}(c^*) = 0$. More usually, there is a tradeoff between S_{D} and $S_{\bar{\mathrm{D}}}$ displayed through the ROC curve, a plot of the true-positive rates versus the false-positive rates, $\{(S_{\bar{\mathrm{D}}}(c), S_{\mathrm{D}}(c)),\ c \in (-\infty, \infty)\}$, or, equivalently, the function $\{(u, \mathrm{ROC}(u)), u \in (0, 1)\} = \{(u, S_{\mathrm{D}}(S_{\bar{\mathrm{D}}}^{-1}(u))), u \in (0, 1)\}$.

Higher values of the true-positive rate $S_{\mathrm{D}}(c)$ obtained by lowering the threshold are achieved at the expense of increasing the false-positive rate $S_{\bar{\mathrm{D}}}(c)$. Biomarkers that better discriminate disease from nondisease have ROC curves that are higher and to the left of the positive unit quadrant. Swets and Picket (1982), Hanley (1989), and, more recently, Pepe (2000b) have discussed the attributes of ROC curves for evaluating diagnostic markers of disease.

Regression models for ROC curves can be used to examine covariates that affect the discriminatory capacity of a biomarker. Covariates can include factors related to the environment in which the biomarker is measured, the protocol for obtaining and processing samples, subject characteristics, and even disease characteristics. It is important to assess factors that influence the performance of a biomarker to optimize use of the biomarker in practice.

We have previously proposed parametric ROC regression models (Pepe 1997, 2000a) of the generalized linear model (GLM) form,

$$\mathrm{ROC}_{\mathbf{x}}(u) = g\{\boldsymbol{\theta}'\mathbf{x} + h_{\boldsymbol{\alpha}}(u)\}, \qquad 0 \leq u \leq 1,$$

where $\mathbf{x}$ denotes covariates, g^{-1} is a link function, and h is a baseline function specified up to parameters $\boldsymbol{\alpha}$. Choices for g and $h_{\boldsymbol{\alpha}}$ used in our applications were $g = \Phi$ and $h_{\boldsymbol{\alpha}}(u) = \alpha_0 + \alpha_1 \Phi^{-1}(u)$, where Φ denotes the cumulative normal distribution function. This corresponds to extending the classic binormal model for the ROC curve (Metz 1986) to include covariates. The baseline function $h_{\boldsymbol{\alpha}}$ essentially defines the location and shape of the ROC curve, whereas $\boldsymbol{\theta}$ quantifies covariate effects. In this article we extend the regression models by allowing arbitrary nonparametric baseline functions for $h_{\boldsymbol{\alpha}}$. It turns out that the extra flexibility is gained at little cost in efficiency of estimating $\boldsymbol{\theta}$.

In Section 2 we describe the regression modeling framework and procedures for estimating $\boldsymbol{\theta}$ and the baseline ROC

Tianxi Cai is Assistant Professor (E-mail: *tcai@u.washington.edu*) and Margaret Sullivan Pepe is Professor (E-mail: *mspepe@u.washington.edu*), Department of Biostatistics, University of Washington, Seattle, WA 98195. The authors thank the editor, the associate editor, and the reviewers for valuable comments. Support for this research was provided by National Institutes of Health grants AI29168 and GM54438.

Journal of the American Statistical Association
December 2002, Vol. 97, No. 460, Theory and Methods
DOI 10.1198/00000000

function. We outline asymptotic distribution theory and procedures for making statistical inference in Section 3. We performed simulation studies to illustrate the advantage gained in terms of robustness of inference by allowing a nonparametric form for h and to determine whether the new semiparametric procedure has reduced efficiency relative to the parametric approach. We summarize results of the simulation studies in Section 4, and present an illustrative example in Section 5. The dataset is derived from a study to evaluate PSA as a biomarker for the early detection of prostate cancer. We give some closing remarks in Section 6.

2. ESTIMATION

Suppose that the data for analysis are organized as N_{D} data records for n_{D} subjects with disease,

$$\{(Y_{ik}, \mathbf{Z}_{ik}, \mathbf{Z}_{\mathrm{D}ik}), k = 1, \ldots, K_i,\ i = 1, \ldots, n_{\mathrm{D}}\},$$

and $N_{\bar{\mathrm{D}}}$ data records for $n_{\bar{\mathrm{D}}}$ subjects without disease,

$$\{(Y_{jl}, \mathbf{Z}_{jl}), l = 1, \ldots, K_j,\ j = n_{\mathrm{D}} + 1, \ldots, n_{\mathrm{D}} + n_{\bar{\mathrm{D}}}\},$$

where each subject may have more than one data record, $N_{\mathrm{D}} = \sum_{i=1}^{n_{\mathrm{D}}} K_i$ and $N_{\bar{\mathrm{D}}} = \sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} K_j$. The covariates denoted by $\mathbf{Z}$ are relevant to both diseased and nondiseased subjects. Examples might include the subject's age or the type of biomarker represented by Y (see Pepe 2000a, sec. 3.3). We include another category of covariates, denoted by $\mathbf{Z}_{\mathrm{D}}$, that are specific to subjects with disease but are not relevant to nondiseased subjects. Examples include be severity of disease or timing of biomarker measurement before onset of clinical symptoms (see Sec. 5).

The ROC curve that compares the biomarker distributions in diseased subjects who have covariates $(\mathbf{Z}, \mathbf{Z}_{\mathrm{D}})$ with nondiseased subjects who have covariate level $\mathbf{Z}$ is modeled as

$$\mathrm{ROC}_{\mathbf{Z},\mathbf{Z}_{\mathrm{D}}}(u) = g\{h_0(u) + \boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}_{\mathrm{D}}'\mathbf{Z}_{\mathrm{D}}\}, \qquad (1)$$

where g is a monotone increasing function mapping $(-\infty, \infty)$ to $(0, 1)$ and h_0 is an unspecified monotone increasing function from $(0, 1)$ to $(-\infty, \infty)$. We have previously noted that, conditional on the covariates $\{\mathbf{Z}, \mathbf{Z}_{\mathrm{D}}\}$ and $D = 1$, the expected value of $I\{Y \geq S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(u)\}$ is $\mathrm{ROC}_{\mathbf{Z},\mathbf{Z}_{\mathrm{D}}}(u) = g\{h_0(u) + \boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}_{\mathrm{D}}'\mathbf{Z}_{\mathrm{D}}\}$. This motivates the following class of estimating equations for $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}, \boldsymbol{\beta}_{\mathrm{D}})$ based on binary indicator variables (Pepe 1997):

$$\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik}, u)\mathbf{X}_{ik}\big[I\big\{Y_{ik} \geq S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(u)\big\} - g\big\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\big\}\big]\,d\hat{v}(u) = 0,$$

where the prespecified constants (a, b) are chosen such that $P\{Y_{11} < S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{11}}(a)\}$ and $P\{Y_{11} > S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{11}}(b)\}$ are positive; to simplify notation, we write $\mathbf{X} = [\mathbf{Z}', \mathbf{Z}_{\mathrm{D}}']'$, so that $\boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}_{\mathrm{D}}'\mathbf{Z}_{\mathrm{D}} = \boldsymbol{\theta}_0'\mathbf{X}$, w is a positive bounded uniformly continuous weight function; and for now we assume that $h_0(\cdot)$ and $S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(\cdot)$ are known and that $\hat{v}(\cdot)$ is a known increasing but possibly data-dependent function. In practice, neither $h_0(\cdot)$ nor $S^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(\cdot)$ is known and also need to be estimated.

Our semiparametric approach to estimation involves two steps. First, for $S_{\bar{\mathrm{D}},\mathbf{Z}}$ we use a semiparametric location model (Pepe 1998; Heagerty and Pepe 1999),

$$S_{\bar{\mathrm{D}},\mathbf{Z}}(c) = S_0(c - \boldsymbol{\gamma}_0'\mathbf{Z}).$$

We estimate the parameter $\boldsymbol{\gamma}_0$ as the solution to

$$\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j} \mathbf{Z}_{jl}(Y_{jl} - \boldsymbol{\gamma}'\mathbf{Z}_{jl}) = 0, \qquad (2)$$

which is denoted by $\hat{\boldsymbol{\gamma}}$, and the survivor function S_0 with the empirical distribution of the residuals

$$\widehat{S}_0(c) = \frac{1}{N_{\bar{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j} I(Y_{jl} - \hat{\boldsymbol{\gamma}}'\mathbf{Z}_{jl} \geq c). \qquad (3)$$

We then estimate the baseline ROC function h_0 and the parameter $\boldsymbol{\theta}_0$ simultaneously as solutions to

$$\sum_{i=1}^{n_{\bar{\mathrm{D}}}}\sum_{k=1}^{K_i} w(X_{ik}, u)\big[I\big\{Y_{ik} \geq \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(u)\big\} - g\big\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\big\}\big] = 0, \quad \text{for } u \in [a, b] \qquad (4)$$

and

$$\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik}, u)\mathbf{X}_{ik}\big[I\big\{Y_{ik} \geq \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(u)\big\} - g\big\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\big\}\big]\,d\hat{v}(u) = 0, \qquad (5)$$

where $\widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}}(u) = \widehat{S}_0^{-1}(u) + \hat{\boldsymbol{\gamma}}'\mathbf{Z}$.

Remark 1. Although biomarkers are typically measured on continuous scales, the methodology is equally applicable to biomarkers with discrete or ordinal distributions. Such data often arise in diagnostic radiology and in psychology studies, where ROC analysis is already a well-accepted approach to evaluating new procedures.

Remark 2. It follows from (4) and the monotone increasing property of the function g that the estimator $\hat{h}$ is a monotone increasing function. In this article we assume that $h_0(\cdot)$ has a continuous first derivative.

Remark 3. In practice, let $a \leq u_1 < u_2 < \cdots < u_L \leq b$ be the set of distinct jump points of $\widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u)$ on $[a, b]$. Also, let $\hat{v}_l = \hat{v}(t_{l+1}) - \hat{v}(t_l)$ and $h_l = h(u_l)$. Then solving the two sets of estimating equations is equivalent to solving the $p + L$ equations

$$\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i} w(X_{ik}, u_l)\big[I\big\{Y_{ik} \geq \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u_l)\big\} - g(\boldsymbol{\theta}'\mathbf{X}_{ik} + h_l)\big] = 0, \quad l = 1, \ldots, L$$

and

$$\sum_{l=1}^{L}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u_l)\mathbf{X}_{ik}\big[I\big\{Y_{ik} \geq \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u_l)\big\} - g(\boldsymbol{\theta}'\mathbf{X}_{ik} + h_l)\big]\hat{v}_l = 0.$$

The Newton–Raphson algorithm can be used to solve the foregoing $p+L$ equations. It is not hard to see that in this case, inverting the $(p+L)\times(p+L)$ Jacobian matrix can be obtained by inverting a diagonal $L\times L$ matrix and a $p\times p$ matrix.

Remark 4. The covariates $\mathbf{Z}$ and $\mathbf{Z}_{\mathrm{D}}$ are assumed to be bounded. If the covariate $\mathbf{Z}$ takes on only a few values, a completely nonparametric approach to estimating $S_{\bar{\mathrm{D}},\mathbf{Z}}$ may be preferred. For example, if the problem is to compare two biomarkers and $\mathbf{Z}$ is an indicator of the type of biomarker quantified by Y, then $\widehat{S_{\bar{\mathrm{D}},\mathbf{Z}}}(\cdot)$ can be estimated separately for each $\mathbf{Z}$ using the empirical survivor function for the corresponding biomarker among the nondiseased subjects.

Remark 5. A parametric formulation for $h_0(u)$, such as $h_0(u)=\alpha_1+\alpha_2\Phi^{-1}(u)$ say, can be accommodated by replacing (4) with an equation identical in form to (5) with $[1\Phi^{-1}(u)]'$ substituted for $\mathbf{X}_{ik}$ in the integrand. Then (4) and (5) are solved simultaneously. (See Pepe (1997) for this parametric approach.)

Remark 6. As indicated by our notation, the data records may be clustered. For example, if multiple biomarkers are measured on a subject or if the same biomarker is measured at different times, then each subject may have several data records in the analysis. The probability distributions and consequent ROC curves are defined in a marginal sense. Therefore, our estimating equations provide valid estimates for clustered data. In the next section we account for correlation between data records in the same cluster in making statistical inference.

Remark 7. The function $\hat{v}(u)$ can be data dependent, but we assume that it converges to a deterministic function uniformly in u. For example, we can choose $\hat{v}(u)$ to be the counting process $\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i} I(Y_{ik}\geq \widehat{S}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}^{-1}(u))$. We can also restrict $\hat{v}(u)$ to be 0 in certain ranges of u that might not be of particular interest. For example, if large false-positive rates are not of particular interest, we might model the ROC curve only over a restricted region, say, $u\leq 20\%$. In that case, we choose $\hat{v}(u)=0$ for $u>.2$. (See Weiand, Gail, James, and James 1989 and Pepe 1998 for further discussion about restricting inference about ROC curves to clinically relevant ranges of false-positive rates.) We let the weight function $w(x,u)$ be 1 in the analyses presented in this article, although more general choices are possible. Note that we do not allow w to be dependent on $\boldsymbol{\theta}$ or h, a condition necessary to guarantee the uniqueness of the solution $\{\hat{\boldsymbol{\theta}},\hat{h}(u)\}$.

Remark 8. The form of the ROC model $\mathrm{ROC}_{\mathbf{x}}(u)=g\{h_0(u)+\boldsymbol{\theta}_0'\mathbf{x}\}$ suggests that the covariate effect is the same at all false-positive rates u. This can be relaxed by including interactions between functions of u and $\mathbf{x}$ in the regression model. In general, we can write

$$\mathrm{ROC}_{\mathbf{x}}(u)=g\{\boldsymbol{\theta}_0'\mathbf{x}+\boldsymbol{\alpha}_1'\eta_1(u)\mathbf{x}+\cdots+\boldsymbol{\alpha}_q'\eta_q(u)\mathbf{x}+h_0(u)\},$$

where $\eta_k(u)$ are specified functions. This is analogous to the standard procedure for relaxing the proportional hazards assumption in the Cox model for survival data, where covariate effects can change with time by including interactions between time and the covariate in the relative risk function. Our estimation procedures are appropriate for the more general model as well.

3. INFERENCE IN LARGE SAMPLES

3.1 Asymptotic Properties of $\hat{\boldsymbol{\theta}}$

Let $\{\hat{\boldsymbol{\theta}},\hat{h}(u)\}$ denote the solution to (4) and (5) and let $\hat{h}(u,\boldsymbol{\theta})$ denote the solution to (4) for any given $\boldsymbol{\theta}$. We show in Appendix A that

Theorem 1. $\{\hat{\boldsymbol{\theta}},\hat{h}(u)\}$ are unique for large n and are consistent for $u\in[a,b]$, where $0<a<b<1$.

To obtain large-sample distributions for $\{\hat{\boldsymbol{\theta}},\hat{h}(u)\}$, we first show in Appendix B that

Lemma 1. $n_{\bar{\mathrm{D}}}^{\frac{1}{2}}[S_{\bar{\mathrm{D}},\mathbf{z}}\{\widehat{S}_{\bar{\mathrm{D}},\mathbf{z}}^{-1}(u)\}-u]$ is asymptotically equivalent to

$$\widetilde{\mathcal{P}}(u,\mathbf{z})=-\frac{n_{\bar{\mathrm{D}}}^{\frac{1}{2}}}{N_{\bar{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j}\Big[I\big\{Y_{jl}-\boldsymbol{\gamma}_0'\mathbf{Z}_{jl}>S_0^{-1}(u)\big\}-u$$
$$+\big\{(\mathbb{Z}_{\bar{\mathrm{D}},1}-\mathbf{z})'\mathbb{Z}_{\bar{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}\big\}(Y_{jl}-\boldsymbol{\gamma}_0'\mathbf{Z}_{jl})\dot{S}_0\big\{S_0^{-1}(u)\big\}\Big],$$

where for any function $f(x)$, $\dot{f}(x)$ denotes $df(x)/dx$, $\mathbb{Z}_{\bar{\mathrm{D}},1}$ is the limit of $\overline{\mathbf{Z}}_{\bar{\mathrm{D}},1}=\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j}\mathbf{Z}_{jl}/n_{\bar{\mathrm{D}}}$, $\mathbb{Z}_{\bar{\mathrm{D}},2}$ is the limit of $\overline{\mathbf{Z}}_{\bar{\mathrm{D}},2}=\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j}\mathbf{Z}_{jl}\mathbf{Z}_{jl}'/n_{\bar{\mathrm{D}}}$, and $\widetilde{\mathcal{P}}(u,\mathbf{z})$ converges in distribution to a Gaussian process.

To obtain interval estimates of specific components of $\boldsymbol{\beta}_0$, in Appendix C we show that

Theorem 2. $n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is asymptotically equivalent to

$$n_{\mathrm{D}}^{-\frac{1}{2}}\mathbb{A}^{-1}\left\{\sum_{i=1}^{n_{\mathrm{D}}}U_i(\boldsymbol{\theta}_0)+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}U_j(\boldsymbol{\theta}_0)\right\},$$

where $\mathbb{A}$ is defined in (A.7),

$$U_i(\boldsymbol{\theta}_0)=\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik},u)\{\mathbf{X}_{ik}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}e_{ik}(u)\,dv(u)$$
$$\text{for } i=1,\ldots,n_{\mathrm{D}},$$

$$U_j(\boldsymbol{\theta}_0)=-\frac{n_{\mathrm{D}}}{N_{\bar{\mathrm{D}}}}\sum_{l=1}^{K_j}\int_a^b\mathbb{B}(u)\mathbb{Z}_{\bar{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}(Y_{jl}-\boldsymbol{\gamma}_0'\mathbf{Z}_{jl})$$
$$\times\dot{S}_0\big\{S_0^{-1}(u)\big\}\,dv(u)\quad\text{for } j=n_{\mathrm{D}}+1,\ldots,n_{\mathrm{D}}+n_{\bar{\mathrm{D}}},$$

$\mathbb{X}(u;\boldsymbol{\theta})$ is the limit of $\overline{\mathbf{X}}(u;\boldsymbol{\theta})$, defined in (A.5), $e_{ik}(u)=I\{Y_{ik}>S_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}^{-1}(u)\}-g\{h_0(u)+\boldsymbol{\theta}_0'\mathbf{X}_{ik}\}$, $v(u)$ is the limit of $\hat{v}(u)$, $a_{ik}(u)$ is defined in (A.8), and $\mathbb{B}(u)$ is the limit of $\widehat{\mathbb{B}}(u)=n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\{\mathbf{X}_{ik}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}a_{ik}(u)(\mathbb{Z}_{\bar{\mathrm{D}},1}-\mathbf{Z}_{ik})'\dot{h}_0(u)$.

Therefore, the distribution of $n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ can be approximated by a normal random vector with mean $\mathbf{0}$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}=n_{\mathrm{D}}^{-1}\widehat{\mathbb{A}}^{-1}(\hat{\boldsymbol{\theta}})\{\sum_{i=1}^{n_{\mathrm{D}}}\widehat{U}_i(\hat{\boldsymbol{\theta}})\widehat{U}_i(\hat{\boldsymbol{\theta}})'+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\widehat{U}_j(\hat{\boldsymbol{\theta}})\widehat{U}_j(\hat{\boldsymbol{\theta}})'\}\widehat{\mathbb{A}}^{-1}(\hat{\boldsymbol{\theta}})$, where $\widehat{\mathbb{A}}^{-1}(\boldsymbol{\theta})$ is defined in

(A.6) and $\widehat{U}_i$ and $\widehat{U}_j$ are obtained by replacing all theoretical quantities in U_i and U_j by their empirical counterparts. One possible approach to obtain an estimate for $\dot{S}_0(\cdot)$ and $\dot{h}_0(\cdot)$ is to use $\hat{\dot{S}}_0(c) = \frac{1}{b_s}\int_{c_1}^{c_2} \mathcal{K}_s(\frac{c-x}{b_s})\, d\widehat{S}_0(x)$ and $\hat{\dot{h}}(u) = \frac{1}{b_h}\int_{u_1}^{u_2} \mathcal{K}_h(\frac{u-x}{b_h})\, d\hat{h}(x)$, where b_s and b_h are some positive bandwidth parameters and the kernel functions $\mathcal{K}_s$ and $\mathcal{K}_h$ are bounded functions with support $[-1, 1]$ and that integrate to 1. The Epanechnikov kernel function $K(x) = .75(1-x^2)\times I(|x| \le 1)$ was used for the kernel estimates of both $\dot{S}_0(\cdot)$ and $\dot{h}_0(\cdot)$ through our numerical studies. The bandwidth was chosen to be $b_s = b_h = 4/\max\{\min(n_{\mathrm{D}}, n_{\overline{\mathrm{D}}})^{4/5}, 50\}$.

3.2 Estimating the Receiver Operating Characteristic Curve

Based on $\{\hat{\boldsymbol{\theta}}, \hat{h}(u)\}$, the ROC curve for a test with covariates $\mathbf{x}$ can be estimated by $\widehat{\mathrm{ROC}}(u; \mathbf{x}) = g\{\hat{\boldsymbol{\theta}}'\mathbf{x} + \hat{h}(u)\}$. To obtain the distribution of $\widehat{\mathrm{ROC}}(u; \mathbf{x})$, in Appendix D we prove the following theorem.

Theorem 3. $Q(u; \mathbf{x}) = n_{\mathrm{D}}^{\frac{1}{2}}[g^{-1}\{\widehat{\mathrm{ROC}}(u; \mathbf{x})\} - g^{-1}\{\mathrm{ROC}(u; \mathbf{x})\}]$ has the same asymptotic distribution as

$$\widetilde{Q}(u; \mathbf{x}) = n_{\mathrm{D}}^{-\frac{1}{2}}\left\{\sum_{i=1}^{n_{\mathrm{D}}} \widetilde{Q}_i(u; \mathbf{x}) + \sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}} \widetilde{Q}_j(u; \mathbf{x})\right\},$$

and $\widetilde{Q}(u; \mathbf{x})$ converges weakly to a zero-mean Gaussian process, where $\widetilde{Q}_i(u; \mathbf{x}) = \{\mathbf{x} - \mathbb{X}(u; \boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}U_i(\boldsymbol{\theta}_0) + \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)e_{ik}(u)/a(u)$,

$$\begin{aligned}\widetilde{Q}_j(u; \mathbf{x}) = {}& \{\mathbf{x} - \mathbb{X}(u; \boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}U_j(\boldsymbol{\theta}_0) \\ & - \frac{n_{\mathrm{D}}}{N_{\overline{\mathrm{D}}}}\dot{h}_0(u)\sum_{l=1}^{K_j}\Bigg[I\{Y_{jl} - \boldsymbol{\gamma}_0{}'\mathbf{Z}_{jl} > S_0^{-1}(u)\} - u \\ & + \frac{a_{\mathbf{z}}(u)'}{a(u)}\mathbb{Z}_{\overline{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}(Y_{jl} - \boldsymbol{\gamma}_0'\mathbf{Z}_{jl})\dot{S}_0\{S_0^{-1}(u)\}\Bigg],\end{aligned}$$

$a(u)$ is the limit of $n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i} a_{ik}(u)$, and $a_{\mathbf{z}}(u)$ is the limit of $n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i} a_{ik}(u)(\mathbb{Z}_{\overline{\mathrm{D}},1} - \mathbf{Z}_{ik})$.

In practice, to approximate the distribution of $\widehat{Q}(u; \mathbf{x})$, we consider the following *resampling* technique (Parzen, Wei, and Ying 1994). First, let

$$\widehat{Q}(u; \mathbf{x}) = n_{\mathrm{D}}^{-\frac{1}{2}}\left\{\sum_{i=1}^{n_{\mathrm{D}}} \widehat{Q}_i(u; \mathbf{x})\mathcal{G}_i + \sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}} \widehat{Q}_j(u; \mathbf{x})\mathcal{G}_j\right\}, \qquad (6)$$

where $\widehat{Q}_i(u; \mathbf{x})$ and $\widehat{Q}_j(u; \mathbf{x})$ are obtained by replacing all of the theoretical quantities in $\widetilde{Q}_i(u; \mathbf{x})$ and $\widetilde{Q}_j(u; \mathbf{x})$ by their empirical counterparts. Conditional on the observations $\{(Y_{ik}, \mathbf{Z}_{ik}, \mathbf{Z}_{\mathrm{D}ik}), k = 1, \ldots, K_i; i = 1, \ldots, n_{\mathrm{D}} + n_{\overline{\mathrm{D}}}\}$, the process $\widehat{Q}(u; \mathbf{x})$ has the same limiting covariance function as that of $\widetilde{Q}(u; \mathbf{x})$. Furthermore, conditional on the data, the process $\widehat{Q}(u; \mathbf{x})$ is *tight* (Billingsley 1986). It follows that the distribution of $Q(u; \mathbf{x})$ can be approximated by the conditional distribution of $\widehat{Q}(u; \mathbf{x})$.

We can then approximate the distribution of $Q(u; \mathbf{x})$ by generating M independent samples of $\{\mathcal{G}_i, i = 1, \ldots, n_{\mathrm{D}} + n_{\overline{\mathrm{D}}}\}$. For the mth sample, we obtain a realization $\hat{q}_{(m)}(u; \mathbf{x})$ of $\widehat{Q}(u; \mathbf{x})$, $m = 1, \ldots, M$. For any given u, we can use

$$\hat{\sigma}^2(u|\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} \hat{q}_{(m)}(u; \mathbf{x})^2$$

to estimate the variance of $Q(u; \mathbf{x})$. Then a $(1-\zeta)$ pointwise confidence interval for $\mathrm{ROC}(u; \mathbf{x})$ is given by

$$g\big[g^{-1}\big\{\widehat{\mathrm{ROC}}(u; \mathbf{x})\big\} \pm c_{\zeta/2}n_{\mathrm{D}}^{-1/2}\hat{\sigma}(u; \mathbf{x})\big],$$

where c_ζ is the 100ζ upper percentile point of the standard normal distribution. To construct a $(1-\zeta)$ simultaneous confidence band for $\{\mathrm{ROC}(u; \mathbf{x}), a \le u \le b\}$, we first find d_ζ such that

$$\mathrm{pr}\left\{\sup_{u\in[a,b]} \frac{|\widehat{Q}(u; \mathbf{x})|}{\hat{\sigma}(u; \mathbf{x})} \le d_\zeta\right\} = 1 - \zeta.$$

Then a $(1-\zeta)$ confidence band for $\mathrm{ROC}(u; \mathbf{x})$ for $a \le u \le b$ is given by

$$g\big[g^{-1}\big\{\widehat{\mathrm{ROC}}(u; \mathbf{x})\big\} \pm d_\zeta n_{\mathrm{D}}^{-1/2}\hat{\sigma}(u; \mathbf{x})\big].$$

The probability and the quantile d_ζ can be approximated with these M realizations of $\widehat{Q}(u; \mathbf{x})$.

4. SIMULATION STUDIES

4.1 Flexible Modeling Confers Robustness

We would expect the newly proposed semiparametric procedure to be more robust than a procedure that specifies a parametric form for $h_0(\cdot)$. In particular, we would expect inference about covariate effects and the baseline ROC curve to be more reliable in the model that does not require specification of a parametric form for $h_0(\cdot)$. To investigate this, we simulate data and fit a misspecified parametric model to these data. For n_{D} diseased subjects, we generate random variables z and ϵ following uniform (0, 10) and standard normal distributions, and let

$$Y_i = \psi(\beta z_i + \epsilon_i) + 2z_i, \qquad i = 1, \ldots, n_{\mathrm{D}}.$$

Similarly, we generate z and ϵ for $n_{\overline{\mathrm{D}}}$ diseased subjects and construct

$$Y_j = 6 + 2z_j + \epsilon_j, \qquad j = n_{\mathrm{D}} + 1, \ldots, n_{\mathrm{D}} + n_{\overline{\mathrm{D}}},$$

where $\psi(x) = 6 + \log\{-6\log\Phi(-x)\}/2$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\beta = 1$. The induced ROC curve is then

$$\mathrm{ROC}(u; T, z) = \Phi[-\psi^{-1}\{6 - \Phi^{-1}(u)\} + \beta z],$$

which follows the GLM model with

$$h_0(u) = -\psi^{-1}\{6 - \Phi^{-1}(u)\}.$$

We fit the following parametric model to the data, using the method of Pepe (1997):

$$\mathrm{ROC}_z(u) = \Phi\{\alpha_0 + \alpha_1\Phi^{-1}(u) + \beta z\},$$

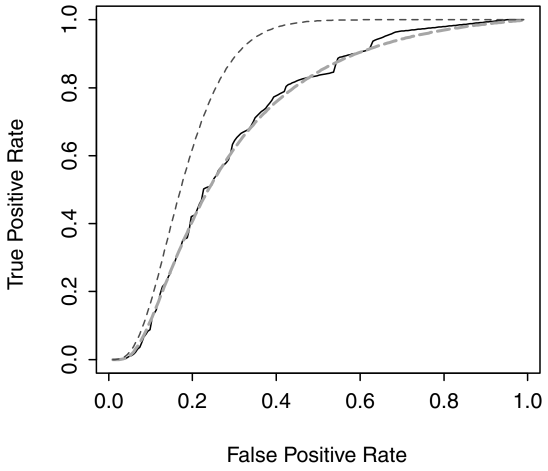

Figure 1. Estimated ROC Curve at Baseline Using the Semiparametric (———) and the Misspecified Parametric Approaches (- - - -) (– – –, Truth).

thus misspecifying the baseline function as having the form $h_0(t) = \alpha_0 + \alpha_1 \Phi^{-1}(u)$. The semiparametric procedure described in this article is also implemented. Figure 1 displays fitted baseline ROC curves (at $z = 0$), based on one simulated dataset. The semiparametric method provides an estimated ROC curve much closer to the truth than the parametric approach, suggesting that it is more robust.

Table 1 presents the empirical bias and mean squared error of $\hat{\boldsymbol{\beta}}$ based on 500 simulated datasets. The parametric estimator is severely biased on average. In contrast, the semiparametric approach yields an estimator with essentially no bias. This example, albeit somewhat extreme, demonstrates that the less-restrictive model assumptions made in the semiparametric approach confer some robustness to the estimates of both the ROC curve and of the covariate effect. This robustness is particularly important given that formal procedures are not yet available for checking the fit of the parametric model. We next investigate whether the added robustness in the semiparametric approach is gained at the expense of reduced statistical efficiency.

4.2 Relative Efficiency

We simulated data to include both a covariate common to diseased and nondiseased subjects and a covariate relevant only for diseased subjects. These are denoted by z and T. In particular, data for diseased subjects were generated from the linear model

$$Y_i = 12 + \beta_t T_i + (2 + \beta_z) z_i + \epsilon_i, \qquad i = 1, \ldots, n_{\mathrm{D}},$$

and data for diseased subjects followed the model

$$Y_j = 10 + 2 z_j + \epsilon_j, \qquad j = n_{\mathrm{D}} + 1, \ldots, n_{\mathrm{D}} + n_{\bar{\mathrm{D}}},$$

where ϵ, T, and z are random variables with probability distribution that are standard normal, exponential (rate = 1) and Bernoulli (probability = .5). This configuration for the data induces the ROC curve

$$\mathrm{ROC}(u; T, z) = \Phi\{\Phi^{-1}(u) + 2 + \beta_t t + \beta_z z\}.$$

We use an ROC curve model of this form for the PSA data analysis (Sec. 5), where z represents subject age and T represents the time between serum sampling for the PSA biomarker and onset of clinical symptoms of cancer.

Table 1. Estimates of β Compared With Its Actual Value, β = 1, Based on a Semiparametric Model and on a Misspecified Parametric Model

	Semiparametric	*Parametric*
Bias	.031	.128
Mean squared error	.019	.042

NOTE: Results are based on 500 simulated datasets of $n_{\mathrm{D}} = n_{\bar{\mathrm{D}}} = 200$.

Table 2. Estimates of Covariate Effects From Correctly Specified Parametric and Semiparametric Models

		Semiparametric		*Parametric*	
$n_{\mathrm{D}} = n_{\bar{\mathrm{D}}}$		$\beta_t = -1$	$\beta_z = 3$	$\beta_t = -1$	$\beta_z = 3$
100	Bias	.082	.375	.055	.320
	Mean squared error	.056	1.522	.047	1.486
200	Bias	.027	.108	.012	.082
	Mean squared error	.019	.184	.017	.170

For each simulated dataset, we obtained point estimates of β_t and β_z with our semiparametric approach and the parametric approach of Pepe (1997). Table 2 presents the bias and mean squared error based on 500 simulations.

The results in Table 2 show that even though estimates from the semiparametric model are not as efficient as those calculated using the fully parametric approach, their efficiency is very close. The empirical efficiency of the semiparametric method relative to the parametric method is 95% for β_z and 90% for β_t at a sample size of 200.

4.3 Asymptotic Inference in Finite Samples

We also conducted simulation studies to examine the validity of the large-sample approximations for making inference in finite sample sizes. We simulated 1,000 sets of data with $n_{\mathrm{D}} = n_{\bar{\mathrm{D}}} = n = 50$, 100, 200, and 400 from the same models described in Section 4.2. Here we let $\beta_t = -1$ and $\beta_z = 2$. Table 3 presents the bias, average of the standard error estimators, the sampling standard error, and the coverage probability of the 95% confidence intervals for β_t and β_z. The standard error estimator are reasonably close to the true sampling standard errors, at least for sample size $n \geq 100$. In addition, for confidence intervals, the empirical coverage probabilities are close to their nominal counterparts. For a small sample, $n = 50$, it appears that the estimated standard errors based on large-sample approximations tend to be smaller than the sampling standard errors. The bootstrap standard error may be used instead when the sample size is small.

Table 3. Bias, Average of the Standard Error Estimator (Ave($\widehat{SE}$)), Standard Error (SE), and the Coverage Probability (Coverage) of the 95% Confidence Interval

	$\beta_t = -1$				$\beta_z = 2$			
$n_D = n_{\bar{D}}$	*Bias*	*Ave($\widehat{SE}$)*	*SE*	*Coverage*	*Bias*	*Ave($\widehat{SE}$)*	*SE*	*Coverage*
50	.101	.256	.316	.914	.245	.604	.828	.908
100	.034	.161	.175	.932	.072	.397	.433	.932
200	.026	.113	.112	.947	.052	.267	.280	.941
400	.014	.078	.078	.945	.010	.183	.186	.945

NOTE: Each entry is based on 1,000 simulation samples.

5. EXAMPLE: EARLY DETECTION OF PROSTATE CANCER WITH SERUM PROSTATE-SPECIFIC ANTIGEN LEVELS

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 in the Beta-Carotene and Retinol Efficacy Trial (CARET) (Thornquist et al. 1993), in an effort to quantify the capacity of PSA for discriminating men with prostate cancer from those without before the onset of clinical symptoms.

Briefly, CARET enrolled more than 12,000 men and randomized them to intervention or placebo to prevent lung cancer. As part of the protocol for the trial, serum was periodically drawn and stored from study participants. All 88 subjects who developed prostate cancer over the course of the trial were included in the PSA case-control study. Serum samples stored before diagnosis with cancer were analyzed for PSA. An age-matched set of 88 control subjects also had their stored serum samples analyzed for PSA. Etzioni, Pepe, Longton, Hu, and Goodman (1999) have provided a complete description of the study design and plots of the data.

Increasing age is associated with increasing serum PSA level and can potentially affect the discriminatory capacity of PSA. Thus we used $z = age$ (years) as a covariate in our ROC model for PSA. In addition, among subjects who develop cancer, it is likely that PSA measured closer to the time of onset of clinical symptoms is more predictive of disease than are measures taken earlier in time. We included a covariate T, defined as the time between the onset of symptoms and the time at which the serum sample was drawn. We then fit the following model to the data:

$$\mathrm{ROC}_{T,z}(u) = \Phi\{h_0(u) + \beta_t T + \beta_z z\}.$$

Using our semiparametric approach, the estimate of β_t is $-.120$ per year with standard error .041, and the coefficient for age, β_z, is estimated as $-.014$ per year of age with standard error .020. The negative coefficient for T implies that discrimination improves as T decreases, that is, when PSA is measured closer to diagnosis. The negative coefficient for age suggests that discrimination is better in younger men than in older men, but the evidence is not conclusive (p value $=$ .484). Figure 2 shows the estimated ROC curves and their 95% confidence bands at different times for patients who were 60 years old when the serum for measuring PSA was obtained. Also shown are curves estimated using the parametric binormal model

$$\mathrm{ROC}_{T,z}(u) = \Phi\{\alpha_0 + \alpha_1 \Phi^{-1}(u) + \beta_t T + \beta_z z\}.$$

Observe that the curves are similar for the parametric and semiparametric methods. The regression coefficients and their estimated standard errors for the parametric method in this example are almost identical to the semiparametric ones, $\hat{\beta}_t = -.119$, $se(\hat{\beta}_t) = .041$ and $\hat{\beta}_z = -.014$, $se(\hat{\beta}_z) = .019$. It appears that the binormal model does fit the data adequately in this example and that the semiparametric methods fit the model with efficiency comparable to that of the fully parametric approach.

6. DISCUSSION

This article extends the parametric ROC regression method of Pepe (1997, 2000a) to a semiparametric approach. The reductions in requirements for model specification and increased robustness are attractive features. Other approaches to ROC regression have been proposed. One popular method is to model the distributions of test results (Tosteson and Begg 1998; Pepe 1998). Covariates whose associations with Y differ in diseased and non-diseased populations (i.e., interactions with D) induce effects on the ROC curve. Another approach is to model a summary measure of the ROC curve and to use derived variables, estimated summary measures, for fitting regression models (Thompson and Zucchini 1989; Dorfman, Berbaum, and Metz 1992). Pepe (1998) contrasted these regression models with those considered in this article that directly model the ROC curve itself. Our preference is for this latter approach, primarily because it directly models the quantities of interest.

Asymptotic distribution theory has been derived. In contrast to earlier work (Pepe 1997), the theory allows clustered data which in practice arises frequently, as evidenced by our two examples. We proposed a simulation method for making inference about ROC curves. This technique avoids the need to derive explicit analytic expressions for variance-covariance processes, which seem intractable in our setting. Moreover, relative to other resampling methods such as the bootstrap, the computational burden is minimal.

The application presented in this article concerns a biomarker for prostate cancer. We used a probit link function in our model,

$$\mathrm{ROC}_x(u) = g\{h_0(u) + \theta x\},$$

where $g = \Phi$. Other choices of link function might be preferred. A logistic link allows the interpretation of $\exp(\theta)$ as

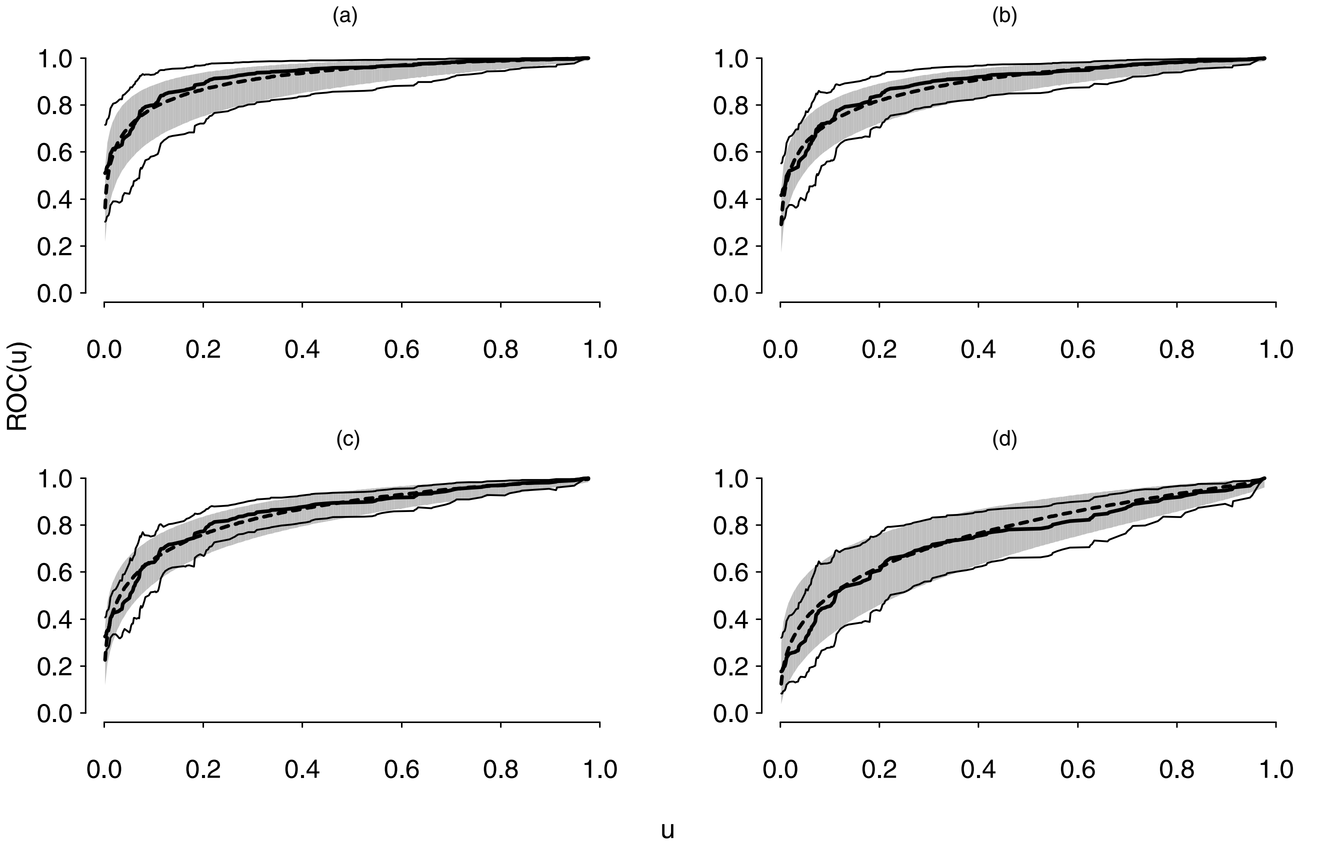

Figure 2. Estimated ROC Curve Using Semiparametric (———) and Parametric (– – –) Approaches for PSA as a Biomarker of Prostate Cancer in 60-Year-Old Men. Shown also are 95% Confidence Bands (thin solid lines, semiparametric; shaded regions, parametric) (a) Serum sample taken at diagnosis; (b) Serum sample taken 2 years before diagnosis; (c) Serum sample taken 4 years before diagnosis; (d) Sample taken 8 years before diagnosis.

an odds ratio, for example. A log link implies that $\exp(\theta)$ is the increase in the true-positive rate per unit increase in x when the false-positive rate is held constant (Pepe 1997). These interpretations do not depend on the baseline function h_0, and this method now enables inference about $\boldsymbol{\theta}$ without specifying a form for h_0.

APPENDIX A: PROOF OF THEOREM 1

The strong law of large numbers ensures consistency of $\hat{\boldsymbol{\gamma}}$ and that $\widehat{S}_0^{-1}(u)$ converges to $S_0^{-1}(u)$ uniformly in $u \in [a, b]$. It follows that $\widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{z}}(u) = \widehat{S}_0^{-1}(u) + \hat{\boldsymbol{\gamma}}'\mathbf{z}$ converges to $S^{-1}_{\bar{\mathrm{D}},\mathbf{z}}(u)$ uniformly in $u \in [a, b]$ and $\mathbf{z} \in D^z_\eta = \{\mathbf{Z} : \|\mathbf{Z}\| \le \eta\}$ for any $\eta \ge 0$, almost surely, as $n \to \infty$. It then follows from the uniform law of large numbers that for any $\eta \ge 0$, $\epsilon > 0$, uniformly in $u \in [a, b]$ and $\boldsymbol{\theta} \in D^\theta_\eta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \eta\}$,

$$\frac{1}{n_{\mathrm{D}}} \sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\big[I\big\{Y_{ik} > \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u)\big\} - g\{h_0(u) + \boldsymbol{\theta}'\mathbf{X}_{ik} - \epsilon\}\big]$$
$$\to k_{\mathrm{D}} E\big[w(\mathbf{X}_{1k}, u)\{g(h_0(u) + \boldsymbol{\theta}_0\mathbf{X}_{1k}) - g(h_0(u) + \boldsymbol{\theta}\mathbf{X}_{1k} - \epsilon)\}\big]$$

almost surely as $n_{\mathrm{D}} \to \infty$ and $n_{\bar{\mathrm{D}}} \to \infty$, where k_{D} is the limit of $\sum_{i=1}^{n_{\mathrm{D}}} K_i / n_{\mathrm{D}}$. It follows that for large n_{D} and $n_{\bar{\mathrm{D}}}$, $u \in [a, b]$, $\boldsymbol{\theta} \in D^\theta_\eta$, and sufficiently large $\boldsymbol{\epsilon}$,

$$\frac{1}{n_{\mathrm{D}}} \sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\big[I\big\{Y_{ik} > \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u)\big\} - g\{h_0(u) + \boldsymbol{\theta}'\mathbf{X}_{ik} - \epsilon\}\big] > 0 \tag{A.1}$$

and

$$\frac{1}{n_{\mathrm{D}}} \sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\big[I\big\{Y_{ik} > \widehat{S}^{-1}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u)\big\} - g\{h_0(u) + \boldsymbol{\theta}'\mathbf{X}_{ik} + \epsilon\}\big] < 0. \tag{A.2}$$

This, coupled with the monotonicity and continuity of g for large n_{D} and $n_{\bar{\mathrm{D}}}$, $u \in [a, b]$, and $\boldsymbol{\theta} \in D^\theta_\eta$, implies that there exists a unique $\hat{h}(u; \boldsymbol{\theta})$ such that

$$\frac{1}{n_{\mathrm{D}}} \sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\big[I\big\{Y_{ik} > \widehat{S}_{\bar{\mathrm{D}},\mathbf{Z}_{ik}}(u)\big\} - g\{\hat{h}(u; \boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{X}_{ik}\}\big] = 0. \tag{A.3}$$

By differentiating both sides of (A.3) with respect to $\boldsymbol{\theta}$, we obtain the identity

$$-\frac{\partial}{\partial \boldsymbol{\theta}} \hat{h}(u; \boldsymbol{\theta}) = \overline{\mathbf{X}}(u; \boldsymbol{\theta}), \tag{A.4}$$

where

$$\overline{\mathbf{X}}(u; \boldsymbol{\theta}) = \frac{\sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\dot{g}\{\hat{h}(u; \boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{X}_{ik}\}\mathbf{X}_{ik}}{\sum_{i=1}^{n_{\mathrm{D}}} \sum_{k=1}^{K_i} w(\mathbf{X}_{ik}, u)\dot{g}\{\hat{h}(u; \boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{X}_{ik}\}}. \tag{A.5}$$

To show the existence and uniqueness of $\hat{\boldsymbol{\theta}}$, we let $V(\boldsymbol{\theta})$ be the left side of (5), with $h_0(u)$ replaced by $\hat{h}(t, \boldsymbol{\theta})$. It follows from (A.4) that

$\frac{1}{n_{\mathrm{D}}}\frac{\partial}{\partial\boldsymbol{\theta}}V(\boldsymbol{\theta}) = -\widehat{\mathbb{A}}(\boldsymbol{\theta})$, where

$$\widehat{\mathbb{A}}(\boldsymbol{\theta}) = \frac{1}{n_{\mathrm{D}}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik}, u)\{\mathbf{X}_{ik} - \overline{\mathbf{X}}(u; \boldsymbol{\theta})\}^{\otimes 2} \times \dot{g}\{\hat{h}(u; \boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{X}_{ik}\}\, d\hat{v}(u), \qquad \text{(A.6)}$$

for any vector or matrix $\mathbf{b}$, $\mathbf{b}^{\otimes 0} = 1$, $\mathbf{b}^{\otimes 1} = \mathbf{b}$, and $\mathbf{b}^{\otimes 2}$ is defined as $\mathbf{b}\mathbf{b}'$. Furthermore, because only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ do (A.1) and (A.2) hold for any $\boldsymbol{\epsilon} > 0$, we have that $\hat{h}(u; \boldsymbol{\theta}_0) \to h_0(u)$ uniformly in $u \in [a, b]$ and $\frac{1}{n_{\mathrm{D}}}\frac{\partial}{\partial\boldsymbol{\theta}}V(\boldsymbol{\theta}_0) \to -\mathbb{A}$, where

$$\mathbb{A} = E\int_a^b \{\mathbf{X}_{ik} - \mathbb{X}(u; \boldsymbol{\theta}_0)\}^{\otimes 2} a_{ik}(u)\, dv(u), \qquad \text{(A.7)}$$

$$a_{ik}(u) = w(\mathbf{X}_{ik}, u)\dot{g}\{h_0(u) + \boldsymbol{\theta}_0'\mathbf{X}_{ik}\}, \qquad \text{(A.8)}$$

and $\mathbb{X}(u; \boldsymbol{\theta})$ is the limit of $\overline{\mathbf{X}}(u; \boldsymbol{\theta})$. It is easy to see that $\mathbb{A}$ is positive definite.

Now, because $\frac{1}{n}V(\boldsymbol{\theta}_0) \to 0$, by the standard inverse function theorem there exists a unique solution $\hat{\boldsymbol{\theta}}$ to the equation $V(\boldsymbol{\theta}) = 0$ in a neighborhood of $\boldsymbol{\theta}_0$. This, coupled with the nonnegativity of $\widehat{\mathbb{A}}(\boldsymbol{\theta})$ for large n, ensures uniqueness of the root of $V(\boldsymbol{\theta}) = 0$ in the entire domain of $\boldsymbol{\theta}$ asymptotically. The foregoing proof also implies that $\hat{\boldsymbol{\theta}}$ is strongly consistent and that $\hat{h}(u; \hat{\boldsymbol{\theta}}) \to h_0(u)$ almost surely uniformly in $u \in [a, b]$.

APPENDIX B: PROOF OF LEMMA 1

By the standard central limit theorem, $n_{\overline{\mathrm{D}}}^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = n_{\overline{\mathrm{D}}}^{-\frac{1}{2}}\mathbb{Z}_{\overline{\mathrm{D}},2}^{-1}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\mathbf{Z}_{jl}(Y_{jl} - \boldsymbol{\gamma}_0'\mathbf{Z}_{jl})$ converges in distribution to a mean $\mathbf{0}$ multivariate normal random variable. It follows from the functional central limit theorem (Pollard 1990) that for any $\eta > 0$,

$$n_{\overline{\mathrm{D}}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\big[I(Y_{jl} - \boldsymbol{\gamma}'\mathbf{Z}_{jl} > c) - S_0\{c + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)'\mathbf{Z}_{jl}\}\big],$$
$$(c, \gamma) \in [S_0^{-1}(b), S_0^{-1}(a)] \times D_\eta^\gamma,$$

converges in distribution to a Gaussian process, where $D_\eta^\gamma = \{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| \le \eta\}$. It then follows from the equicontinuity (Pollard 1990) of the foregoing process and the consistency of $\hat{\boldsymbol{\gamma}}$ that

$$\sup_{c\in[S_0^{-1}(b), S_0^{-1}(a)]}\left| n_{\overline{\mathrm{D}}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\big[I(Y_{jl} - \hat{\boldsymbol{\gamma}}'\mathbf{Z}_{jl} > c) - S_0\{c + (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'\mathbf{Z}_{jl}\}\big] - n_{\overline{\mathrm{D}}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\{I(Y_{jl} - \boldsymbol{\gamma}_0'\mathbf{Z}_{jl} > c) - S_0(c)\}\right| \to 0 \qquad \text{(B.1)}$$

in probability. This implies that

$$n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\big\{\widehat{S}_0(c) - S_0(c)\big\} \approx \frac{n_{\overline{\mathrm{D}}}^{\frac{1}{2}}}{N_{\overline{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\big[I(Y_{jl} - \boldsymbol{\gamma}_0{}'\mathbf{Z}_{jl} > c) - S_0(c) + S_0\{c + (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'\mathbf{Z}_{jl}\} - S_0(c)\big]$$
$$\approx \frac{n_{\overline{\mathrm{D}}}^{\frac{1}{2}}}{N_{\overline{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\Big\{I(Y_{jl} - \boldsymbol{\gamma}_0{}'\mathbf{Z}_{jl} > c) - S_0(c) + \dot{S}_0(c)\big(\mathbb{Z}_{\overline{\mathrm{D}},1}'\mathbb{Z}_{\overline{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}\big)(Y_{jl} - \boldsymbol{\gamma}_0'\mathbf{Z}_{jl})\Big\}.$$

From the functional central limit theorem, we see that $n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\{\widehat{S}_0(c) - S_0(c)\}$ converges in distribution to a mean $\mathbf{0}$ Gaussian process. It then follows from a Taylor series expansion and the stochastic equicontinuity of $n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\{\widehat{S}_0(c) - S_0(c)\}$ that

$$n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\big[S_{\overline{\mathrm{D}},\mathbf{z}}\{\widehat{S}_{\overline{\mathrm{D}},\mathbf{z}}^{-1}(u)\} - u\big] \approx n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\big[S_0\{\widehat{S}_0^{-1}(u)\} - u\big] + n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\dot{S}_0\{\widehat{S}_0^{-1}(u)\}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'\mathbf{z} \approx \widetilde{\mathcal{P}}(u, \mathbf{z}),$$

where

$$\widetilde{\mathcal{P}}(u, \mathbf{z}) = -\frac{n_{\overline{\mathrm{D}}}^{\frac{1}{2}}}{N_{\overline{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\Big[I\big\{Y_{jl} - \boldsymbol{\gamma}_0{}'\mathbf{Z}_{jl} > S_0^{-1}(u)\big\} - u + \dot{S}_0\{S_0^{-1}(u)\}\big\{(\mathbb{Z}_{\overline{\mathrm{D}},1} - \mathbf{z})'\mathbb{Z}_{\overline{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}\big\}(Y_{jl} - \boldsymbol{\gamma}_0'\mathbf{Z}_{jl})\Big].$$

The functional central limit theorem implies that $\widetilde{\mathcal{P}}(u, \mathbf{z})$ converges in distribution to a Gaussian process and hence that $n_{\overline{\mathrm{D}}}^{\frac{1}{2}}\{S_{\overline{\mathrm{D}},\mathbf{z}}(\widehat{S}_{\overline{\mathrm{D}},\mathbf{z}}^{-1}(u)) - u\}$ converges in distribution to a Gaussian process.

APPENDIX C: PROOF OF THEOREM 2

By the consistency of $\hat{\boldsymbol{\theta}}$ and a Taylor series expansion of $V(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$, we obtain

$$n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \approx \widehat{\mathbb{A}}^{-1}(\boldsymbol{\theta}_0)n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0). \qquad \text{(C.1)}$$

Note that

$$n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0) \approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b \big[w(\mathbf{X}_{ik}, u)\mathbf{X}_{ik}\hat{e}_{ik}(u) + \mathbf{X}_{ik}\{\hat{h}(u, \boldsymbol{\theta}_0) - h_0(u)\}a_{ik}(u)\big]\, dv(u),$$

where $\hat{e}_{ik}(u) = I\{Y_{ik} > \widehat{S}_{\overline{\mathrm{D}},\mathbf{Z}_{ik}}^{-1}(u)\} - g\{h_0(u) + \boldsymbol{\theta}_0'\mathbf{X}_{ik}\}$. Using a Taylor series expansion of $\hat{h}(u; \boldsymbol{\theta}_0)$ around $h_0(u)$,

$$n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0) \approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik}, u)\left\{\mathbf{X}_{ik} - \frac{\sum_{\iota=1}^{n_{\mathrm{D}}}\sum_{\kappa=1}^{K_\iota}a_{\iota\kappa}(u)\mathbf{X}_{\iota\kappa}}{\sum_{\iota=1}^{n_{\mathrm{D}}}\sum_{\kappa=1}^{K_\iota}a_{\iota\kappa}(u)}\right\} \times \hat{e}_{ik}(u)\, dv(u).$$

It follows from a Taylor series expansion and (B.1) that

$$n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}w(\mathbf{X}_{ik}, u)\hat{e}_{ik}(u) \approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}w(\mathbf{X}_{ik}, u)e_{ik}(u) + p_{10}^{\frac{1}{2}}n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}a_{ik}(u)\dot{h}_0(u)\widetilde{\mathcal{P}}(u, \mathbf{Z}_{ik}),$$

where $p_{10} = n_{\mathrm{D}}/n_{\overline{\mathrm{D}}}$. This, coupled with the uniform convergence of $\overline{\mathbf{X}}(u; \boldsymbol{\theta}_0)$ and the weak convergence of $n_{\mathrm{D}}^{\frac{1}{2}}\{S_{\overline{\mathrm{D}},\mathbf{z}}(\widehat{S}_{\overline{\mathrm{D}},\mathbf{z}}^{-1}(u)) - u\}$, ensures that

$$n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0) \approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}U_i(\boldsymbol{\theta}_0) + \hat{p}_{10}^{\frac{1}{2}}n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b\{\mathbf{X}_{ik} - \mathbb{X}(u; \boldsymbol{\theta}_0)\} \times a_{ik}(u)\dot{h}_0(u)\widetilde{\mathcal{P}}(u, \mathbf{Z}_{ik})\, dv(u).$$

Because $\sup_{u\in[a,b]}\left|n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\{\mathbf{X}_{ik}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}a_{ik}(u)\right|\to 0$ almost surely, we have

$$\hat{p}_{10}^{\frac{1}{2}}n_{\mathrm{D}}^{-1}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\int_a^b\{\mathbf{X}_{ik}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}a_{ik}(u)\dot{h}_0(u)\widetilde{\mathcal{P}}(u,\mathbf{Z}_{ik})\,dv(u)$$
$$\approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}U_j(\boldsymbol{\theta}_0).$$

It follows from the almost sure convergence of $\widehat{\mathbb{A}}(\boldsymbol{\theta}_0)\to\mathbb{A}$ that

$$n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\approx n_{\mathrm{D}}^{-\frac{1}{2}}\mathbb{A}^{-1}\left\{\sum_{i=1}^{n_{\mathrm{D}}}U_i(\boldsymbol{\theta}_0)+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}U_j(\boldsymbol{\theta}_0)\right\},$$

From the central limit theorem, we see that the distribution of $n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ can be approximated by a normal vector with mean 0 and covariance matrix

$$n_{\mathrm{D}}^{-1}\mathbb{A}^{-1}\left\{\sum_{i=1}^{n_{\mathrm{D}}}U_i(\boldsymbol{\theta}_0)U_i(\boldsymbol{\theta}_0)'+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}U_j(\boldsymbol{\theta}_0)U_j(\boldsymbol{\theta}_0)'\right\}\mathbb{A}^{-1}.$$

APPENDIX D: PROOF OF THEOREM 3

Note that $Q(u;\mathbf{x})=n_{\mathrm{D}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\mathbf{x}+n_{\mathrm{D}}^{\frac{1}{2}}\{\hat{h}(u;\hat{\boldsymbol{\theta}})-h_0(u)\}$. By a Taylor series expansion of $\hat{h}(u;\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$, we obtain

$$n_{\mathrm{D}}^{\frac{1}{2}}\{\hat{h}(u;\hat{\boldsymbol{\theta}})-h_0(u)\}$$
$$\approx-\mathbb{X}(u;\boldsymbol{\theta}_0)'\mathbb{A}^{-1}n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0)+n_{\mathrm{D}}^{\frac{1}{2}}\{\hat{h}(u;\boldsymbol{\theta}_0)-h_0(u)\}.$$

Furthermore, by expanding the estimating function in (4), evaluated at $\{\boldsymbol{\theta}_0,\hat{h}(u;\boldsymbol{\theta}_0)\}$, around $h_0(u)$, it follows that

$$\begin{aligned}Q(u;\mathbf{x})&\approx\{\mathbf{x}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0)+n_{\mathrm{D}}^{\frac{1}{2}}\{\hat{h}(u;\boldsymbol{\theta}_0)-h_0(u)\}\\&\approx\{\mathbf{x}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}n_{\mathrm{D}}^{-\frac{1}{2}}V(\boldsymbol{\theta}_0)+\frac{n_{\mathrm{D}}^{-\frac{1}{2}}}{a(u)}\\&\quad\times\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}\left\{w(\mathbf{X}_{ik},u)e_{ik}(u)+n_{\overline{\mathrm{D}}}^{-\frac{1}{2}}a_{ik}(u)\dot{h}_0(u)\widetilde{\mathcal{P}}(u,\mathbf{Z}_{ik})\right\}.\end{aligned}$$

Therefore, $Q(u;\mathbf{x})$ has the same asymptotic distribution as

$$\widetilde{Q}(u;\mathbf{x})=n_{\mathrm{D}}^{-\frac{1}{2}}\left\{\sum_{i=1}^{n_{\mathrm{D}}}\widetilde{Q}_i(u;\mathbf{x})+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\widetilde{Q}_j(u;\mathbf{x})\right\}.$$

To show that $\widetilde{Q}(u;\mathbf{x})$ converges weakly to a zero-mean Gaussian process, let

$$Q_1=n_{\mathrm{D}}^{-\frac{1}{2}}\left\{\sum_{i=1}^{n_{\mathrm{D}}}U_i(\boldsymbol{\theta}_0)+\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}U_j(\boldsymbol{\theta}_0)\right\},$$

$$Q_2(u)=n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\sum_{k=1}^{K_i}w(\mathbf{X}_{ik},u)e_{ik}(u),$$

and

$$\begin{aligned}Q_3(u)=-\frac{n_{\mathrm{D}}^{\frac{1}{2}}}{N_{\overline{\mathrm{D}}}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\overline{\mathrm{D}}}}\sum_{l=1}^{K_j}\Bigg[&I\{Y_{jl}-\boldsymbol{\gamma}_0{}'\mathbf{Z}_{jl}>S_0^{-1}(u)\}-u\\&+\frac{a_{\mathbf{z}}(u)'}{a(u)}\mathbb{Z}_{\overline{\mathrm{D}},2}^{-1}\mathbf{Z}_{jl}(Y_{jl}-\boldsymbol{\gamma}_0'\mathbf{Z}_{jl})\dot{S}_0(S_0^{-1}(u))\Bigg].\end{aligned}$$

Then $\widetilde{Q}(u;\mathbf{x})=\{\mathbf{x}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}Q_1+Q_2(u)/a(u)+\dot{h}_0(u)Q_3(u)$. It is straightforward to show that for any finite number of points $\{u_1,\ldots,u_{m_1}\}$, the joint distribution of $\{\widetilde{Q}(u_k;\mathbf{x}),k=1,\ldots,m_1\}'$ is asymptotically normal with mean 0. To prove that $\widetilde{Q}(u;\mathbf{x})$ is tight (Billingsley 1986), because $\{\mathbf{x}-\mathbb{X}(u;\boldsymbol{\theta}_0)\}'\mathbb{A}^{-1}$ and $a(u)$ and $\dot{h}_0(u)$ are nonrandom functions, it is sufficient to show that Q_1, $Q_2(u)$, and $Q_3(u)$ are tight. Now because Q_1 does not involve u, Q_1 is tight. The tightness of $Q_2(u)$ follows from some basic properties of empirical processes (Shorack and Wellner 1986, p. 109). The functional central limit theorem implies the weak convergence of $\widetilde{\mathcal{P}}(u,\mathbf{z})$ and hence the tightness of $Q_3(u)$. The finite-dimensional convergence and tightness of $\widetilde{Q}(u;\mathbf{x})$ imply that $\widetilde{Q}(u;\mathbf{x})$ converges to a mean 0 Gaussian process.

[Received April 2001. Revised March 2002.]

REFERENCES

Billingsley, P. (1986), *Probability and Measure* (2nd ed.), New York: Wiley.

Dorfman, D. D., Berbaum, K. S., and Metz, C. E. (1992), "Receiver Operating Characteristic Rating Analysis: Generalization to the Population of Readers and Patients With the Jackknife Method," *Statistics in Radiology*, 27, 723–731.

Elmer-Dewitt, P., Lemonick, M., Park, A., Nash, M. (2001), "Medicine: The Future of Drugs," *Time*, 157, 56–102.

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.

Hanley, H. A. (1989), "Receiver Operating Characteristic (ROC) Methodology: The State of the Art," *Clinical Reviews in Diagnostic Imaging*, 29, 307–335.

Heagerty, P. J., and Pepe, M. S. (1999), "Semiparametric Estimation of Regression Quantiles With Application to Standardizing Weight for Height and Age in U.S. Children," *Applied Statistics*, 48, 533–551.

Metz, C. E. (1986), "ROC Methodology in Radiologic Imaging," *Investigative Radiology*, 21, 720–733.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994), "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81, 341–350.

Pepe, M. S. (1997), "A Regression Modelling Framework for Receiver Operating Characteristic Curves in Medical Diagnostic Testing," *Biometrika*, 84, 595–608.

——— (1998), "Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results," *Biometrics*, 54, 124–135.

——— (2000a), "An Interpretation for the ROC Curve and Inference Using GLM Procedures," *Biometrics*, 56, 352–359.

——— (2000b), "Receiver Operating Characteristic Methodology," *Journal of the American Statistical Association*, 95, 308–311.

Pepe, M. S., Etzioni, R., Feng, Z. D., Potter, J. D., Thompson, M. L., Thornquist, M., Winget, M., Yasui, Y. (2001), "Phases of Biomarker Development for Early Detection of Cancer," *Journal of the National Cancer Institute*, 93, 1054–1061.

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

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.

Srivastava, S., and Kramer, B. S. (2000), "Early Detection Cancer Research Network," *Laboratory Investigation*, 80, 1147–1148.

Swets, J. A., and Pickett, R. M. (1982), *Evaluation of Diagnostic Systems: Methods From Signal Detection Theory*, New York: Academy Press.

Thompson, M. L., and Zucchini, W. (1989), "On the Statistical Analysis of ROC Curves," *Statistics in Medicine*, 8, 1277–1290.

Thornquist, M. D., Omenn, G. S., Goodman, G. E., Grizzle, J. E., Rosenstock, L., Barnhart, S., Anderson, G. L., Hammar, S., Balmes, J., Cherniack, M., Cone, J., Cullen, M. R., Glass, A., Keogh, J. P., Meyskens, F., Valanis, B., Williams, J. H. (1993), "Statistical Design and Monitoring of the Carotene and Retinol Efficacy Trial," *Controlled Clinical Trials*, 14, 308–324.

Tosteson, A. N., and Begg, C. B. (1988), "A General Regression Methodology for ROC Curve Estimation," *Medical Decision Making*, 8, 204–215.

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.