Statistica Sinica **18**(2008), 947-965

A SEMIPARAMETRIC APPROACH FOR THE COVARIATE SPECIFIC ROC CURVE WITH SURVIVAL OUTCOME

Xiao Song[1] and Xiao-Hua Zhou[2,3]

[1] *University of Georgia,*
[2] *University of Washington and* [3] *Puget Sound Health Care System*

Abstract: The receiver operating characteristic (ROC) curve has been extended to survival data recently, including the nonparametric approach by Heagerty, Lumley and Pepe (2000) and the semiparametric approach by Heagerty and Zheng (2005) using standard survival analysis techniques based on two different time-dependent ROC curve definitions. However, both approaches do not involve covariates other than the biomarker and cannot be used to estimate the ROC curve adjusted for covariates. To account for the covariate effect, we propose a joint model approach which assumes that the hazard of failure depends on the biomarker and the covariates through a proportional hazards model and that the biomarker depends the covariates through a semiparametric location model. We propose semiparametric estimators for covariate-specific ROC curves corresponding to the two time-dependent ROC curve definitions, respectively. We show that the estimators are consistent and converge to Gaussian processes. In the case of no covariates, the estimators are demonstrated to be more efficient than the Heagerty-Lumley-Pepe estimator and the Heagerty-Zheng estimator via simulation studies. In addition, the estimators can be easily extended to other survival models. We apply these estimators to an HIV dataset.

Key words and phrases: Location model, proportional hazards model, receiver operating characteristic curve, survival analysis.

1. Introduction

The receiver operating characteristic (ROC) curve is a popular tool to assess the effect of biomarkers in screening and predicting disease. A biomarker can be a single variable or a composite score of several variables. The ROC curve has recently shown promises in the identification of biomarkers using high dimensional microarray data (Pepe, Longton, Anderson and Schummer (2003)). The curve can be viewed as a plot of the true versus false positive rates among all possible thresholds for classifying disease and nondisease patients. One appealing property of the ROC curve is that it provides a common scale for comparing the accuracy of biomarkers, which may be measured in different units, in distinguishing two states of a binary outcome. Various approaches have been proposed for

estimating the ROC curve (Zhou, Obuchowski and McClish (2002, Chaps. 4 and 5) and Pepe (2003, Chap. 5)). When covariates have an impact on the accuracy of a biomarker, it is important to account for them (Zhou, Obuchowski and McClish (2002), Chapter 8; Pepe (2003, Chap. 6)). This leads to the covariate-specific ROC curve.

Recently, the ROC curve has been extended to survival times to evaluate the accuracy of biomarkers in classifying subjects based on time to an event, such as time to progression to a disease. Heagerty and Zheng (2005) give a systematic review of such time-dependent ROC curves. Basically, the true positive rate (TPR) and false positive rate (FTR) of a ROC curve can be extended in two different ways for survival time: the TPR is generalized to incident TPR and cumulative TPR, and the FPR is generalized to dynamic FPR and static FPR. Heagerty, Lumley and Pepe (2000) proposed a nonparametric approach for the time-dependent ROC curve based on the incident TPR and the dynamic FPR, using the Kaplan-Meier estimator of the survival distribution and the empirical distribution estimator of the biomarker. Heagerty and Zheng (2005) took a semiparametric approach for the time-dependent ROC curve based on the cumulative TPR and the dynamic FPR, using a proportional hazards model for a linear combination of several variables as the biomarker. Both the Heagerty-Lumley-Pepe and Heagerty-Zheng approaches can be used to evaluate and compare biomarkers in classifying subjects based on their survival times (Heagerty and Zheng (2005)); the former is useful in distinguishing subjects failing by a given time and those failing after this time, and the latter is useful in distinguishing subjects failing at a given time and those failing after this time. However, these two approaches do not include covariates, variables other than those used in defining the biomarker, in the ROC curve, which may be important in assisting classification. An example is the HIVNET 012, a randomized clinical trial to compare nevirapine (200mg at labor onset, and 2mg/kg for babies within 72 hours of birth) and zidovudine (600mg at labor onset, 300mg every three hours until delivery, and 4mg/kg orally twice daily for babies for seven days) for prevention of mother-to-child transmission of HIV-1 (Jackson et al. (2003)). HIV-1 infected pregnant women in Kampala, Uganda, were recruited between November 1997 and April 1999, with 313 assigned nevirapine and 313 zidovudine. Two possible biomarkers, the maternal HIV-1 RNA and CD4 count, were measured at baseline. It is of interest to evaluate the capacities of the biomarkers in the classification of babies based on their time to HIV infection or death, which may be different for the two treatment groups. Thus it may be important to adjust for the treatment in constructing the ROC curves.

In this paper, we consider two types of time-dependent ROC curves for survival data adjusted for covariates. They correspond to the ROC curves used

by Heagerty, Lumley and Pepe (2000) and Heagerty and Zheng (2005), respectively, in the case of no covariates. We allow the biomaker to be composed of several variables, as considered by Heagerty and Zheng (2005). Moreover, covariates other than those variables contained in the biomarker can be adjusted in constructing the ROC curves. We assume that the survival time depends on the biomarker and covariates through a proportional hazards model, and that the biomarker depends on the covariates through a semiparametric location model. Semiparametric estimators are proposed for the time-dependent ROC curves. We show that these estimators are consistent and converge to Gaussian processes. This approach can be easily extended to other survival models, as discussed in Section 7.

The paper is organized as follows. We define the covariate specific time-dependent ROC curves in Section 2, and derive the estimators in Section 3. Asymptotic properties are given in Section 4. We assess the finite sample performance of the estimators by simulation in Section 5, and apply the method to the HIVNET 012 data described above in Section 6. The paper concludes with discussions in Section 7.

2. Definition

Let Y be a biomarker, T the survival time, C the censoring time, and X a vector of covariates that may affect T. The observed survival data are $V = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Suppose that larger values of Y are associated with greater hazards; otherwise, Y can be recoded if necessary to achieve this. The ROC curve for survival time is defined based on TPR and FPR by analogy to that for a binary outcome. To define the covariate specific time-dependent ROC curve, we first define the conditional TPR and FPR given the covariate X. Specifically, for $X = x$, the cumulative TPR, the incident TPR and the (dynamic) FPR at time t are defined, respectively, as

$$\begin{aligned}
\mathrm{TPR}_C(y; t, x) &= P(Y > y | T \leq t, X = x), \\
\mathrm{TPR}_I(y; t, x) &= P(Y > y | T = t, X = x), \\
\mathrm{FPR}(y; t, x) &= P(Y > y | T > t, X = x).
\end{aligned}$$

Then we define the cumulative ROC curve as

$$\mathrm{ROC}_C(v; t, x) = \mathrm{TPR}_C\left\{\mathrm{FPR}^{-1}(v; t, x); t, x\right\},$$

and the incident ROC curve as

$$\mathrm{ROC}_I(v; t, x) = \mathrm{TPR}_I\left\{\mathrm{FPR}^{-1}(v; t, x); t, x\right\}.$$

Both the cumulative and the incident time-dependent ROC curves can be used to evaluate and compare the accuracy of biomarkers in classifying subjects

based on their survival times while adjusting for covariates; the former is useful in distinguishing subjects failing by a given time and those failing after this time, while the latter is useful in distinguishing subjects failing at a given time and those failing after this time. The Heagerty-Lumley-Pepe approach deals with the cumulative ROC curve without covariates, and the Heagerty-Zheng approach estimates the incident ROC curve without covariates.

3. Estimation

With some simple algebra, TPR_C, TPR_I and FPR can be written as

$$\mathrm{TPR}_C(y;t;x) = \frac{\int_y^\infty \{1 - S(t|u,x)\}\, dP(Y \le u|X = x)}{\int_{-\infty}^\infty \{1 - S(t|u,x)\}\, dP(Y \le u|X = x)}, \tag{3.1}$$

$$\mathrm{TPR}_I(y;t,x) = \frac{\int_y^\infty f(t|u,x) dP(Y \le u|X = x)}{\int_{-\infty}^\infty f(t|u,x) dP(Y \le u|X = x)}, \tag{3.2}$$

$$\mathrm{FPR}(y;t;x) = \frac{\int_y^\infty S(t|u,x) dP(Y \le u|X = x)}{\int_{-\infty}^\infty S(t|u,x) dP(Y \le u|X = x)}, \tag{3.3}$$

where $S(t|y,x) = P(T \ge t|Y = y, X = x)$ is the conditional survival distribution function given $Y = y$ and $X = x$, and $f(t|y,x) = -dS(t|y,x)/dt$ is the corresponding conditional survival density.

To estimate these quantities, we assume a proportional hazard model for the survival time,

$$\lambda(t|Y,X) = \lambda_0(t) \exp\left(\beta_0 Y + \gamma_0^T X\right), \tag{3.4}$$

and a semiparametric location model for the biomarker,

$$P(Y \le y|X) = H(y - \alpha_0^T X), \tag{3.5}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, and $H(\cdot)$ is an unspecified distribution function. Suppose that the observed data $\{(V_i, \Delta_i, Y_i, X_i) : i = 1, \ldots, n\}$ are independent and identically distributed samples from (V, Δ, Y, X).

The estimator of α_0, say $\hat{\alpha}$, can be obtained by solving

$$\sum_{i=1}^n (Y_i - \alpha^T X_i) X_i = 0, \tag{3.6}$$

and $H(y)$ can be estimated by

$$\hat{H}(y, \hat{\alpha}) = n^{-1} \sum_{i=1}^n I\left(Y_i - \hat{\alpha}^T X_i \le y\right).$$

Under (3.4), we can write the survival function $S(t|y,x) = \exp\{-\Lambda_0(t) \exp(\beta y + \gamma^T x)\}$, where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is the cumulative baseline hazard function.

Let $N_i(u) = I(V_i \le u, \Delta_i = 1)$ be the counting process and $R_i(u) = I(V_i \ge u)$ be the at risk process. An estimator of $S(t|y,x)$ is

$$\hat{S}(t|y,x) = \exp\left\{-\hat{\Lambda}_0(t)\exp\left(\hat{\beta}y + \hat{\gamma}^T x\right)\right\},$$

where $(\hat{\beta}, \hat{\gamma})$ is the maximum partial likelihood estimator of (β_0, γ_0) that solves the partial score equation at time L,

$$\sum_{i=1}^{n}\int_0^L \left\{(Y_i, X_i^T)^T - \frac{\sum_{j=1}^n R_i(u)(Y_i, X_i^T)^T \exp\{\beta Y_i + \gamma^T X_i\}}{\sum_{j=1}^n R_i(u)\exp\{\beta Y_i + \gamma^T X_i\}}\right\} dN_i(u) = 0,$$

and $\hat{\Lambda}_0(t)$ is the Breslow estimator of $\Lambda_0(t)$ given by

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n}\int_0^t \frac{dN_i(u)}{\sum_{j=1}^n R_i(u)\exp\left\{\hat{\beta}Y_i + \hat{\gamma}^T X_i\right\}}.$$

The density $f(t|y,x)$ can be estimated by $\hat{f}(t|y,x) = -\partial\hat{S}(t|y,x)/\partial t$. Substituting the estimators of $S(t|u,x)$, $f(t|u,x)$ and $P(Y \le u|X = x)$ into (3.1)–(3.3), we obtain estimators of TPR_C, TPR_I and FPR:

$$\begin{aligned}\widehat{\mathrm{TPR}}_C(y;t,x) &= \frac{\int_y^\infty \left\{1 - \hat{S}(t|u,x)\right\} d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})}{\int_{-\infty}^\infty \left\{1 - \hat{S}(t|u,x)\right\} d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})} \\ &= \frac{\sum_{i=1}^n \left[1 - \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\}\right] I\left\{Y_i - \hat{\alpha}(X_i - x) \ge y\right\}}{\sum_{i=1}^n \left[1 - \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\}\right]}, \qquad (3.7)\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{TPR}}_I(y;t,x) &= \frac{\int_y^\infty \hat{f}(t|u,x)\hat{S}(t|u,x)d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})}{\int_{-\infty}^\infty \hat{f}(t|u,x)\hat{S}(t|u,x)d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})} \\ &= \frac{\sum_{i=1}^n \exp\left\{\hat{\beta}Y_i - \hat{\alpha}(X_i - x)\right\} \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\} I\left\{Y_i - \hat{\alpha}(X_i - x) \ge y\right\}}{\sum_{i=1}^n \exp\left\{\hat{\beta}Y_i - \hat{\alpha}(X_i - x)\right\} \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\}}, \qquad (3.8)\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{FPR}}(y;t;x) &= \frac{\int_y^\infty \hat{S}(t|u,x)d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})}{\int_{-\infty}^\infty \hat{S}(t|u,x)d\hat{H}(u - \hat{\alpha}x, \hat{\alpha})} \\ &= \frac{\sum_{i=1}^n \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\} I\left\{Y_i - \hat{\alpha}(X_i - x) \ge y\right\}}{\sum_{i=1}^n \hat{S}\left\{t|Y_i - \hat{\alpha}(X_i - x), x\right\}}. \qquad (3.9)\end{aligned}$$

Thus the estimators of $\mathrm{ROC}_C(v;t)$ and $\mathrm{ROC}_I(v;t)$ are $\widehat{\mathrm{ROC}}_C(v;t,x) = \widehat{\mathrm{TPR}}_C\{\widehat{\mathrm{FPR}}^{-1}(v;t,x);t,x\}$ and $\widehat{\mathrm{ROC}}_I(v;t,x) = \widehat{\mathrm{TPR}}_I\{\widehat{\mathrm{FPR}}^{-1}(v;t,x);t,x\}$, respectively. Note that both estimators can be used in the case of no covariate by

setting $\hat{\alpha}$ in (3.7)–(3.9) to be 0. For valid estimation, t should be less than the maximum follow-up time.

4. Asymptotic Properties

Let $\mathcal{Y}$ and $\mathcal{X}$ be the supports of Y and X, respectively. We derive the asymptotic properties of the estimators under some regularity conditions given in Appendix A.

Lemma 1. *Given $x \in \mathcal{X}$, $n^{1/2}\{\hat{H}(y - \hat{\alpha}^T x, \hat{\alpha}) - H(y - \alpha_0^T x)\}$, as a process in y, converges to a mean zero Gaussian process on $\mathcal{Y}$ with covariance given in Appendix B.*

Lemma 2. *Given $(x,t) \in \mathcal{X} \times [0, L]$, $n^{1/2}\{\hat{S}(t|\cdot, x) - S(t|\cdot, x)\}$ converges to a mean zero Gaussian process on $\mathcal{Y}$ with covariance given in Appendix B.*

The proofs are sketched in Appendix B.

Using these lemmas, we show in Appendix C the following theorem.

Theorem 1. *Given $(x,t) \in \mathcal{X} \times [0, L]$, $n^{1/2}\{\widehat{FPR}(\cdot; t, x) - FPR(\cdot; t, x)\}$, $n^{1/2}\{\widehat{TPR}_C(\cdot; t, x) - TPR_C(\cdot; t, x)\}$, and $n^{1/2}\{\widehat{TPR}_I(\cdot; t, x) - TPR_I(\cdot; t, x)\}$ converge to mean zero Gaussian processes on $\mathcal{Y}$, with covariances given in Appendix C.*

The asymptotic properties of the ROC curves then follow from Theorem 1 by the functional delta method (van der Vaart and Wellner (2000, Chap. 3.9)).

Theorem 2. *Given $(x,t) \in \mathcal{X} \times [0, L]$, $n^{1/2}\{\widehat{ROC}_C(\cdot; t, x) - ROC_C(\cdot; t, x)\}$ and $n^{1/2}\{\widehat{ROC}_I(\cdot; t, x) - ROC_I(\cdot; t, x)\}$ converge to mean zero Gaussian processes on $[p, q]$ with covariances given in Appendix C, where p and q are defined in condition H in Appendix C.*

The covariance formulas for these processes contain density functions. Smoothing techniques are needed to compute the standard errors based on these formulas. Alternatively, we may compute the standard errors and confidence bands using the bootstrap method.

5. Simulation Studies

To assess the performance of the estimators, we conduct simulations under the following scenarios.

We first consider the simple case of no covariate. The hazard of failure depends on the biomarker only through the proportional hazards model. The true regression coefficient is $\beta_0 = 1$, and the baseline hazard is constant at 0.1. The censoring distribution is the exponential distribution with mean 30 truncated at 20, leading to a censoring rate of 34%. For estimation of the cumulative ROC curve, we compare the estimator $\widehat{\mathrm{ROC}}_C$ with the Heagerty-Lumley-Pepe

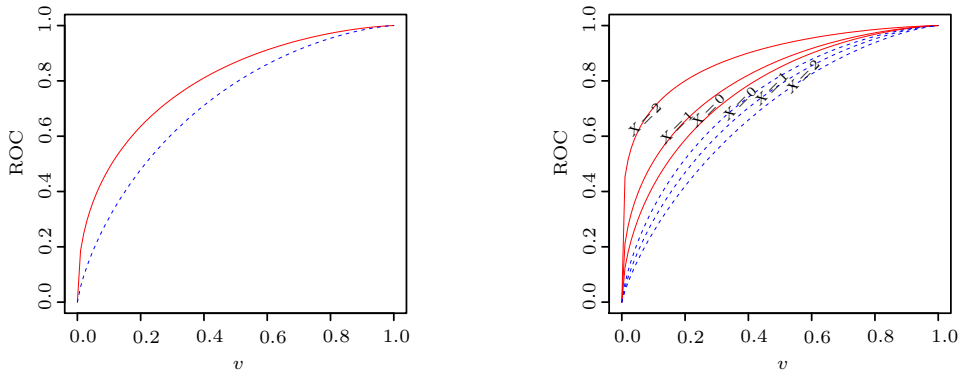

Figure 1. True ROC curves at $t = 5$ in simulation. Solid line, cumulative ROC; dashed line, incident ROC.

estimator; for estimation of the incident ROC curve, we compare the estimator $\widehat{\text{ROC}}_I$ with the Heagerty-Zheng estimator. The smoothing bandwidth for the Heagerty-Lumley-Pepe approach is taken to be $n^{-1/3}/3$. We consider the ROC curves at $t = 5$ with the true curves shown in Figure 1(a). For the sample sizes $n = 300$ and 600, we generate 500 simulated data sets. The ROC curves are estimated at $v = 0.1, 0.3, 0.5, 0.7, 0.9$. For all these estimators, the standard errors are computed by the bootstrap method, using 100 resampled data sets, as the standard deviation of the ROC estimates from the resampled data sets at the same v. The 95% Wald confidence intervals are constructed using the bootstrap standard errors. The results are shown in Tables 1 and 2 for ROC_C and for ROC_I, respectively. All estimators show negligible bias. The bootstrap standard errors track the empirical standard deviations well, and the coverage probabilities are close to the nominal level for all the semiparametric estimators. However, they may be a little below the nominal level for the nonparametric Heagerty-Lumley-Pepe estimator when v is close to 0 or 1. The semiparametric estimator $\widehat{\text{ROC}}_C$ has smaller bias and is more efficient than the nonparametric Heagerty-Lumley-Pepe estimator, as we expect. Interestingly, $\widehat{\text{ROC}}_I$ also achieves smaller bias and better efficiency than the Heagerty-Zheng estimator, although they are both semiparametric estimators.

Next we consider the case when the hazard of failure depends on the marker Y and a single covariate X through the proportional hazards model. The true regression coefficients are $\beta_0 = 1$ and $\gamma_0 = 0.5$, and the baseline hazard is constant at 0.1. The covariate X is generated from a normal distribution with mean 1 and variance 1, and the marker Y is generated from a conditional normal distribution with mean X and variance 1. The censoring time is generated in the same way as described above, with the censoring rate being 33%. We estimate the ROC

curves for $X = 0, 1, 2$ at $t = 5$ and $v = 0.1, 0.3, 0.5, 0.7, 0.9$, with the true curves shown in Figure 1(b). The Heagerty-Lumley-Pepe estimator and the Heagerty-Zheng estimator are not applicable in this case. We use $\widehat{\mathrm{ROC}}_C$ to estimate the cumulative ROC curve and $\widehat{\mathrm{ROC}}_I$ to estimate the incident ROC curve. The results are shown in Tables 3 and 4. Both estimators work well.

Table 1. Simulation results for the cumulative ROC in the case of no covariate. HLP: Heagerty-Lumley-Pepe estimator; B: bias; SD: empirical standard deviation across simulated data sets; SE: average of estimated standard errors; CP: coverage probability of the 95% confidence interval.

			$n = 300$				$n = 600$			
v	ROC_C	Method	B	SD	SE	CP	B	SD	SE	CP
0.1	0.4841	$\widehat{\mathrm{ROC}}_C$	−0.0019	0.0397	0.0386	0.934	−0.0027	0.0254	0.0272	0.958
		HLP	−0.0362	0.0746	0.0753	0.904	−0.0385	0.0555	0.0536	0.864
] 0.3	0.7362	$\widehat{\mathrm{ROC}}_C$	−0.0019	0.0292	0.0284	0.926	−0.0021	0.0190	0.0201	0.964
		HLP	−0.0219	0.0565	0.0584	0.920	−0.0222	0.0415	0.0408	0.928
0.5	0.8677	$\widehat{\mathrm{ROC}}_C$	−0.0020	0.0186	0.0181	0.932	−0.0014	0.0121	0.0128	0.958
		HLP	−0.0122	0.0432	0.0416	0.938	−0.0122	0.0288	0.0293	0.940
0.7	0.9463	$\widehat{\mathrm{ROC}}_C$	−0.0014	0.0094	0.0092	0.936	−0.0010	0.0061	0.0065	0.964
		HLP	−0.0058	0.0272	0.0260	0.916	−0.0052	0.0176	0.0182	0.960
0.9	0.9903	$\widehat{\mathrm{ROC}}_C$	−0.0007	0.0024	0.0024	0.958	−0.0004	0.0016	0.0017	0.956
		HLP	−0.0020	0.0102	0.0097	0.880	−0.0021	0.0074	0.0071	0.924

Table 2. Simulation results for the incident ROC in the case of no covariate. HZ: Heagery-Zheng estimator; B, bias; SD: empirical standard deviation across simulated data sets; SE: average of estimated standard errors; CP: coverage probability of the 95% confidence interval.

			$n = 300$				$n = 600$			
v	ROC_I	Method	B	SD	SE	CP	B	SD	SE	CP
0.1	0.3042	$\widehat{\mathrm{ROC}}_I$	−0.0010	0.0200	0.0196	0.932	−0.0007	0.0133	0.0136	0.958
		HZ	−0.0082	0.0273	0.0275	0.932	−0.0045	0.0196	0.0193	0.936
0.3	0.6088	$\widehat{\mathrm{ROC}}_I$	−0.0020	0.0240	0.0237	0.948	−0.0017	0.0162	0.0166	0.946
		HZ	−0.0050	0.0285	0.0284	0.934	−0.0034	0.0197	0.0199	0.956
0.5	0.7935	$\widehat{\mathrm{ROC}}_I$	−0.0023	0.0192	0.0189	0.944	−0.0013	0.0130	0.0132	0.960
		HZ	−0.0027	0.0218	0.0212	0.932	−0.0016	0.0146	0.0149	0.950
0.7	0.9129	$\widehat{\mathrm{ROC}}_I$	−0.0017	0.0114	0.0112	0.944	−0.0010	0.0076	0.0079	0.962
		HZ	−0.0024	0.0128	0.0123	0.934	−0.0013	0.0082	0.0086	0.970
0.9	0.9837	$\widehat{\mathrm{ROC}}_I$	−0.0011	0.0034	0.0034	0.942	−0.0006	0.0023	0.0023	0.960
		HZ	−0.0013	0.0037	0.0036	0.944	−0.0007	0.0024	0.0025	0.960

Table 3. Simulation results for the cumulative ROC in the case of a single covariate. B: bias; SD: empirical standard deviation across simulated data sets; SE: average of estimated standard errors; CP: coverage probability of the 95% confidence interval.

			$n = 300$				$n = 600$			
X	v	ROC_C	B	SD	SE	CP	B	SD	SE	CP
0	0.1	0.4217	−0.0026	0.0407	0.0398	0.936	−0.0014	0.0268	0.0280	0.954
	0.3	0.7001	−0.0019	0.0325	0.0322	0.944	−0.0003	0.0219	0.0226	0.956
	0.5	0.8494	−0.0012	0.0213	0.0211	0.944	−0.0002	0.0144	0.0148	0.952
	0.7	0.9391	−0.0009	0.0108	0.0109	0.952	−0.0003	0.0074	0.0076	0.950
	0.9	0.9890	−0.0005	0.0027	0.0027	0.954	−0.0001	0.0018	0.0019	0.950
1	0.1	0.5091	−0.0023	0.0395	0.0391	0.952	−0.0002	0.0269	0.0275	0.958
	0.3	0.7509	−0.0016	0.0281	0.0279	0.948	−0.0001	0.0193	0.0196	0.956
	0.5	0.8753	−0.0011	0.0174	0.0174	0.940	−0.0003	0.0119	0.0122	0.952
	0.7	0.9494	−0.0010	0.0087	0.0088	0.958	−0.0003	0.0060	0.0061	0.946
	0.9	0.9908	−0.0007	0.0023	0.0023	0.948	−0.0002	0.0015	0.0016	0.960
2	0.1	0.7030	−0.0005	0.0368	0.0372	0.952	0.0005	0.0270	0.0258	0.934
	0.2	0.8594	−0.0014	0.0210	0.0215	0.952	−0.0005	0.0159	0.0149	0.936
	0.3	0.9315	−0.0017	0.0120	0.0122	0.966	−0.0005	0.0090	0.0084	0.932
	0.5	0.9724	−0.0014	0.0059	0.0059	0.960	−0.0004	0.0043	0.0040	0.924
	0.9	0.9950	−0.009	0.0017	0.0016	0.902	−0.0004	0.0011	0.0011	0.924

Table 4. Simulation results for the incident ROC in the case of a single covariate. B: bias; SD: empirical standard deviation across simulated data sets; SE: average of estimated standard errors; CP: coverage probability of the 95% confidence interval.

			$n = 300$				$n = 600$			
X	v	ROC_I	B	SD	SE	CP	B	SD	SE	CP
0	0.1	0.3408	−0.0024	0.0269	0.0271	0.942	−0.0007	0.0189	0.0187	0.956
	0.3	0.6440	−0.0019	0.0280	0.0284	0.956	−0.0002	0.0199	0.0196	0.938
	0.5	0.8173	−0.0011	0.0205	0.0208	0.954	0.0000	0.0145	0.0143	0.944
	0.7	0.9251	−0.0008	0.0113	0.0114	0.942	−0.0002	0.0079	0.0079	0.938
	0.9	0.9863	−0.0005	0.0031	0.0031	0.958	−0.0001	0.0021	0.0021	0.948
1	0.1	0.2960	−0.0019	0.0181	0.0185	0.960	−0.0003	0.0129	0.0127	0.944
	0.3	0.5991	−0.0013	0.0229	0.0232	0.956	−0.0001	0.0160	0.0160	0.952
	0.5	0.7863	−0.0009	0.7854	0.0187	0.954	−0.0001	0.0128	0.0129	0.950
	0.7	0.9093	−0.0013	0.0111	0.0113	0.954	−0.0005	0.0078	0.0079	0.942
	0.9	0.9828	−0.0011	0.0036	0.0035	0.942	−0.0004	0.0024	0.0024	0.944
2	0.1	0.2545	−0.0013	0.0150	0.0154	0.944	−0.0007	0.0104	0.0106	0.964
	0.2	0.5505	−0.0033	0.0218	0.0225	0.950	−0.0016	0.0156	0.0157	0.946
	0.3	0.7492	−0.0047	0.0199	0.0206	0.936	−0.0019	0.0144	0.0144	0.948
	0.5	0.8877	−0.0046	0.0143	0.0143	0.936	−0.0018	0.0101	0.0100	0.938
	0.9	0.9772	−0.0039	0.0061	0.0057	0.848	−0.0017	0.0038	0.0038	0.926

6. Application

We applied these approaches to the HIVNET 012 data described in Section 1. It is thought that the maternal HIV-1 RNA and CD4 count may contain information in predicting the time to HIV infection or death of a child. The indicator for the treament, $X = I(\text{treatment} = \text{nevirapine})$, might have impact on the capacity of the two biomarkers. Here the survival time is the time to HIV infection or death of a child. There were 89 events in the nevirapine group, and 60 events in the zidovudine group. We estimated the cumulative and incident ROC curves for the two biomarkers adjusted for the treatment using the proposed estimators.

To ensure the validity of these estimators, we checked the proportional hazards assumption using the method in Therneau and Grambsch (2000, Chap. 6.2). We considered two proportional hazards models: one including log transformed maternal HIV-1 RNA and X; the other including log transformed CD4 count and X. Use of log transformations on HIV-1 RNA and CD4 counts is standard in the medical literature. There were no evidences against the proportional hazards assumptions.

To compare the accuracy of the two biomarkers in distinguishing subjects failing by a given time t and those failing after t, we estimated the cumulative ROC curves using the estimator $\widehat{\text{ROC}}_C$ for $X = 0, 1$ at $t = 0.5, 1, 1.5$ and 2 years. The 95% Wald confidence intervals were computed by the bootstrap method using 100 resampled data sets. The results are shown in Figure 2. Considering the FPR to be less than 15%, HIV-1 RNA seems to be a better biomarker than CD4 count for subjects taking nevirapine, but CD4 count seems to be a better marker in the zidovudine group.

To compare the accuracy of the two biomarkers in distinguishing subjects failing at t and those failing after t, we estimated the incident ROC curves using the estimator $\widehat{\text{ROC}}_I$ adjusting for $X = 0, 1$ with the same choices of time t. The results are presented in Figure 3. In contrast to their effects of classification of failures by or after these times, CD4 count seems to be a better biomarker than HIV-1 RNA for subjects taking nevirapine, but HIV-1 RNA seems to be a better marker in the zidovudine group when the FPR is less than 15%.

7. Discussion

We have proposed semiparametric estimators for the cumulative ROC curve and the incident ROC curve for survival data that may adjust for covariate effects. The proposed estimators are consistent and converge to Gaussian processes. These approaches work well in the case of moderate sample sizes. In the case of no covariate, these estimators are more efficient than the Heagerty-Lumley-Pepe method and the Heagerty-Zheng method, respectively.

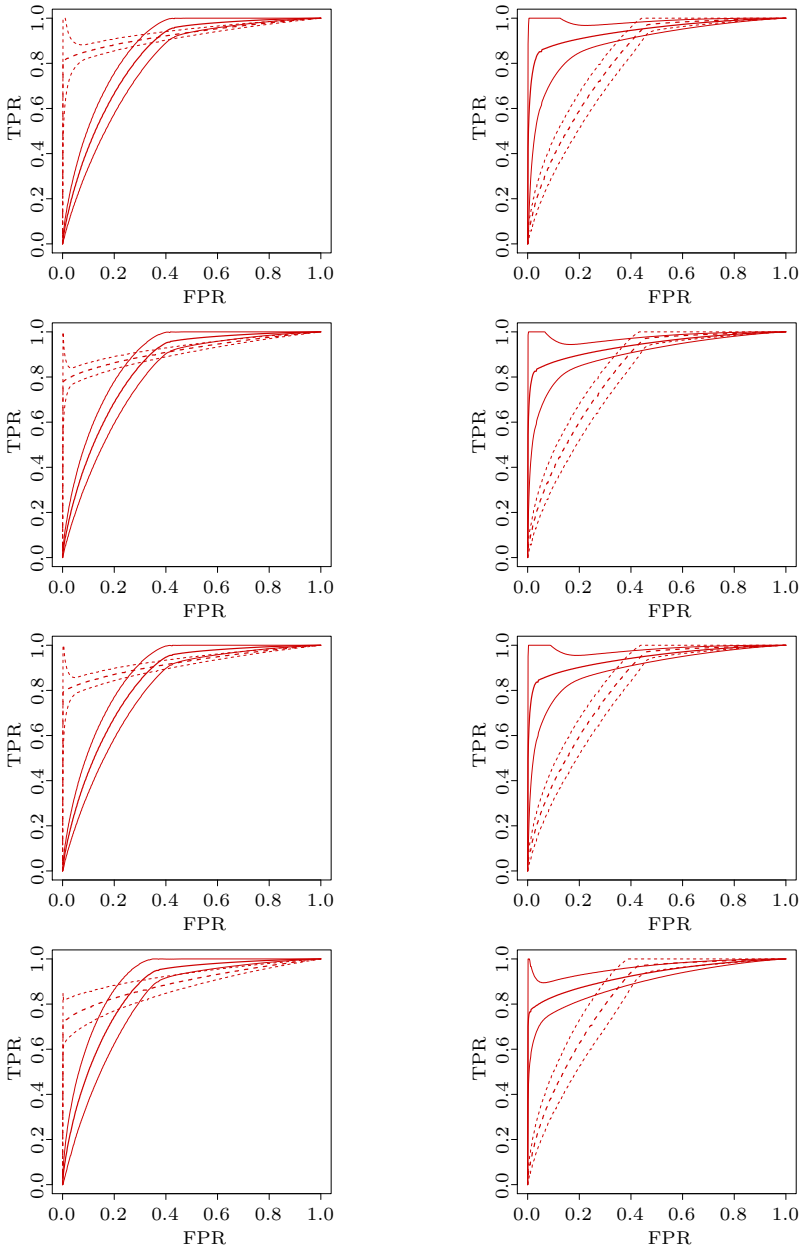

Figure 2. Estimated cumulative ROC curves adjusted for treatment for the HIVNET 012 data. The plots are, from the top, for $t = 0.5, 1.0, 1.5$ and 2, respectively. Left plots, zidovudine; Right plots, nevirapine; ROC curves for the maternal HIV-1 RNA, solid lines; ROC curves for the maternal CD4 count, dashed line; 95% pointwise confidence intervals are shown with the intermediate curves, the estimates themselves are shown with the center curves.

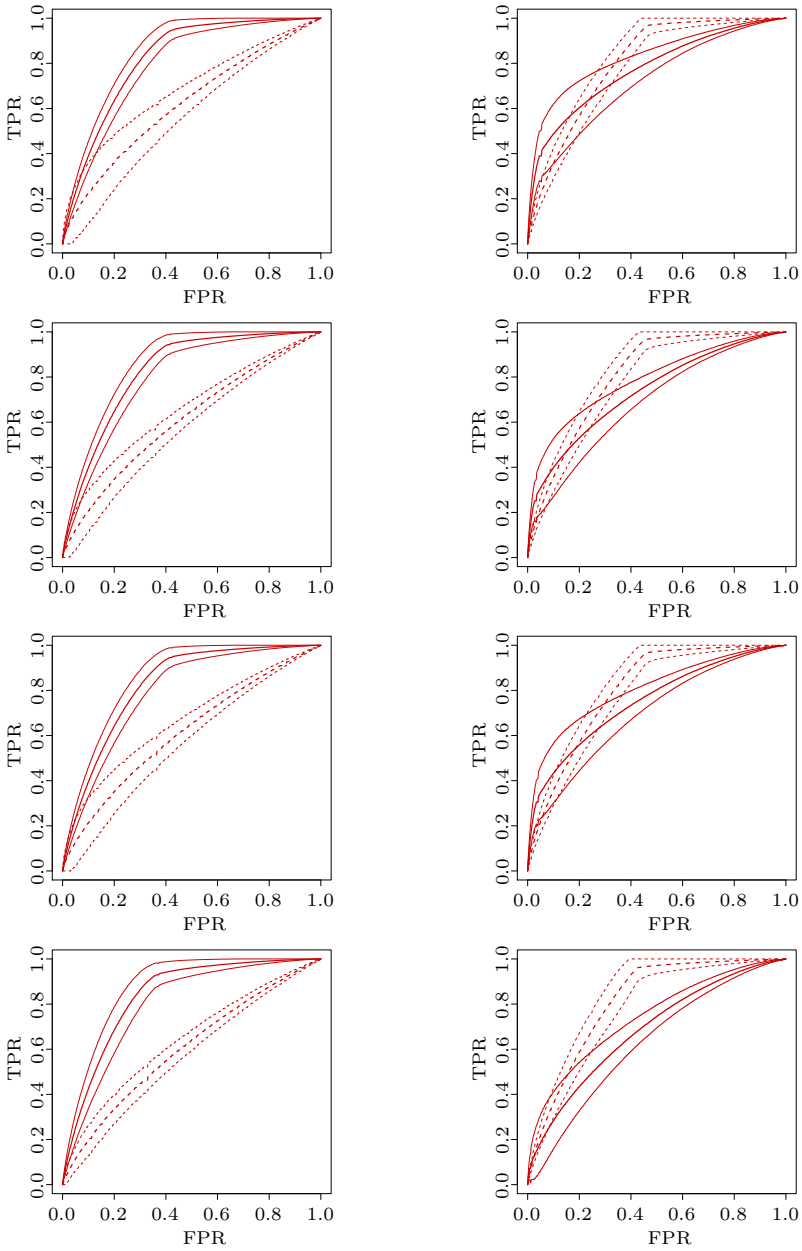

Figure 3. Estimated incident ROC curves adjusted for treatment for the HIVNET 012 data. The plots are, from the top, for $t = 0.5, 1.0, 1.5$ and 2, respectively. Left plots, zidovudine; Right plots, nevirapine; ROC curves for the maternal HIV-1 RNA, solid lines; ROC curves for the maternal CD4 count, dashed line; 95% pointwise confidence intervals are shown with the intermediate curves, the estimates themselves are shown with the center curves.

Cai, Pepe, Lumley, Zheng and Jenny (2006) estimated the time-dependent ROC curve based on the cumulative TPR and static FPR, assuming standard binary regression models for the cumulative TPR and the static FPR, and a proportional hazards model for the censored distribution. Loosely speaking, Cai et al.'s approach assumes models for the conditional distributions $P(Y|T, X)$ and $P(T|X)$, while our approach assumes models for the conditional distributions $P(T|Y, X)$ and $P(Y|X)$, where $P(A|B)$ denotes the conditional distribution of A given B. Our approach provides an alternative way to estimate the time-dependent ROC curve adjusted for covariate. Compared to Cai et al.'s approach, our approach has the advantage of simple computation; the estimating equations (3.4) and (3.6) are much easier to solve than the estimating equations (3.3) and (3.4) in Cai, Pepe, Lumley, Zheng and Jenny (2006).

We assume a standard proportional hazards model for the hazard given the biomarker and covariates, and a semiparametic location model for the conditional distribution of the biomarker given the covariates. The consistency of the proposed estimators depend on the correct specification of these models. However, the approach can be easily adapted to more flexible models. For example, we can use other survival models, such as the stratified proportional hazards model, the accelerated failure time model, and the transformation model, as long as we can obtain consistent estimators for the survival distributions. The nonparametric transformation model (Song, Ma, Huang and Zhou (2006)) may be an attractive extension as it includes most popular survival models as special cases, such as the proportional hazards model and the accelerated failure time model. The semiparametric location model is used only for estimating the conditional distribution $P(Y \leq y|X = x)$. We can remove this assumption and estimate $P(Y \leq y|X = x)$ by kernel smoothing method when the number of covariates is small and the sample size is relatively large, since the kernel method may not work well otherwise.

A common usage of the ROC curve is for the comparison of markers. In this paper, we focus on estimation of the covariate specific ROC. The estimated ROC curves from different biomarkers provide informal comparison for the biomarkers. Formal comparison of the ROC curves can be pursued based on the areas under the ROC curves (AUCs) similar to the case for binary ROC curves (Pepe (2003, Chap. 5)). This is beyond the scope of this paper and is to be investigated in the future.

Acknowledgements

This research is supported in part by NIH grants RO1EB005829 (Zhou) and U01-AI46702 (Song). The views expressed in this article are those of the authors

and do not necessarily reflect the position or policy of the Department of Veterans Affairs.

Appendix A: Regularity Conditions

Let $\eta = (\beta, \gamma^T)^T$, $\hat{\eta} = (\hat{\beta}, \hat{\gamma}^T)^T$, $z = (y, x^T)^T$, $Z = (Y, X^T)^T$, $W = (V, \Delta, Z)$, and

$$Q_{bi}(\eta; W_i, s) = R_i(s) Z_i^{\otimes b} \exp\left(\eta^T Z_i\right), \quad b = 0, 1, 2,$$

where $a^{\otimes b} = 1, a, aa^T$ for $b = 0, 1, 2$ respectively. Suppose $\eta_0 = (\beta_0, \gamma_0^T)^T$ is an internal point in a compact set, and write $\mathcal{N}(\eta_0)$ as a neighborhood of η_0.

To derive the asymptotic properties, we assume the following regularity conditions.

A. T and C are independent given Z.

B. $P(V \geq L) > 0$ for a constant $L > 0$.

C. $E(Z^T Z) < \infty$, $E\left\{\sup_{\eta \in \mathcal{N}(\eta_0)} Z^T Z \exp\left(\eta^T Z_i\right)\right\} < \infty$.

D. $\Sigma(\eta_0) = \int_0^\tau \left\{\frac{EQ_2(\eta_0; W_i, s)}{EQ_0(\eta_0; W_i, s)} - \frac{EQ_1^2(\eta_0, W_i, s)}{EQ_0^2(\eta_0; W_i, s)}\right\} EdN(s)$ is positive definite.

E. For $(y, x) \in \mathcal{Y} \times \mathcal{X}$, $S(t|y, x)$ is an absolutely continuous function for $t \in [0, L]$.

F. $H(u)$ is bounded and has bounded first- and second-order derivatives $H'(u)$ and $H''(u)$ for $u \in (-\infty, +\infty)$.

G. $\Gamma = E(XX^T)$ is positive definite.

H. The conditional densities

$$f_1(y; t, x) = -\frac{d\mathrm{TPR}_C(y; t, x)}{dy} = \frac{\{1 - S(t|y, x)\} H'(y - \alpha_0^T x)}{\int_{-\infty}^{\infty} \{1 - S(t|u, x)\} dP(Y \leq u | X = x)},$$

$$f_1^*(y; t, x) = -\frac{d\mathrm{TPR}_I(y; t, x)}{dy} = \frac{\{f(t|y, x) S(t|y, x)\} H'(y - \alpha_0^T x)}{\int_{-\infty}^{\infty} f(t|u, x) S(t|u, x) dP(Y \leq u | X = x)},$$

$$f_0(y; t, x) = -\frac{d\mathrm{FPR}(y; t, x)}{dy} = \frac{S(t|y, x) H'(y - \alpha_0^T x)}{\int_{-\infty}^{\infty} S(t|u, x) dP(Y \leq u | X = x)}$$

exist, and $f_0(y; t, x)$ is positive for $y \in \left[F^{(-1)}(p) - \varepsilon, F^{(-1)}(q) + \varepsilon\right]$ for some constants p and q, $0 < p < q < 1$, and $\varepsilon > 0$.

Appendix B: Proofs of Lemma 1 and Lemma 2

Proof of Lemma 1. Note that the semiparametric location model is equivalent to $Y = \alpha_0^T X + e$, where e has the distribution function $H(\cdot)$. Under Conditions C and F, the least square estimator $\hat{\alpha}$ is consistent and asymptotically normal with

$$n^{\frac{1}{2}}(\hat{\alpha} - \alpha_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \Gamma^{-1} \left(X_i Y_i - X_i X_i^T \alpha_0 \right) + o_p(1). \tag{B.1}$$

Let $\mathcal{N}(\alpha_0)$ be a compact neighborhood of α_0. Letting $B(\alpha) = EH\{y+\alpha^T(X-x) - \alpha_0^T X\}$, by the Functional Central Limit Theorem, $n^{1/2}[\hat{H}(y - \alpha^T x, \alpha) - B(\alpha)]$ converges to a mean zero Gaussian process on $(y,\alpha) \in \mathcal{Y} \times \mathcal{N}(\alpha_0)$. It follows from the equicontinuity of the foregoing process and the consistency of $\hat{\alpha}$, that

$$\sup_{y \in \mathcal{Y}} \left| n^{\frac{1}{2}} \left[\hat{H}(y - \hat{\alpha}^T x, \hat{\alpha}) - B(\hat{\alpha}) \right] - n^{\frac{1}{2}} \left[\hat{H}(y - \alpha_0^T x, \alpha_0) - H\left\{ y - \alpha_0^T x \right\} \right] \right| = o_p(1).$$

This implies that

$$\begin{aligned} & n^{\frac{1}{2}} \left\{ \hat{H}(y - \hat{\alpha}^T x, \hat{\alpha}) - H(y - \alpha_0^T x) \right\} \\ & \quad = n^{\frac{1}{2}} \left\{ \hat{H}(y - \alpha_0^T x, \alpha_0) - H(y - \alpha_0^T x) \right\} + n^{\frac{1}{2}} \left\{ B(\hat{\alpha}) - H(y - \alpha_0^T x) \right\} + o_p(1), \end{aligned} \tag{B.2}$$

uniformly for $y \in \mathcal{Y}$. Under condition F, the second term in (B.2) can be written by a Taylor series expansion as

$$H'(y - \alpha_0^T x) \left\{ E(X) - x \right\}^T n^{-\frac{1}{2}} (\hat{\alpha} - \alpha_0) + o_p(1),$$

uniformly for $y \in \mathcal{Y}$. This, together with (B.1), implies that

$$n^{\frac{1}{2}} \left\{ \hat{H}(y - \hat{\alpha}^T x, \hat{\alpha}) - H(y - \alpha_0^T x) \right\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} h(\alpha_0, y, x; Z_i) + o_p(1),$$

where

$$\begin{aligned} h(\alpha_0, y, x; Z_i) = \big[& \left\{ I\left(Y_i - \alpha_0^T X_i \le y - \alpha_0^T x \right) - H(y - \alpha_0^T x) \right\} \\ & + H'(y - \alpha_0^T x) \left\{ E(X) - x \right\}^T \Gamma^{-1} \left(X_i Y_i - X_i X_i^T \alpha_0 \right) \big]. \end{aligned}$$

Thus $n^{1/2}\{\hat{H}(y - \hat{\alpha}^T x, \hat{\alpha}) - H(y - \alpha_0^T x)\}$ converges to a Gaussian process $\mathcal{H}(y; x)$ with covariance

$$\operatorname{cov}\left\{ \mathcal{H}(y_1; x), \mathcal{H}(y_2; x) \right\} = \operatorname{cov}\left\{ h(\alpha_0, y_1, x; Z_i), h(\alpha_0, y_2, x; Z_i) \right\}.$$

Proof of Lemma 2. Under Conditions A–E, $n^{1/2}(\hat{\eta} - \eta)$ is normal, and $n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\}$ converges to a Gaussian process (Andersen and Gill (1982)).

Thus $n^{-1/2}\{\hat{S}(t|\cdot,x) - S(t|\cdot,x)\}$, as a functional differentiable with respect to (η, Λ_0), converges to a Gaussian process $\mathcal{S}(\cdot;t,x)$ on $\mathcal{Y}$ by the functional delta method. Specifically, by the functional Taylor expansion, with some algebra, we can show that

$$n^{\frac{1}{2}}\left\{\hat{S}(t|y,x) - S(t|y,x)\right\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi(\eta_0,t,y,x,W_i) + o_p(1),$$

where

$$\xi(\eta_0,t,y,x,W_i) = S(t|y,x)\exp\left\{\eta_0^T z\right\}\left(\delta(t,\eta_0,W_i)\right.$$
$$\left. - \left\{\int_0^t \frac{EQ_1(\eta_0;W_i,s)\,EdN(s)}{E^2Q_0(\eta_0;W_i,s)} - \Lambda_0(t)\begin{pmatrix} y \\ x \end{pmatrix}\right\} g(\eta_0,W_i)\right),$$

$$\delta(t,\eta_0,W_i) = \int_0^t \frac{dN_i(s)EQ_0(\eta_0;W_i,s) - R_i(s)\exp\left\{\eta_0^T Z_i\right\}EdN(s)}{E^2Q_0(\eta_0;W_i,s)},$$

$$g(\eta;W_i) = \Sigma^{-1}(\eta_0)\int_0^L\left\{Z_i - \frac{EQ_1(\eta_0;W_i,s)}{EQ_0(\eta_0;W_i,s)}\right\}\left\{dN_i(s) - \frac{Q_{0i}(\eta_0;W_i,s)}{EQ_0(\eta_0;W_i,s)}EdN(s)\right\}.$$

Thus $\mathrm{cov}\{\mathcal{S}(y_1;t,x),\mathcal{S}(y_2;t,x)\} = \mathrm{cov}\{\xi(\eta_0,t,y_1,x,W_i),\xi(\eta_0,t,y_2,x,W_i)\}$.

Appendix C: Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1. Since FPR is differentiable as a composite functional of $(\eta,\Lambda_0,H,\alpha)$, by the functional delta method, $n^{1/2}\{\widehat{\mathrm{FPR}}(\cdot;t,x) - \mathrm{FPR}(\cdot;t,x)\}$ converges to a Gaussian process $\mathcal{F}(\cdot;t,x)$ with mean zero on $\mathcal{Y}$. To derive the asymptotic covariance, using the functional Taylor expansion, we have

$$n^{\frac{1}{2}}\left\{\widehat{\mathrm{FPR}}(y;t,x) - \mathrm{FPR}(y;t,x)\right\}$$
$$= n^{\frac{1}{2}}\left[\int_{-\infty}^{\infty} S(t|u,x)dH(u-\alpha_0^T x)\right]^{-1}\left[\int_y^{\infty}\left\{\hat{S}(t|u,x) - S(t|u,x)\right\}dH(u-\alpha_0^T x)\right.$$
$$\left. + \int_y^{\infty} S(t|u,x)d\left\{\hat{H}(u-\hat{\alpha}^T x,\hat{\alpha}) - H(u-\alpha^T x)\right\}\right]$$
$$-n^{\frac{1}{2}}\left[\int_{-\infty}^{\infty} S(t|u,x)dH(u-\alpha_0^T x)\right]^{-2}\int_y^{\infty} S(t|u,x)dH(u-\alpha_0^T x)$$
$$\times\left[\int_{-\infty}^{\infty}\left\{\hat{S}(t|u,x) - S(t|u,x)\right\}dH(u-\alpha_0^T x)\right.$$
$$\left. + \int_{-\infty}^{\infty} S(t|u,x)d\left\{\hat{H}(u-\hat{\alpha}^T x,\hat{\alpha}) - H(u-\alpha_0^T x)\right\}\right] + o_p(1)$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \varpi_i(y;t,x) + o_p(1),$$

where

$$\begin{aligned}\varpi_i(y;t,x) = &\left[\int_{-\infty}^{\infty} S(t|u,x)dH(u-\alpha_0^T x)\right]^{-1} \\ &\times\left[\int_y^{\infty}\xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x)+\int_y^{\infty}S(t|u,x)dh(\alpha_0,u,x;Z_i)\right] \\ &-\left[\int_{-\infty}^{\infty} S(t|u,x)dH(u-\alpha_0^T x)\right]^{-2}\int_y^{\infty}S(t|u,x)dH(u-\alpha_0^T x) \\ &\times\left[\int_{-\infty}^{\infty}\xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x)+\int_{-\infty}^{\infty}S(t|u,x)dh(\alpha_0,u,x;Z_i)\right].\end{aligned}$$

Thus $\operatorname{cov}\{\mathcal{F}(y_1;t,x),\mathcal{F}(y_2;t,x)\} = \operatorname{cov}\{\varpi_i(y_1;t,x),\varpi_i(y_2;t,x)\}$.

Similarly, we can show that $n^{1/2}\{\widehat{\mathrm{TPR}}_C(\cdot;t,x) - \mathrm{TPR}_C(\cdot;t,x)\}$ and $n^{1/2}\{\widehat{\mathrm{TPR}}_I(\cdot;t,x) - \mathrm{TPR}_I(\cdot;t,x)\}$ converge to Gaussian processes $\mathcal{T}_C(\cdot;t,x)$ and $\mathcal{T}_I(\cdot;t,x)$ on $\mathcal{Y}$, respectively, with mean zero,

$$\begin{aligned}\operatorname{cov}\{\mathcal{T}_C(y_1;t,x),\mathcal{T}_C(y_2;t,x)\} &= \operatorname{cov}\{\zeta_i(y_1;t,x),\zeta_i(y_2;t,x)\}, \\ \operatorname{cov}\{\mathcal{T}_I(y_1;t,x),\mathcal{T}_I(y_2;t,x)\} &= \operatorname{cov}\{\zeta_i^*(y_1;t,x),\zeta_i^*(y_2;t,x)\},\end{aligned}$$

where

$$\begin{aligned}\zeta_i(y;t,x) = &\left[\int_{-\infty}^{\infty}\{1-S(t|u,x)\}\,dH(u-\alpha_0^T x)\right]^{-1} \\ &\times\left[\int_y^{\infty}-\xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x)+\int_y^{\infty}\{1-S(t|u,x)\}\,dh(\alpha_0,u,x;Z_i)\right] \\ &-\left[\int_{-\infty}^{\infty}\{1-S(t|u,x)\}\,dH(u-\alpha_0^T x)\right]^{-2}\int_y^{\infty}\{1-S(t|u,x)\}\,dH(u-\alpha_0^T x) \\ &\times\left[-\int_{-\infty}^{\infty}\xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x)+\int_{-\infty}^{\infty}\{1-S(t|u,x)\}\,dh(\alpha_0,u,x;Z_i)\right],\end{aligned}$$

$$\begin{aligned}\zeta_i^*(y;t,x) = &\left[\int_{-\infty}^{\infty}S(t|u,x)dH(u-\alpha_0^T x)\right]^{-1} \\ &\times\left[\int_y^{\infty}u\exp(\beta_0 u)\,S(t|u,x)dH(u-\alpha_0^T x)g(\eta;Z_i)\right. \\ &\left.+\int_y^{\infty}\xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x)+\int_y^{\infty}S(t|u,x)dh(\alpha_0,u,x;Z_i)\right]\end{aligned}$$

$$
\begin{aligned}
&- \left[\int_{-\infty}^{\infty} S(t|u,x)dH(u-\alpha_0^T x)\right]^{-2} \int_{y}^{\infty} S(t|u,x)dH(y-\alpha_0^T x) \\
&\times \left[\int_{-\infty}^{\infty} u\exp(\beta_0 u)\, S(t|u,x)dH(u-\alpha_0^T x)g(\eta,Z_i)\right. \\
&\left. + \int_{-\infty}^{\infty} \xi(\eta_0,t,u,x,Z_i)dH(u-\alpha_0^T x) - \int_{-\infty}^{\infty} S(t|u,x)dh(\alpha_0,u,x;Z_i)\right].
\end{aligned}
$$

Proof of Theorem 2. Since $\mathrm{ROC}_C(v;t,x)$ is a composite functional of $S(t|y,x)$ and $H(\hat{\alpha}, y-\hat{\alpha}x)$, under Assumption H, by the functional delta method, $n^{1/2}\{\widehat{\mathrm{ROC}}_C(\cdot;t;x) - \mathrm{ROC}_C(\cdot;t,x)\}$ converges to a Gaussian process $\mathcal{G}_C(\cdot;t,x)$ on $[p,q]$. Specifically, using the functional Taylor expansion, we have

$$
n^{\frac{1}{2}}\left\{\widehat{\mathrm{ROC}}_C(v;t,x) - \mathrm{ROC}_C(v;t,x)\right\} = n^{\frac{1}{2}}\sum_{i=1}^{n}\varphi_i(v;t,x) + o_p(1),
$$

where

$$
\varphi_i(v;t,x) = \zeta_i\left\{\mathrm{FPR}^{-1}(v;t,x)\right\} - f_1\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}\frac{\varpi_i\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}}{f_0\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}}.
$$

Thus $\mathrm{cov}\{\mathcal{G}_C(v_1;t,x), \mathcal{G}_C(v_2;t,x)\} = \mathrm{cov}\{\varphi_i(v_1;t,x), \varphi_i(v_2;t,x)\}$.

Similarly, we can show that $n^{1/2}\{\widehat{\mathrm{ROC}}_I(\cdot;t;x) - \mathrm{ROC}_I(\cdot;t,x)\}$ converges to a Gaussian process $\mathcal{G}_I(\cdot;t,x)$ on $[p,q]$ with mean zero and covariance $\mathrm{cov}\{\mathcal{G}_I(v_1;t,x), \mathcal{G}_I(v_2;t,x)\} = \mathrm{cov}\{\varphi_i^*(v_1;t,x), \varphi_i^*(v_2;t,x)\}$, where

$$
\begin{aligned}
&\varphi_i^*(v;t,x) \\
&= n^{\frac{1}{2}}\sum_{i=1}^{n}\zeta_i^*\left\{\mathrm{FPR}^{-1}(v;t,x)\right\} - f_1^*\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}\frac{\varpi_i\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}}{f_0\left\{\mathrm{FPR}^{-1}(v;t,x);t,x\right\}}.
\end{aligned}
$$

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Amer. Statist.* **10**, 1100-1120.

Cai, T., Pepe, M. S., Lumley, T., Zheng, Y. and Jenny, N. S. (2006). The sensitivity and specificity of markers for event times. *Biostatistics* **7**, 182-197.

Heagerty, P. J., Lumley, T. and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337-344.

Heagerty P. J. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92-105.

Jackson, J. B., Musoke, P., Fleming T., Guay, L. A., Bagenda, D., Allen M., Nakabiito, C., Sherman, P. B., Owor, M., Ducar C., Deseyve M., Mwatha, A., Emel L., Duefield C., Mirochnick M., Fowler M. G., Mofenson L., Miotti P., Gigliotti M., Bray D. and Mmiro F. (2003). Intrapatrum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampals, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *Lancet* **362**, 859-868.

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

Pepe, M. S., Longton, G., Anderson, G. L. and Schummer, M. (2003). Selecting differentially expressed genes from microarray experiments. *Biometrics* **59**,133-142.

Song, X., Ma, S., Huang, J. and Zhou, X. H. (2006). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics*, in press.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Zhou, X-H., Obuchowski, N. A. and McClish, D. K. (2002). *Statistical Methods in Diagnostic Medicine*. Wiley, New York.

Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA 30602, U.S.A.

E-mail: xsong@uga.edu

HSR&D Center of Excellence, VA Puget Sound Health Care System. Department of Biostatistics, University of Washington #1400 Met Park, West 1100 Olive Way, Seattle WA 98101, U.S.A.

E-mail: azhou@u.washington.edu

(Received May 2006; accepted November 2006)