Statistics
in Medicine

Research Article

Received 1 September 2011, Accepted 7 March 2012 Published online 22 June 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5386

Nonparametric receiver operating characteristic-based evaluation for survival outcomes

Xiao Song,[a*†] Xiao-Hua Zhou[b] and Shuangge Ma[c]

For censored survival outcomes, it can be of great interest to evaluate the predictive power of individual markers or their functions. Compared with alternative evaluation approaches, approaches based on the time-dependent receiver operating characteristics (ROC) rely on much weaker assumptions, can be more robust, and hence are preferred. In this article, we examine evaluation of markers' predictive power using the time-dependent ROC curve and a concordance measure that can be viewed as a weighted area under the time-dependent area under the ROC curve profile. This study significantly advances from existing time-dependent ROC studies by developing nonparametric estimators of the summary indexes and, more importantly, rigorously establishing their asymptotic properties. It reinforces the statistical foundation of the time-dependent ROC-based evaluation approaches for censored survival outcomes. Numerical studies, including simulations and application to an HIV clinical trial, demonstrate the satisfactory finite-sample performance of the proposed approaches. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: time-dependent ROC; concordance measure; inverse-probability-of-censoring weighting; marker evaluation; survival outcomes

1. Introduction

In biomedical studies with survival outcomes, it is of great interest to evaluate the predictive capacity of markers or their functions. An example is the HIVNET 012, a randomized clinical trial conducted to compare nevirapine (200 mg at labor onset and 2 mg/kg for babies within 72 h of birth) and zidovudine (600 mg at labor onset, 300 mg every 3 h until delivery, and 4 mg/kg orally twice daily for babies for 7 days) for prevention of mother-to-child transmission of HIV-1 [1]. HIV-1 infected pregnant women in Kampala, Uganda, were recruited between November 1997 and April 1999, with 313 assigned nevirapine and 313 zidovudine. Follow-up were continued until July 2004. We are interested in assessing the predictive capacity of baseline HIV-1 RNA and CD4 count on the time to HIV infection or death of a child.

There are multiple approaches that can be used for marker evaluation. One is to use the estimation measure (e.g., estimated regression coefficient or its significance level) as a proxy of prediction capacity. Another approach is to extend the notion of the proportion of variation explained by the markers, which has been well developed in linear regression, to survival analysis [2, 3]. Survival outcomes can also be transformed to categorical outcomes. A representative example is the 5-year survival status in cancer studies. With the categorical outcomes so obtained, many evaluation techniques described in [4] can be adopted. The aforementioned approaches can be limited in that they rely on specific (semi-)parametric assumptions and hence are subject to model mis-specification. The approaches that rely on the creation of categorical outcomes can be less informative and hence demand larger sample sizes. Moreover, such approaches have difficulties dealing with censored survival times.

[a]*Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA, U.S.A.*
[b]*Puget Sound Health Care System and University of Washington, Seattle, WA, U.S.A.*
[c]*School of Public Health, Yale University, New Haven, CT, U.S.A.*
**Correspondence to: Xiao Song, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Paul Coverdell Center, Room 129c, Athens, GA 30602, U.S.A.*
[†]*E-mail: xsong@uga.edu*

In this article, we consider evaluating the predictive capacity of markers using the time-dependent approaches based on receiver operating characteristics (ROC). The foundation of the time-dependent ROC has been laid in [5–8] and others. This approach is a natural extension of the binary ROC approaches developed in diagnosis studies [4]. It evaluates and compares markers on the basis of summary indexes such as the area under the ROC curve (AUC), with a larger AUC indicating a more powerful marker. A significant advantage of the ROC-based approaches is that they require minimal model assumptions and hence can be more robust.

Literature review suggests that the existing time-dependent ROC-based approaches may have the following limitations. First, they may focus on a specific time point as opposed to the whole time range. Time-dependent ROC curves usually vary over time. Although it is valuable to evaluate and compare markers at one specific time point, it may be more desirable to have a summary measure that can evaluate the overall predictive capacity across the whole time range or within a specific time interval. Second, they may rely on specific model assumptions. For example, Heagerty and Zheng estimated the time-dependent ROC based on Cox type models [6]; Chambless and Diao proposed an estimator for AUC based on a proportional hazards model [9]; Cai *et al.* estimated the time-dependent ROC under the assumption of generalized linear models for the true positive rate and false positive rate [7]. Third, in many existing studies, they have not rigorously established the asymptotic properties, which may prevent more extensive usage of those approaches. For example, Chambless and Diao proposed a nonparametric estimator for the cumulative AUC by recursive calculation at the ordered event times mimicking the Kaplan–Meier estimator, but they did not derive the asymptotic distribution of the estimator [9]. Compared with the kernel estimator proposed in [10], our proposed estimator for the cumulative AUC does not require kernel smoothing.

This study has been motivated by the effectiveness of time-dependent ROC-based approaches and their drawbacks described previously. In this article, we study a concordance measure τ, which is built on the time-dependent ROC of a marker and provides an effective summary measure (over time). It provides a useful alternative to the c-index [11] and can be more informative. Heagerty and Zheng first proposed the concordance measure in [6]. That study may be limited in that it depends on the assumption of Cox type models and no asymptotic properties are derived. We further extend the concordance measure to a specific time interval, which can be useful in assessing the marker predicability in the interval. In addition, for time-dependent ROCs and the concordance measure, we developed nonparametric estimators, which rely on weaker assumptions and hence are more robust. The computational simplicity and robustness of the proposed estimators may make them preferred in practice. Moreover, for the various estimators, we rigorously establish their asymptotic properties, which strengthens the statistical foundation of the time-dependent ROC-based approaches and provides more insights into their properties.

The rest of the article is organized as follows. In Section 2, we describe the data structure and various evaluation measures. We proposed the nonparametric estimators in Section 3 and established their asymptotic properties. We presented numerical studies, including simulations and application to the HIVNET 012, in Sections 4 and 5. The paper concludes with discussions in Section 6. We provided technical details in the Appendix.

2. Data structure and time-dependent receiver operating characteristic curves

For subject $i(= 1, \ldots, n)$, let X_i denote the value of a continuous marker and T_i denote the survival time. Without loss of generality, assume that a larger value of X is more likely associated with a shorter survival time.

We proposed two types of time-dependent ROC curves, which have different definitions of the true positive rate, for survival outcomes [6, 8]. We define the *cumulative time-dependent ROC curve* at time t as

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

where $\mathrm{FPR}(x;t) = P(X > x|T > t)$ is the false positive rate, $\mathrm{TPR}^C(x;t) = P(X > x|T \leqslant t)$ is the cumulative true positive rate, and $\mathrm{FPR}^{-1}(v;t)$ is the vth percentile for subjects surviving beyond time t, $v \in [0, 1]$. In contrast, we define the *incident time-dependent ROC curve* at time t as

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

where $\mathrm{TPR}^I(x;t) = P(X > x|T = t)$ is the incident true positive rate for subjects dying at time t. Both the cumulative and incident ROC curves can be used to evaluate and compare the accuracy of markers. The former is useful in distinguishing subjects failing *by* a given time and those failing after it, whereas the latter is useful in distinguishing subjects failing *at* a given time and those failing after it.

With the time-dependent ROC curves, we consider the AUCs (area under curve) as summary measures. Particularly, the AUC for the cumulative ROC curve is

$$AUC^C(t) = P(X_i > X_j|T_i \leqslant t, T_j > t), \quad i \neq j,$$

and the AUC for the incident ROC curve is

$$AUC^I(t) = P(X_i > X_j|T_i = t, T_j > t), \quad i \neq j.$$

The AUCs for time-dependent ROC curves are functions of time. Although the AUC profiles provide useful information, it may be more preferable to have a single index to summarize the performance of a marker and facilitate the comparison of markers. Specifically, we consider the concordance measure, which is defined as

$$\tau = P\left\{X_i > X_j|T_i < T_j\right\}, \quad i \neq j.$$

for marker X. The concordance measure is related to the incident ROC through the formula

$$\tau = \int_0^{+\infty} w(t)\mathrm{AUC}^I(t)\mathrm{d}t,$$

where $w(t) = 2f(t)S(t)$, and $f(t)$ and $S(t)$ are respectively the density and survival functions of the survival time T [6]. Thus, the concordance measure can be viewed as a weighted average of the incident AUC over time. Note that τ is invariant under monotone increasing transformation of X.

τ is an overall measure for the whole time range. In practice, we usually restrict our attention to an interval (L_1, L_2). For example, in cancer studies, 5-year survival with $L_1 = 0$ and $L_2 = 5$ can be more meaningful than survival over the whole time range. In this consideration, we propose the concordance measure restricted to the interval (L_1, L_2)

$$\tau^{L_1,L_2} = P\left\{X_i > X_j|T_i < T_j, L_1 < T_i < L_2\right\}.$$

This measure is useful when we evaluate and compare the markers on different time intervals. τ^{L_1,L_2} is related to the incident ROC curve via

$$\tau^{L_1,L_2} = \int_{L_1}^{L_2} w^{L_1,L_2}(t)\mathrm{AUC}^I(t)\mathrm{d}t,$$

where $w^{L_1,L_2}(t) = 2f(t)S(t)/\int_{L_1}^{L_2} 2f(u)S(u)\mathrm{d}u$. For a noninformative marker, $\tau^{L_1,L_2} = 0.5$. When τ^{L_1,L_2} is greater than 0.5, the marker is informative in classification of subjects on the time interval (L_1, L_2).

3. Estimation

We propose nonparametric estimators for the various time-dependent ROC measures and establish their properties. The estimators have similar forms. We start from the estimator for the concordance measure because it is the simplest.

In practice, the survival time T may be right censored by a censoring time C. Thus, the observed survival data for subject i are $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leqslant C_i)$. We assume that T and C are independent. This assumption may be relaxed as discussed in Section 6.

3.1. Concordance measure

Note that the concordance measure τ can be written as

$$\tau = \frac{P\left(X_i > X_j, T_i < T_j\right)}{P(T_i < T_j)}.$$

If $P(C \geqslant T) > 0$, adopting the inverse-probability-of-censoring weighting (IPCW) technique [12], by iterated expectations, we can show that $E\left\{\Delta_i I(V_i < V_j)/S_C^2(T_i)\right\} = P(T_i < T_j)$, where $S_C(t)$ is the survival function of the censoring time C. In addition, if C is independent of X, we have $E\left\{\Delta_i I(X_i > X_j, V_i < V_j/S_C^2(T_i))\right\} = P(X_i > X_j, T_i < T_j)$. Thus, we propose estimating τ by

$$\hat{\tau} = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C^2(V_i)} I(X_i > X_j, V_i < V_j)}{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C^2(V_i)} I(V_i < V_j)},$$

where $\hat{S}_C(t)$ is the Kaplan–Meier estimator of $S_C(t)$. Similarly, we estimate τ^{L_1,L_2} by

$$\hat{\tau}^{L_1,L_2} = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C^2(V_i)} I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)}{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C^2(V_i)} I(V_i < V_j, L_1 < V_i < L_2)}.$$

We refer to these estimators as IPCW estimators hereafter.

To derive the asymptotic properties, we assume $S_C(\min(T, L_2)) = P(C \geqslant \min(T, L_2)) > 0$. For the overall concordance measure with $L_2 = \infty$, this requires that the censoring time does not have a shorter support than the survival time. However, this assumption is relaxed for the restricted concordance measure when L_2 is selected such that $P(C > L_2) > 0$. It can be shown that $\hat{\tau}^{L_1,L_2}$ is a consistent estimator of τ^{L_1,L_2}, and

$$n^{1/2}\left(\hat{\tau}^{L_1,L_2} - \tau^{L_1,L_2}\right) = n^{-1/2} \sum_{i=1}^n \varphi_i^{L_1,L_2} + o_p(1). \tag{1}$$

Hence, $n^{1/2}\left(\hat{\tau}^{L_1,L_2} - \tau^{L_1,L_2}\right)$ is asymptotically normally distributed. We provided the definition of $\varphi_i^{L_1,L_2}$ in the Appendix.

For a vector a, define $a^{\otimes 2} = aa^T$. The variance of $\hat{\tau}^{L_1,L_2}$ can be estimated by $n^{-1}\sum_{i=1}^n \left(\hat{\varphi}_i^{L_1,L_2} - \overline{\hat{\varphi}^{L_1,L_2}}\right)^{\otimes 2}$, where $\overline{\hat{\varphi}^{L_1,L_2}} = n^{-1}\sum_{i=1}^n \hat{\varphi}_i^{L_1,L_2}$, and we obtained $\hat{\varphi}_i^{L_1,L_2}$ by substituting the unknown quantities with estimates in $\varphi_i^{L_1,L_2}$. We provided more details in the Appendix.

3.2. *Comparison of markers*

When comparing the predictability of two markers X_{k_1} and X_{k_2} in the interval $[L_1, L_2]$, we can use the difference of the concordance measures $\tau_{k_1}^{L_1,L_2}$ and $\tau_{k_2}^{L_1,L_2}$ and estimate it by $\hat{\tau}_{k_1}^{L_1,L_2} - \hat{\tau}_{k_2}^{L_1,L_2}$. The asymptotical normality of the estimator follows from (1). We can also compare the predictability of two markers in two intervals (L_1, L_2) and (L_1^*, L_2^*), say, using $\tau_{k_1}^{L_1,L_2} - \tau_{k_2}^{L_1^*,L_2^*}$. This can be easily extended to compare multiple markers in multiple intervals. By analogy, we can compare the markers at a given time t based on $AUC^C(t)$ or $AUC^I(t)$.

3.3. *Area under the receiver operating characteristic curve*

Consider the cumulative and incident AUCs. Because AUCs are the integrals of ROC curves, we may estimate an AUC by replacing the corresponding ROC curve by its estimate. A disadvantage of this approach is that it requires estimating the ROC curve in the whole interval [0,1].

As an alternative, assuming $P(C \geqslant t) > 0$, an estimator of $AUC^C(t)$ can be obtained in analogy to the derivation of the estimator of τ. We propose estimating $AUC^C(t)$ with the IPCW estimator

$$\widehat{AUC}^C(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)} I(X_i > X_j, V_i < t, V_j > t)}{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)} I(V_i < t, V_j > t)}. \tag{2}$$

Estimation of the incident AUC is more complicated because it involves conditioning on an incident time point rather than time intervals. Using the kernel smoothing technique, we propose estimating $AUC^I(t)$ with the IPCW estimator

$$\widehat{AUC}^I(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)} K_h(V_i - t) I(X_i > X_j, V_j > t)}{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)} K_h(V_i - t) I(V_j > t)},$$

where $K_h(t) = h^{-1}K(h^{-1}t)$ with $K(\cdot)$ being a kernel density and h being the bandwidth. We use R-fold cross-validation to select the bandwidth h. Specifically, the dataset is divided in R equal-sized subsets. Let $\widehat{AUC}^I_{-r}(t)$ be the estimate of $AUC^I(t)$ with the rth subset D_r deleted ($r = 1, \ldots, R$). We choose h by minimizing

$$Q(h;t) = \sum_{r=1}^R \sum_{i \in D_r} \sum_{j \in D_r} \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)} K(h^{-1}(V_i - t)) I(V_j > t) \{I(X_i > X_j) - \widehat{AUC}^I_{-r}(t)\}^2.$$

The asymptotic properties of the aforementioned estimators can be derived by arguments similar to $\hat{\tau}^{L_1,L_2}$, although it is more involved for $\widehat{AUC}^I(t)$ because of the involvement of kernel smoothing. Specifically, it can be shown that (i) $\widehat{AUC}^C(t)$ is a consistent estimator of $AUC^C(t)$, and

$$n^{1/2}\left\{\widehat{AUC}^C(t) - AUC^C(t)\right\} = n^{-1/2}\sum_{i=1}^n \varphi_i^C + o_p(1), \tag{3}$$

where φ_i^C is given in the Appendix. It follows that $\widehat{AUC}^C(t)$ is normally distributed. The asymptotic variance can be estimated by replacing the unknown quantities with their estimates. (ii) $\widehat{AUC}^I(t)$ is a consistent estimator of $AUC^I(t)$, and

$$(nh)^{1/2}\left\{\widehat{AUC}^I(t) - AUC^I(t)\right\} - (nh^5)^{1/2}b = (nh)^{1/2}\left\{n^{-1}\sum_{i=1}^n \varphi_i^I\right\} + o_p(1), \tag{4}$$

where φ_i^I and b are given in the Appendix. It follows that $(nh)^{1/2}\left\{\widehat{AUC}^I(t) - AUC^I(t)\right\} - (nh^5)^{1/2}b$ is asymptotically normally distributed. Although we can estimate the variance of $\widehat{AUC}^I(t)$ using a plug-in approach, this estimator does not work well for moderate sample size in our simulation. Thus, we take an alternative approach, that is, resampling [13]. Specifically, given the observed data, we compute

$$\widehat{AUC}^{I*}(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}^*_C(V_i)\hat{S}^*_C(t)} K_h(V_i - t) I(X_i > X_j, V_j > t) Z_i Z_j}{\sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\hat{S}^*_C(V_i)\hat{S}^*_C(t)} K_h(V_i - t) I(V_j > t) Z_i Z_j}, \tag{5}$$

where $\{Z_1, \ldots Z_n\}$ are independently generated from a positive distribution with both mean and variance equal to one, and

$$\hat{S}^*_C(t) = \hat{S}_C(t) \exp\left\{-n^{-1}\sum_{i=1}^n \int_0^t \frac{(Z_i - 1)\mathrm{d}M_i(u)}{n^{-1}\sum_{j=1}^n Y_j(u)}\right\}.$$

Let $\widehat{AUC}^{I*}(t) = \left(\widehat{AUC}^{I*}_1(t), \ldots \widehat{AUC}^{I*}_K(t)\right)^T$. Then conditional on the observed data, the unconditional distribution of $n^{-1/2}\left\{\widehat{AUC}^I(t) - AUC^I(t)\right\}$ can be approximated by the conditional distribution of $n^{-1/2}\left\{\widehat{AUC}^{I*}(t) - \widehat{AUC}^I(t)\right\}$. Therefore, the standard error of $\widehat{AUC}^I(t)$ can be estimated by the standard deviation of $G = \left\{\widehat{AUC}^{I*}_r(t) : r = 1, \ldots R\right\}$, where each $\widehat{AUC}^{I*}_r(t)$ is computed from (5) on the basis of a set of $\{Z_1, \ldots Z_n\}$.

Statistics in Medicine

Table I. Simulation results for estimating cumulative AUC and incident AUC at $t = 0.3$ and 0.5.

			$n = 200$				$n = 400$			
t		True	Bias	SD	SE	CP	Bias	SD	SE	CP
0.3	$\text{AUC}_1^C(t)$	0.848	−0.001	0.030	0.030	0.940	0.001	0.021	0.021	0.939
	$\text{AUC}_2^C(t)$	0.760	0.003	0.039	0.037	0.920	0.002	0.025	0.026	0.957
	Diff	0.088	−0.005	0.052	0.050	0.935	−0.001	0.035	0.035	0.956
	$\text{AUC}_1^I(t)$	0.735	0.002	0.048	0.052	0.941	0.003	0.039	0.042	0.947
	$\text{AUC}_2^I(t)$	0.661	0.006	0.058	0.057	0.917	0.005	0.043	0.047	0.946
	Diff	0.074	−0.004	0.081	0.090	0.960	−0.002	0.065	0.072	0.961
0.5	$\text{AUC}_1^C(t)$	0.853	−0.002	0.031	0.032	0.949	0.000	0.023	0.022	0.941
	$\text{AUC}_2^C(t)$	0.763	0.002	0.041	0.040	0.928	0.001	0.028	0.028	0.943
	Diff	0.090	−0.004	0.055	0.054	0.945	−0.001	0.039	0.038	0.948
	$\text{AUC}_1^I(t)$	0.721	−0.003	0.072	0.071	0.905	0.003	0.056	0.059	0.923
	$\text{AUC}_2^I(t)$	0.649	−0.001	0.080	0.081	0.913	−0.004	0.063	0.066	0.934
	Diff	0.072	−0.002	0.123	0.125	0.915	0.007	0.095	0.103	0.945

AUC, area under the receiver operating characteristic curve; Diff, $\text{AUC}_1^C(t) - \text{AUC}_2^C(t)$; SD, empirical standard deviation; SE, average of standard errors; CP, coverage probability of 95% Wald CI.

4. Simulation

We considered two markers (X_1, X_2), which had a joint normal distribution with the same mean 2, variance 1, and correlation 0.2. The survival time depended on the two markers through a proportional hazards model with the regression coefficients equal to (1.5,1) and the baseline hazard being a constant 0.01. We generated the censoring time from an exponential distribution with mean $\mu_C = 1$, leading to a censoring rate of 44%. For sample size $n = 200$ and 400, we generated 1000 simulated datasets.

We estimated the cumulative AUCs and the incident AUCs for the two markers at $t = 0.3$ and 0.5. For each AUC, we also estimated the difference between the two markers. The Epanechnikov kernel was used for estimation of the incident AUC with $h = n^{-1/3}$. We estimated the standard error of the cumulative AUC on the basis of the asymptotic variance, whereas we estimated the standard error for the incident AUC using the resampling method. Table I suggests that the IPCW estimates perform well with negligible bias. The empirical standard deviations track the standard errors reasonably well. The coverage probabilities are mostly close to the nominal level; they are a bit low for the incident AUC estimator when the sample size is small but improve when the sample size increases. The standard errors of the incident AUC estimates are larger than those of the cumulative AUC estimates. An intuitive explanation is that by ignoring the estimation of the censoring distribution, the estimation of the cumulative AUC includes all subjects, whereas the estimation of the incident AUC only includes subjects who are at risk after $t - h$.

We estimated the concordance measure $\tau_1^{L_1,L_2}, \tau_2^{L_1,L_2}$ and their difference for $L_1 = 0$ and $L_2 = 1, 3, +\infty$. We computed the standard errors and 95% Wald CIs on the basis of the asymptotic properties as described in Section 3.1. The results are shown in Table II. We observed similar satisfactory performance as in Table I.

For the cumulative AUC, we also compared the IPCW estimator with Chambless and Diao's nonparametric estimator [9] and Chiang and Chung's kernel estimator [10]. Because no variance estimator is available for Chambless and Diao's estimator, we only compared the bias and empirical standard deviation of the two estimators (Table I, supplementary materials).[‡] It seems that Chiang–Chung's estimator has the largest bias. Our estimator performs similarly to Chambless–Diao's estimator with slightly smaller bias and slightly larger standard deviation.

To evaluate the effect of censoring rate, we conducted simulations with various μ_C (Tables II and III, supplementary materials). As expected, the performance of the proposed estimators tends to improve with smaller empirical standard deviation and better coverage probabilities as the censoring rate decreases. The instant AUC estimator generally requires more observed events to ensure reasonably

[‡]*Supporting information may be found in the online version of this article.*

Table II. Simulation results for estimating the concordance measure.

				$n = 200$				$n = 400$			
L_1	L_2		True	Bias	SD	SE	CP	Bias	SD	SE	CP
0	1	$\tau_1^{L_1,L_2}$	0.774	−0.001	0.021	0.022	0.949	0.001	0.015	0.015	0.948
		$\tau_2^{L_1,L_2}$	0.699	0.002	0.026	0.025	0.933	0.001	0.017	0.017	0.956
		Diff	0.074	−0.003	0.033	0.033	0.949	0.000	0.023	0.023	0.957
0	3	$\tau_1^{L_1,L_2}$	0.767	0.000	0.022	0.023	0.960	0.001	0.016	0.016	0.948
		$\tau_2^{L_1,L_2}$	0.692	0.002	0.026	0.027	0.946	0.001	0.017	0.018	0.967
		Diff	0.074	−0.002	0.035	0.036	0.962	−0.001	0.024	0.025	0.961
0	$+\infty$	τ_1	0.764	0.003	0.022	0.024	0.951	0.004	0.017	0.017	0.939
		τ_2	0.689	0.005	0.026	0.027	0.939	0.004	0.017	0.019	0.957
		Diff	0.074	−0.002	0.035	0.037	0.967	0.000	0.025	0.026	0.961

Diff, $\tau_1^{L_1,L_2} - \tau_2^{L_1,L_2}$; SD, empirical standard deviation; SE, average of standard errors; CP, coverage probability of 95% Wald CI.

well performance. We also conducted simulation studies under the accelerated failure time model, or when $0.32\log(X_1)$ and $0.32\log(X_2)$ has the same joint normal distribution as X_1 and X_2 in the previous setting or when the censoring time has a uniform distribution on the interval (0,1.63). We observed similar results were observed (Tables IV–IX, supplementary materials).

5. Application

We applied the approaches to the nevirapine arm in the HIVNET 012. The survival time was the time to HIV infection or death of a child. The two markers were HIV-1 RNA and CD4 count measured at the baseline, which were available among 301 women with 60 events. For each marker, we estimated the cumulative and incident AUCs within 1.1 year, which is the 95th percentile of the observed failure times. We used the Epanechnikov kernel for estimation of the incident AUC with the bandwidth selected by 10-fold cross-validation. Figure 1 shows the estimated cumulative AUC curves for both markers and their difference. The estimated cumulative AUC of HIV-1 RNA seems to increase over time, whereas the estimated cumulative AUC for CD4 count is close to 0.6 over time. The estimated cumulative AUC of HIV-1 RNA is larger than CD4 across the time, indicating that HIV-1 RNA may be a better marker for classifying whether the subjects have a survival time less than time t in [0,1], although the difference of

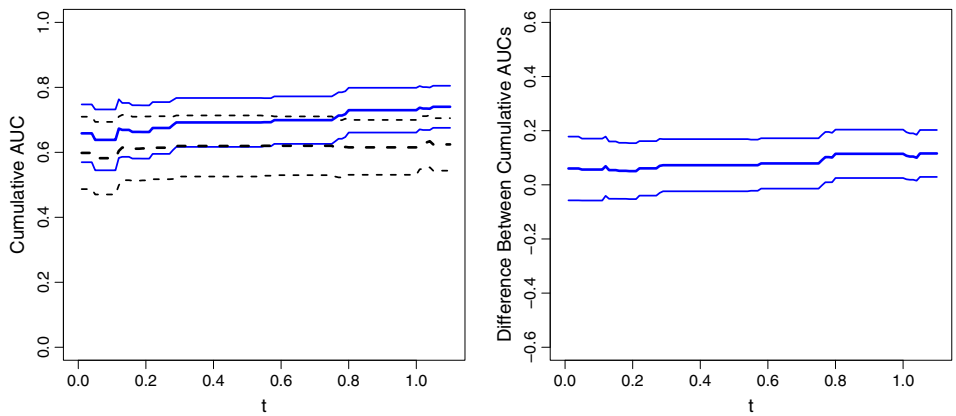

Figure 1. Estimated cumulative area under the receiver operating characteristic curve (AUC) profiles. Left plot: solid line, HIV-1 RNA; dashed line, CD4. Right plot, the difference between the two AUC profiles. In the plots, 95% pointwise CIs are shown with the outer curves; the estimates themselves are shown with the center curves.

the AUC curves is not significant before $t \approx 0.8$. Figure 2 shows the corresponding incident AUC curves ($h = 0.31$). The estimated incident AUC of HIV-1 RNA seems to increase before 0.7 year and then start to decrease, whereas the estimated incident AUC for CD4 is close to 0.6, with the former uniformly larger than the latter. This also implies that HIV-1 RNA may be a better marker. The variation of the estimated incident AUC curves seems to be larger than the variation of the cumulative AUC across the time, which conforms to what we have observed in the simulation studies.

We also estimated the concordance measure for HIV-1 RNA and CD4 count within 1.1 year. Table III shows the results for the concordance measures. For both markers, the estimated concordance measures are larger than 0.5, so are the lower limits of the 95% CIs. This indicates that the maternal HIV-1 RNA and CD4 count are informative markers in classifying subjects based on the survival time. The difference between the concordance measure of HIV-1 RNA and that of CD4 is equal to 0.100 with the 95% CI not covering zero. Thus, maternal HIV-1 RNA seems to be a better marker than maternal CD4 count.

We were also interested in comparing the early and late predication capacities of HIV-1 RNA and CD4 count. The concordance measures were calculated restricted to the intervals [0,0.1] and [1,1.1]. The results are shown in Table IV. The difference between the two intervals for maternal HIV-1 RNA is marginally significant, whereas it is not significant for maternal CD4. This indicates that maternal HIV-1 RNA may have better late predication.

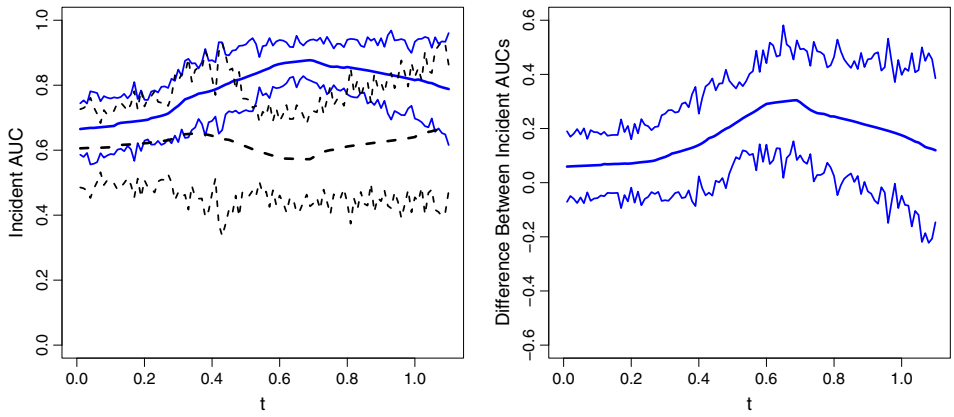

Figure 2. Estimated incident area under the receiver operating characteristic curve (AUC) profiles. Left plot: solid line, HIV-1 RNA; dashed line, CD4. Right plot, the difference between the two AUC profiles. In the plots, 95% pointwise CIs are shown with the outer curves; the estimates themselves are shown with the center curves.

Table III. Estimation of the concordance measure within 1.1 year for the HIVNET 012 data.

	Est	SE	CI
HIV-1 RNA	0.710	0.031	(0.649,0.772)
CD4	0.610	0.038	(0.535,0.686)
RNA – CD4	0.100	0.040	(0.020,0.179)

Est, Estimate; SE, standard error; CI, 95% Wald CI.

Table IV. Estimation of the concordance measure within intervals [0,0.1] and [1,1.1].

	$\tau^{0,0.1}$		$\tau^{1,1.1}$		$\tau^{0,0.1} - \tau^{1,1.1}$	
	Est	SE	Est	SE	Est	SE
HIV-1 RNA	0.635	0.047	0.780	0.061	−0.145	0.075
CD4	0.577	0.056	0.672	0.115	−0.095	0.128

Est, Estimate; SE, standard error.

6. Discussion

6.1. Connections with existing measures

The concordance measure τ is related to the c-index [11]. In the case of no tied survival times, the c-index is

$$\hat{C} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i I(X_i < X_j, V_i < V_j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i I(V_i < V_j)}.$$

Thus, it differs from $1-\hat{\tau}$ only in weights. When there is no censoring, $1-\hat{\tau}$ is equivalent to $\hat{C}$. However, there is fundamental difference between these two estimators in the presence of censoring. More specifically, under the assumption that censoring is independent of (T, X), $1-\hat{\tau}$ estimates $1-\tau = P(X_i < X_j | T_i < T_j)$, whereas the c-index estimates $C = E(S_C^2(T_i)I(X_i < X_j, T_i < T_j))/E(S_C^2(T_i)I(X_i < X_j, T_i < T_j))$. This suggests that C depends on the censoring distribution whereas τ does not. Because usually our interest is on the survival time, τ seems more appropriate. This is especially true when we compare markers over different time intervals with possibly different censoring distributions. In Table X in the supplementary materials, we compared τ and $1-C$ under various censoring distributions. This set of simulations clearly indicates the censoring-dependent nature of the c-index and suggests that the proposed concordance can be a more preferable measure. Following the asymptotic studies on $\hat{\tau}^{L_1,L_2}$, we may establish the asymptotic properties of $\hat{C}$ as described in the supplementary materials.

Pencina and D'Agostino provided a similar estimate for C that includes all pairs of subjects in which at least one is not censored, whereas $\hat{C}$ includes only the pairs of which the shorter survival times are not censored [14]. The asymptotic properties are only shown under certain semiparametric models such as the proportional hazards model and the accelerated failure time model. Our simulation studies indicates that the Wald CI of our proposed IPCW estimator is comparable with the CI (13) in [14] (not shown).

τ is also related to Kendall's tau defined as $\tau^{\#} = P((X_i - X_j)(T_i > T_j) > 0) - P((X_i - X_j)(T_i - T_j) < 0)$ for (X, T). In fact, $\tau = 0.5(1-\tau^{\#})$ for independent and identically distributed observations. Various estimators of Kendall's tau are available when neither variable is censored [15] or when both variables are censored survival times [16]. Beaudoin *et al.* proposed estimators when one variable is censored survival time and the other variable is an uncensored marker [17]. Their estimators involved kernel smoothing estimation of the conditional distribution of the survival distribution. But they have derived no asymptotic properties.

6.2. Limitations and possible extensions

Techniques based on receiver operating characteristics have been extensively used in biomedical studies for evaluating markers and/or their functions. In this article, we focus on the time-dependent ROC, which is a natural extension of the simple ROC for binary data and can be used to evaluate the predictive power on survival outcomes. This study significantly advances from existing ones by developing nonparametric estimators for various ROC-related measures and rigorously establishing the asymptotic properties. For simplicity of notation, we make the assumption of independent censoring. When the censoring depends on the markers, a (semi)parametric model will be needed to describe such dependence, and then the IPCW estimators will be applicable with minor modifications. We conducted numerical studies, and these demonstrate satisfactory performance of the IPCW estimators. We have also examined several other simulation settings and obtained similar satisfactory results (omitted).

For simplicity of notation, we have been focusing on single markers in this study. The IPCW estimators are directly applicable to composite markers (e.g., linear combinations of markers). In addition, for binary data, Ma and Huang [18] and other studies have shown that it is possible to generate optimal combinations of markers using ROC. We expect that the time-dependent ROC can also be used for such a purpose. In practice, the capacities of markers may depend on other covariates. For example, in the HIVNET 012 study, the effects of maternal HIV-1 RNA and CD4 may be different for the two treatment groups. It is useful to investigate the extension of the summary measures to covariate adjusted

time-dependent ROC curves [8] so that such important covariate effects can be adjusted. The aforementioned extensions demand significant methodological development and will be pursued in separate studies.

Appendix

Let $Y_i(u) = P(V_i \geqslant u)$ be the at risk process, $N_i(u) = I(V_i \leqslant u, \Delta_i = 0)$ be the counting process for the censoring time, $\Lambda(u) = -\log S_C(u)$ be the cumulated hazard function for the censoring time, and $M_i(u) = N_i(u) - \int_0^u Y_i(s)\mathrm{d}\Lambda(s)$. Let $W_i = (X_i, V_i, \Delta_i)$.

Proof of asymptotic properties of $\hat{\tau}^{L_1,L_2}$

The consistency of $\hat{\tau}^{L_1,L_2}$ follows from

$$\begin{aligned}
&n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\Delta_i}{S_C^2(V_i)}I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)\\
&\quad\xrightarrow{p} E\{I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)\},\\
&n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\Delta_i}{S_C^2(V_i)}I(V_i < V_j, L_1 < V_i < L_2) \xrightarrow{p} E\{I(V_i < V_j, L_1 < V_i < L_2)\},
\end{aligned}$$

and

$$\sup_u |\hat{S}_C(u) - S_C(u)| \to 0.$$

Let

$$\begin{aligned}
\varphi_i^{L_1,L_2} &= h_{i1}^{L_1,L_2} + h_{i2}^{L_1,L_2},\\
h_{i1}^{L_1,L_2} &= \int \frac{\left\{q_1^{L_1,L_2}(u) - q_2^{L_1,L_2}(u)\tau\right\}}{E\{Y_i(u)\}E\left\{I(T_i < T_j, L_1 < T_i < L_2)\right\}}\mathrm{d}M_i(u),\\
h_{i2}^{L_1,L_2} &= \frac{q_3^{L_1,L_2}(W_i)}{E\left\{I(T_i < T_j, L_1 < T_i < L_2)\right\}} - \frac{\tau^{L_1,L_2}q_4^{L}(W_i)}{E\left\{I(T_i < T_j, L_1 < T_i < L_2)\right\}},\\
q_1^{L_1,L_2}(u) &= E\left\{\frac{2\Delta_i I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)}{S_C^2(V_i)}I(V_i \geqslant u)\right\},\\
q_2^{L_1,L_2}(u) &= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{2\Delta_i I(V_i < V_j, L_1 < V_i < L_2)}{S_C^2(V_i)}I(V_i \geqslant u)\right\},\\
q_3^{L_1,L_2}(W_i) &= E\left\{\frac{\Delta_i}{S_C^2(V_i)}I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)\right.\\
&\qquad + \frac{\Delta_j}{S_C^2(V_j)}I(X_j > X_i, V_j < V_i, L_1 < V_j < L_2)\\
&\qquad \left.-2E\left\{I(X_i > X_j, T_i < T_j, L_1 < T_i < L_2)\right\}|W_i\right\}, \quad \text{and}\\
q_4^{L_1,L_2}(W_i) &= E\left\{\frac{\Delta_i}{S_C^2(V_i)}I(V_i < V_j, L_1 < V_i < L_2) + \frac{\Delta_j}{S_C^2(V_j)}I(V_j < V_i, L_1 < V_j < L_2)\right.\\
&\qquad \left.-2E\left\{I(T_i < T_j, L_1 < T_i < L_2)\right\}|W_i\right\}.
\end{aligned}$$

We consider $(L_1, L_2) = (0, \infty)$, as the arguments are essentially the same for any time interval (L_1, L_2). We drop the superscript $(0, \infty)$ for simplicity of notations. Note that $n^{1/2}(\hat{\tau} - \tau)$ can be written as $I_1 + I_2$, where $I_s = I_{s1} + I_{s2}$ for $s = 1, 2$ and

$$I_{11}=\frac{n^{-3/2}\sum_i\sum_j\frac{\Delta_i}{\hat{S}_C^2(V_i)}I(X_i>X_j,V_i<V_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{\hat{S}_C^2(V_i)}I(V_i<V_j)}-\frac{n^{-3/2}\sum_i\sum_j\frac{\Delta_i}{\hat{S}_C^2(V_i)}I(X_i>X_j,V_i<V_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)},$$

$$I_{12}=\frac{n^{-3/2}\sum_i\sum_j\frac{\Delta_i}{\hat{S}_C^2(V_i)}I(X_i>X_j,V_i<V_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)}-\frac{n^{-3/2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(X_i>X_j,V_i<V_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)},$$

$$I_{21}=\frac{n^{-3/2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(X_i>X_j,V_i<V_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)}-\frac{n^{1/2}P(X_i>X_j,T_i<T_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)},$$

$$I_{22}=\frac{n^{1/2}P(X_i>X_j,T_i<T_j)}{n^{-2}\sum_i\sum_j\frac{\Delta_i}{S_C^2(V_i)}I(V_i<V_j)}-\frac{n^{1/2}P(X_i>X_j,T_i<T_j)}{P(T_i<T_j)}.$$

Using a martingale representation of $n^{1/2}(\hat{S}_C-S_C)$ [19] and the uniform convergence of

$$q_{2n}(x)=n^{-2}\sum_i\sum_j\frac{2\Delta_i I(V_i<V_j)}{S^2(V_i)}I(V_i\geqslant u)$$

to $q_2(u)$, we can show that

$$\begin{aligned}I_{11}&=-\tau n^{-3/2}\sum_i\sum_j\frac{2\Delta_i I(V_i<V_j)}{S^2(V_i)}\left\{\frac{\hat{S}_C(V_i)-S(V_i)}{\hat{S}_C(V_i)}\right\}+o_p(1)\\&=-\tau n^{-1}\sum_r\int\frac{q_2(u)\mathrm{d}M_r(u)}{E\{Y_r(u)\}}+o_p(1),\end{aligned}$$

Similarly, we can show that

$$I_{12}=\frac{1}{P(T_i<T_j)}n^{-1}\sum_r\int\frac{q_1(u)\mathrm{d}M_r(u)}{E\{Y_r(u)\}}+o_p(1).$$

Using properties of U-statistics [20], we can show that

$$\begin{aligned}I_{21}&=-\frac{1}{P(T_i<T_j)}n^{-1/2}\sum_i q_4(W_i)+o_p(1),\\I_{22}&=\frac{\tau}{P(T_i<T_j)}n^{-1/2}\sum_i q_3(W_i)+o_p(1).\end{aligned}$$

Therefore,

$$n^{1/2}(\hat{\tau}-\tau)=n^{-1/2}\sum_{i=1}^n\varphi_i+o_p(1).\tag{6}$$

When estimating the variance of $\hat{\tau}^{L_1,L_2}$, we have the following definition:

$$\begin{aligned}\hat{\varphi}_i^{L_1,L_2}&=\hat{h}_{i1}^{L_1,L_2}+\hat{h}_{i2}^{L_1,L_2},\\\hat{h}_{i1}^{L_1,L_2}&=\int\frac{\left\{\hat{q}_1^{L_1,L_2}(u)-\tau\hat{q}_2^{L_1,L_2}(u)\right\}}{n^{-1}\sum_{j=1}^n Y_j(u)\xi^{L_1,L_2}}\mathrm{d}\hat{M}_i(u),\\\hat{h}_{i2}^{L_1,L_2}&=\frac{q_3^{L_1,L_2}(W_i)-\tau^{L_1,L_2}q_4^{L_1,L_2}(W_i)}{\xi^{L_1,L_2}},\end{aligned}$$

$$\xi^{L_1,L_2} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\Delta_i I(V_i < V_j, L_1 < V_i < L_2)}{\hat{S}_C^2(V_i)},$$

$$\hat{q}_1^{L_1,L_2}(u) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{2\Delta_i I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)}{\hat{S}_C^2(V_i)} I(V_i \geqslant u),$$

$$\hat{q}_2^{L_1,L_2}(u) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{2\Delta_i I(V_i < V_j, L_1 < V_i < L_2)}{\hat{S}_C^2(V_i)} I(V_i \geqslant u),$$

$$\begin{aligned}\hat{q}_3^{L_1,L_2}(W_i) = n^{-1}\sum_{j=1}^{n}\Bigg\{&\frac{\Delta_i}{\hat{S}_C^2(V_i)} I(X_i > X_j, V_i < V_j, L_1 < V_i < L_2)\\ &+\frac{\Delta_j}{\hat{S}_C^2(V_j)} I(X_j > X_i, V_j < V_i, L_1 < V_j < L_2)\Bigg\}\\ &-2n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{\Delta_r}{\hat{S}_C^2(V_r)} I(X_r > X_s, V_r < V_s, L_1 < V_r < L_2),\end{aligned}$$

$$\begin{aligned}\hat{q}_4^{L_1,L_2}(W_i) = n^{-1}\sum_{j=1}^{n}\Bigg\{&\frac{\Delta_i}{\hat{S}_C^2(V_i)} I(V_i < V_j, L_1 < V_i < L_2) + \frac{\Delta_j}{\hat{S}_C^2(V_j)} I(V_j < V_i, L_1 < V_i < L_2)\Bigg\}\\ &-2n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{\Delta_r}{\hat{S}_C^2(V_r)} I(V_r < V_s, L_1 < V_r < L_2),\end{aligned}$$

and $d\hat{M}_i(u) = dN_i(u) - Y_i(u)d\hat{\Lambda}(u)$ with

$$\hat{\Lambda}(u) = \sum_{i=1}^{n}\int_0^u \frac{\mathrm{d}N_i(s)}{\sum_{j=1}^{n} Y_j(s)}$$

being the Nelson–Aalen estimator of $\Lambda(u)$.

Proof of asymptotic properties of $\widehat{AUC}^C$

The derivation of the asymptotic properties of $\widehat{AUC}^C$ is similar to those of $\hat{\tau}^{L_1,L_2}$, and hence, the details are omitted. In expression (3),

$$\begin{aligned}\varphi_i^C &= h_{i1}^C + h_{i2}^C,\\ h_{i1}^C &= \int \frac{\{q_1^C(u) - q_2^C(u)\tau\}}{E\{Y_i(u)\}E\{I(T_i < t, T_j > t)\}}\mathrm{d}M_i(u),\\ h_{i2}^C &= \frac{q_3^C(W_i) - AUC^C q_4(W_i)}{E\{I(T_i < t, T_j > t)\}},\end{aligned}$$

$$q_1^C(u) = \lim_{n\to\infty} n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\Delta_i I(X_i > X_j, V_i \leqslant t, V_j > t)}{S_C(V_i)S_C(t)}\{I(t \geqslant u) + I(V_i \geqslant u)\},$$

$$q_2^C(u) = \lim_{n\to\infty} n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\Delta_i I(V_i \leqslant t, V_j > t)}{S_C(V_i)S_C(t)}\{I(t \geqslant u) + I(V_i \geqslant u)\},$$

$$\begin{aligned}q_3^C(W_i) = E\Bigg\{&\frac{\Delta_i}{S_C(V_i)S_C(t)} I(X_i > X_j, V_i \leqslant t, V_j > t) + \frac{\Delta_j}{S_C^2(V_j)} I(X_j > X_i, V_j \leqslant t, V_i > t)\\ &-2E\{I(X_i > X_j, T_i \leqslant t, T_j > t)\}|W_i\Bigg\},\end{aligned}$$

and

$$q_4^C(W_i) = E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}I(V_i \leqslant t, V_j > t) + \frac{\Delta_j}{S_C^2(V_j)}I(V_j \leqslant t, V_i > t) - 2E\left\{T_i \leqslant t, T_j > t\right\}|W_i\right\}.$$

Proof of asymptotic properties of $\widehat{AUC}^I$

Suppose that $K(u)$ is a bounded symmetric density, $h \to 0$, $nh \to \infty$, and $(nh)^5 = O(1)$. Note that

$$(nh)^{1/2}\left(\widehat{AUC}^I - AUC^I\right) = I_1^I + I_2^I + I_3^I, \tag{7}$$

where $I_s^I = I_{1s}^I + I_{2s}^I$ for $s = 1, 2$, and

$$I_{11}^I = \frac{n^{-3/2}h^{1/2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)}K_h(V_i - t)I(V_j > t)} - \frac{n^{-3/2}h^{1/2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)},$$

$$I_{12}^I = \frac{n^{-3/2}h^{1/2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{\hat{S}_C(V_i)\hat{S}_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)} - \frac{n^{-3/2}h^{1/2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)},$$

$$I_{21}^I = \frac{n^{-3/2}h^{1/2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)} - \frac{n^{1/2}h^{1/2}E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)\right\}}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)},$$

$$I_{22}^I = \frac{n^{1/2}h^{1/2}E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)\right\}}{n^{-2}\sum_{i=1}^n\sum_{j=1}^n \frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)} - \frac{n^{1/2}h^{1/2}E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)\right\}}{E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)\right\}},$$

$$I_3^I = \frac{n^{1/2}h^{1/2}E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)\right\}}{E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(V_j > t)\right\}} - n^{1/2}h^{1/2}P(X_i > X_j | T_j > t, T_i = t).$$

By the local U-process theory [21], we can show that

$$I_1^I + I_2^I = n^{-1/2}h^{1/2}\sum_{i=1}^n \varphi_i^I + o_p(1), \tag{8}$$

where

$$\varphi_i^I = h_{i1}^I + h_{i2}^I,$$
$$h_{i1}^I = \frac{\{q_1^I(W_i) - q_2^I(W_i)AUC^I\}}{p(T_j > t, T_i = t)E\{Y_i(u)\}},$$
$$h_{i2}^I = \frac{q_3^I(W_i) - AUC^I q_4(W_i)}{p(T_j > t, T_i = t)},$$
$$q_1^I(W_i) = E\left[\frac{3}{3!}\sum_{(i,j,r)\in\Gamma} g_1^I(W_i, W_j, W_r)\middle| W_i\right],$$
$$g_1^I(W_i, W_j, W_r) = \frac{K_h(V_i - t)\Delta_i I(X_i > X_j, V_j > t)}{S_C(V_i)S_C(t)}\int [I(t \geq u) + I(V_i \geq u)]\,\mathrm{d}M_r(u),$$
$$q_2^I(W_i) = E\left[\frac{3}{3!}\sum_{(i,j,r)\in\Gamma} g_2^I(W_i, W_j, W_r)\middle| W_i\right],$$
$$g_2^I(W_i, W_j, W_r) = \frac{K_h(V_i - t)\Delta_i I(V_j > t)}{S_C(V_i)S_C(t)}\int [I(t \geq u) + I(V_i \geq u)]\,\mathrm{d}M_r(u),$$
$$q_3^I(W_i) = E\left\{\frac{K_h(V_i - t)\Delta_i}{S_C(V_i)S_C(t)}I(X_i > X_j, V_j > t) + \frac{K_h(V_j - t)\Delta_j}{S_C^2(V_j)}I(X_j > X_i, V_i > t)\right.$$
$$\left. -2p(X_i > X_j, T_j > t, T_i = t)\middle| W_i\right\},$$
$$q_4^I(W_i) = E\left\{\frac{K_h(V_i - t)\Delta_i}{S_C(V_i)S_C(t)}I(T_i \leq t, T_j > t) + \frac{K_h(V_j - t)\Delta_j}{S_C^2(V_j)}I(T_j \leq t, T_i > t)\right.$$
$$\left. -2p(T_j > t, T_i = t)\middle| W_i\right\},$$
$$p(X_i > X_j, T_j > t, T_i = t) = \lim_{h\to 0}(2h)^{-1}P(X_i > X_j, T_j > t, t - h < T_i < t + h),$$
$$p(T_j > t, T_i = t) = \lim_{h\to 0}(2h)^{-1}P(T_j > t, t - h < T_i < t + h).$$

Next, we consider I_3^I. With some algebra, we can show that

$$E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i - t)I(X_i > X_j, V_j > t)\right\}$$
$$= E\left[I(X_i > X_j, T_j > t)\int K(w)p_{T|X}(t + hw; X_i)\mathrm{d}w\right],$$

where $p_{T|X}(t; X)$ is the conditional density of T given X. Using Taylor's expansion,

$$p_{T|X}(t + hw; X_i) = p_{T|X}(t; X_i) + hwp'_{T|X}(t; X_i) + h^2w^2p''_{T|X}(t; X_i) + o_p(h^2),$$

and noting that $K(w)$ is symmetric, we have

$$\int K(w)p_{T|X}(t + hw)\mathrm{d}w = p_{T|X}(t; X_i) + h^2p''_{T|X}(t; X_i)\int K(w)w^2\mathrm{d}w + o_p(h^2).$$

Thus,

$$E\{K_h(T_i - t)I(X_i > X_j, T_j > t)\}$$
$$= E[I(X_i > X_j, T_j > t)p_{T|X}(t; X_i)]$$
$$+ h^2E\left[I(X_i > X_j, T_j > t)p''_{T|X}(t; X_i)\right]\int K(w)w^2\mathrm{d}w + o_p(h^2).$$

Similarly, we can show that

$$E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i-t)I(V_j>t)\right\}$$
$$=E\left[I(T_j>t)p_T(t)\right]+h^2E\left[I(X_i>X_j,T_j>t)p_T''(t)\right]\int K(w)w^2\mathrm{d}w+o_p(h^2),$$

where $p_T(t)$ is the density of T. Hence,

$$\frac{E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i-t)I(X_i>X_j,V_j>t)\right\}}{E\left\{\frac{\Delta_i}{S_C(V_i)S_C(t)}K_h(V_i-t)I(V_j>t)\right\}}-\frac{E\left[I(X_i>X_j,T_j>t)p_{T|X}(t;X_i)\right]}{E\left[I(T_j>t)p_T(t)\right]}$$
$$=h^2b+o_p(h^2),$$

where

$$b=\int K(w)w^2\mathrm{d}w\left\{E\left[I(T_j>t)p_T(t)\right]\right\}^{-2}$$
$$\times\left\{E\left[I(T_j>t)p_T(t)\right]E\left[I(X_i>X_j,T_j>t)p_{T|X}''(t;X_i)\right]\right.$$
$$\left.-E\left[I(X_i>X_j,T_j>t)p_{T|X}(t;X_i)\right]E\left[I(X_i>X_j,T_j>t)p_T''(t)\right]\right\}.$$

It can be shown that

$$\frac{E\left[I(X_i>X_j,T_j>t)p_{T|X}(t;X_i)\right]}{E\left[I(T_j>t)p_T(t)\right]}=P(X_i>X_j|T_j>t,T_i=t).$$

Hence,

$$I_3^I=(nh^5)^{1/2}b+o_p((nh^5)^{1/2}). \tag{9}$$

Combining (7), (8) and (9), we have

$$(nh)^{1/2}\left(\widehat{AUC}^I-AUC^I\right)-(nh^5)^{1/2}b=n^{-1/2}h^{1/2}\sum_{i=1}^n\varphi_i^I+o_p(1).$$

Note that $E(\varphi_i^I)=0, E(h_{i1}^I)=0, E(h_{i2}^I)=0,$

$$var\left\{n^{-1/2}h^{1/2}\sum_{i=1}^n\varphi_i^I\right\}=hE\left(h_{i1}^I\right)^2+hE(h_{i2}^I)^2.$$

It can be shown that

$$hE(h_{i1}^I)^2\to 0,$$
$$hE(h_{i2}^I)^2\to\Omega,$$

where

$$\Omega=\frac{\Omega_1+\left\{AUC^I\right\}^2\Omega_2-2AUC^I\Omega_3}{E^2\left\{I(T_j>t)\right\}},$$
$$\Omega_1=\frac{\int K^2(s)\mathrm{d}s}{S_C^3(t)}E\left\{\left(E\left[I(X_i>X_j,V_j>t)\middle|X_i\right]\right)^2p_{T|X}(t;X_i)\right\},$$
$$\Omega_2=\frac{\int K^2(s)\mathrm{d}s}{S_C^3(t)}\left(E\left[I(V_j>t)\right]\right)^2p_T(t),$$
$$\Omega_3=\frac{\int K^2(s)\mathrm{d}s}{S_C^3(t)}E\left[I(X_i>X_j,V_j>t)p_{T|X}(t;X_i)\right].$$

This implies that $n^{-1/2}h^{1/2}\sum_{i=1}^n h_i^I\to 0$ in probability. Then we need to verify the Lindeberg condition by proving

$$n^{-1}h\sum h_{i2}^{I2}I\left(n^{-1/2}h^{1/2}\left|h_{i2}^{I}\right|>\varepsilon\right)\to 0.$$

It suffices to show that

$$n^{-1}h\sum\left\{q_3^I(W_i)\right\}^2 I\left(n^{-1/2}h^{1/2}\left|q_3^I(W_i)\right|>\varepsilon\right)\to 0,$$
$$n^{-1}h\sum\left\{q_4^I(W_i)\right\}^2 I\left(n^{-1/2}h^{1/2}\left|q_4^I(W_i)\right|>\varepsilon\right)\to 0.$$

The first expression follows from that for some $M>0$,

$$P\left[n^{-1/2}h^{1/2}\left|q_3^I(W_i)\right|>\varepsilon\right]\leqslant P\left[n^{-1/2}h^{1/2}h^{-1}M>\varepsilon\right]=P\left[n^{-1/2}h^{-1/2}M>\varepsilon\right]\to 0.$$

Therefore, by the central limit theorem, $n^{-1/2}h^{1/2}\sum_{i=1}^{n}\varphi_i^I$ converges to a normal distribution with mean 0 and variance equal to Ω.

Acknowledgements

The research is partially supported by a UGARF grant from University of Georgia (Song), by NIH grant R21CA152460 (Song), and by the Department of Veterans Affairs project grant #XVA 61-036 (Song and Zhou). The views expressed in this article are those of the authors and do not necessarily represent the views of NIH or the Department of Veterans Affairs.

References

1. Jackson JB, Musoke P, Fleming T, Guay LA, Bagenda D, Allen M, Nakabiito C, Sherman J, Bakaki P, Owor M, Ducar C, Deseyve M, Mwatha A, Emel L, Duefield C, Mirochnick M, Fowler MG, Mofenson L, Miotti P, Gigliotti M, Bray D, Mmiro F. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *Lancet* 2003; **362**:859–868.
2. Graf E, Schmoor C, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine* 1999; **18**:2529–2545.
3. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics* 2000; **56**:249–255.
4. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: New York, 2003.
5. Heagerty PJ, Lumley T, Pepe MS. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**:337–344.
6. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**:92–105.
7. Cai T, Pepe MS, Lumley T, Zheng Y, Jenny NS. The sensitivity and specificity of markers for event times. *Biostatistics* 2006; **7**:182–197.
8. Song X, Zhou XH. A semiparametric approach for the covariate specific ROC curve with survival outcome. *Statistica Sinica* 2008; **18**:947–965.
9. Chambless LE, Diao G. The sensitivity and specificity of markers for event times. *Statistics in Medicine* 2006; **25**:3474–3486.
10. Chiang CT, Hung H. Non-parametric estimation for time-dependent AUC. *Journal of Statistical Planning and Inference* 2010; **140**:1162–1174.
11. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Jounal of the American Medical Association* 1982; **247**:2543–2546.
12. Robins JM, Rotnitxky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**:846–866.
13. Cai T, Cheng S. Robust combination of multiple diagnostic tests for classifying censored event times. *Biostatistics* 2008; **9**:216–233.
14. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 2004; **23**:2109–2123.
15. Kendall MG. *Rank Correlation Methods*. Charles Griffin: London, 1970.
16. Wang W, Wells MT. Estimation of Kendalls tau under censoring. *Statistica Sinica* 2000; **10**:1199–1215.
17. Beaudoin D, Duchesne T, Genest C. Improving the estimation of Kendall's tau when censoring affects only one of the variables. *Computational Statistics and Data Analysis* 2007; **51**:5743–5764.
18. Ma S, Huang J. Regularized ROC method for disease classification and biomarker selection with microarray data. *Bioinformatics* 2005; **21**:4356–4362.
19. Gill RD. *Censoring and Stochastic Integrals*. Mathematisch Centrum: Amsterdam, 1980.
20. Serfling RJ. *Approximation Theorems of Mathematical Statistics*. Springer-Verlag: New York, 1967.
21. Giné E, Mason DW. On local U-statistic processes and the estimation of densities of functions of several sample variables. *Annals of Statistics* 2007:1105–1145.