Published in final edited form as:

Biometrics. 2012 June ; 68(2): 388–396. doi:10.1111/j.1541-0420.2011.01671.x.

Evaluating Prognostic Accuracy of Biomarkers under Competing Risk

Yingye Zheng^{*}, Tianxi Cai[†], Yuying Jin[‡], and Ziding Feng^{*}

^{*}Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109

[‡]Department of biostatistics, University of Washington, WA 98195

[†]Department of biostatistics, Harvard School of Public Health, Boston, MA 02115

Summary

To develop more targeted intervention strategies, an important research goal is to identify markers predictive of clinical events. A crucial step towards this goal is to characterize the clinical performance of a marker for predicting different types of events. In this manuscript, we present statistical methods for evaluating the performance of a prognostic marker in predicting multiple competing events. To capture the potential time-varying predictive performance of the marker and incorporate competing risks, we define time- and cause-specific accuracy summaries by stratifying cases based on causes of failure. Such definition would allow one to evaluate the predictive accuracy of a marker for each type of event and compare its predictiveness across event types. Extending the nonparametric crude cause-specific ROC curve estimators by Saha and Heagerty (2010), we develop inference procedures for a range of cause-specific accuracy summaries. To estimate the accuracy measures and assess how covariates may affect the accuracy of a marker under the competing risk setting, we consider two forms of semiparametric models through the cause-specific hazard framework. These approaches enable a flexible modeling of the relationships between the marker and failure times for each cause, while efficiently accommodating additional covariates. We investigate the asymptotic property of the proposed accuracy estimators and demonstrate the finite sample performance of these estimators through simulation studies. The proposed procedures are illustrated with data from a prostate cancer prognostic study.

Keywords

Biomarker evaluation; Cause-specific Hazard; Competing risk; Negative predictive value; Positive predictive value; Receiver Operating Characteristics Curve (ROC curve); Survival analysis

1. Introduction

Prognostic markers that provide accurate and individualized prediction of disease progression are fundamental for successful disease management and treatment selection. For many patients with prostate cancer, the disease progresses slowly over many years and does not pose a threat to the patient's life. Recommending these patients for invasive treatment may result in unnecessary side effects and significant morbidity. Therefore it is of great

^{*}yzheng@fhcrc.org.

Supplementary Materials

Web Appendices, Tables and Figures referenced in Sections 4, 5, and 6 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

clinical importance to develop a biomarker-based test to identify patients with the more aggressive form of the disease, i.e., those who are likely to die specifically from prostate cancer in future times, before treatment decision is made.

To evaluate the prognostic performance of a marker for an event of interest, prospective studies are often conducted such that the marker is measured at baseline and patients are followed over time for the development of the event. Such evaluations are challenging. First, the event status of subjects can change over time: subjects who are initially event free might eventually develop the event. Second, the risk of developing the event given the marker value might change over time. Third, the event time of interest might be censored due to loss to follow-up or termination of the study. Furthermore, patients might develop other types of potentially related events whose occurrence could preclude the event of interest from occurring or being observed. Comprehensive statistical tools to address all these issues would be of great value to the field of biomarker research.

In recent years, extending the traditional cross-sectional concept of diagnostic accuracy to accommodate the additional dimension of time has been an area of active research (Heagerty et al., 2000; Cai et al., 2006; Zheng and Heagerty, 2007; Zheng et al., 2010). See Pepe et al. (2008) for a review on such extensions. In this setting disease status is generally considered as a function of the disease ascertainment at time t , $D(t)$, and accuracy summaries are time-dependent functions. Following the tradition of medical diagnostic research, there are two main approaches to describing the accuracy of a marker Y measured at baseline: the retrospective and prospective summaries of accuracy. Two commonly used retrospective measures for quantifying the accuracy of Y in predicting $D(t)$ are the time-dependent true positive fraction (TPF) and false positive fraction (FPF), respectively defined as

$$\text{TPF}_t(c) = P\{Y \geq c | D(t) = 1\}, \text{FPF}_t(c) = P\{Y \geq c | D(t) = 0\}.$$

The ROC curve evaluated at a FPF of p , $\text{ROC}_t(p) = \text{TPF}_t\{\text{FPF}_t^{-1}(p)\}$, is often used to summarize the trade-off between TPF and FPF with varying cut-off values, and the area under the curve (AUC) provides a global summary of the discriminatory capacity of the marker. Two prospective measures are the positive predictive value (PPV) and negative predictive value (NPV) functions, respectively defined as

$$\text{PPV}_t(v) = P\{D(t) = 1 | H(Y) \geq v\} \text{NPV}_t(v) = P\{D(t) = 0 | H(Y) < v\},$$

where $H(Y)$ is the cumulative density function of Y . A time-dependent PPV curve at a predictive time t is then defined as a plot of $\text{PPV}_t(v)$ versus v for v in an open interval of $(0, 1)$ (Moskowitz and Pepe, 2004; Zheng et al., 2008). Statistical procedures for estimating these quantities have been considered in the literature, with focus on accommodating censoring in failure time data and the time-varying effect of markers (Zheng et al., 2010). ROC curve analysis has also been extended to non-censored multiple classes (Li and Fine, 2008). However most current methods do not consider competing risk events and have only been developed for standard univariate event time settings.

The analysis of time-to-event data is complicated by the presence of censoring and/or dependent competing risks, both of which occur frequently in prospective studies. While administrative censoring is a nuisance and should not impact definitions for predictive accuracy, competing risk events are real phenomena that occur in the population and therefore should impact how accuracy measures are defined (Pepe et al., 2008). In fact, characterizing the predictive accuracy of a marker for different types of events is an

important scientific objective. Understanding whether a marker is highly predictive for one particular event but not others could potentially lead to more rational and cost-effective use of specific medications or treatment strategies. For example, for patients diagnosed with prostate cancer (PC), invasive procedure for treating PC should only be given to individuals with an aggressive form of the disease (i.e., those who are at risk of dying *from* rather than dying *with* prostate cancer). Therefore the most important outcome when considering PC-related treatment is PC-specific death. Deaths from other causes are competing risk events. Simply censoring competing risk events will introduce a type of informative censoring, and may lead to distorted risk estimates (Prentice et al., 1978; Pepe and Mori, 1993) and subsequently biased accuracy estimators. For estimating accuracy summaries, a key question needs to be addressed: should subjects who died of other causes be considered cases or controls? We focus primarily on the discrimination between subjects who develop a specific event of interest by a given time t and those who have not yet experienced any event by t . Such an analysis essentially relates to the estimation of the cumulative incidence, which is identifiable even under the competing risks setting (Tsiatis, 1975). To examine whether a marker can be used to predict a particular type of event, we propose to compare its accuracy for this event with that for other types of events.

Statistical procedures for making inference about accuracy measures under the competing risks setting are not well developed. Recently, Saha and Heagerty (2010) defined cause-specific ROC curves by stratifying cases at time t (i.e., these with $D(t) = 1$) by event types and proposed estimation procedures for these quantities. Their methods focus primarily on the accuracy estimation for a single marker without allowing additional covariates that may affect the accuracy. In this manuscript, we develop semiparametric procedures for making inference about covariate-specific time-dependent accuracy summaries that accommodate competing risk events. Compared with a nonparametric approach, semiparametric procedures often are more efficient with correctly specified models. We consider two survival models, one is based on the widely applied cause-specific Cox proportional hazards (PH) model (Prentice et al., 1978), which models hazard rates of each event type under a proportional hazard model. However, the validity of the model relies on the assumption that the effects of a marker are constant and linear over time for *all* types of event times. To relax such an assumption, we propose a cause-specific ‘smoothed’ Cox regression model to allow for the potential time-varying and non-linear marker effect on the risks. In Section 2, we define time-dependent prognostic accuracy summaries in the presence of competing risk events. In Section 3, we describe estimation procedures for these accuracy summaries under the two classes of models. We comment on the theoretical properties of the proposed interval estimation procedures in Section 4. To evaluate the performance of the proposed estimators, we report results from simulation studies in Section 5. Illustrations of our procedure with a prostate cancer study are detailed in Section 6. We conclude in Section 7 with some remarks.

2. Time-dependent Cause-specific Accuracy

2.1 Notation

Let Y denote the baseline marker measurement potentially predictive of the occurrence time of J distinct events, denoted by $\{\mathcal{T}_1, \dots, \mathcal{T}_J\}$, and $\mathbf{Z}$ denote a p -dimensional vector of baseline covariates that may affect the accuracy of Y . In the presence of competing risks, only the time to the occurrence of the first event is potentially observable. Thus, without censoring, for $\{\mathcal{T}_1, \dots, \mathcal{T}_J\}$, one observes $T = \min(\mathcal{T}_1, \dots, \mathcal{T}_J)$ and the event type index η , which takes value j if $T = \mathcal{T}_j$. With the additional complication from censoring, observable quantities for event times are $(X, \Delta, \Delta\eta)$, where $X = \min(T, C)$, $\Delta = \mathbb{I}(T < C)$ and C is the censoring time. Suppose data for analysis consist of n independent and identically distributed copies of $\mathbb{D}_i = (X_i, \Delta_i, \Delta_i\eta_i, Y_i, \mathbf{Z}_i)$ for $i = 1, \dots, n$. For censoring, we require the

standard assumption that C_j is independent of T_j conditional on Y_j and $\mathbf{Z}_j$. By convention we assume that larger values of the marker Y are associated with higher risks at each levels of $\mathbf{Z}$.

2.2 Conditional Accuracy Summaries under Competing Risk

In the absence of competing risks, different versions of time-dependent accuracy measures can be considered in a variety of medical settings (Heagerty and Zheng, 2005). For example, cases (i.e., $D(t) = 1$) are *cumulative* if they are defined as those with $T \leq t$, *incident* if $T = t$, for a clinical event time T . In addition, controls (i.e., $D(t) = 0$) are *dynamic* if they are defined as $T > t$ and *static* if $T > \tau$, for a fixed time τ . In this manuscript, we illustrate our methods with cumulative cases and dynamic controls, but note that the methods can be easily extended to other versions of measures. In addition, we focus on an evaluation that is conditional on covariate value $\mathbf{z}$. The conditional accuracy measures characterize a marker's capacity to discriminate between subjects with $T \leq t$, versus those with $T > t$, within the sub-population with covariates $\mathbf{z}$. These quantities provide useful tools for evaluating the clinical utilities of a marker in classifying patients with potential good and bad outcomes at a future time, and also pin-pointing the factors that affect the performance of the biomarker under evaluation.

In the presence of competing risks, specific to a given time t and the j th event type, Saha and Heagerty (2010) defined cases as those who developed the j th event by t and controls as those who are free of any event by t , i.e.

$$\text{Case } j: T \leq t, \eta = j; \text{Control: } T > t.$$

Following such a definition, among subjects with covariate $\mathbf{z}$, the TPF function of Y specific to event type $\eta = j$ at time t is

$$\text{TPF}_t^{(j)}(c; \mathbf{z}) = P\{Y \geq c | T \leq t, \eta = j, \mathbf{Z} = \mathbf{z}\};$$

while the FPF of Y is defined based on those subjects with $T > t$ across all event types:

$$\text{FPF}_t(c; \mathbf{z}) = P\{Y \geq c | T > t, \mathbf{z}\}.$$

A cause-specific ROC curve for event type j conditional on $\mathbf{z}$ is defined as

$\text{ROC}_t^{(j)}(p; \mathbf{z}) = \text{TPF}_t^{(j)}\{\text{FPF}_t^{-1}(p; \mathbf{z}); \mathbf{z}\}$. In parallel, given covariate $\mathbf{z}$, the PPV of Y specific to event type j at time t is

$$\text{PPV}_t^{(j)}(v; \mathbf{z}) = P\{T \leq t, \eta = j | H(Y) > v, \mathbf{Z} = \mathbf{z}\},$$

while the NPV is defined based on those subjects with $T > t$ across all event types:

$$\text{NPV}_t(v; \mathbf{z}) = P\{T > t | H(Y) > v, \mathbf{Z} = \mathbf{z}\}.$$

Conditional on $\mathbf{Z} = \mathbf{z}$, if Y is useful for discriminating cases who will develop event type j within time t since baseline from those who do not, one would expect $\text{ROC}_t^{(j)}(p; \mathbf{z})$ to be

higher but $\text{ROC}_t^{(j)}(p; \mathbf{z})$ to be lower for $j' < j$. Similarly, the slope of $\text{PPV}_t^{(j)}(v; \mathbf{z})$ is steeper relative to those of $\text{PPV}_t^{(j')}(v; \mathbf{z})$ for other competing risk events.

There are practical applications where other “case” and “control” definitions may be useful. In some applications those who die of other causes prior to t may be considered as controls, for example when these subjects experienced an event that is unlikely to be related to the event of interest. Signaling them as “positive” may be considered a medical error. Thus, the cases and controls can be defined as

$$\text{Case: } T \leq t, \eta = j; \text{Control: } (T > t) \cup (T \leq t \cap \eta \neq j).$$

Here to establish the accuracy of Y for predicting the j th event type, one only need to consider the corresponding ROC curve, $\text{ROC}_t^{*(j)}(p; \mathbf{z}) = \text{TPF}_t^{(j)} \left[\{\text{FPF}_t^{(-j)}(p; \mathbf{z})\}^{-1}; \mathbf{z} \right]$, where

$$\text{FPF}_t^{(-j)}(c; \mathbf{z}) = P\{Y \geq c | (T > t) \cup (T \leq t \cap \eta \neq j), \mathbf{Z} = \mathbf{z}\}.$$

Similarly, for prospective accuracy we can consider jointly $\text{PPV}_t^{(j)}(v, \mathbf{z})$ and

$$\text{NPV}_t^{(-j)}(v; \mathbf{z}) = P\{(T > t) \cup (T \leq t \cap \eta \neq j) | H(Y) > v, \mathbf{Z} = \mathbf{z}\}.$$

Note that $P\{(T > t) \cup (T \leq t \cap \eta \neq j)\} \equiv 1 - P\{T \leq t, \eta = j\}$ is the probability of surviving event j up to time t . With such definitions, one can directly evaluate the marker's performance in separating subjects who experience event of interest by time t and all other subjects by inspecting a single ROC curve, or PPV/NPV curve for the event of interest.

3. Estimation

When all the components of $\mathbf{Z}$ are discrete, one may estimate the cause-specific conditional ROC curves based on the nonparametric procedures developed by Saha and Heagerty (2010) using observations in $\{\mathbf{D}_i: 1 \leq i \leq n, \mathbf{Z}_i = \mathbf{z}\}$. However, fully nonparametric approaches are not feasible when $\mathbf{Z}$ contains continuous covariates, especially if $p \geq 3$. Here, we propose flexible semiparametric procedures to accommodate more general structures of $\mathbf{Z}$. To this end, we note that the accuracy measures of interest can be written as

$$\text{TPF}_t^{(j)}(c; \mathbf{z}) = \frac{\int_c^\infty \mathcal{F}_{y, \mathbf{z}}^{(j)}(t) dH_{\mathbf{z}}(y)}{\int_{-\infty}^\infty \mathcal{F}_{y, \mathbf{z}}^{(j)}(t) dH_{\mathbf{z}}(y)}, \quad (3.1)$$

$$\text{FPF}_t(c; \mathbf{z}) = \frac{\int_c^\infty \{1 - \sum_{j=1}^J \mathcal{F}_{y, \mathbf{z}}^{(j)}(t)\} dH_{\mathbf{z}}(y)}{\int_{-\infty}^\infty \{1 - \sum_{j=1}^J \mathcal{F}_{y, \mathbf{z}}^{(j)}(t)\} dH_{\mathbf{z}}(y)}, \text{ and} \quad (3.2)$$

$$\text{FPF}_t^{(-j)}(c; \mathbf{z}) = \frac{\int_c^\infty \{1 - \mathcal{F}_{y, \mathbf{z}}^{(j)}(t)\} dH_{\mathbf{z}}(y)}{\int_{-\infty}^\infty \{1 - \mathcal{F}_{y, \mathbf{z}}^{(j)}(t)\} dH_{\mathbf{z}}(y)}, \quad (3.3)$$

where $\mathcal{F}_{y,z}^{(j)}(t) = P(T \leq t, \eta = j | Y = y, \mathbf{Z} = \mathbf{z})$ is the conditional cumulative incidence function for event type j , $H_z(y) = P(Y \leq y | \mathbf{Z} = \mathbf{z})$. Similarly, for cause-specific prospective accuracy, we have

$$\text{PPV}_t^{(j)}(v; \mathbf{z}) = \frac{\int_{H_z^{-1}(v)}^{\infty} \mathcal{F}_{y,z}^{(j)}(t) dH_z(y)}{1 - v}, \text{ and} \quad (3.4)$$

$$\text{NPV}_t^{(j)}(v; \mathbf{z}) = \frac{\int_{-\infty}^{H_z^{-1}(v)} \{1 - \sum_{j=1}^J \mathcal{F}_{y,z}^{(j)}(t)\} dH_z(y)}{v}, \text{ or} \quad (3.5)$$

$$\text{NPV}_t^{(-j)}(v; \mathbf{z}) = \frac{\int_{-\infty}^{H_z^{-1}(v)} \{1 - \mathcal{F}_{y,z}^{(j)}(t)\} dH_z(y)}{v}. \quad (3.6)$$

To make inference about these accuracy measures, we first note that $H_z(y)$ can be estimated nonparametrically based on empirical distribution functions of Y when $\mathbf{Z}$ is discrete or semiparametrically by imposing flexible models such as the location-scale model for continuous $\mathbf{Z}$. See Heagerty and Pepe (1999) and Zheng et al. (2010) for details of such approaches. We let $\hat{H}_z(y)$ denote the estimator of $H_z(y)$. The other key quantity here is the conditional cumulative incidence function for cause j , $\mathcal{F}_{y,z}^{(j)}(t)$. We detail two approaches for estimating this quantity below.

Estimation with Cause-specific Cox Regression Models

The standard analysis involves modeling cause-specific hazard functions of different causes of failure under a Cox PH assumption. This model could be equivalently specified, via the conditional *cumulative* cause-specific hazard form, as $\Lambda_{yz}^{(j)}(t) = \Lambda_0^{(j)}(t) \exp\{\alpha Y + \mathbf{Z}^\top \beta^{(j)}\}$, where $\Lambda_0^{(j)}(t)$ is the baseline cause-specific cumulative hazard function for event type j (Prentice et al., 1978). These cumulative hazard functions relate to $\mathcal{F}_{y,z}^{(j)}(t)$ through

$\mathcal{F}_{y,z}^{(j)}(t) = \int_0^t \mathcal{S}_{y,z}(u) \Lambda_{yz}^{(j)}(du)$, where $\mathcal{S}_{y,z}(u) = \exp\{-\sum_{j=1}^J \Lambda_{yz}^{(j)}(u)\}$ is the overall survival function. Under the cause-specific PH model, one may obtain a consistent estimate of

$\mathcal{F}_{y,z}^{(j)}(t)$, $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ based on the standard partial likelihood, treating event times of other types of event as censored observations. Combining $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ along with $\hat{H}_z(v)$, we may construct plug in estimators of the cause-specific accuracy summaries,

$\text{TPF}_t^{(j)}(v; \mathbf{z})$, $\text{FPF}_t(v; \mathbf{z})$, $\text{FPPF}_t^{(-j)}(v; \mathbf{z})$, $\text{PPV}_t^{(j)}(v; \mathbf{z})$, $\text{NPV}_t(v; \mathbf{z})$, $\text{NPV}_t^{(-j)}(v; \mathbf{z})$, based on equations (3.1), (3.2), (3.3), (3.4)(3.5), (3.6), respectively. For example, an estimator for $\text{PPV}_t^{(j)}(v; \mathbf{z})$ is $\hat{\text{PPV}}_t^{(j)}(v; \mathbf{z}) = (1 - v)^{-1} \int_{\hat{H}_z^{-1}(v)}^{\infty} \hat{\mathcal{F}}_{y,z}^{(j)}(t) d\hat{H}_z(y)$. The estimation procedures for ROC curve in the presence of competing risk follow similarly, i.e., $\text{ROC}_t^{(j)}(p; \mathbf{z}) = \text{TPF}_t^{(j)}\{\text{FPF}_t^{-1}(p; \mathbf{z}); \mathbf{z}\}$, or $\text{ROC}_t^{*(j)}(p; \mathbf{z}) = \text{TPF}_t^{(j)}\left[\{\text{FPPF}_t^{(-j)}(p; \mathbf{z})\}^{-1}; \mathbf{z}\right]$.

Estimation with Cause-specific Smoothed Cox Regression Model

The PH model we considered above requires the PH assumption to hold for the effects of both Y and Z on $\mathcal{T}$ across *all* types of failure, which is much stronger than that in the setting without competing risks (Fine, 2001). When the marker effect is nonlinear or time-varying,

the resulting estimator $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ may be biased and thus lead to biased estimates of the accuracy summaries. To overcome such difficulties, more flexible models such as the smoothed Cox regression model (Dabrowska, 1997) would be particularly appealing. We propose to develop estimation procedures by considering a cause-specific smoothed cumulative hazard function $\Lambda_{y,z}^{(j)}(t) = \Lambda_{0y}^{(j)}(t) \exp(\mathbf{Z}^\top \beta^{(j)})$. The model assumes a parametric functional form only for effects of Z on cause-specific hazards but leaves the effect of Y unspecified. This allows for possibly nonlinear and time-varying effect for the marker while simultaneously accounting for additional covariates. To obtain an estimator for $\beta^{(j)}$, we propose to maximize a smoothed cause-specific profile likelihood and consider $\hat{\beta}^{(j)} = \arg\max_{\beta} C^{(j)}(\beta)$, where

$$C^{(j)}(\beta^{(j)}) = 1/n \sum_{i=1}^n \int \int \{\beta^{(j)\top} \mathbf{z}_i - \log \hat{\pi}_y^{(j)}(t, \beta^{(j)})\} d\hat{N}_y^{(j)}(t) dy,$$

$\hat{N}_y^{(j)}(s) = n^{-1} \sum_{i=1}^n K_h(Y_i - y) N_i^{(j)}(s)$, $\hat{\pi}_y^{(j)}(s, \beta^{(j)}) = n^{-1} \sum_{i=1}^n K_h(Y_i - y) I(X_i \geq s) \exp(\mathbf{z}_i^\top \beta^{(j)})$ with $N_i^{(j)}(s) = I(X_i \leq s, \eta_i = j, \Delta_i = 1)$, and $K_h(\cdot) = K\{(y - Y_i)/h\}/h$. Here K is a symmetric kernel density function and the bandwidth $h = O(n^{-\nu})$ with $\nu \in (1/5, 1/2)$. We then propose to obtain an estimator for the smoothed cumulative incidence function as

$$\hat{\mathcal{F}}_{y,z}^{(j)}(t) = \int_0^t \hat{\mathcal{S}}_{y,z}(u) \frac{\exp(\mathbf{z}^\top \hat{\beta}^{(j)}) d\hat{N}_y^{(j)}(s)}{\hat{\pi}_y^{(j)}(s, \hat{\beta}^{(j)})}, \text{ with } \hat{\mathcal{S}}_{y,z}(u) = \exp \left\{ - \int_0^u \sum_{j=1}^J \frac{\exp\{\mathbf{z}^\top \hat{\beta}^{(j)}\} d\hat{N}_y^{(j)}(s)}{\hat{\pi}_y^{(j)}(s, \hat{\beta}^{(j)})} \right\}.$$

Plugging in $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ and $\hat{H}_z(y)$ to (3.1), (3.2), (3.3), (3.4)(3.5), (3.6) yields estimators of the proposed accuracy summaries $\widehat{\text{TPF}}_t^{(j)}(v; \mathbf{z})$, $\widehat{\text{FPF}}_t^{(j)}(v; \mathbf{z})$, $\widehat{\text{FPF}}_t^{(-j)}(v; \mathbf{z})$, $\widehat{\text{PPV}}_t^{(j)}(v; \mathbf{z})$, $\widehat{\text{NPV}}_t^{(j)}(v; \mathbf{z})$, and $\widehat{\text{NPV}}_t^{(-j)}(v; \mathbf{z})$, respectively. For example, $\widehat{\text{PPV}}_t^{(j)}(v; \mathbf{z})$ can be estimated as

$$\widehat{\text{PPV}}_t^{(j)}(v; \mathbf{z}) = (1 - v)^{-1} \int_{\hat{H}_z^{-1}(v)}^{\infty} \hat{\mathcal{F}}_{y,z}^{(j)}(t) d\hat{H}_z(y).$$

4. Inference

We studied the asymptotic properties of the proposed cause-specific accuracy summary estimators. Similar to the estimation, the key step to our theoretical development is the establishment of asymptotic theory for the cumulative incidence function $\mathcal{F}_{y,z}^{(j)}(t)$. Under the cause-specific PH model, Cheng et al. (1998) provided procedures for making inference about the cumulative incidence function based on $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$. For the smoothed cause-specific PH model, the asymptotic derivations for $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ can be obtained by following similar arguments given in Dabrowska (1997) and Zheng et al. (2010) under the smoothed PH model without competing risks. Specifically, in Appendix A of Supplementary Materials available online, we first provide asymptotic expansion for smoothed cumulative incidence

function $\hat{\mathcal{F}}_{y,z}^{(j)}(t)$ and the bivariate function $\hat{\mathcal{F}}_{y,z}^{(j)}(t, \mathbf{c}; \mathbf{z}) = \int_{\mathbf{c}}^{\infty} \hat{\mathcal{F}}_{y,z}^{(j)}(t) d\hat{H}_z(y)$. We show that

process $\mathcal{W}_{\mathcal{F}}^{(j)}(t, c; \mathbf{z}) = n^{\frac{1}{2}} \{ \hat{\mathcal{F}}^{(j)}(t, c; \mathbf{z}) - \mathcal{F}^{(j)}(t, c; \mathbf{z}) \}$ is asymptotically equivalent to a sum of independent and identically distributed (i.i.d) terms, $n^{-\frac{1}{2}} \sum_{i=1}^n \xi_{\mathcal{F}_i}^{(j)}(t, c; \mathbf{z})$, which converges in distribution to a Gaussian process. Subsequently the results, or the asymptotic results from Cheng et al. (1998) can be used as building blocks for establishing the asymptotic normality of various accuracy summary estimators based on either smoothed or ordinary cause-specific Cox proportional hazards models, respectively. We show in Appendix B that for a generic shorthand $\mathcal{A}$ representing either $\text{TPF}_t^{(j)}$, FPF_t , $\text{FPF}_t^{(-j)}$, $\text{ROC}_t^{(j)}$, $\text{PPV}_t^{(j)}$, NPV_t , or $\text{NPV}_t^{(-j)}$ the processes $\mathcal{W}_{\mathcal{A}}(\cdot; \mathbf{z}) = n^{\frac{1}{2}} \{ \hat{\mathcal{A}}(\cdot; \mathbf{z}) - \mathcal{A}(\cdot; \mathbf{z}) \}$ is asymptotically equivalent to a sum of independent and identically distributed (i.i.d) terms, $n^{-\frac{1}{2}} \sum_{i=1}^n \zeta_{\mathcal{A}_i}(\cdot, \mathbf{z})$, where $\zeta_{\mathcal{A}_i}(\cdot, \mathbf{z})$ is defined in Appendix B for each of the accuracy summary measures. Furthermore, we can show by the functional central limit theorem (Pollard, 1990) that the process $\mathcal{W}_{\mathcal{A}}(\cdot; \mathbf{z})$ converges weakly to a mean zero Gaussian process in (t, c, p, v) . To construct point-wise and simultaneous confidence intervals for cause-specific ROC curve and PPV, NPV curves, one can follow a resampling procedure that was proposed by Park and Wei (2003) using the theoretical results derived here. Since theoretical variances estimated empirically require density estimation, bootstrap procedure can be used as an alternative approach for variance calculation. We employ the bootstrap-based variance estimators in our simulation studies.

5. Simulation

We conduct numerical studies to assess the performance of the estimators under two scenarios. We consider the sample size $n = 2000$ and the results are summarized based on 2000 replications for each configuration. We first consider the case when the hazards of all events depend on the marker Y and a single covariate Z through the PH models of the form $\Lambda^{(1)}(t) = t \exp\{\log(5)Y + \log(1.5)Z\}/10$, for event of interest, and $\Lambda^{(2)}(t) = t \exp\{\log(1.25)Y + \log(0.9)Z\}/10$ for the competing risk events. We generate the censoring time from a uniform distribution with mean 25. This yields a censoring rate of 25% approximately. The covariate Z is a binary random variable from a Bernoulli distribution with probability equal to 0.5. The marker Y is generated from a truncated normal with $Y = \text{sign}(Y_0) \min(3, |Y_0|)$ and $Y_0 \sim N(0.5Z, 1)$. We calculate ROC curve values for $Z = 0, 1$ at $t = 3$, and for $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and NPV and PPV values for $v = 0.1, 0.3, 0.5, 0.7, 0.9$. We present results only for estimators based on the first definition of cases and controls specified in Section 2, and leave the results for estimators using the second definition in Appendix C online. We evaluate the performance of $\text{ROC}_t^{(j)}(p; \mathbf{z})$ estimators: $\tilde{\text{ROC}}_t^{(j)}(p; \mathbf{z})$ using the cause-specific Cox regression model and $\widehat{\text{ROC}}_t^{(j)}(p; \mathbf{z})$ using the smoothed cause-specific Cox regression model. The smoothing bandwidth for $\widehat{\text{ROC}}_t^{(j)}(p; \mathbf{z})$ is taken to be $h = n^{-1/3} \text{var}(Y)^{1/2}$. Results are presented in Table 1 for retrospective accuracy measures and Table 2 for prospective accuracy measures. As expected, estimators based on the cause-specific Cox regression model has little bias, and estimators based on the cause-specific smoothed Cox model also show negligible bias (all less than 5%). For all estimators, the estimated standard errors based on bootstrap track the empirical standard errors well. The coverage probabilities of pointwise confidence intervals are close to the nominal 95% level in most of the cases. We found in a few situations, most at the tail parts, estimators based on smoothed Cox regression models could have coverages below 95%, which are mostly due to the fact that standard errors are quite small relative to the absolute values of bias.

We consider a second scenario with Y_i generated from a conditional lognormal model, $Y_i = \exp(0.5Z_i)\rho_i$, where ρ and Z are random variables from standard log-normal and Bernoulli with probability equal to 0.5, respectively. In addition, $\mathcal{T}_{1i}$ the time to an event of interest,

is simulated with a hazard function of the form $\lambda^{(1)}(t) = \exp\{\log(t+1)Y + 1.5Z\}$, whereas $\mathcal{T}_{2h}$, the time to a competing risk event, follows an exponential model with $\lambda^{(2)}(t) = 0.2 \exp(1.15Y + 0.85Z)$. In addition to the two estimators used under the first scenario, we also consider a mixed approach, where we estimate $\mathcal{F}_{y,z}^{(1)}(t)$ from a cause-specific smoothed Cox model but use a cause-specific Cox model for estimating $\mathcal{F}_{y,z}^{(2)}(t)$. Since the effect of Y on T_1 varies as a function of t , the procedure assuming PH for both types of events would likely fail, and the mixed approach would be unbiased and more efficient than the procedure using smoothed cause-specific Cox models for both types of events. As shown in Tables 3 and 4, indeed substantial bias can be observed for estimators derived under the PH assumptions, whereas the estimators employing the more flexible assumptions have negligible bias. An estimator with mixed (and correct) assumptions in general has the smallest mean square error (MSE) for estimating both prospective and retrospective accuracy. The gain in efficiency is especially pronounced for accuracy summaries for competing risk events. Hence there is an advantage for considering more flexible modeling approaches as proposed here to achieve robustness. This aspect is important especially when formal model checking for survival regression model is not a routine practice.

6. Example: PSA for Prostate Cancer Prognosis

In the Pacific Northwest Prostate Cancer SPORE (Specialized Program of Research Excellence), a cohort of patients diagnosed with prostate cancer (PC) are under follow-up for clinical outcomes. Information on biomarkers such as serum prostate-specific antigen (PSA) at diagnosis, tumor grade (Gleason score) and other clinical variables have been collected. One of the goals of our prostate cancer research is to identify novel markers, used in combination with other existing markers and clinical factors, for improved predictions of PC-specific mortality so that treatment options can be tailored to specific individuals. The SPORE cohort consists of 752 incident PC cases who were diagnosed at ages 40–64 years in 1993–1996 and who are followed for disease progression and survival. The median followup time among patients alive is 12 years. We include 671 patients with PSA measures prior to prostate cancer primary treatment in our analysis. By the time the data are analyzed, 144 patients are deceased, with 54 deaths are due to PC, and a substantial number of subjects, 90 of them, died from other causes. This leads to a competing risk situation since the focus of our research is on death due to PC. Since PSA has been used as a prognostic marker in the past, it is a crucial first step to evaluate the predictive performance of PSA-based criteria for PC specific death. Such evaluation could shed light on the clinical utility of PSA for guiding treatment decisions since only subjects who are likely to die of PC should be recommended for aggressive treatment of PC. As the predictiveness of PSA may depend on other characteristics, we will also evaluate how the accuracy measures relate to tumor grade, categorized as low (Gleason score ≤ 6), moderate (Gleason score 3+4), and high (Gleason score 4+3 and > 7). We first consider a cause-specific Cox PH model,

$\Lambda_{y,z}^{(j)}(t) = \Lambda_0^{(j)}(t) \exp\{\alpha^{(j)} \log(\text{PSA}) + \beta_1^{(j)} I(\text{grade} = \text{moderate}) + \beta_2^{(j)} I(\text{grade} = \text{High})\}$, with $j = 1$ and 2 indicating PC-specific deaths and death due to other causes, respectively. Under such a model, pre-treatment PSA (logarithm transformed throughout this section) is a strong predictor for PC mortality, with a hazard ratio of 2.03 (95% CI: [1.70, 2.41]), comparing two individuals with the same tumor grade but 1 unit difference in PSA level on the log scale. We further calculated cause-specific summaries stratified by tumor grade for both classes of accuracy, using methods described in Section 3. We first consider summary measures based on defining cases by event types. This leads to separate ROC or PPV curves for PC-specific death and death due to other causes, respectively. As shown in Figure 1, PSA, as its name implied, has good discriminatory power for separating people who died due to PC versus these who are still alive $t_0 = 10$ years after diagnosis. For example, AUC of ROC curves at t_0 among subjects with high grade tumor is 0.76, (95% CI: [0.69, 0.83]), however the

corresponding AUC for other causes of death is 0.54 (95% CI: [0.45, 0.63]). For comparison, we also consider plotting accuracy based on death due to all-causes. It appeared that such an outcome, without acknowledging competing risk situation, underestimated the clinical performance of PSA, with an AUC of 0.65 (95% CI: [0.57, 0.72]).

Since NPVs for all groups do not appear to vary much over the entire range of PSA values we focus presentation below only on PPV. The covariate-specific PPV curves for v ranging from 0.05 to 0.95 from the proposed procedures are displayed in Figure 1 (bottom panels). Similar to ROC curves, we compare PPV curves for people who experience PC-specific death within $t_0 = 10$ years to PPV curves for people who experienced other types of death within 10 years. In addition, PPV curves for 10-year mortality without regard to failure types are also plotted. We observe that PPV curve increases relatively more steeply for PC-specific death outcome but to a much less degree for all causes of death, especially among a subgroup population with high grade tumor. In contrast, the averaged 10-year mortality risk for other causes of death varies very little across a wide range of PSA values used as positivity criteria for defining a positive test. For example, among patients with high grade tumor, $\widehat{PPV}_{10}$ is 0.24 with 95% CI: [0.16, 0.32] at $v = 0.1$; and 0.59 with 95% CI: [0.39, 0.75] at $v = 0.9$. Such information may be helpful for urologists to make decision on treatment options. For example, in practice it may be sensible to only select patients for more aggressive treatment if they are at high risk for short-term PC specific death. If 'high' risk is gauged by an average risk of 20% for 10-year PC mortality, then based on cause-specific PPV curves in Figure 1, we would opt to give the treatment to all patients with high grade tumor, and only those patients with moderate tumor type and a PSA value above the top 20% in the population, since $\widehat{PPV}_{10}$ is about 0.2 when $v = 0.8$ for this group.

Since PSA does not predict other causes of death, it is reasonable to consider those who died from other causes before 10 years as 'controls' when evaluating the performance of PSA. We therefore also calculated accuracy summaries of PSA according to our second definitions of cases and controls described in Section 2. The estimates are very similar to the estimates for PC-specific death outcome obtained according to the first definition (Web Appendix D).

7. Discussion

In this manuscript we present statistical methods for characterizing time-dependent classification accuracy summaries of a prognostic marker when there are potentially many causes of failure. To incorporate competing risk, we give two definitions of cause-specific accuracy summaries. In the first definition cases are stratified by causes of failure and controls are a common group of patients who remain free of any events at the prediction time. Such a definition would allow one to evaluate the accuracy of a marker for each type of event, and the effectiveness of a marker for selecting the targeted event is gauged by a contrast of the accuracy for the event to that of other competing risk events. The second definition simply classifies patients experience the event of interest as cases by the prediction time and treats people who experience competing risk events or who have not experienced any events by that time as controls. The adequacy of these definitions depends on specific clinical settings and the nature of clinical events considered. For example, if a biomarker is used to help deciding whether a surgery is needed for a subject with prostate cancer, then flagging positive an individual died due to cardiovascular disease may lead to unnecessary surgical procedure. In this case the second definition, treating these individual as controls, can be considered. In practice it is prudent to first calculate accuracy summaries that stratify by event types to understand how marker performs for different events. If the marker is not sensitive for competing risk events or the competing risk events are not related to the event of interest, one may consider accuracy summaries based on the second

definition to make a direct inference regarding the markers performance on the targeted event.

Prior to applying a marker in clinical setting, it is often important to characterize its performance in different subpopulation. We therefore focus on developing procedures for making inference about conditional cause-specific accuracy summaries, extending the non-parametric cause-specific ROC curve estimation (Saha and Heagerty, 2010) for the setting without covariates. Our simulations indicate that the standard cause-specific PH model, albeit efficient, can induce substantial bias in all accuracy summaries when the PH assumption does not hold. Although the methods presented in this paper focus on two classes of classification accuracy, they can be readily extended to other summaries such as the integrated discrimination improvement index (Pencina et al., 2008).

A key quantity for deriving our time-dependent accuracy summaries is the cumulative incidence function. Our proposed procedures were derived through the most commonly used PH framework by modeling the covariate effects on the hazard rates. Alternatively, one may direct model the covariate effects on cumulative incidence functions as in Fine and Gray (1999) and Scheike et al. (2008). For example, Scheike et al. (2008) proposed binomial regression models for $P(T \leq t, \eta = k | Y, \mathbf{Z})$ using an inverse probability weighted method to account for censoring. Such an approach would lead to computationally simpler and more flexible cause-specific accuracy summaries.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant U01-CA86368, P01-CA053996, R01-GM085047 and R01-GM079330 awarded by the National Institutes of Health.

References

- Cai T, Pepe MS, Zheng Y, Lumley T, Jenney N. The sensitivity and specificity of markers for event time. *Biostatistics*. 2006; 7:182–197. [PubMed: 16079162]
- Cheng S, Fine J, Wei L. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998; 54:219–228. [PubMed: 9544517]
- Dabrowska D. Smoothed Cox regression. *The Annals of Statistics*. 1997; 25:1510–1540.
- Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics*. 2001; 2:85–97. [PubMed: 12933558]
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509.
- Heagerty P, Zheng Y. Survival model accuracy and roc curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
- Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–344. [PubMed: 10877287]
- Heagerty PJ, Pepe MS. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Applied Statistics*. 1999; 48:533–551.
- Li J, Fine J. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics*. 2008; 9:566. [PubMed: 18304996]
- Moskowitz C, Pepe M. Quantifying and comparing the accuracy of binary biomarkers when predicting a failure time outcome. *Statistics in medicine*. 2004; 23:1555–1570. [PubMed: 15122736]
- Park Y, Wei LJ. Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika*. 2003; 90:717–23.

- Pencina M, D'Agostino R Sr, D'Agostino R Jr, Vasan R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*. 2008; 27:157–172. [PubMed: 17569110]
- Pepe M, Mori M. *Statistics in Medicine*. 1993; 12:737–751. [PubMed: 8516591]
- Pepe M, Zheng Y, Jin Y, Huang Y, Parikh C, Levy W. Evaluating the ROC performance of markers for future events. *Lifetime Data Analysis*. 2008; 14:86–113. [PubMed: 18064569]
- Pollard, D. *Empirical processes: theory and applications*. Institute of Mathematical Statistics; 1990.
- Prentice R, Kalbeisch J, Peterson A Jr, Flournoy N, Farewell V, Breslow N. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34:541–554. [PubMed: 373811]
- Saha P, Heagerty P. Time-Dependent Predictive Accuracy in the Presence of Competing Risks. *Biometrics*. 2010:999–1011. [PubMed: 20070296]
- Scheike T, Zhang M, Gerds T. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*. 2008; 95:205.
- Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*. 1975; 72:20. [PubMed: 1054494]
- Zheng Y, Cai T, Pepe M, Levy W. Time-dependent predictive values of prognostic biomarkers with failure time outcome. *Journal of the American Statistical Association*. 2008; 103:362–368. [PubMed: 19655041]
- Zheng Y, Cai T, Stanford J, Feng Z. Semiparametric Models of Time-Dependent Predictive Values of Prognostic Biomarkers. *Biometrics*. 2010; 66:50–60. [PubMed: 19397579]
- Zheng Y, Heagerty P. Prospective accuracy for longitudinal markers. *Biometrics*. 2007; 63:332–341. [PubMed: 17688486]

Figure 1.

Time-dependent ROC curves (top panel) and PPV curves (bottom panel) of PSA by different values of Gleason score at $t = 10$ year after diagnosis in the prostate cancer SPORC study. Solid lines: cause-specific curves for prostate cancer specific death; dashed lines: cause-specific curves for death due to other causes; dotted lines: curves for all causes of death. Shaded area: 95% confidence simultaneous band for the cause-specific curve.

Table 1

Simulation results: percent of bias ((Estimate-True)/True×100), standard errors×10³ (Sse), estimated standard errors×10³ (Ase) of area under the ROC curve ($AUC_t^j(z)^*$) and $ROC_t^j(u;z)$ under PH models for different types of events.

		$AUC_t^1(z)$		$ROC_t^1(u;z)$		$AUC_t^2(z)$		$ROC_t^2(u;z)$	
	v	0.1	0.5	0.9	0.9	0.1	0.5	0.9	0.9
z = 0									
Cox	True	0.88	0.67	0.94	0.99	0.62	0.24	0.66	0.95
	Bias(%)	-0.10	-0.38	-0.03	-0.02	-0.19	-0.63	-0.19	-0.13
	Sse	0.64	1.70	0.49	0.05	1.52	1.71	1.98	0.60
	Ase	0.64	1.70	0.49	0.04	1.48	1.67	1.92	0.59
	Cov	0.95	0.94	0.95	0.99	0.94	0.94	0.94	0.94
	Smooth								
Bias(%)	-0.81	-2.58	-0.31	-0.10	-1.08	-4.86	-0.92	-0.39	
Sse	0.92	2.51	1.03	0.20	2.00	2.57	3.15	1.31	
Ase	0.94	2.56	1.09	0.19	1.98	2.6	3.13	1.31	
Cov	0.91	0.91	0.96	0.90	0.93	0.91	0.94	0.95	
z = 1									
Cox	True	0.90	0.73	0.95	0.99	0.67	0.32	0.70	0.95
	Bias(%)	-0.22	-0.48	-0.06	0.00	-0.46	-1.38	-0.38	-0.09
	Sse	0.51	1.37	0.4	0.05	1.46	1.94	1.79	0.51
	Ase	0.54	1.43	0.42	0.04	1.43	1.91	1.75	0.49
	Cov	0.95	0.95	0.95	0.97	0.94	0.93	0.94	0.94
	Smooth								
Bias(%)	-1.37	-3.72	-0.37	-0.04	-2.05	-8.49	-1.45	-0.26	
Sse	0.76	2.32	0.82	0.12	1.83	3.05	2.77	0.97	
Ase	0.77	2.38	0.84	0.12	1.85	3.13	2.80	0.97	
Cov	0.88	0.80	0.95	0.94	0.89	0.86	0.94	0.95	

* : j = 1 for event of interest, j=2 for competing risk event.

Table 2

Simulation results: percent of bias ((Estimate-True)/True×100), standard errors×10³ (Sse), estimated standard errors×10³ (Ase) of $PPV_t^j(v;z)^*$ and $NPV_t^j(v;z)$ under PH regression models for different types of events.

NPV _t (v; z)													
PPV _t ¹ (v; z)				PPV _t ² (v; z)									
v	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.9
z = 0													
Cox													
True	0.35	0.55	0.89	0.20	0.18	0.09	0.80	0.69	0.54				
Bias(%)													
Sse	-0.15	-0.20	-0.27	-0.67	-0.29	-0.30	-0.20	0.23	0.22				
Ase	1.45	2.02	1.54	1.16	1.24	1.17	1.89	1.58	1.56				
Cov	1.48	2.08	1.60	1.18	1.26	1.17	1.86	1.56	1.57				
Smooth													
Bias(%)	-1.15	-1.34	-2.50	-1.65	-1.41	-1.58	-0.49	0.40	0.76				
Sse	1.46	2.12	2.13	1.17	1.40	1.83	3.25	1.79	1.57				
Ase	1.48	2.16	2.22	1.19	1.41	1.87	3.24	1.81	1.58				
Cov	0.94	0.94	0.87	0.94	0.94	0.94	0.95	0.94	0.94				
z = 1													
Cox													
True	0.60	0.83	0.97	0.14	0.10	0.03	0.76	0.55	0.34				
Bias(%)													
Sse	-0.31	-0.30	-0.34	-0.29	-0.11	-1.44	-0.06	0.20	0.35				
Ase	1.54	1.42	0.51	0.99	0.89	0.52	1.88	1.86	1.44				
Cov	1.56	1.45	0.67	0.98	0.89	0.52	1.85	1.89	1.46				
Smooth													
Bias(%)	-1.52	-1.62	-1.98	-0.90	-0.67	5.38	0.00	0.79	2.29				
Sse	1.56	1.65	1.21	1.00	1.08	1.20	2.65	1.97	1.44				
Ase	1.58	1.64	1.29	0.99	1.07	1.20	2.69	2.03	1.47				
Cov	0.91	0.88	0.80	0.94	0.94	0.93	0.95	0.95	0.93				

*: j = 1 for event of interest, j=2 for competing risk event.

Table 3

Simulation Results: % bias and mean squared errors (MSE) for estimates of $AUC_t^j(z)^*$ and $ROC_t^j(v|z)$ under a nonPH regression model for the event of interest.

v	AUC _t ¹ (z)		ROC _t ¹ (v z)		AUC _t ² (z)		ROC _t ² (v z)		
	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
z = 0									
True	0.62	0.27	0.65	0.94	0.61	0.25	0.63	0.93	
% bias									
Cox	-5.0	-18.3	-5.2	-0.9	-3.8	-15.9	-3.5	-0.6	
Smooth	-2.3	-8.7	-2.0	-0.6	-1.5	-5.8	-1.3	-0.4	
Mix	-2.2	-8.3	-2.0	-0.6	-1.1	-4.7	-0.8	-0.3	
MSE*10 ⁵									
Cox	105	264	122	7	60	180	60	4	
Smooth	48	103	87	11	80	127	177	25	
Mix	47	99	86	10	17	47	22	2	
z = 1									
% bias									
Cox	-8.3	-24.9	-8.5	-1.4	-3.4	-13.7	-3.1	-0.4	
Smooth	-1.4	-4.3	-1.2	-0.3	-1.5	-4.0	-1.6	-0.3	
Mix	-1.7	-5.3	-1.5	-0.4	0.2	1.2	0.1	0.1	
MSE*10 ⁵									
Cox	350	996	402	18	69	234	66	3	
Smooth	34	120	45	4	71	167	146	16	
Mix	38	132	49	4	25	73	28	1	

* : j = 1 for event of interest, j=2 for competing risk event.

Table 4

Simulation results: estimates (Est.), standard errors (SE), mean squared errors (MSE) of $PPV_t^j(v; z)^*$ and $NPV_t(v; z)$ under a nonPH regression model for the event of interest.

v	PPV _t ¹ (v; z)					PPV _t ² (v; z)					NPV _t (v; z)				
	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
z=0															
True	0.49	0.55	0.75	0.09	0.10	0.13	0.54	0.51	0.47						
%bias															
Cox	-3.5	-6.2	-10.3	1.5	0.7	-5.3	-2.5	-1.5	1.1						
Smooth	-2.7	-3.7	-6.8	-0.4	-0.4	1.7	-0.6	0.7	1.7						
Mix	-2.7	-3.7	-6.7	0.6	0.9	2.4	-0.5	0.6	1.6						
MSE*10 ⁵															
Cox	55	146	653	9	10	25	48	35	30						
Smooth	43	81	359	9	13	46	115	48	35						
Mix	43	81	340	8	10	30	102	47	34						
z=1															
true	0.2	0.26	0.53	0.12	0.14	0.24	0.81	0.79	0.75						
%bias															
Cox	5.5	-2.2	-15.9	-3.5	-5.0	-12	-4.1	-4.0	-3.1						
Smooth	-0.4	-1.3	-4.2	-1.3	-2.5	-3.5	-0.7	-0.4	-0.2						
Mix	-1.3	-2.5	-7.1	1.4	1.8	5.3	-0.6	-0.4	-0.2						
MSE*10 ⁵															
Cox	25	22	800	11	19	146	131	120	72						
Smooth	14	30	277	11	22	160	38	20	19						
Mix	15	33	344	11	17	116	23	17	19						

* : j = 1 for event of interest, j=2 for competing risk event.