Published in final edited form as:

Biom J. 2011 March ; 53(2): 294–307. doi:10.1002/bimj.201000150.

Incorporating short-term outcome information to predict long-term survival with discrete markers

Layla Parast^{*,1}, Su-Chun Cheng², and Tianxi Cai¹

¹Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

²Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA

Abstract

In disease screening and prognosis studies, an important task is to determine useful markers for identifying high-risk subgroups. Once such markers are established, they can be incorporated into public health practice to provide appropriate strategies for treatment or disease monitoring based on each individual's predicted risk. In the recent years, genetic and biological markers have been examined extensively for their potential to signal progression or risk of disease. In addition to these markers, it has often been argued that short-term outcomes may be helpful in making a better prediction of disease outcomes in clinical practice. In this paper we propose model-free non-parametric procedures to incorporate short-term event information to improve the prediction of a long-term terminal event. We include the optional availability of a single discrete marker measurement and assess the additional information gained by including the short-term outcome. We focus on the semi-competing risk setting where the short-term event is an intermediate event that may be censored by the terminal event while the terminal event is only subject to administrative censoring. Simulation studies suggest that the proposed procedures perform well in finite samples. Our procedures are illustrated using a data set of post-dialysis patients with end-stage renal disease.

Keywords

Biomarkers; Disease prognosis; Risk prediction; Survival analysis

1 Introduction

A common goal in disease screening and prognosis studies is to determine useful markers for identifying high-risk individuals. Such markers could be used in clinical practice as bases for selecting optimal prevention or treatment strategies. The recent advancement in science and technology has led to the discovery of many genetic and biological markers that hold great potential to signal progression or risk of disease (Bauer *et al.*, 2007; Pepe and Thompson, 2000; Wilson *et al.*, 1998). In addition to these markers, it has often been argued that incorporating short-term outcomes may further improve the prediction accuracy. Much recent progress has been made to investigate the predictive potential of short-term outcomes

Conflict of Interest

The authors have declared no conflict of interest.

in the clinical literature. For example, the extent to which short-term outcomes can predict long-term outcomes for pediatric polytrauma patients was examined by Van der Sluis *et al.* (1997). Weisner *et al.* (2003) demonstrated a clear association between short-term and long-term treatment success for alcohol addiction. Cortese and Andersen (2010) and Lee *et al.* (2002) examined how the development of graft versus host disease affects the survival of patients who received bone marrow or stem cell transplantation for acute leukemia. To provide potential treatment options over time and long-term trajectories for various types of diseases, one may develop prediction rules for long-term outcomes by incorporating short-term response into therapy. Therefore, it may be of clinical interest to include information about a short-term outcome in addition to marker measurements to improve prediction of long-term survival.

Most existing methods for incorporating censored short-term outcome information in predicting long-term survival focus on modeling the disease process and are derived under semi-parametric models in a multi-state survival setting (Hatteville *et al.*, 2002; Kay, 1986; Klein *et al.*, 1994; Putter *et al.*, 2006). Datta *et al.* (2000) provide non-parametric estimators for the stage-occupation probabilities of the three-stage irreversible illness-death model in which an individual may die with or without ever experiencing the short-term event of interest. Under a semi-competing risk setting with no covariates, Fine *et al.* (2001) proposed a model-based approach to estimate the marginal distribution of the long-term terminal event. Gray (1994) studied the potential gain in power for testing the difference in the overall survival between two groups from using short-term disease progression outcome information. Van Houwelingen (2007) proposed a landmark prediction procedure to estimate survival using information about whether or not the event has occurred by a given landmark time point. However, this procedure only considers the case when the short-term and long-term outcomes are the same clinical event. Van Houwelingen and Putter (2008) compared this landmark procedure and multi-state modeling using time-dependent covariates in an application to acute lymphoid leukemia data. Recently, Parast and Cai (2010) provided procedures for constructing and evaluating landmark prediction rules for a long-term outcome among patients who survived past a landmark point without experiencing a short-term event.

In this paper, we propose model-free non-parametric procedures to predict the risk of failure for the long-term outcome given the event time of the short-term outcome and information about an additional discrete marker. The detailed estimation procedures are described in Section 2 and the inference procedures are given in Section 3. Simulation results given in Section 4 suggest that the proposed procedures perform well in finite samples. Our procedures are illustrated using a data set of post-dialysis patients with end-stage renal disease (ESRD).

2 Estimation

2.1 Risk prediction given survival to a landmark time

Let T_L and T_S denote event times for the long-term and short term outcomes, respectively. We consider the semi-competing risk setting under which T_L is a terminal event subject only to administrative censoring C , while T_S may be censored by T_L if T_L occurs before T_S . In other words, T_L can only be censored by the non-informative censoring time C , not by T_S . For example, T_L may represent time to death and T_S may represent time to hospitalization. Our goal is to predict T_L using information on T_S and a discrete marker Z measured at baseline. For example, Z may be a binary marker indicating the presence of a gene mutation previously linked to the terminal event of interest.

We focus on prediction of T_L among subjects who have not experienced the terminal event by some landmark time t_0 defined as $\Omega_{T_L > t_0} = \{T_L > t_0\}$. In particular, we aim to estimate the risk of failure within τ from t_0 for subjects in $\Omega_{T_L > t_0}$. Note that this can be interpreted as the probability that one's residual life from time t_0 is less than τ years. Moreover, if short-term outcome information is used, the prediction should only use information about T_S up to the landmark point t_0 . We first consider an estimate which does not use any T_S or Z information denoted by $P_{t_0, \tau}^{(0)} = P(T_L \leq t_0 + \tau | T_L > t_0)$. One may also consider an estimate that uses only baseline marker information, Z , denoted by $P_{t_0, \tau}^{(1)}(z) = P(T_L \leq t_0 + \tau | T_L > t_0, Z = z)$. When T_S and T_L are observed for all patients, we estimate $P_{t_0, \tau}^{(0)}$ and $P_{t_0, \tau}^{(1)}(z)$ using estimators similar to those proposed by Parast and Cai (2010):

$$\begin{aligned}\widehat{P}_{t_0, \tau}^{(0)} &= \frac{\sum_i I(t_0 < T_{Li} \leq t_0 + \tau)}{\sum_i I(T_{Li} > t_0)}, \\ \widehat{P}_{t_0, \tau}^{(1)}(z) &= \frac{\sum_i I(t_0 < T_{Li} \leq t_0 + \tau, Z_i = z)}{\sum_i I(T_{Li} > t_0, Z_i = z)}.\end{aligned}$$

We now consider estimating the risk of failure within τ years from t_0 given information about T_S and Z for subjects in $\Omega_{T_L > t_0}$. Therefore, we are interested in estimating

$$\begin{aligned}P_{t_0, \tau}^{(2)}(T_{S,Z}) &= P(T_L \leq t_0 + \tau | T_L > t_0, \min(T_S \leq t_0), I(T_S = t_0), Z) \\ &= P_{t_0, \tau, T_S, Z} I(T_S \leq t_0) + P_{t_0, \tau, Z} I(T_S > t_0),\end{aligned}$$

where

$$P_{t_0, \tau, t_s, z} = P(T_L \geq t_0 + \tau | T_L > t_0, T_S = t_s, Z = z) \quad \text{if } T_S = t_s \leq t_0, \quad (1)$$

$$P_{t_0, \tau, z} = P(T_L \geq t_0 + \tau | T_L > t_0, T_S > t_0, Z = z) \quad \text{if } T_S > t_0. \quad (2)$$

The quantity $P_{t_0, \tau, t_s, z}$ focuses on subjects whose T_S occurs prior to the landmark point t_0 while $P_{t_0, \tau, z}$ focuses on subjects whose T_S has not yet occurred by t_0 .

We first focus on estimation of $P_{t_0, \tau, t_s, z}$ for any given $t_s \leq t_0$ and note that $P_{t_0, \tau, t_s, z} = P(t_0 < T_L \leq t_0 + \tau, Z = z | T_S = t_s) / P(T_L > t_0, Z = z | T_S = t_s)$. To estimate $P_{t_0, \tau, t_s, z}$ for subjects with $T_S = t_s$ when T_L and T_S are observed for all patients, we propose a kernel-based estimator

$$\widehat{P}_{t_0, \tau, t_s, z} = \frac{\sum_i I(T_{Si} \leq t_0) I(t_0 < T_{Li} \leq t_0 + \tau, Z_i = z) K_h\{\psi(T_{Si}) - \psi(t_s)\}}{\sum_i I(T_{Si} \leq t_0) I(T_{Li} > t_0, Z_i = z) K_h\{\psi(T_{Si}) - \psi(t_s)\}},$$

where $K_h(x) = h^{-1} K(x/h)$, $K(\cdot)$ is a known smooth symmetric kernel density function with a bounded support, $\psi: [0, \infty) \rightarrow (-\infty, \infty)$ is a given transformation function and the bandwidth $h > 0$ is assumed to be $O(n^{-\nu})$, for $1/5 < \nu < 1/2$. An appropriate transformation $\psi(\cdot)$ to implement in smoothing could potentially improve estimation (Park *et al.*, 1997; Wand *et al.*, 1991). In our numerical examples, we let $K(\cdot)$ to be the standard normal density function and $\psi(x) = \log(x)$.

For subjects with $T_s > t_0$ we predict their risk by conditioning on $I(T_s > t_0)$ and Z . Similar to the estimator given in Parast and Cai (2010), when T_L and T_s are observed for all patients we estimate $P_{t_0, \tau, z}$ for subjects with $T_s > t_0$ by

$$\widehat{P}_{t_0, \tau, z} = \frac{\sum_i I(t_0 < T_{Li} \leq t_0 + \tau, Z_i = z, T_{si} > t_0)}{\sum_i I(T_{Li} > t_0, Z_i = z, T_{si} > t_0)}.$$

We now have a procedure for estimating $P_{t_0, \tau}^{(2)}(T_s, Z)$ for all subjects in $\Omega_{T_L > t_0}$:

$$\widehat{P}_{t_0, \tau}^{(2)}(T_s, Z) = \widehat{P}_{t_0, \tau, T_s, Z} I(T_s \leq t_0) + \widehat{P}_{t_0, \tau, Z} I(T_s > t_0).$$

2.2 Estimation in the presence of censoring

Due to censoring, T_L and T_s are only potentially observable as $X_L = \min(T_L, C)$, $X_s = \min(T_s, C)$, $\delta_L = I(T_L \leq C)$ and $\delta_s = I(T_s \leq C)$. We assume that $P(X_L > t_0 + \tau, X_s > t_0) > 0$. For simplicity, we assume that C is independent of (T_L, T_s, Z) with a survival function $G(\cdot)$ and identically distributed Potentially observable data for analysis is composed of n independent random vectors $\{(X_{Li}, \delta_{Li}, X_{si}, \delta_{si}, Z_i), i = 1, \dots, n\}$. Due to semi-competing risk, T_s not observable if $T_L < T_s$ and observable otherwise. Nevertheless, the observed data have sufficient information about T_s and T_L for the estimation of (1) and (2). The conditional risk (1) only involves the subgroup with $T_s \leq t_0 < T_L$. Note that this subgroup excludes those who would experience T_L before T_s and includes only those with complete T_s information. Thus, individuals in this subgroup would not be subject to informative censoring of T_L by T_s . The probability in (2) involves the subgroup with $(T_s, T_L) > t_0$ and can be inferred based on the subset of patients with $(X_s, X_L) > t_0$, which is always observable.

When censoring is present, we propose incorporating inverse probability weighting (IPW) which reweights complete observations with the probability of being observable (Rotnitzky and Robins, 2005; Robins and Rotnitzky, 1992). Specifically, we estimate $P_{t_0, \tau}^{(0)}$ and $P_{t_0, \tau}^{(1)}(z)$ by:

$$\begin{aligned} \widehat{P}_{t_0, \tau}^{(0)} &= \frac{\sum_i \widehat{W}_{Li} I(t_0 < X_{Li} \leq t_0 + \tau)}{\sum_i \widehat{W}_{Li} I(X_{Li} > t_0)}, \\ \widehat{P}_{t_0, \tau}^{(1)}(z) &= \frac{\sum_i \widehat{W}_{Li} I(t_0 < X_{Li} \leq t_0 + \tau, Z_i = z)}{\sum_i \widehat{W}_{Li} I(X_{Li} > t_0, Z_i = z)}, \end{aligned}$$

where

$$\widehat{W}_{Li} = \frac{I(t_0 < X_{Li} \leq t_0 + \tau) \delta_{Li}}{\widehat{G}(X_{Li})} + \frac{I(X_{Li} > t_0 + \tau)}{\widehat{G}(t_0 + \tau)}$$

and $\widehat{G}(\cdot)$ is the Kaplan–Meier estimator of $G(\cdot)$. Similarly, to estimate $P_{t_0, \tau}^{(2)}(T_s, Z)$ when censoring is present, we propose the following estimator:

$$\widehat{P}_{t_0, \tau}^{(2)}(T_s, Z) = \widehat{P}_{t_0, \tau, T_s, Z} I(T_s \leq t_0) + \widehat{P}_{t_0, \tau, z} I(T_s > t_0),$$

where

$$\widehat{P}_{t_0, \tau, t_s, z} = \frac{\sum_i \widehat{W}_{Li} I(T_{Si} \geq t_0) I(t_0 < X_{Li} \leq t_0 + \tau, Z_i = z) K_h\{\psi(T_{Si}) - \psi(t_s)\}}{\sum_i \widehat{W}_{Li} I(T_{Si} \geq t_0) I(X_{Li} > t_0, Z_i = z) K_h\{\psi(T_{Si}) - \psi(t_s)\}}, \quad (3)$$

$$\widehat{P}_{t_0, \tau, z} = \frac{\sum_i \widehat{W}_{Li} I(t_0 < X_{Li} \leq t_0 + \tau, Z_i = z, T_{Si} > t_0)}{\sum_i \widehat{W}_{Li} I(X_{Li} > t_0, Z_i = z, T_{Si} > t_0)}$$

with $\widehat{W}_{Li}$ as previously defined. Note that $X_{Li} > t_0$ implies $C_i > t_0$ and thus when $X_{Li} > t_0$ and $X_{Si} \leq t_0$, then $T_{Si} = X_{Si}$ and $I(T_{Si} \leq t_0)$ is observable. In addition, it can be shown that $E_{\Omega_{T_L > t_0}}[I(C_i > t_0) W_{Li} | T_{Si}, T_{Si}, Z_i] = 1$ where

$$W_{Li} = \frac{I(t_0 < X_{Li} \leq t_0 + \tau) \delta_{Li}}{G(X_{Li})} + \frac{I(X_{Li} > t_0 + \tau)}{G(t_0 + \tau)}.$$

As for any non-parametric functional estimation problem, the choice of the smoothing parameter h in $\widehat{P}_{t_0, \tau, t_s, z}$ is crucial. To obtain an appropriate h we propose to use a standard K -fold cross-validation procedure. Specifically, we randomly split the data into K subsets of equal size denoted by $\{\mathcal{J}_k, k=1, \dots, K\}$. For each k , with a given h we estimate $\widehat{P}_{t_0, \tau, t_s, z}$ using all observations not in $\mathcal{J}_k$ to calculate the appropriately weighted mean-squared using (3). We then use all observations in $\mathcal{J}_k$ to calculate the appropriately weighted mean-squared error

$$\text{MSE}(k) = \sum_{i \in \mathcal{J}_k} \widehat{W}_{Li} I(X_{Li} > t_0, T_{Si} \leq t_0) \left[I(X_{Li} \leq t_0 + \tau) - \widehat{P}_{t_0, \tau, T_{Si}, Z_i} \right]^2. \quad (4)$$

Let h_{opt} be the minimizer of $\sum_{k=1}^K \text{MSE}(k)$. We expect the order of the bandwidth estimator to be $n^{-1/5}$ resulting in an estimator in which the bias² = $O_p(h^{-4}) = O_p(n^{-4/5})$ and the variance = $O_p((nh)^{-1}) = O_p(n^{-4/5})$ (Fan and Gijbels, 1996; Wand and Jones, 1995). Due to the difficulty in estimating the bias, one may wish to undersmooth and obtain an estimator whose variance dominates bias. In practice, one can choose $h = h_{\text{opt}} n^{-c_0}$, for some $c_0 \in (0, 3/10)$ to obtain the final bandwidth for estimation.

2.3 Information gained by incorporating short-term outcome

To examine the incremental value gained from using the short-term outcome T_s information, we aim to compare the predicted individual risk for subjects using (T_s, Z) versus using Z alone. We propose to compare these estimated risks using two different measures: the Brier score (Brier, 1950; Gerds and Schumacher, 2006; Lawless and Yuan, 2010; Redelmeier *et al.*, 1991) and the area under the Receiver Operating Characteristic (ROC) curve (Begg,

1991; Hanley, 1989; Swets and Pickett, 1982). In our case the Brier score is measuring the mean-squared distance between $I(X_{Li} \leq t_0 + \tau)$ and the predicted risk among $\Omega_{T_{Li} > t_0}$. Let

$$\widehat{BS}_{t_0, \tau}^{(2)} = \frac{\sum_i \widehat{W}_{Li} \left[I(X_{Li} > t_0 + \tau) - \widehat{P}_{t_0, \tau}^{(2)}(T_{si}, Z_i) \right]}{\sum_i \widehat{W}_{Li} I(X_{Li} > t_0)} \quad (5)$$

be the Brier score for our proposed procedure, which incorporates T_s and Z information. For the prediction rule using Z only and the prediction rule using no information, the Brier score can be obtained by replacing $\widehat{P}_{t_0, \tau}^{(2)}(T_{si}, Z_i)$ in (5) with $\widehat{P}_{t_0, \tau}^{(1)}(Z_i)$ and $\widehat{P}_{t_0, \tau}^{(0)}$, respectively, to

obtain $\widehat{BS}_{t_0, \tau}^{(1)}$ and $\widehat{BS}_{t_0, \tau}^{(0)}$. One can directly compare the performance of different prediction estimates simply by comparing the modified Brier scores. However, a clearer interpretation may be gained by defining the increase in the proportion of explained variation in estimate

A over B as $\text{IPEV}_{A/B} = 1 - \widehat{BS}_{t_0, \tau}^{(A)} / \widehat{BS}_{t_0, \tau}^{(B)}$, where $\widehat{BS}_{t_0, \tau}^{(A)} \leq \widehat{BS}_{t_0, \tau}^{(B)}$. Therefore to determine whether we have gained additional information to aid in risk prediction by using short-term outcome information, one may examine $\text{IPEV}_{2/1}$.

An alternative measure for evaluating prediction rules is to consider the area under the ROC curve (AUC), which can be interpreted as the probability that a prediction rule will assign a higher risk to patient i with $t_0 < T_{Li} \leq t_0 + \tau$ than to patient j with $T_{Lj} > t_0 + \tau$. Using IPW we propose to estimate the AUC for $P_{t_0, \tau}^{(2)}(T_s, Z)$ as

$$\widehat{AUC}_{t_0, \tau}^{(2)} = \frac{\sum_{i: X_{Li} > t_0} \sum_{j: X_{Lj} > t_0} I \left[P_{t_0, \tau}^{(2)}(T_{si}, Z_i) > P_{t_0, \tau}^{(2)}(T_{sj}, Z_j) \right] \widehat{W}_{Li} \widehat{W}_{Lj} I(X_{Li} \geq t_0 + \tau) I(X_{Lj} > t_0 + \tau)}{\sum_{i: X_{Li} > t_0} \sum_{j: X_{Lj} > t_0} \widehat{W}_{Li} \widehat{W}_{Lj} I(X_{Li} \leq t_0 + \tau) I(X_{Lj} > t_0 + \tau)} \quad (6)$$

Similarly, for the prediction rule using Z only and the prediction rule using no information, the AUC can be obtained by replacing $\widehat{P}_{t_0, \tau}^{(2)}(T_{si}, Z_i)$ in (6) with $\widehat{P}_{t_0, \tau}^{(1)}(Z_i)$ and $\widehat{P}_{t_0, \tau}^{(0)}$,

respectively, to obtain $\widehat{AUC}_{t_0, \tau}^{(1)}$ and $\widehat{AUC}_{t_0, \tau}^{(0)}$. Note that AUC is a retrospective discrimination measure that is free of disease prevalence while the Brier score is a prospective predictive accuracy measure that depends on the absolute risk (Cook, 2007; Gerds *et al.*, 2008; Pencina *et al.*, 2008). Examining both prediction performance measures is useful for identifying optimal scoring systems.

2.4 Residual risk prediction given the occurrence of short-term event

Depending on the event, if the time of the short-term event is known for individual i , $T_{si} = t_s$, one may be more interested in estimating risk of failure within τ years from t_s , rather than τ years from t_0 . Let $P_{t_s, \tau, t_s, z} = P(T_{Li} \leq t_s + \tau | T_{Li} > t_s, T_{si} = t_s, Z = z)$ be the subject-specific residual risk i.e. the risk that the long-term event will occur τ years from t_s given it has not yet occurred by time t_s along with baseline marker information. For example, in a population of patients with ESRD we may be interested in $P_{t_s, \tau, t_s, z}$ when T_s is hospitalization and T_L is death. An individual patient would clearly be interested in their probability of residual survival from the time of hospitalization. However in a setting where T_s is the onset of diabetes, T_L is the occurrence of a cardiovascular event, and t_0 is a clinician chosen landmark time, one would be interested in $t_0 + \tau$ rather than $t_s + \tau$, though it may still be useful to incorporate time of diabetes onset in the estimation procedure. Estimation of $P_{t_s, \tau, t_s, z}$ follows

similarly from (3) with $t_0 = t_s$ and we denote the resulting estimator as $\widehat{P}_{t_s, \tau, t_s, z}$. Note that only information up to t_s is used for estimating predicted risk of an event before $t_s + \tau$.

3 Inference

To make inference about $\widehat{P}_{t_0, \tau, t_s, z}$ and $\widehat{P}_{t_s, \tau, t_s, z}$, asymptotic properties may be derived. It can be shown using the uniform consistency and the root-n convergence rate of $\widehat{G}(\cdot)$ (Kalbfleisch and Prentice, 2002) that the uncertainty of $\widehat{G}(\cdot)$ in $\widehat{W}_{Li}$ is negligible. Therefore, the proposed estimators are asymptotically equivalent to the estimators obtained by replacing $\widehat{W}_{Li}$ with W_{Li} . Subsequently, one may use the same arguments as given by Bickel and Rosenblatt (1973), Dabrowska (1989) and Cai *et al.* (2010) to show that

$\sup_{t_0, \tau, z, t_s} |\widehat{P}_{t_0, \tau, t_s, z} - P_{t_0, \tau, t_s, z}| = O_p \left\{ (nh)^{-1/2} \log(n) \right\}$. Furthermore, one can show that $\sqrt{nh} (\widehat{P}_{t_0, \tau, t_s, z} - P_{t_0, \tau, t_s, z})$ is asymptotically equivalent to a sum of i.i.d terms and thus converges in distribution to a normal random variable. Asymptotic properties for $\widehat{P}_{t_s, \tau, t_s, z}$ can similarly be shown.

To construct confidence intervals for the predicted risks, we propose a perturbation-resampling method (Cai *et al.*, 2005; Park and Wei, 2003; Tian *et al.*, 2007) to approximate the distributions of the aforementioned estimators. Specifically, let $\{V_i^{(b)}, i=1, \dots, n\}$ for $b = 1, \dots, B$ be $n \times B$ independent copies of a positive random variable V from a known distribution with unit mean and unit variance. For $T_s = t_s$, let

$$\widehat{P}_{t_0, \tau, t_s, z}^{(b)} = \frac{\sum_i \widehat{W}_{Li}^{(b)} I(X_{si} \leq t_0) I(t_0 \leq X_{Li} \leq t_0 + \tau, Z_i = z) K_h \{ \psi(T_{si}) - \psi(t_s) \} V_i^{(b)}}{\sum_i \widehat{W}_{Li}^{(b)} I(X_{si} \leq t_0) I(X_{Li} > t_0, Z_i = z) K_h \{ \psi(T_{si}) - \psi(t_s) \} V_i^{(b)}},$$

where

$$\widehat{W}_{Li}^{(b)} = \frac{I(t_0 < X_{Li} \leq t_0 + \tau) \delta_{Li}}{\widehat{G}^{(b)}(X_{Li})} + \frac{I(X_{Li} > t_0 + \tau)}{\widehat{G}^{(b)}(t_0 + \tau)}$$

and $\widehat{G}^{(b)}$ is the Kaplan–Meier estimator of $G(\cdot)$ with weights $V_i^{(b)}$. Then one can estimate the variance of $\widehat{P}_{t_0, \tau, t_s, z}$ as the empirical variance of $\{\widehat{P}_{t_0, \tau, t_s, z}^{(b)}, b=1, \dots, B\}$. To construct a $100(1-2\alpha)\%$ confidence interval (CI) one can either use the empirical α th and $(100-\alpha)$ th quantiles of the perturbed samples or a normal approximation.

Similarly, using similar arguments as given by Cai *et al.* (2010) and Tian *et al.* (2009) it can be shown that the unconditional distribution of the risk comparison measures can be approximated by the conditional distribution of the corresponding perturbed estimates given the data. For example, the unconditional distribution of $\widehat{AUC}_{t_0, \tau}^{(2)}$ can be approximated by the conditional distribution of the perturbed estimates denoted by $\widehat{AUC}_{t_0, \tau}^{(2)(b)}$ for $b = 1, \dots, B$ given the data. It follows that the variance of $\widehat{AUC}_{t_0, \tau}^{(2)}$ can be estimated by the empirical variance of $\widehat{AUC}_{t_0, \tau}^{(2)(b)}$ for $b = 1, \dots, B$ and a $100(1-2\alpha)\%$ confidence for $\widehat{AUC}_{t_0, \tau}^{(2)}$ can be obtained either using the normal CI or the empirical quantiles of the perturbed samples. To make inference

about $AUC_{t_0,\tau}^{(2)} - AUC_{t_0,\tau}^{(1)}$, let $\widehat{AUC}_{t_0,\tau}^{(CV)(k)}$ denote the cross-validated estimate of $AUC_{t_0,\tau}^{(k)}$ for $k = 1, 2$. Then inference about $AUC_{t_0,\tau}^{(2)} - AUC_{t_0,\tau}^{(1)}$ can be based on $\widehat{AUC}_{t_0,\tau}^{(CV)(2)} - \widehat{AUC}_{t_0,\tau}^{(CV)(1)}$ and the empirical variance of $\left\{B^{-1} \sum_{i=1}^n \left\{ \widehat{AUC}_{t_0,\tau}^{(2)(b)} - \widehat{AUC}_{t_0,\tau}^{(1)(b)} \right\}, b=1, \dots, B\right\}$. A confidence interval for $IPEV_{2/1}$ can be similarly constructed.

4 Simulations

To examine finite-sample performance of the proposed procedures, we conducted simulation studies under various scenarios and focused on estimation of $P_{t_0, \tau, t_s, z}$ and $P_{t_s, \tau, t_s, z}$. The covariate Z was generated using a Bernoulli distribution with a success probability 0.5 and bivariate log-normal models were used to generate (T_s, T_L) , which depended on Z . In each setting, we generated 1000 realizations of size n for $n = 5000$ and focused on estimation of $P_{t_0, \tau, t_s, z}$ and $P_{t_s, \tau, t_s, z}$ for $z = 1$. The smoothing parameter h was obtained using the 3-fold cross-validation scheme and $c_0 = 1/10$. In particular, we let h equal to $h_{opt} n^{-1/10}$ where h_{opt} is the minimizer of (4) and $n = \sum_{i=1}^{5000} I(X_{Li} > t_0)$. The standard error estimates were obtained based on 500 perturbations each.

In setting (a), the low event rate setting, the survival times were generated from

$$\begin{pmatrix} T_s \\ T_L \end{pmatrix} = \exp(-0.5Z_i + \varepsilon) \quad \text{with } \varepsilon \sim N\left\{\begin{pmatrix} 1.2 \\ 3.4 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.2 \end{pmatrix}\right\}.$$

For illustration, we chose t_0 to be year 2 and τ to be 8 years. Therefore, among patients who have experienced the short-term event before year 2, $T_s \leq 2$, but have not experienced the long-term event by year 2, $T_L > 2$, we wish to estimate their probability of survival past year 10. Under these conditions, $P(T_s \leq t_0) = 0.40$ and $P(T_L > t_0) = 0.95$ and $P(t_0 < T_L \leq t_0 + \tau | T_L > t_0) = 0.20$. We consider the case when $C = \min(\exp\{W\}; 35)$ where $W \sim N(2, 2)$ yielding a censoring rate of approximately 32% for T_s and 73% for T_L . With this censoring pattern, $P(X_s \leq t_0, \delta_s = 1) = 0.34$ and $P(X_L > t_0) = 0.74$ and $P(t_0 < X_L \leq t_0 + \tau, \delta_L = 1 | X_L > t_0) = 0.15$. Estimates of $P_{t_0, \tau, t_s, z}$ for $t_0 = 2$, $\tau = 8$, $z = 1$ are shown for multiple values of t_s in Table 1(a). Values of t_s in the table were chosen as the 20th, 40th, 60th and 80th percentiles of $T_s < t_0$. We also considered estimation of $P_{t_s, \tau, t_s, z}$ for $z = 1$, the probability that the long-term event will occur within 8 years of t_s given the short-term event occurred at some time, t_s , before year 2, $T_s = t_s \leq 2$, and $T_s > t_s$. These estimates are shown for multiple values of t_s in Table 2(a). All estimates have slight bias. The standard errors estimated from the resampling method are close to the empirical standard errors and the empirical coverage levels are close to their nominal level. However, bias estimates tend to be high and empirical coverage levels tend to be low for earlier values of t_s .

For setting (b), the moderate event rate setting, the survival times were generated from

$$\begin{pmatrix} T_s \\ T_L \end{pmatrix} = \exp(-0.5Z_i + \varepsilon) \quad \text{with } \varepsilon \sim N\left\{\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.2 \end{pmatrix}\right\}.$$

Under these conditions, $P(T_s \leq t_0) = 0.48$ and $P(T_L > t_0) = 0.96$ and $P(t_0 < T_L \leq t_0 + \tau | T_L > t_0) = 0.31$. We consider the case when C was generated from an exponential distribution with rate 0.05, yielding a censoring rate of approximately 15% for

T_s and 55% for T_L . With this censoring pattern, $P(X_s \leq t_0, \delta_s=1)=0.46$ and $P(X_L > t_0)=0.88$ and $P(t_0 < X_L \leq t_0 + \tau, \delta_L=1 | X_L > t_0)=0.25$. We again consider both $P_{t_0, \tau, t_s, z}$ and $P_{t_s, \tau, t_s, z}$ for $z = 1$ in this setting. The results are shown in Tables 1(b) and 2(b). All point estimators have negligible bias. The standard errors estimated from the resampling method are again close to the empirical standard errors and the empirical coverage levels are close to their nominal level.

In both settings we examined the value of using the additional short-term outcome information in $\widehat{P}_{t_0, \tau}^{(2)}(T_s, Z)$ using the procedures discussed in Section 2.3. To overcome potential biases due to overfitting, in practice one should consider corresponding cross-validated estimates. Here we use a 3-fold cross-validation procedure. Brier score estimates are shown in Tables 3(a) and (b). In the low event rate setting, the proportion of explained variation in predicting long-term $\tau + t_0$ survival increased by 0.0975 after incorporating short-term outcome information compared to using Z information only. In the higher event rate setting, the proportion of explained variation in predicting long-term $\tau + t_0$ survival increased by 0.1122 after incorporating short-term outcome information compared to using Z information only. AUC estimates are shown in Tables 4(a) and (b). As expected, a model using no information gives an AUC of 0.5. In the low event rate setting including the short-term outcome information results in a 0.1331 increase in AUC compared to using Z information only. In the higher event rate setting including the short-term outcome information results in a 0.1336 increase in AUC compared to using Z information only. In both situations, it appears advantageous to use short-term event information to aid in prediction.

5 Example

We illustrate our proposed procedures using a data set of patients with ESRD who had initiated maintenance peritoneal or hemodialysis from the United States Renal Data System (USRDS). The USRDS consists of a total of 463 240 patients from 18 regional areas who initiated dialysis between 2000 and 2004. Our analysis focused on the subset of patients without a previous kidney transplant from the six-state New England area. This data set consists of 9785 patients with individual information about time to death or kidney transplant, time to first hospitalization, age, serum creatinine and erythropoietin use at initiation of dialysis, gender and race. Time at risk was measured from the date of entry, defined as the earliest outpatient date on or after the date of initiation of maintenance peritoneal or hemodialysis to the first hospitalization date, if available, and the censoring date at transplantation, death, loss to follow-up, or December 31, 2005.

We consider T_s as the time to first hospitalization and T_L as the time to death or kidney transplant. There were a total of 5480 hospitalizations, 5448 deaths, 111 transplantations and the 1-year hospitalization rate was about 0.52; the 1-year death or transplant rate was about 0.33. We aim to make a prediction of 1-year survival among those who survived 6 months using the proposed procedures incorporating time of hospitalization occurring before 6 months along with a single discrete baseline marker. Among the 7679 patients who survived past 6 months, 37% were hospitalized within 6 months of initiating dialysis and 63.3% did not have a hospitalization before 6 months. Figure 1 shows the probabilities of interest, $P_{t_0, \tau, t_s, z}$ and $P_{t_s, \tau, t_s, z}$, for a patient with $T_s = t_s = 24$ days.

For illustration, we estimate the quantities of interest using erythropoietin use as the baseline binary marker of interest. Erythropoietin is a hormone produced by the kidney that promotes the formation of red blood cells in the bone marrow. About 90% of patients on dialysis experience anemia as a result of erythropoietin deficiency (Adamson and Eschbach, 1998). Anemia contributes substantially to morbidity and mortality among these patients, and the

introduction of erythropoietin therapy in 1989 significantly improved the quality of life and reduced the need for red-cell transfusions (Eschbach *et al.*, 1989). In this data set, about 45% of patients use erythropoietin.

Figure 2 displays $\widehat{P}_{t_0, \tau}^{(2)}(t_s, z) = \widehat{P}_{t_0, \tau, t_s, z}$, the proposed estimator using both time of hospitalization and erythropoietin use information over time for patients hospitalized within the first 6 months. The estimated risk of a hospitalized subject having an event within 1 year from 6 months ranges from 0.285 to 0.364 and increases slightly over t_s regardless of erythropoietin use. Using marker information only, $\widehat{P}_{t_0, \tau}^{(1)}(z)$, the predicted risk of having an event between 6 months and 1.5 years is equal to 0.2139 for patients using erythropoietin and 0.2531 for patients not using erythropoietin regardless of hospitalization time. Note that $\widehat{P}_{t_0, \tau}^{(1)}(z)$ is the estimated predicted risk for all patients who survived past 6 months whereas $\widehat{P}_{t_0, \tau, t_s, z}$ is for patients who survived past 6 months and were hospitalized before 6 months. For patient who were not hospitalized within 6 months, $\widehat{P}_{t_0, \tau}^{(2)}(t_s, z) = \widehat{P}_{t_0, \tau, z} = 0.1595$ for patients using erythropoietin and 0.1937 for those not using erythropoietin.

Estimates for $\widehat{P}_{t_0, \tau, t_s, z}$ and $\widehat{P}_{t_s, \tau, t_s, z}$ at multiple values of t_s are shown in Table 5 with corresponding standard error estimates from the perturbation-resampling method. As similarly seen from Fig. 2, the estimated risk increases slightly over t_s regardless of erythropoietin use. Looking at residual survival, the estimated risk of a subject using erythropoietin having an event within 1 year from the time of hospitalization increases and then decreases over t_s while the predicted risk for a subject not using erythropoietin decreases over time. For example, a subject using erythropoietin who is hospitalized at 24 days but survives past 24 days would have a predicted risk of 0.4373 of having an event within 1 year from hospitalization time. If this same subject survives past 6 months, their predicted risk of having an event between 6 months and 1.5 years would be 0.3088.

Brier score estimates with corresponding values of the increase in explained variation and AUC estimates are shown in Table 6. To overcome potential biases due to overfitting, we consider a 3-fold cross-validation procedure. Standard errors were estimated using the perturbation-resampling method. With erythropoietin use as the discrete marker of interest, the proportion of explained variation in predicting long-term survival increases by 0.0283 (95% CI: 0.0197, 0.0368) after incorporating time of hospitalization information compared to using erythropoietin use only. In addition, the AUC increases by 0.0836 (95% CI: 0.0674, 0.0998) indicating a gain in predictive ability after incorporating short-term outcome information.

6 Discussion

In this paper, we propose non-parametric procedures for predicting long-term survival based on a discrete marker along with information on a short-term disease outcome. We consider risk prediction for both the residual life since a given landmark point t_0 and the residual life since the occurrence of the short-term outcome. When the short-term outcome T_s is highly correlated with the long-term outcome T_{∞} , one may expect that the inclusion of short-term outcome information could significantly improve the prediction accuracy. We quantify the incremental value of T_s based on the percent reduction in prediction error as measured by the Brier score and the increase in the overall prediction accuracy as measured by the AUC. Other types of prediction performance measures such as the overall mis-classification rate could be considered. For simplicity, we assumed independent censoring with C $\perp$ Z. The proposed procedures could easily be incorporated to allow for marker-dependent censoring

by replacing $\widehat{G}$ in $\widehat{W}_{L_i}$ with $\widehat{G}_z$ obtained as the Kaplan–Meier estimator of $P(C > t | Z = z)$ for the subgroup with $Z_i = Z$.

It is important note that the proposed methods hold for any choice of t_0 and τ . In practice, one may consider the proposed estimators across a range of t_0 and τ as a basis for selecting an appropriate time point such that the prediction model is most accurate. For example, suppose t_0 is pre-selected by a clinician, such as a required 1-month appointment with a nephrologist for patients on dialysis. One could make simultaneous inference using the perturbation-resampling method as described, to guide the choice of τ . This would be clinically meaningful as it would provide insight on what time point/period the short-term outcome information is most useful in prediction.

The proposed procedures could be extended to incorporate a continuous marker using multidimensional smoothing. However, even two-dimensional smoothing requires a large sample size in the presence of censoring (Dabrowska, 1987). This lack of power is even more pronounced in our setting due to the reduced effective sample size given the t_0 landmark. Alternatively, one could consider more parametric methods as previously posed in the literature. However, risk estimates from these model-based procedures may have poor accuracy if the model assumptions fail to hold. Improving prediction performance in this setting warrants further research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government. The authors are grateful to the Editor and the reviewers for their insightful comments on the article. Support for this research was provided by National Institutes of Health grants T32GM074897 (Parast) and is R01-GM079330 (Cai).

References

- Adamson J, Eschbach J. Erythropoietin for end-stage renal disease. *New England Journal of Medicine*. 1998; 339:625–627. [PubMed: 9718384]
- Bauer K, Brown M, Cress R, Parise C, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the California cancer registry. *Cancer*. 2007; 109:1721–1728. [PubMed: 17387718]
- Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine*. 1991; 10:1887–1895. [PubMed: 1805316]
- Bickel P, Rosenblatt M. On some global measures of the deviations of density function estimates. *The Annals of Statistics*. 1973; 1:1071–1095.
- Brier G. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*. 1950; 78:1–3.
- Cai T, Tian L, Uno H, Solomon S, Wei L. Calibrating parametric subject-specific risk estimation. *Biometrika*. 2010; 97:389–404. [PubMed: 23049123]
- Cai T, Tian L, Wei LJ. Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika*. 2005; 92:619–632.
- Cook N. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115:928–935. [PubMed: 17309939]

- Cortese G, Andersen P. Competing risks and time-dependent covariates. *Biometrical Journal*. 2010; 52:138–158. [PubMed: 20029852]
- Dabrowska D. Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*. 1987; 14:181–197.
- Dabrowska D. Uniform consistency of the kernel conditional Kaplan–Meier estimate. *The Annals of Statistics*. 1989; 17:1157–1167.
- Datta S, Satten G, Datta S. Nonparametric estimation for the three-stage irreversible illness-death model. *Biometrics*. 2000; 56:841–847. [PubMed: 10985224]
- Eschbach J, Abdulhadi M, Browne J, Delano B, Downing M, Egrie J, Evans R, Friedman E, Graber S, Haley N, Korbet S, Krantz S, Lundin AP, Nissenson A, Ogden D, Paganini E, Rader B, Rutsky E, Stivelman J, Stone W, Teschan P, Van Stone J, Van Wyck D, Zuckerman K, Adamson J. Recombinant human erythropoietin in anemic patients with end-stage renal disease. *Annals of Internal Medicine*. 1989; 111:992–1000. [PubMed: 2688507]
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and its Applications*. Chapman & Hall/CRC; London, Boca Raton, FL: 1996.
- Fine J, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88:907–919.
- Gerds T, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical Journal*. 2008; 50:457–479. [PubMed: 18663757]
- Gerds T, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*. 2006; 48:1029–1040. [PubMed: 17240660]
- Gray R. A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika*. 1994; 81:527–539.
- Hanley HA. Receiver Operating Characteristic (ROC) methodology: the state of the art. *Clinical Reviews in Diagnostic Imaging*. 1989; 29:307–335.
- Hatleville L, Mahe C, Hill C. Prediction of the long-term survival in breast cancer patients according to the present oncological status. *Statistics in Medicine*. 2002; 21:2345–2354. [PubMed: 12210619]
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Wiley; New York: 2002.
- Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*. 1986; 42:855–865. [PubMed: 2434150]
- Klein J, Keiding N, Copelan E. Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine*. 1994; 12:2315–2332. [PubMed: 8134735]
- Lawless J, Yuan Y. Estimation of prediction error for survival models. *Statistics in Medicine*. 2010; 29:262–274. [PubMed: 19882678]
- Lee S, Klein J, Barrett A, Ringden O, Antin J, Cahn J, Carabasi M, Gale R, Giral S, Hale G, Ilhan O, McCarthy P, Socie G, Verdonck L, Weisdorf D, Horowitz M. Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse. *Blood*. 2002; 100:406–414. [PubMed: 12091329]
- Parast, L.; Cai, T. Landmark prediction of survival; Harvard University Biostatistics Working Paper Series 123; 2010; <http://www.bepress.com/harvardbiostat/paper123/>
- Park B, Kim W, Ruppert D, Jones M, Signorini D, Kohn R. Simple transformation techniques for improved non-parametric regression. *Scandinavian Journal of Statistics*. 1997; 24:145–163.
- Park Y, Wei LJ. Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika*. 2003; 90:717–723.
- 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, Thompson M. Combining diagnostic test results to increase accuracy. *Biostatistics*. 2000; 1:123–140. [PubMed: 12933515]
- Putter H, van der Hage J, de Bock G, Elgalta R, van de Velde C. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal*. 2006; 48:366–380. [PubMed: 16845902]

- Redelmeier D, Bloch D, Hickam D. Assessing predictive accuracy: how to compare Brier scores. *Journal of Clinical Epidemiology*. 1991; 44:1141–1146. [PubMed: 1941009]
- Robins J, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology, Methodological Issues*. 1992:297–331.
- Rotnitzky A, Robins J. Inverse probability weighted estimation in survival analysis. *Encyclopedia of Biostatistics*. 2005; 4:2619–2625.
- Swets, JA.; Pickett, RM. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press; New York: 1982.
- Tian L, Cai T, Goetghebeur E, Wei L. Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*. 2007; 94:297–311.
- Tian L, Cai T, Wei L. Identifying subjects who benefit from additional information for better prediction of the outcome variables. *Biometrics*. 2009; 65:894–902. [PubMed: 18945268]
- Van der Sluis C, Kingma J, Eisma W, Ten Duis H. Pediatric polytrauma: short-term and long-term outcomes. *The Journal of Trauma: Injury, Infection, and Critical Care*. 1997; 43:501–506.
- Van Houwelingen H. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*. 2007; 34:70–85.
- Van Houwelingen H, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis*. 2008; 14:447–463. [PubMed: 18836831]
- Wand, M.; Jones, M. *Kernel Smoothing*. Chapman & Hall/CRC; London, Boca Raton, FL: 1995.
- Wand M, Marron J, Ruppert D. Transformations in density estimation. *Journal of the American Statistical Association*. 1991; 86:343–353.
- Weisner C, Thomas Ray G, Mertens J, Satre D, Moore C. Short-term alcohol and drug treatment outcomes predict long-term outcome. *Drug and Alcohol Dependence*. 2003; 71:281–294. [PubMed: 12957346]
- Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998; 97:1837–1847. [PubMed: 9603539]

Figure 1.

Visual display of probabilities of interest, (A) P_{t_0, τ, t_s, Z_i} and (B) P_{t_s, τ, t_s, Z_i} for a patient with $T_{s_i} = t_s = 24$ days for $t_0 = 6$ months and $\tau = 1$ year. (A) Given patient i is hospitalized 24 days after initiation of dialysis, $T_{s_i} = t_s = 24$, but survives past 6 months, P_{t_0, τ, t_s, Z_i} is the probability of death within the shaded region. (B) Given patient i is hospitalized 24 days after initiation of dialysis, $T_{s_i} = t_s = 24$, but survives past 24 days, P_{t_s, τ, t_s, Z_i} is the probability of death within the shaded region.

Figure 2.

Estimates of $\hat{P}_{t_0, \tau, t_s, z} = \hat{P}(T_L \leq t_0 + \tau | T_L > t_0, T_s = t_s, Z = z)$ over time among patients with $T_s = t_s \leq t_0$ for those using erythropoietin ($z = 1$, solid black thick line) and for those not using erythropoietin ($z = 0$, solid black thin line) with $t_0 = 6$ months (183 days) and $\tau = 1$ year. Note that using erythropoietin information only, $\hat{P}(T_L \leq t_0 + \tau | T_L > t_0, Z = z)$ is equal to 0.2139 and 0.2531 for patients using erythropoietin and not using erythropoietin, respectively.

Table 1

Bias estimates for $\widehat{P}_{t_0, \tau, t_s, \bar{z}}$ with corresponding averages of the standard error estimates from the perturbation–resampling method (ASE), empirical standard errors (ESE), and empirical 95% coverage levels (coverage) from the (a) low event rate setting and (b) moderate event rate setting. Values of t_s were chosen as the 20th, 40th, 60th and 80th percentiles of $T_{\bar{z}} \leq t_0$.

t_s	$t_{0.20}$	$t_{0.40}$	$t_{0.60}$	$t_{0.80}$
(a) <i>Low event rate setting</i>				
Bias	−0.0025	−0.0029	−0.0009	−0.0001
ASE	0.0496	0.0408	0.0347	0.0306
ESE	0.0494	0.0423	0.0361	0.0315
Coverage	0.9440	0.9270	0.9450	0.9480
(b) <i>Moderate event rate setting</i>				
Bias	−0.0054	−0.0050	−0.0027	−0.0024
ASE	0.0460	0.0382	0.0331	0.0303
ESE	0.0488	0.0379	0.0337	0.0298
Coverage	0.9420	0.9440	0.9540	0.9510

Table 2

Bias estimates for $\widehat{P}_{t_s, \tau, t_s, z}$ where $t_s < t_0$ with corresponding averages of the standard error estimates from the perturbation-resampling method (ASE), empirical standard errors (ESE), and empirical 95% coverage levels (coverage) from the (a) low event rate setting and (b) moderate event rate setting. Values of t_s were chosen as the 20th, 40th, 60th and 80th percentiles of $T_s \leq t_0$.

	$t_{0.20}$	$t_{0.40}$	$t_{0.60}$	$t_{0.80}$
<i>(a) Low event rate setting</i>				
Bias	−0.0014	−0.0009	0.0000	0.0007
ASE	0.0514	0.0425	0.0365	0.0325
ESE	0.0547	0.0443	0.0382	0.0325
Coverage	0.9510	0.9390	0.9510	0.9510
<i>(b) Moderate event rate setting</i>				
Bias	−0.0031	−0.0049	−0.0030	−0.0019
ASE	0.0428	0.0361	0.0318	0.0297
ESE	0.0437	0.0356	0.0321	0.0308
Coverage	0.9520	0.9510	0.9540	0.9450

Table 3

Estimated apparent and cross-validated (CV) average and standard deviation estimates for the Brier score using T_S , Z information ($BS_{t_0, \tau}^{(2)}$), Z information only ($BS_{t_0, \tau}^{(1)}$), and no information ($BS_{t_0, \tau}^{(0)}$) and the increase in proportion of explained variation (IPEV) from the (a) low event rate setting and (b) moderate event rate setting with $t_0 = 2$ years and $\tau = 8$ years.

	Apparent		CV (3-fold)	
	Average	SD	Average	SD
(a) Low event rate setting				
$BS_{t_0, \tau}^{(2)}$	0.1393	0.0046	0.1930	0.0066
$BS_{t_0, \tau}^{(1)}$	0.1582	0.0046	0.2139	0.0065
$BS_{t_0, \tau}^{(0)}$	0.1623	0.0047	0.2192	0.0066
IPEV _{2/0}	0.1418	0.0158	0.1193	0.0164
IPEV _{1/0}	0.0250	0.0063	0.0241	0.0064
IPEV _{2/1}	0.1197	0.0149	0.0975	0.0156
(b) Moderate event rate setting				
$BS_{t_0, \tau}^{(2)}$	0.1810	0.0038	0.2045	0.0045
$BS_{t_0, \tau}^{(1)}$	0.2076	0.0035	0.2303	0.0040
$BS_{t_0, \tau}^{(0)}$	0.2134	0.0033	0.2366	0.0037
IPEV _{2/0}	0.1518	0.0134	0.1358	0.0143
IPEV _{1/0}	0.0273	0.0060	0.0266	0.0060
IPEV _{2/1}	0.1280	0.0117	0.1122	0.0124

Table 4

Estimated apparent and cross-validated (CV) average and standard deviation estimates for the AUC using T_g , Z information ($AUC_{t_0, \tau}^{(2)}$), Z information only ($AUC_{t_0, \tau}^{(1)}$), and no information ($AUC_{t_0, \tau}^{(0)}$) from the (a) low event rate setting and (b) moderate event rate setting with $t_0 = 2$ years and $\tau = 8$ years.

	Apparent		CV (3-fold)	
	Average	SD	Average	SD
(a) <i>Low event rate setting</i>				
$AUC_{t_0, \tau}^{(2)}$	0.7389	0.0140	0.7306	0.0147
$AUC_{t_0, \tau}^{(1)}$	0.5974	0.0125	0.5975	0.0125
$AUC_{t_0, \tau}^{(0)}$	0.5000	0.0000	0.5000	0.0000
(b) <i>Moderate event rate setting</i>				
$AUC_{t_0, \tau}^{(2)}$	0.7290	0.0105	0.7224	0.0112
$AUC_{t_0, \tau}^{(1)}$	0.5889	0.0100	0.5888	0.0100
$AUC_{t_0, \tau}^{(0)}$	0.5000	0.0000	0.5000	0.0000

Estimates of $\widehat{P}_{t_0, \tau, t_s, z}$ and $\widehat{P}_{t_s, \tau, t_s, z}$ with corresponding standard error estimates from the perturbation-resampling method (SE) when $t_0 = 6$ months, $\tau = 1$ year and using erythropoietin versus not using erythropoietin as the single discrete marker. Values of t_s were chosen as the 20th, 40th, 60th and 80th percentiles of $T_s \leq t_0$.

Table 5

t_s	Using erythropoietin				Not using erythropoietin			
	10 days	24 days	49 days	97 days	10 days	24 days	49 days	97 days
$\widehat{P}_{t_0, \tau, t_s, z} = P(T_L \mid t_0 + \tau \mid T_L > t_0, T_S = t_s, t_0, Z = z)$								
Estimate	0.2865	0.3088	0.3312	0.3375	0.3273	0.3323	0.3419	0.3565
SE	0.0217	0.0178	0.0169	0.0182	0.0197	0.0162	0.0148	0.0165
$\widehat{P}_{t_s, \tau, t_s, z} = P(T_L \mid t_s + \tau \mid T_L > t_s, T_S = t_s, t_0, Z = z)$								
Estimate	0.4091	0.4373	0.4491	0.3907	0.5678	0.5200	0.4698	0.4247
SE	0.0275	0.0240	0.0217	0.0221	0.0250	0.0212	0.0191	0.0193

Table 6

Apparent and cross-validated (CV) estimates and corresponding standard errors from the perturbation-resampling method (SE) for Brier score and AUC using hospitalization time and marker information ($BS_{t_0, \tau}^{(2)}$, $AUC_{t_0, \tau}^{(2)}$), marker information only ($BS_{t_0, \tau}^{(1)}$, $AUC_{t_0, \tau}^{(1)}$), and no information ($BS_{t_0, \tau}^{(0)}$, $AUC_{t_0, \tau}^{(0)}$) and the increase in proportion of explained variation (IPEV) using ESRD patient data set with erythropoietin use as the single discrete marker with $t_0 = 6$ months and $\tau = 1$ year.

	Apparent	CV (3-fold)	SE
<i>Brier score</i>			
$BS_{t_0, \tau}^{(2)}$	0.1736	0.1741	0.0025
$BS_{t_0, \tau}^{(1)}$	0.1791	0.1792	0.0026
$BS_{t_0, \tau}^{(0)}$	0.1795	0.1796	0.0026
$IPEV_{2/0}$	0.0330	0.0303	0.0045
$IPEV_{1/0}$	0.0021	0.0020	0.0011
$IPEV_{2/1}$	0.0309	0.0283	0.0044
<i>AUC</i>			
$AUC_{t_0, \tau}^{(2)}$	0.6141	0.6111	0.0080
$AUC_{t_0, \tau}^{(1)}$	0.5272	0.5275	0.0070
$AUC_{t_0, \tau}^{(0)}$	0.5000	0.5000	0.0000