Biometrics 64, 1–18 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2008

ESTIMATING SUBJECT-SPECIFIC DEPENDENT COMPETING RISK PROFILE WITH CENSORED EVENT TIME OBSERVATIONS

Yi Li[1,*], Lu Tian[2,], and L.J. Wei[1,***]**

[1] Department of Biostatistics, Harvard University, USA

[2] Department of Health Research & Policy, Stanford University, USA

**email:* yili@jimmy.harvard.edu

***email:* lutian@stanford.edu

****email:* wei@hsph.harvard.edu

Summary: In a longitudinal study, suppose that the primary endpoint is the time to a specific event. This response variable, however, may be censored by an independent censoring variable or by the occurrence of one of several dependent competing events. For each study subject, a set of baseline covariates is collected. The question is how to construct a reliable prediction rule for the future subject's profile of all competing risks of interest at a specific time point for risk-benefit decision makings. In this paper, we propose a two-stage procedure to make inferences about such subject-specific profiles. For the first step, we use a parametric model to obtain a univariate risk index score system. We then estimate consistently the average competing risks for subjects which have the same parametric index score via a nonparametric function estimation procedure. We illustrate this new proposal with the data from a randomized clinical trial for evaluating the efficacy of a treatment for prostate cancer. The primary endpoint for this study was the time to prostate cancer death, but had two types of dependent competing events, one from cardiovascular death and the other from death of other causes.

Key words: Local likelihood function; Nonparametric function estimation; Perturbation-resampling method; Risk index score.

1. INTRODUCTION

Consider a longitudinal clinical study whose primary endpoint is the time to a specific clinical event. However, this event time is possibly censored by an independent censoring variable or by the occurrence of one of several *dependent* competing events. For example, in a randomized clinical trial to evaluate the efficacy of estrogen diethylstilbestrol (DES) for treating stage 3 or 4 prostate cancer, 242 patients were randomly assigned to two high dose groups ($\geqslant 1$ mg/day) and 241 subjects were assigned to two low dose groups ($\leqslant 0.2$ mg/day) (Byar and Green, 1980; Cheng et al., 1998). The primary endpoint for the study is the time to prostate cancer death. At the end of the study, there were 48, 78 and 34 deaths due to prostate cancer, cardiovascular diseases and other causes in the high dose groups. For the low dose groups, the corresponding numbers of deaths are 77, 61 and 46, respectively. With respect to the overall survival, the high dose groups appeared to be superior to the low dose groups. Furthermore, the treatment with high doses of DES reduced the prostate cancer death. However, there was a serious concern about its potential fatal cardiovascular-related toxicity.

To quantify the "pure" treatment effect for prostate cancer in the presence of possibly dependent competing risks is a rather challenging task, if not impossible (Tsiatis, 1975). The risk-benefit decision makings on the proper usage of DES should depend on the entire profile of all competing risks, not solely on the prostate cancer mortality. Moreover, since the choice of balancing the risk and benefit is rather subject-specific, it is important to know how to utilize the future patient's "baseline" characteristics to predict such an individual-level competing risk profile.

A classical method of handling dependent competing risk problem is to model the so-called cause-specific hazard function for the primary endpoint via the Cox proportional hazards model (Cox, 1972). However, it is not clear how to utilize this technique to make survival

predictions (Kalbfleisch and Prentice, 1980; Pepe and Mori, 1993). A useful alternative to deal with competing risks is to consider the cumulative incidence functions (Benichou and Gail, 1990; Gaynor et al., 1993; Gelman et al., 1990; Korn and Dorey, 1992; Goldhirsch et al., 1994). Recently Cheng et al. (1998) and Fine and Gray (1999) modeled the cumulative incidence function with the subject's covariates, for example, via a Cox-type model. Further novel procedures along this line have been studied, for example, by Fine (2001), Klein and Andersen (2005), Klein (2006) and Scheike et al. (2008). Another fruitful class of parametric or semi-parametric methods is to consider latent failure time modeling (Kalbfleisch and Prentice, 2002; Lawless, 2003; Andersen et al., 2002; Li et al., 2007) to analyze the competing risk data. The validity of predicting the competing risk profiles based on a parametric or semi-parametric model is heavily dependent on the adequacy of the fitted model.

In this paper, we are interested in constructing subject-level predictions of all dependent competing risks of interest at a specific time point, or a set of time points. When, for each subject, more than one baseline covariate is involved, a purely nonparametric function estimation procedure for the above event rates may not perform well even with relatively large samples. Here, we consider the case that there is a primary event of interest and construct a two-stage procedure. For the first step, we use a parametric or semi-parametric model to create a univariate risk index score predictive to the event rate of the primary interest. We then use a nonparametric function estimation method to make joint inferences about the average competing risks for subjects with the same index score. The new proposal is illustrated with the data from the above DES study.

For the case with only one risk category involved, Cai et al. (2010) utilized a similar two stage procedure for predicting the mean risk of subjects who have the same parametric risk score. Other novel semi-parametric methods for predicting risk of a single event with a high dimensional covariate vector have been proposed, for example, by Bair and Tibshirani

(2004) using the supervised principal component approach. In the present paper, we took an approach of using a local multinomial likelihood for nonparametric smoothing technique at the second stage, which is a non-trivial generalization from Cai et al. (2010). We conducted an extensive numerical study to examine the performance of the new procedure compared with a one-step semi-parametric method, for example, using the generalized additive models. The new proposal appears to be superior to its one-step counterparts with respect to the mean squared error criterion.

2. CONSISTENT ESTIMATION FOR MEAN COMPETING RISKS OF SUBJECTS WITH THE SAME PARAMETRIC RISK SCORE

Suppose that there are K distinct types of possibly dependent competing events. For a random subject in the study, let $\tilde{T}$ be the study time period from the study entry to the first time point at which one out of these K events occurs. Let ϵ be a random variable whose possible values are $\{1, \cdots, K\}$. If $\epsilon = k$, Type k event is observed at $\tilde{T}$. Also, let U be the subject's "baseline" covariate vector. Furthermore, suppose that we are interested in the K conditional event rates at a specific time point t_0, that is,

$$\pi_k(U) = \mathrm{pr}(\tilde{T} \leqslant t_0, \epsilon = k \mid U), \ k = 1, \cdots, K. \tag{2.1}$$

In practice, $\tilde{T}$ is often censored by an independent continuous variable C with an unknown survival distribution $G(\cdot)$. Assume that C is independent of $\tilde{T}$ and U. Let $T = \min(\tilde{T}, C)$ and $\Delta = I(\tilde{T} = T)$, where $I(\cdot)$ is the indicator function. Also, let $\{(\tilde{T}_i, C_i, \epsilon_i, U_i), i = 1, \cdots, n\}$ be n independent copies of $(\tilde{T}, C, \epsilon, U)$. The problem is how to make inference about (2.1) based on the incomplete event time observations $\{(T_i, \Delta_i\epsilon_i, U_i), i = 1, \cdots, n\}$. Unfortunately, if the dimension of U is greater than one, any existing nonparametric regression estimator for (2.1) may not perform well even when the sample size n is large and the event rates are not extremely low or high. Instead of estimating such fine subject-level event rates (2.1),

a feasible, practical alternative is to construct a *univariate* parametric risk index system based on U and group the study subjects with respect to this scoring system. Then using a univariate nonparametric function estimation procedure, one may estimate consistently these K average competing event rates for each stratum whose subjects have the same index score.

To construct a univariate scoring system, we consider the case that there is a primary event of interest for the study, say, the event corresponding to $\epsilon = 1$. Let X, a $p \times 1$ vector, be a function of U and the first component of X is one. Let X_i be the counterpart of X from $U_i, i = 1, \cdots, n$. Consider a parametric *working* model for the primary event rate:

$$\pi_1(U) = g(\beta' X), \tag{2.2}$$

where g is a known strictly increasing, smooth function, for example, the anti-logit function, and β is a $p \times 1$ vector of unknown parameters. Without censoring, one may use the maximum likelihood estimator or a simple estimating function such as

$$n^{-1} \sum_{i=1}^{n} X_i \{ I(T_i \leqslant t_0, \epsilon = 1) - g(\beta' X_i) \} \tag{2.3}$$

to estimate β.

In the presence of independent right censoring, one may modify (2.3) by adjusting censoring. One possible modification is

$$R(\beta) = n^{-1} \sum_{i=1}^{n} \frac{w_i}{\hat{G}(T_i \wedge t_0)} X_i \{ I(T_i \leqslant t_0, \epsilon = 1) - g(\beta' X_i) \}, \tag{2.4},$$

where $w_i = I(T_i \wedge t_0 \leqslant C_i) = I(T_i \leqslant t_0)\Delta_i + I(T_i \geqslant t_0)$ and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator for $G(\cdot)$. This generalization has been studied, for example, by Zheng et al. (2007) and Uno et al. (2008) in different settings. Heuristically, for a large sample size n, conditional on $\tilde{T}$ and U, the expected value of $w_i / \hat{G}(T_i \wedge t_0)$ is one. This implies that for large n, $R(\beta) \approx$ (2.3). Therefore, asymptotically one would expect that a root $\hat{\beta}$ to $R(\beta) = 0$ is free of the study-specific censoring distribution $G(\cdot)$. It is important to note that under rather mild conditions, $\hat{\beta}$ converges to a finite value β_0 even when the model (2.2) is not correctly specified (Uno et

al., 2008). This stability property, coupled with the fact that β_0 is free of the study-specific censoring distribution, is essential for developing our inference procedures. Note that if the model (2.2) is correctly specified, $g(\hat{\beta}'X)$ would be a consistent estimator for (2.1).

Now, consider a future subject from the same study population, whose U and X are U^0 and X^0 with potential, but unobservable $(\tilde{T}, \epsilon)' = (\tilde{T}^0, \epsilon^0)'$. Let $\hat{\beta}'X^0 = v$, a given constant. We are interested in estimating the following $(K-1)$ average event rates at time t_0 :

$$\text{pr}(\tilde{T}^0 \leqslant t_0,\ \epsilon^0 = k \ \mid\ \hat{\beta}'X^0 = v),\ k = 1, \cdots, K-1, \tag{2.5}$$

where the probability is with respect to the future observation $(U^0, \tilde{T}^0, \epsilon^0)$ as well as the observed data $\{(T_i, \Delta_i\epsilon_i, U_i), i = 1, \cdots, n\}$, from which $\hat{\beta}$ is estimated. Note that the probabilities in (2.5) depend on the sample size n and are convergent to the following conditional probabilities

$$\eta_k(v) = \text{pr}(\tilde{T}^0 \leqslant t_0,\ \epsilon^0 = k \ \mid\ \beta_0'X^0 = v),\ k = 1, \cdots, K-1, \tag{2.6}$$

as $n \to \infty$. Also note that (2.6) is the set of the multinomial cell probabilities for future subjects whose limiting risk score is v. For the non-censored case, let us consider a nonparametric estimation procedure for $\eta(v) = \{\eta_1(v), \cdots, \eta_{K-1}(v)\}'$ via a localized multinomial likelihood function. Specifically, let $Y_{ik} = I(T_i \leqslant t_0, \epsilon_i = k)$ for $k = 1, \ldots, K-1$, and $\hat{\beta}'X_i = \hat{V}_i$. For notational ease, write $p_k = \eta_k(v)$, the probability of failing with cause k prior to time t_0 given score v, for $k = 1, \cdots, K-1$. Then, a kernelized log-likelihood function for $\eta(v)$, expressed with the unknown parameter vector $p = (p_1, \cdots, p_{K-1})'$ is

$$\sum_{i=1}^{n} K_h(\hat{V}_i - v) \sum_{k=1}^{K-1} \log \left\{ p_k^{Y_{ik}} (1 - \sum_{k=1}^{K-1} p_k)^{1 - \sum_{k=1}^{K-1} Y_{ik}} \right\}, \tag{2.7}$$

where $p_k \geqslant 0$, $\sum_{k=1}^{K-1} p_k \leqslant 1$, $K_h(s) = K(s/h)/h$ for a symmetric standard kernel function $K(\cdot)$ with a finite support and h is the smooth parameter.

In the presence of censoring, we add a weight function $w_i/\hat{G}(T_i \wedge t_0)$ in front of $K_h(\cdot)$ in

(2.7). The resulting log-likelihood is

$$\sum_{i=1}^{n} \frac{w_i}{\hat{G}(T_i \wedge t_0)} K_h(\hat{V}_i - v) \sum_{k=1}^{K-1} \log \left\{ p_k^{Y_{ik}} (1 - \sum_{k=1}^{K-1} p_k)^{1-\sum_{k=1}^{K-1} Y_{ik}} \right\}. \tag{2.8}$$

An estimator for $\eta(v) = (\eta_1(v), \ldots, \eta_{K-1})$ can be obtained by maximizing (2.8) with respect to p's with the above constraints.

The performance of this nonparametric local estimator may be improved by replacing p_k of each summand in (2.8) by

$$\frac{\exp\{a_k + b_k(\hat{V}_i - v)\}}{1 + \sum_{i=1}^{K-1} \exp\{a_k + b_k(\hat{V}_i - v)\}},$$

where $a = (a_1, \cdots, a_{K-1})'$ and $b = (b_1, \cdots, b_{K-1})'$ are unknown vectors of parameters. Here, the rational is to use a linear function $a_k + b_k(V - v)$ to approximate $\log\{\eta_k(V)/\eta_K(V)\}$ in a small neighborhood of v (Fan and Gijbels, 1996). The resulting log-likelihood function is

$$\ell(a, b; v) = \sum_{i=1}^{n} \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \sum_{k=1}^{K-1} \Bigg(Y_{ik}\{a_k + b_k(\hat{V}_i - v)\} - \log[1 + \sum_{k=1}^{K-1} \exp\{a_k + b_k(\hat{V}_i - v)\}] \Bigg). \tag{2.9}$$

Let $\hat{a}$ and $\hat{b}$ be the maximizers for $\ell(a, b; v)$ with respect to a and b. Also, let $\hat{\eta}_k(v)$ be

$$\exp(\hat{a}_k)/[1 + \sum_{k=1}^{K-1} \exp(\hat{a}_k)], \quad k = 1, \cdots, K-1. \tag{2.10}$$

3. CONSTRUCTING POINTWISE AND SIMULTANEOUS CONFIDENCE INTERVALS FOR $\eta_k(\cdot)$ OVER THE RISK SCORE

We can show that when $h = O(n^{-\nu}), 1/5 < \nu < 1/2$, $\hat{\eta}_k(v)$ is consistent estimator for $\eta_k(v), k = 1, \cdots, K-1$ (see web based appendix A). Moreover, the joint distribution of

$$\{(nh)^{1/2}[f\{\hat{\eta}_k(v)\} - f\{\eta_k(v)\}], k = 1, \cdots, K-1\} \tag{3.1}$$

can be approximated by a multivariate normal with mean 0 and covariance matrix $\Sigma(v)$, where $f(\cdot) : [0, 1] \to [-\infty, \infty]$ is a given smooth, strictly increasing function. In this paper, we let $f(\cdot)$ be the logit function.

To estimate the covariance matrix $\Sigma(v)$ associated with (3.1), we utilize a perturbation-resampling procedure which is similar to a wild bootstrapping method (Mammen, 1993) and has been successfully applied to many interesting inference problems, especially in survival analysis (Gilbert et al, 2004; Tian et al. 2005). Specifically, let $\{B_i, i = 1, \cdots, n\}$ be a random sample from the unit exponential. Let $a^* = \{a_1^*, \cdots, a_{K-1}^*\}'$ be the minimizer of $\ell^*(a, b; v)$, a perturbed version of (2.9), where

$$\ell^*(a,b;v) = \sum_{i=1}^{n} B_i \frac{w_i K_h(V_i^* - v)}{G^*(T_i \wedge t_0)} \left(\sum_{k=1}^{K-1} Y_{ik}\{a_k + b_k(V_i^* - v)\} - \log[1 + \sum_{k=1}^{K-1} \exp\{a_k + b_k(V_i^* - v)\}] \right).$$

Here, $G^*(\cdot)$ and V_i^* are the perturbed counterparts of $\hat{G}(\cdot)$ and $\hat{V}_i$, respectively, i.e.,

$$G^*(t) = \exp\left[-\sum_{i=1}^{n} \int_0^t \frac{B_i d\{I(T_i \leqslant s, \Delta_i = 0)\}}{\sum_{j=1}^{n} B_j I(T_j \geqslant s)} \right],$$

$V_i^* = X_i'\beta^*$ and β^* is the solution to the perturbed estimating equation of (2.4)

$$\sum_{i=1}^{n} \frac{B_i w_i}{G^*(T_i \wedge t_0)} X_i\{I(T_i \leqslant t_0, \epsilon = 1) - g(X_i'\beta)\} = 0.$$

Furthermore, let the corresponding perturbed $\eta_k^*(v)$ be

$$\exp(a_k^*)/\{1 + \sum_{k=1}^{K-1} \exp(a_k^*)\}, k = 1, \cdots, K-1.$$

In the web based appendix B, we show that the covariance matrix $\Sigma(v)$ can be consistently estimated by $\hat{\Sigma}(v)$, the expectation of $(nh)[f\{\eta^*(v)\} - f\{\hat{\eta}(v)\}][f\{\eta^*(v)\} - f\{\hat{\eta}(v)\}]'$ (conditional on the observed data), where $f\{\eta^*(v)\} = (f\{\eta_1^*(v)\}, \cdots, f\{\eta_{K-1}^*(v)\})'$ and $f\{\hat{\eta}(v)\} = (f\{\hat{\eta}_1(v)\}, \cdots, f\{\hat{\eta}_{K-1}(v)\})'$. Noting that $\{G^*(\cdot), V_1^*, \cdots, V_n^*\}$ can be replaced by $\{\hat{G}(\cdot), \hat{V}_1, \cdots, \hat{V}_n\}$ in the perturbation without effecting the asymptotical distributions of $(nh)^{1/2}[f\{\eta^*(v)\} - f\{\hat{\eta}(v)\}]$, since the differences between $G^*(\cdot)$ and $\hat{G}(\cdot)$ as well as between V_i^* and $\hat{V}_i$ are in the order of $O_p(n^{-1/2})$ which is smaller than $f\{\eta^*(v)\} - f\{\hat{\eta}(v)\} = O_p\{(nh)^{-1/2}\}$. However, our experience suggests that the simultaneous perturbation on $\hat{G}(\cdot)$ and $\hat{V}_i$ often can improve the finite-sample performance of the proposed resampling method.

To obtain an approximation to $\hat{\Sigma}(v)$ for a given data set, we generate a large number,

M, of independent realizations from $\{B_i, i = 1, \cdots, n\}$. For each realization, we obtain a realization of $f\{\eta^*(v)\}$. With M such independent realizations, one may use the standard sample covariance matrix estimate $\tilde{\Sigma}(v)$ or a robust version thereof to estimate $\Sigma(v)$. This, coupled with the normal approximation to the distribution of $f\{\hat{\eta}(v)\}$, provides confidence intervals for $f\{\eta_k(v)\}$. A two sided $(1-\alpha)$ confidence interval for $\eta_k(v)$ is

$$f^{-1}[f\{\hat{\eta}_k(v)\} \pm z_{(1-\alpha/2)}(nh)^{-1/2}\tilde{\sigma}_k(v)], \tag{3.2}$$

where $f(\cdot)$ is the logit function, $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution and $\tilde{\sigma}_k(v)$ is the standard error estimate from the kth diagonal element of $\tilde{\Sigma}(v)$. Note that joint confidence regions for $\{\eta_k(v), k = 1, \cdots, K-1\}$ can also be obtained by considering a sup-type statistic: $\sup_{k=1,\cdots,K-1}|\hat{\eta}_k(v) - \eta_k(v)|$ to choose the cutoff point for the confidence intervals (3.2).

To construct a $(1-\alpha)$ simultaneous confidence band for $\eta_k(v)$ over a pre-specified interval $\mathcal{I}$ of v, we cannot use use the conventional method based on a sup-statistic, $\sup_{v\in\mathcal{I}}\tilde{\sigma}_k^{-1}(v)|(nh)^{1/2}\{\hat{\eta}_k(v) - \eta_k(v)\}|$ due to the fact that as a process in v, the limiting distribution of $(nh)^{1/2}\{\hat{\eta}_k(v) - \eta_k(v)\}$ does not exist. On the other hand, one may utilize the strong approximation argument given in Bickel and Rosenblatt (1973) to show that the appropriately scaled sup of a specific transformation of $\hat{\eta}_k(v)$ converges to a proper random variable in distribution. In practice, a $(1-\alpha)$ simultaneous confidence band for $\{\eta_k(v), v \in \mathcal{I}\}$ is

$$f^{-1}[f\{\hat{\eta}_k(v)\} \pm c_\alpha(nh)^{-1/2}\tilde{\sigma}_k(v)], \tag{3.3}$$

where c_α is obtained via the following equation:

$$\mathrm{pr}(\sup_{v\in\mathcal{I}}\tilde{\sigma}_k^{-1}(v)|(nh)^{1/2}[f\{\eta_k^*(v)\} - f\{\hat{\eta}_k(v)\}]| < c_\alpha) = 1-\alpha,$$

and $\{\eta_k^*(v), v \in \mathcal{I}\}$ is obtained by the above perturbation-resampling method with the same set of $\{B_i, i = 1, \cdots, n\}$. The justification of adequacy of this approximation is given in web based appendix B. Note that unlike the pointwise confidence interval estimation for $\eta_k(v)$,

it does not seem trivial to generalize the above simultaneous confidence interval estimation for all $k = 1, \cdots, K-1$.

Like any typical nonparametric function estimation problem, it is important to know how to choose the smooth parameter h in practice. Here, we propose a $J-$fold cross-validation method to choose an optimal h value. To this end, we first randomly divide the entire data set D into J mutually exclusive, roughly equally sized subsets, say, $D_1, \cdots, D_J$. Let the set of observations in D, but not in D_j, be denoted by $D_{(-j)}$, $j = 1, \cdots, J$. We construct the scoring system using $\hat{\beta}_{(-j)}$ estimated with the observations in $D_{(-j)}$. Next, for a fixed h value, let the corresponding nonparametric estimator for $\eta_k(v)$ be $\hat{\eta}_{(k,-j)}(v)$. With these subject-specific risk estimates, we compute the log-likelihood function with observations in D_j:

$$\ell_j(h) = -\sum_{l \in D_j} \frac{w_l}{\hat{G}(T_l \wedge t_0)} \left[\sum_{k=1}^{K} Y_{lk} \log\{\hat{\eta}_{(k,-j)}(\hat{V}_{(l,-j)})\} \right], \tag{3.4}$$

where $\hat{\eta}_{(K,-j)}(v) = 1 - \sum_{k=1}^{K-1} \hat{\eta}_{(k,-j)}(v)$ and $\hat{V}_{(l,-j)} = \hat{\beta}'_{(-j)} X_l, l \in D_j$. Now, let $\ell_{cv}(h) = \sum_{j=1}^{J} \ell_j(h)$. We choose the maximizer h_{op} of $\ell_{cv}(h)$ as an "optimal" choice of the smooth parameter h.

It follows from the argument in Härdle et al. (1988), we expect that the above h_{op} is in the order of $O_p(n^{-1/5})$. To ensure that the validity of the aforementioned large sample properties for $\hat{\eta}_k(v)$, one may choose a smooth parameter $h = h_{op} \times n^{-\xi}$ where $0 < \xi < 3/10$. In practice, we find that the resulting nonparametric estimator performs well with $\xi = 0.1$.

4. AN ILLUSTRATIVE EXAMPLE

We use the new proposal to study a subset of the data from the DES trial discussed in the Introduction section. This data set consists of patient-level observations from the high DES dose groups. There were 242 patients in these groups with a median followup time of 63 months. Here, $\tilde{T}$ is the time from randomization to one of $K = 4$ competing events, and $\epsilon = 1$, for prostate cancer death; $= 2$, for cardiovascular related death; and $= 3$, for other

causes of death; = 4, for surviving beyond t_0. At the end of the study, there were 48, 78 and 34 patients died due to prostate cancer, heart diseases and other causes, respectively. The baseline covariate vector U includes Age (AG), weight index (WT), performance rating (PF), history of cardiovascular disease (HX), serum hemoglobin (HG), size of primary lesion (SZ) and Gleason score (SG). Since this data set was analyzed in the past using a discretized coding system for the covariates due to an easy clinical interpretation (Byar and Green, 1980; Cheng et al., 1998), we followed the same system in our analysis. For convenience to readers, the coding for covariates is summarized in Table 1.

[Table 1 about here.]

First, we consider a case for predicting the subject-level relatively long term competing risks. To this end, let $t_0 = 5$ (years). Since the primary endpoint of the study is the time to prostate cancer death, we fitted the data with a working model (2.2) by letting $X = (1, U)'$ and g be the anti-logit function. The point estimate $\hat{\beta}$ for β via (2.4) is given in Table 2 (a). Although only WY, SZ and SG are statistically significant with this *working* model, we used the entire covariate vector U to build the risk scoring $V = \hat{\beta}'X$. In Figure 1(a), we present a smoothed density estimate of $\hat{V}$, which is a bimodal function. Most study subjects are clustered around $\hat{V} = -4.5$ and -0.9.

[Table 2 about here.]

To estimate $\eta_k(v), k = 1, 2, 3$, we let the kernel $K_h(\cdot)$ for $\hat{\eta}_k(v)$ be the standard Epanechnikov function. The smoothing parameter h was chosen by minimizing $\ell_{cv}(h)$ defined in Section 3 with a 10-fold cross-validation procedure. This results in $h = 0.97$. Lastly we let the 2nd and 98th percentiles of the empirical distribution of $\hat{V}$ be the the boundary points of $\mathcal{I}$. We then constructed pointwise and simultaneous confidence intervals for $\{\eta_k(v), v \in \mathcal{I}\}$ with $M = 1000$ realizations of the random sample from the unit exponential for the perturbation-resampling procedure. In Figure 1(b), for the prostate cancer 5-year

mortality rate estimation, we present the point estimates $\{\hat{\eta}_1(v), v \in \mathcal{I}\}$ with the solid curve, and the 0.95 pointwise intervals (enclosed by dotted curves) and simultaneous band (gray shaded zone). For example, the estimated average prostate cancer mortality rate for patients with an index score of -4.5 was 0.012 with a 95% simultaneous confidence interval of (0.0006,0.17) and a 95% pointwise confidence interval of (0.002,0.05), while the estimated average prostate cancer mortality rate for patients with an index score of -0.9 was 0.35 with a 95% simultaneous confidence interval of (0.27,0.41) and a 95% pointwise confidence interval of (0.30,0.38). In Figure 1(c)(d), we present their counterparts with respect to cardiovascular disease related death and death from other causes.

[Figure 1 about here.]

Note that the 5-year rate from "other causes" is rather flat over v. On the other hand, patients with low risk scores (< -2), the prostate cancer death rates are low. However, the CV mortality rates are high. Therefore, for this group of future patients treated by DES high doses, one would closely monitor the patients' CV functions. For patients with high risk score (> -2), it seems that a high dose DES may not be a good choice for treating prostate cancer.

Now, suppose that we are also interested in predicting a short term competing risk profile. To this end, we let $t_0 = 2$ (years). We fitted the data with a parametric working model (2.2). Here, $X = (1, U)'$. The estimated regression coefficients are given in Table 2(b). Note that these estimates appear to be markedly different from those for the case with $t_0 = 5$ (years), suggesting that the risk score system may depend on the time point of interest. Using the same setting as that for the above long term competing risk prediction problem, the resulting smoothing parameter value h is 1.29. The corresponding profiles for the dependent risks are given in Figure 2. For the present case, the mortality rates for CV death or "other causes"

death are relatively flat over the entire index score. On the other hand, it appears that the high dose DES works well for patients whose risk scores are lower than -2.

[Figure 2 about here.]

5. A NUMERICAL STUDY FOR EXAMINING THE PERFORMANCE OF THE NEW PROCEDURE

We conducted an extensive numerical study to examine the performance of our proposal under practical settings. Instead of using the proposed two stage procedure, one may adapt a generalized additive model (GAM) approach for multinomial outcome data (Yee and Wild, 1996; Yee, 2010) to estimate the cell probability $\pi_k(U)$ in the presence of censoring. We found that our procedure generally performs better than such a semi-parametric method. For example, in one of simulation settings, we mimicked the prostatic cancer example and considered the case with $K = 4$ and a 3×1 vector of covariates, $U = (U_{(1)}, U_{(2)}, U_{(3)})'$. For each subject, we first generated the correlated "latent" times D_1, D_2, and D_3 to three distinct deaths (the fourth cell is for survivors beyond t_0) via the following log-liner model:

$$\begin{aligned} \log D_1 &= \alpha_1 U_{(1)} + \alpha_2 U_{(2)} + e_1, \\ \log D_2 &= \alpha_1 U_{(2)} + \alpha_2 U_{(3)} + e_2, \\ \log D_3 &= \alpha_1 U_{(3)} + \alpha_2 U_{(1)} + e_3, \end{aligned}$$

where the random error (e_1, e_2, e_3) follows a trivariate normal distribution with means being 0, variances being 1 and the covariances being $\rho = 0.5$, and the baseline covariates $U_{(1)}, U_{(2)}$ and $U_{(3)}$ are independent standard normals. Then $\tilde{T}$ is the minimum of D_1, D_2 and D_3 and the event indicator ϵ is defined accordingly.

The censoring time C was assumed to be Unif$[0, \xi]$, where ξ was chosen to yield a certain pre-specified censoring level with the corresponding parameter values α's in the above model.

For each simulation setting, 10,000 simulated data sets were generated to examine the performance of our procedure.

With each simulated data set and a fixed t_0, the estimate of $\pi_k(U_i)$ for the ith subject using our procedure is $\hat{\eta}_k(\hat{\beta}' X_i)$, where $X_i = (1, U_i)'$, and $\hat{\beta}$ and $\hat{\eta}_k(\cdot)$ were obtained via (2.4) and (2.10), respectively. We calculated 10,000 mean squared errors

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\{\hat{\eta}_k(\hat{\beta}' X_i) - \pi_k(U_i)\}^2.$$

and then used the resulting sample mean, the empirical MSE (EMSE) to measure the performance of our procedure.

For the GAM approach, we used the following conventional working model:

$$\log\{\pi_k(U)/\pi_4(U)\} = g_{0k} + g_{1k}(U_{(1)}) + g_{2k}(U_{(2)}) + g_{3k}(U_{(3)}), k = 1, 2, 3,$$

where g_{0k} is a unknown constant, and the functions g_{lk} are completely unspecified, but $E\{g_{lk}(U_{(l)})\} = 0, l = 1, 2, 3$. To fit the censored data, we adapted the GAM methodology for multinomial data developed by Yee and Wild (1996). Specifically, we multiplied the weight matrix in their objective function [denoted by W_i in eq. (6) of Yee and Wild (1996)] by a factor of $w_i/\hat{G}(T_i \wedge t_0)$ [defined below (2.4)] to adjust for right censoring. Then, we used a B-spline with 3 degrees of freedom for each function $g_{lk}(\cdot)$ via the VGAM algorithm (Yee, 2010). With 10,000 simulated data sets, we calculated the corresponding EMSE with the GAM procedure.

In Table 3, we report the EMSEs under various sample sizes n, α's in modeling the latent times, and ξ for censoring. For all the cases, our procedure has smaller EMSEs than the GAM-based method. Note that the GAM, though flexible, is only a working model. Therefore, we do not expect the resulting estimators for the cell probabilities to be consistent.

[Table 3 about here.]

6. REMARKS

It is important to note that in this paper the index scoring system is constructed based on the contrast between the primary event rate and the *average* of all other competing risks at a specific time point via (2.2). In general, it is difficult, if not impossible, to create a univariate scoring system for grouping the subjects, which is sensitive to differentiating subject-level risks of all causes. On the other hand, for some specific situations, one may be able to construct a "sharper" index score. For example, in the DES study, since we are particularly concerned about the fatal cardiovascular risks with the high DES dose treatment, for each subject a modified score may be defined as a contrast of two univariate scores, one is $\hat{\beta}'X$ utilized in this article, and the other one is derived by modeling the CV death rate $\pi_2(U)$ via (2.2).

In this paper, we are interested in estimating the competing risks at a fixed time point (or a set of time points). We find that in general, for a subject with a covariate vector U, its score index for predicting long term risks can be quite different from that for short term risks. If a single score system is needed without a specific set of time points of interest, one may fit the data with a Cox-type model for the conditional cumulative incidence function (Fine and Gray, 1999; Cheng et al., 1998), say, for example, of the time to prostate cancer death in the DES example. The resulting risk estimates $\hat{\eta}_k(v), k = 1, \cdots, K-1$, in (2.5) are functions of time t. It would be interesting to examine the properties of these estimates as processes of t for a fixed risk index v. Cheng et al. (1998) proposed parametric counterparts of such estimators, but their estimators are likely biased when the models are not correctly specified.

7. Supplementary Materials

Web Appendices A and B, referenced in Section 3 are available under the Paper Information link at the Biometrics website http: www.tibs.org/biometrics.

ACKNOWLEDGEMENTS

The authors thank the Editor (Professor Thomas A. Louis), the AE and one anonymous referee for their insightful suggestions. The work was partially supported by the NIH grants.

REFERENCES

Andersen, P.K. (2002) Competing risks as a multi-state model. *Statistical Methods in Medical Research,* 11, 203-215.

Bair, E. and Tibshirani, R. (2004) Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data. *PLos*, 2 (4), e108. doi:10.1371/journal.pbio.0020108.

Bair, E., Hastie, T., Paul, D. and Tibshirani, R. (2006) Prediction by supervised principal components. *JASA*, 101, 119-137.

Benichou, J. and Gail, M. (1990) Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. *Biometrics,* 46, 991-1003.

Bickel, P. and Rosenblatt, M. (1973) On Some Global Measures of the Deviations of Density Function Estimates *Annals of Statistics,* 1, 1071-1095.

Byar, D.P. and Green, S.B. (1980) The choice of treatment for cancer patients based on covariate information. *Bulletin du Cancer*, 67, 477-90.

Cai, T., Tian, L., Uno, H., Solomon, S. and Wei, L.J. (2010) Calibrating Parametric Subject-specific Risk Estimation. *Biometrika*, in press.

Cheng, S.C., Fine, J. and Wei, L.J. (1998) Prediction of Cumulative Incidence Function under the Proportional Hazards Model. *Biometrics,* 54, 219-228.

Cox, D. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological),* 34, 187-220.

Fan, J. and Gijbels, I. (2006) *Local polynomial modelling and its applications*, CRC Press, Boca Raton, Florida.

Fine, J. (2001) Regression modeling of competing crude failure probabilities. *Biostatistics,* 2, 85-97.

Fine, J. and Gray, R. (1999) A Proportional Hazards Model for the Subdistribution of a Competing Risk. *JASA*, 94, 496-514.

Gaynor, J.J., Feuer, E., Tan, C., Wu, D., Little, C. and Straus, D. On the Use of Cause-Specific Failure and Conditional Failure Probabilities: Examples From Clinical Oncology Data. *Journal of the American Statistical Association,* 88, 400-409.

Gelman, R., Gelber, R., Henderson, I. C., Coleman, C. N., and Harris, J. R. (1990). Improved methodology for analyzing local and distant recurrence. *Journal of Clinical Oncology*, 8, 548-555.

Gilbert, P. B., Wei, L. J., Kosorok, M. R. and Clemens, J. D. (2004). Simultaneous Inferences on the Contrast of Two Hazard Functions with Censored Observations *Biometrics,* 58, 773-780.

Goldhirsch, A., Gelber, R. D., Price, K. N., et al. (1994). Effect of systemic adjuvant treatment of first sites of breast cancer relapse. *Lancet,* 343, 377-381.

Härdle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? with discussion. *J. Amer. Statist. Assoc.,* 83, 86-101.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons.

Klein, J.P. and Andersen, P. (2005) Regression modeling of competing risks data based on pseudovalues of the cumulative incidence. *Biometrics,* 61, 223-229

Klein, J. (2006) Modelling competing risks in cancer studies. *Statistics in Medicine*, 25, 1015-1034.

Korn, E. and Dorey, F. (1992) Applications of crude incidence curves. *Statistics in Medicine*, 11, 813-829.

Lawless, J. (2003) *Statistical models and methods for lifetime data,* John Wiley & Sons, Inc., New Jersey.

Li, Y., Tiwari, R., Guha, S. (2007) Mixture Cure Survival Models with Dependent Censoring. *Journal of the Royal Statistical Society: Series B*, 69, 285-306.

Mammen, E. (1993) Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *Annals of Statistics,* 21, 255-285.

Pepe, M. and Mori, M. (1993) Kaplan - meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, 12, 737-51.

Scheike, T., Zhang, M. and Gerds, T. (2008) Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95, 205-220.

Tian, L., Zucker, D. and Wei, L.J. (2005) On the Cox Model With Time-Varying Regression Coefficients. *Journal of the American Statistical Association,* 100, 172-183.

Tian, L., Cai, T., Goetghebeur, E., and Wei, L.J. (2007) Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika,* 94, 297-311.

Tusnady, G. (1977) A remark on the approximation of the sample distribution function in the multidimensional case. *Periodica Mathematica Hungarica,* 8, 53-55.

Tsiatis, A. (1975) A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences,* 72, 20-22.

Uno, H., Cai, T., Tian, T. and Wei, L.J. (2007) Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models. *Journal of the American Statistical Association,* 102, 527-537.

Yee, T.W. and Wild, C.J. (1996) Vector generalized additive models. *Journal of the Royal Statistical Society, Series B*, 58, 481-493.

Yee, T. W. (2010) The VGAM package for categorical data analysis. *Journal of Statistical Software,* 32, 1-34,

Zheng Y. and Heagerty, P. (2007) Prospective Accuracy for Longitudinal Markers. *Biometrics,* 63, 332-341.

Received July 2009.

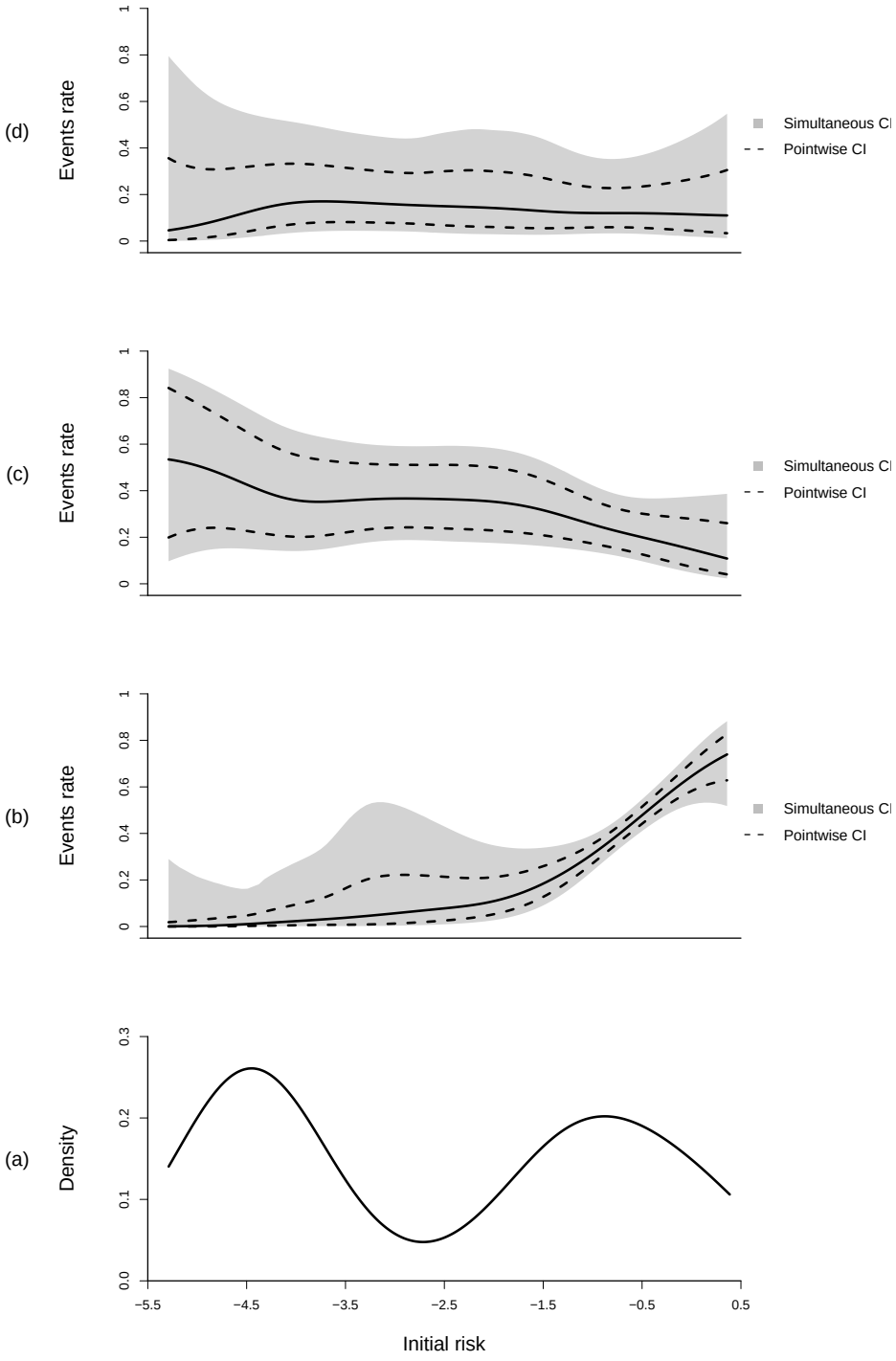

Figure 1. Consistent estimates (solid curve), pointwise 0.95 confidence intervals (enclosed by dotted curves) and simultaneous intervals (gray area) for various risks $\eta_k(v)$ at t_0 =5 years: (a) The density function for the index score; (b) Inference for $\eta_1(v)$; (c) inference for $\eta_2(v)$; (d) Inference for $\eta_3(v)$.

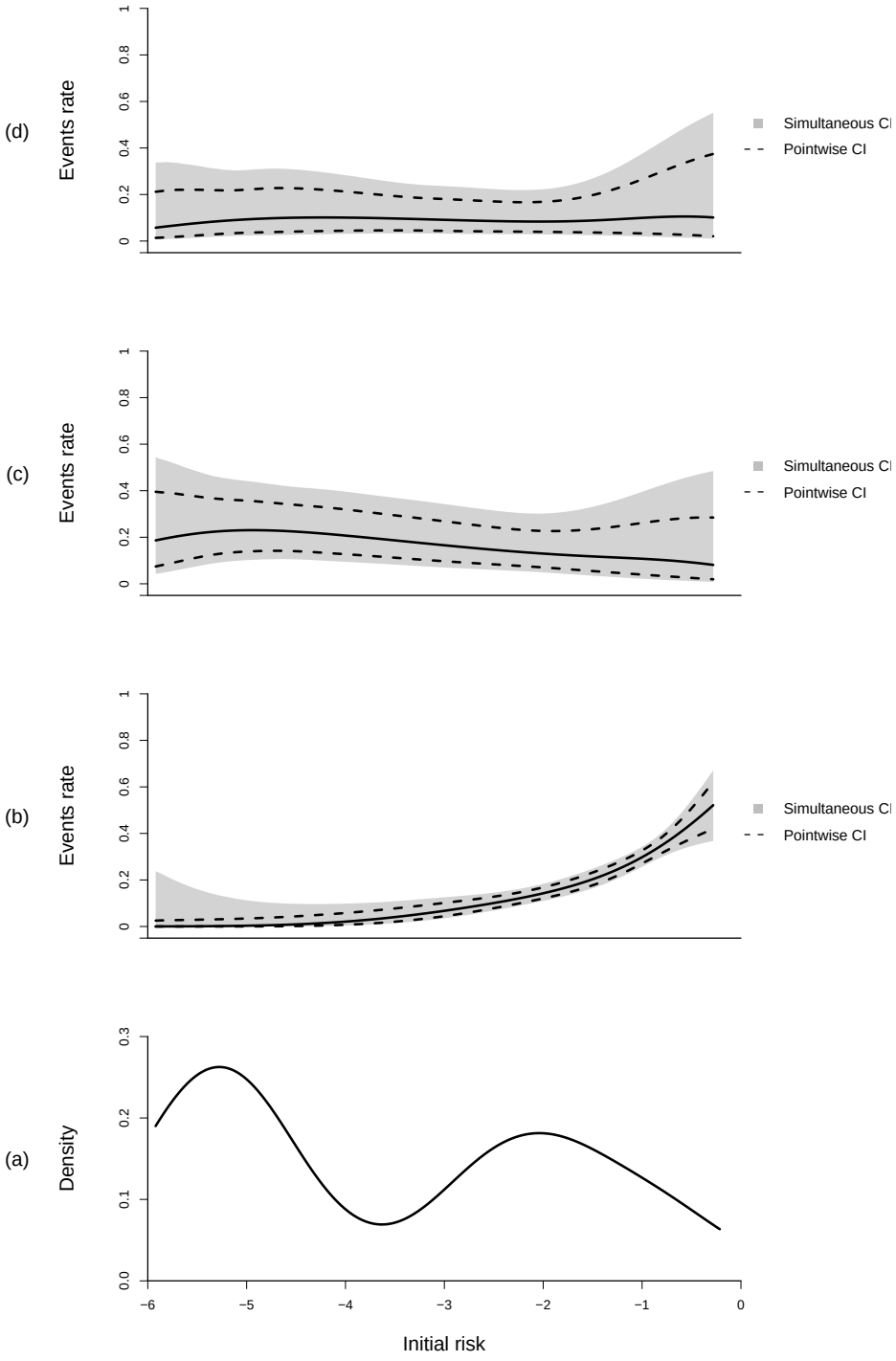

Figure 2. Consistent estimates (solid curve), pointwise 0.95 confidence intervals (enclosed by dotted curves) and simultaneous intervals (gray area) for various risks $\eta_k(v)$ at t_0 =2 years: (a) The density function for the index score; (b) Inference for $\eta_1(v)$; (c) inference for $\eta_2(v)$; (d) Inference for $\eta_3(v)$.

Table 1
Coding of the covariates for the prostate cancer data

	value		
Variable	0	1	2
AG	$<$ 75 years	75-79 years	$\geqslant$ 80 years
WT	$\geqslant$ 100	80-99	$<$ 80
PF	Normal	Limited	
HX	No	Yes	
HG	$\geqslant$ 12g/100ml	9-11.9 g/100ml	$<$9 g/100ml
SZ	$<$ 30 cm^2	$\geqslant$ 30 cm^2	
SG	$\geqslant$ 10	$>$ 10	

Table 2
Regression Coefficient Estimates for Model (2.2) with the data from the high dose groups

(a) Time point $t_0 = 5$ years

Coefficient	Estimate	Std. Error	p-value
Intercept	-4.64	0.79	< 0.01
AG	-0.07	0.31	0.81
WT	0.66	0.37	0.07
PF	0.56	0.61	0.35
HX	-0.56	0.46	0.23
HG	0.46	0.42	0.27
SZ	1.76	0.50	< 0.01
SG	3.37	0.75	< 0.01

(b) Time point $t_0 = 2$ years

Coefficient	Estimate	Std. Error	p-value
Intercept	-5.87	1.12	< 0.01
AG	-0.18	0.39	0.63
WT	0.74	0.40	0.06
PF	-0.15	0.69	0.82
HX	0.29	0.54	0.58
HG	1.19	0.45	< 0.01
SZ	1.12	0.55	0.045
SG	3.25	1.05	< 0.01

Table 3

Empirical mean squared errors for the new method and a procedure based on generalized additve models with parameters α's for estimating $\{\pi_k(\cdot)\}$ at time point t_0 with sample size n and censoring distribution $U(0,\xi)$

$\alpha_1 = \alpha_2 = 1$ $(n = 100)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.125	0.204	0.052	0.128	0.202	0.064
GAM glogit	0.191	0.344	0.089	0.219	0.276	0.106

$\alpha_1 = \alpha_2 = 1$ $(n = 200)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.116	0.181	0.050	0.118	0.195	0.055
GAM glogit	0.222	0.267	0.073	0.192	0.277	0.084

$\alpha_1 = \alpha_2 = 1$ $(n = 400)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.110	0.161	0.048	0.116	0.179	0.046
GAM glogit	0.185	0.250	0.074	0.164	0.258	0.068

$\alpha_1 = 0.5$ $\alpha_2 = 2.5$ $(n = 100)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.130	0.186	0.059	0.119	0.203	0.055
GAM glogit	0.228	0.342	0.093	0.181	0.303	0.102

$\alpha_1 = 0.5$ $\alpha_2 = 2.5$ $(n = 200)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.119	0.174	0.051	0.112	0.201	0.050
GAM glogit	0.183	0.254	0.076	0.179	0.278	0.088

$\alpha_1 = 0.5$ $\alpha_2 = 2.5$ $(n = 400)$

(t_0, ξ)	$(2.7, 15)$	$(5, 15)$	$(1.25, 15)$	$(2.7, 25)$	$(5, 25)$	$(1.25, 25)$
Proposed Method	0.117	0.166	0.046	0.112	0.180	0.044
GAM glogit	0.173	0.242	0.077	0.176	0.242	0.076