

Estimating Subject-Specific Dependent
Competing Risk Profile with Censored Event
Time Observations

Yi Li* Lu Tian†
L. J. Wei‡

*Harvard University and Dana Farber Cancer Institute, yili@jimmy.harvard.edu

†Stanford University School of Medicine, lutian@stanford.edu

‡Harvard University, wei@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper100>

Copyright ©2009 by the authors.

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

Yi Li*, Lu Tian** and L.J. Wei*

*Department of Biostatistics, Harvard University

**Department Health Research & Policy, Stanford University

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.

Keywords: 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 (≥ 1 mg/day) and 241 subjects were assigned to two low dose groups (≤ 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. 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. Note that a similar approach has been taken along this line for predicting the individual-level single event rate without the presence of competing risks by Cai et al. (2008).

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 $\{1, \dots, K\}$. 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,

$$\pi_k(U) = \text{pr}(\tilde{T} \leq t_0, \epsilon = k \mid U), \quad k = 1, \dots, K. \quad (2.1)$$

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 $\{(\tilde{T}_i, C_i, \epsilon_i, U_i), i = 1, \dots, n\}$ 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 $\{(T_i, \Delta_i \epsilon_i, U_i), i = 1, \dots, n\}$. 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 $U_i, i = 1, \dots, n$. Consider a parametric *working* model for the primary event rate:

$$\pi_1(U) = g(\beta'X), \quad (2.2)$$

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

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

to estimate β .

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

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

where $w_i = I(T_i \wedge t_0 \leq C_i) = I(T_i \leq t_0)\Delta_i + I(T_i \geq t_0)$ 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) for handling various problems in survival analysis. 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 approximately 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, the estimate $g(\hat{\beta}'X)$ 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 :

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

where the probability is with respect to the future observation $(U^0, \tilde{T}^0, \epsilon^0)$ as well as the observed data $\{(T_i, \Delta_i \epsilon_i, U_i), i = 1, \dots, n\}$, 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

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

as $n \rightarrow \infty$. 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 $\eta(v) = \{\eta_1(v), \dots, \eta_{K-1}(v)\}'$ via a localized multinomial likelihood function. Specifically, let $Y_{ik} = I(T_i \leq t_0, \epsilon_i = k)$ for $k = 1, \dots, K - 1$, and $\hat{\beta}'X_i = \hat{V}_i$. Then, a kernelized log-likelihood function for $\eta(v)$, expressed with the unknown parameter vector $p = (p_1, \dots, p_{K-1})'$ is

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

where $p_k \geq 0$, $\sum_{k=1}^{K-1} p_k \leq 1$, $K_h(s) = K(s/h)/h$ 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

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

An estimator for $\eta(v)$ 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 $a = (a_1, \dots, a_{K-1})'$ and $b = (b_1, \dots, b_{K-1})'$ are unknown vectors of parameters. Here, the rationale 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 . The resulting log-likelihood function is

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

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

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

In Appendix A, we show that when $h = O(n^{-\nu})$, $1/5 < \nu < 1/2$, $\hat{\eta}_k(v)$ is consistent estimator for $\eta_k(v)$, $k = 1, \dots, K-1$. Moreover, the joint distribution of

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

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

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

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 $\{B_i, i = 1, \dots, n\}$ be a random sample from the unit exponential. Let $a^* = \{a_1^*, \dots, a_{K-1}^*\}'$ 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, where

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

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

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

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

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

It follows from similar arguments given in Cai et al. (2008) that conditional on the data, the limiting distribution of

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

is the same as the unconditional distribution of (3.1). The covariance matrix $\Sigma(v)$ is identical to the limit of $\hat{\Sigma}(v)$, the conditional expected value

of $(nh)[f\{\eta^*(v)\} - f\{\hat{\eta}(v)\}][f\{\eta^*(v)\} - f\{\hat{\eta}(v)\}]'$ (conditional on the data), where $f\{\eta^*(v)\} = (f\{\eta_1^*(v)\}, \dots, f\{\eta_{K-1}^*(v)\})'$ and $f\{\hat{\eta}(v)\} = (f\{\hat{\eta}_1(v)\}, \dots, f\{\hat{\eta}_{K-1}(v)\})'$.

To obtain an approximation to $\hat{\Sigma}(v)$ for a given data set, we generate a large number, M , of independent realizations from $\{B_i, i = 1, \dots, n\}$. 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

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

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 k th diagonal element of $\tilde{\Sigma}(v)$. Note that joint confidence regions for $\{\eta_k(v), k = 1, \dots, K - 1\}$ can also be obtained by considering a sup-type statistic: $\sup_{k=1, \dots, K-1} |\hat{\eta}_k(v) - \eta_k(v)|$ to choose the cutoff point for the confidence intervals (3.2).

Note that since the convergence rate for $\hat{\eta}(v)$ is slower than square root n , in theory, there is no need to consider the sampling variation from $\hat{\beta}$ and $\hat{G}(\cdot)$ for constructing asymptotically valid confidence intervals. However, we find via extensive empirical studies that utilization of such perturbed versions of these statistics can substantially improve the finite sample performance of the nonparametric function estimation.

To construct a $(1 - \alpha)$ simultaneous confidence band for $\eta_k(v)$ over a pre-specified interval \mathcal{I} of v , we cannot 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

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

where c_α is obtained via the following equation:

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

and $\{\eta_k^*(v), v \in \mathcal{I}\}$ is obtained by the above perturbation-resampling method with the same set of $\{B_i, i = 1, \dots, n\}$. The justification of adequacy of this approximation is given in 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 $k = 1, \dots, K - 1$.

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, \dots, D_J . Let the set of observations in D , but not in D_j , be denoted by $D_{(-j)}$, $j = 1, \dots, J$. We construct the scoring system based on observations in $D_{(-j)}$ based on the estimate $\hat{\beta}_{(-j)}$. 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 :

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

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 Hardle 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. EXAMPLE

We illustrate the new proposal with 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.

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 V , which is a bimodal function. Most study subjects are clustered around $V = -4.5$ and -0.9 .

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. Moreover, to make the smooth parameter in the order of slightly larger than $n^{-0.2}$, we multiply the above optimal value by $10^{-0.1}$. This results in $h = 0.97$. Lastly we let the 2nd and 98th percentiles of the empirical distribution of 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.

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 .

5. 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 $\hat{\eta}_k(v), k = 1, \dots, K - 1$, 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.

REFERENCES

- Andersen, P.K. (2002) Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11, 203-215.
- 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. (2008) Calibrating Parametric Subject-specific Risk Estimation, revised for Biometrika.
- 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.
- 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.

Hardle, 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.

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.

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

Appendix A: The Asymptotical Properties of $\hat{\eta}(v)$

Recalling that $\hat{\beta}$ is a solution from the estimating equation

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

it follows from the similar arguments used in Tian et al.(2007) that $\hat{\beta}$ converges to a deterministic limit β_0 and

$$\hat{\beta} - \beta_0 = n^{-1} \xi_i + o_p(n^{-1})$$

where β_0 is the solution of $r(\beta) = 0$,

$$\begin{aligned} \xi_i = & [E\{\dot{g}(\beta_0' X_i) X_i^{\otimes 2}\}]^{-1} \left(X_i \{I(\tilde{T}_i \leq t_0, \epsilon = 1) - g(\beta_0' X_i)\} \right. \\ & \left. - \int_0^{t_0} \mathbb{K}(X_i \{I(\tilde{T}_i \leq t_0, \epsilon = 1) - g(\beta_0' X_i)\}, u) \frac{dM_i^C(u)}{G(T_i \wedge t_0)} \right), \end{aligned}$$

$\mathbb{K}(W, u) = W - E\{WI(\tilde{T} \wedge t_0 \geq u)\}/\text{pr}(\tilde{T} \wedge t_0 \geq u)$ for any random vector W and $M_i^C(u)$ is the martingale process associated with the censoring time C_i . Let $V_i = \beta_0' X_i$ and $\hat{V}_i = \hat{\beta}' X_i$. With slightly abuse of notation, we let $\{\hat{a}(v)', \hat{b}(v)'\}'$ be the maximizer of

$$\sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \left(\sum_{k=1}^{K-1} Y_{ik} \{a_k + b_k(\hat{V}_i - v)\} - \log \left[1 + \sum_{k=1}^{K-1} \exp\{a_k + b_k(\hat{V}_i - v)\} \right] \right),$$

and then it is the solution to the estimating equation

$$\hat{S}(a, b; v) = \{\hat{S}'_1(a, b; v), \dots, \hat{S}'_{K-1}(a, b; v)\}' = 0$$

where

$$\hat{S}'_k(a, b; v) = n^{-1} \sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \begin{pmatrix} 1 \\ \hat{V}_i - v \end{pmatrix} \left\{ Y_{ik} - \frac{\exp\{a_k + b_k(\hat{V}_i - v)\}}{1 + \sum_{k=1}^{K-1} \exp\{a_k + b_k(\hat{V}_i - v)\}} \right\}.$$

To study the asymptotical properties of $\hat{a}(v)$, we let $\hat{\Delta}_a(v) = \{\hat{a}_1(v) - m_1(v), \dots, \hat{a}_{K-1}(v) - m_{K-1}(v)\}'$ and $\hat{\Delta}_b(v) = h\{\hat{b}_1(v) - \dot{m}_1(v), \dots, \hat{b}_{K-1}(v) - \dot{m}_{K-1}(v)\}$, where $m_j(v) = \log\{\eta_j(v)/\eta_K(v)\}$ and $\dot{m}_j(v) = dm_j(v)/dv, j = 1, \dots, K - 1$. Therefore, $\{\hat{\Delta}_a(v)', \hat{\Delta}_b(v)'\}'$ is the solution to the estimating equation

$$\hat{Q}(\Delta_a, \Delta_b; v) = \{\hat{Q}'_1(\Delta_a, \Delta_b; v), \dots, \hat{Q}'_{K-1}(\Delta_a, \Delta_b; v)\}' = 0$$

where $\hat{Q}_k(\Delta_a, \Delta_b; v)$ is

$$n^{-1} \sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \begin{pmatrix} 1 \\ (\hat{V}_i - v)/h \end{pmatrix} \left\{ Y_{ik} - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)}} \right\}$$

$\Delta_a = (\Delta_{a_1}, \dots, \Delta_{a_{K-1}})'$, $\Delta_b = (\Delta_{b_1}, \dots, \Delta_{b_{K-1}})'$ and $\bar{m}_k(u, v) = m_k(v) + \dot{m}_k(v)(u - v)$. Following the similar arguments used in Cai et al. (2008), one may show that the changes in $\hat{Q}_k(\Delta_a, \Delta_b; v)$ by replacing $\hat{G}(\cdot)$ and \hat{V}_i by $G(\cdot)$ and V_i , respectively, are asymptotically negligible. Let $Q_k(\Delta_a, \Delta_b; h, v)$ be

$$n^{-1} \sum_{i=1}^n \frac{w_i K_h(V_i - v)}{G(T_i \wedge t_0)} \begin{pmatrix} 1 \\ (V_i - v)/h \end{pmatrix} \left\{ Y_{ik} - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(V_i - v)/h + \bar{m}_k(V_i, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(V_i - v)/h + \bar{m}_k(V_i, v)}} \right\},$$

and write difference $\hat{Q}_k(\Delta_a, \Delta_b; v) - Q_k(\Delta_a, \Delta_b; v)$ as

$$\begin{aligned} & -n^{-1} \sum_{i=1}^n \{\hat{G}(T_i \wedge t_0) - G(T_i \wedge t_0)\} \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)G(T_i \wedge t_0)} \begin{pmatrix} 1 \\ (\hat{V}_i - v)/h \end{pmatrix} \\ & \left\{ Y_{ik} - \frac{\exp\{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)\}} \right\} + \int_{v-h}^{v+h} K_h(s - v) \\ & d\mathbf{P}_n \left[I(s < \hat{\beta}'X) \begin{pmatrix} 1 \\ (\hat{\beta}'X - v)/h \end{pmatrix} \left\{ Y_k - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(\hat{\beta}'X - v)/h + \bar{m}_k(\hat{\beta}'X, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(\hat{\beta}'X - v)/h + \bar{m}_k(\hat{\beta}'X, v)}} \right\} \right. \\ & \left. - I(s < \beta'_0 X) \begin{pmatrix} 1 \\ (\beta'_0 X - v)/h \end{pmatrix} \left\{ Y_k - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(\beta'_0 X - v)/h + \bar{m}_k(\beta'_0 X, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(\beta'_0 X - v)/h + \bar{m}_k(\beta'_0 X, v)}} \right\} \right] \\ & \times \frac{w}{G(T \wedge t_0)}, \end{aligned}$$

which is bounded by

$$\begin{aligned} & \sup_t |\hat{G}(t) - G(t)| \times n^{-1} \sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0) G(T_i \wedge t_0)} \left(\frac{1}{(\hat{V}_i - v)/h} \right) \left\{ Y_{ik} \right. \\ & \left. - \frac{\exp\{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\Delta_{a_k} + \Delta_{b_k}(\hat{V}_i - v)/h + \bar{m}_k(\hat{V}_i, v)\}} \right\} + O_p(h^{-1}n^{-1/2}) \times \\ & \left(\mathbf{G}_n \left[I(s < \hat{\beta}'X) \left(\frac{1}{(\hat{\beta}'X - v)/h} \right) \left\{ Y_k - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(\hat{\beta}'X - v)/h + \bar{m}_k(\hat{\beta}'X, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(\hat{\beta}'X - v)/h + \bar{m}_k(\hat{\beta}'X, v)}} \right\} \right. \right. \\ & \left. \left. - I(s < \beta'_0 X) \left(\frac{1}{(\beta'_0 X - v)/h} \right) \left\{ Y_k - \frac{e^{\Delta_{a_k} + \Delta_{b_k}(\beta'_0 X - v)/h + \bar{m}_k(\beta'_0 X, v)}}{1 + \sum_{k=1}^{K-1} e^{\Delta_{a_k} + \Delta_{b_k}(\beta'_0 X - v)/h + \bar{m}_k(\beta'_0 X, v)}} \right\} \right] \right) \times \\ & \left. \frac{w}{G(T \wedge t_0)} + O_p(h^2 + n^{-1/2}) \right) = O_p\{n^{-1/2} + (nh^2)^{-3/4 + \delta_0} + (nh)^{-1}\}, \end{aligned}$$

for some small $\delta_0 > 0$, where \mathbf{P}_n and \mathbf{P} are the expectation operator with respect to the empirical distribution of $\{(T_i, \Delta_i, \epsilon_i, U_i), i = 1, \dots, n\}$ and the distribution of (T, Δ, ϵ, U) , respectively, and $\mathbf{G}_n = n^{1/2}(\mathbf{P}_n - \mathbf{P})$. Furthermore, since $Q(\Delta_a, \Delta_b; v) = \{Q_1(\Delta_a, \Delta_b; v), \dots, Q_{K-1}(\Delta_a, \Delta_b; v)\}'$ can be written as sum of n identically distributed independent random functions, it follows from the standard arguments that it uniformly converges to $q(\Delta_a, \Delta_b; v) = \{q_1(\Delta_a, \Delta_b; v), \dots, q_{K-1}(\Delta_a, \Delta_b; v)\}'$, where

$$q_k(\Delta_a, \Delta_b; v) = \begin{pmatrix} g_0(v) \int K(x) \left[\eta_k(v) - \frac{\exp\{\Delta_{a_k} + \Delta_{b_k}x + m_k(v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\Delta_{a_k} + \Delta_{b_k}x + m_k(v)\}} \right] dx \\ g_0(v) \int xK(x) \left[\eta_k(v) - \frac{\exp\{\Delta_{a_k} + \Delta_{b_k}x + m_k(v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\Delta_{a_k} + \Delta_{b_k}x + m_k(v)\}} \right] dx \end{pmatrix},$$

and $g_0(\cdot)$ is the density function of the random variable $\beta'_0 X$. Since $(\Delta'_a, \Delta'_b)' = (0', 0)'$ is the unique solution of $q(\Delta_a, \Delta_b; v) = 0$. $\hat{\Delta}_a(v)$ and $\hat{\Delta}_b(v)$ converge to zero uniformly in v , assuming that the ‘‘slope’’ matrix of $q(\Delta_a, \Delta_b; v)$ is nonsingular. Coupled with the consistency of $\hat{\Delta}_a$ and $\hat{\Delta}_b$, the Taylor series expansion can be used to show that

$$\hat{\Delta}_a(v) = \mathbf{A}(v)^{-1} n^{-1} \sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \begin{pmatrix} Y_{i1} - \frac{\exp\{\bar{m}_1(\hat{V}_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(\hat{V}_i, v)\}} \\ \dots \\ Y_{i(K-1)} - \frac{\exp\{\bar{m}_{K-1}(\hat{V}_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(\hat{V}_i, v)\}} \end{pmatrix} + O_p\{(nh)^{-1/2}\},$$

where

$$\mathbf{A}(u) = g_0(v) \begin{pmatrix} \eta_1(u)\{1 - \eta_1(u)\} & -\eta_1(u)\eta_2(u) & \cdots & -\eta_1(u)\eta_{K-1}(u) \\ -\eta_2(u)\eta_1(u) & \eta_2(u)\{1 - \eta_2(u)\} & \cdots & -\eta_2(u)\eta_{K-1}(u) \\ \cdots & \cdots & \cdots & \cdots \\ -\eta_{K-1}(u)\eta_1(u) & -\eta_{K-1}(u)\eta_2(u) & \cdots & \eta_{K-1}(u)\{1 - \eta_{K-1}(u)\} \end{pmatrix}.$$

Again, following the similar arguments in Cai et al. (2008), one may show that

$$n^{-1} \sum_{i=1}^n \frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \left\{ Y_{ik} - \frac{\exp\{\bar{m}_k(\hat{V}_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(\hat{V}_i, v)\}} \right\}$$

can be approximated by

$$n^{-1} \sum_{i=1}^n \frac{w_i K_h(V_i - v)}{G(T_i \wedge t_0)} \left\{ Y_{ik} - \frac{\exp\{\bar{m}_k(V_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(V_i, v)\}} \right\}$$

uniformly in v up to an order of $o_p\{(nh)^{-1/2}\}$ for $h = n^{-\nu}$, $\nu \in (1/5, 1/2)$.

Noting that the consistent estimator for $\eta_k(v)$ is

$$\hat{\eta}_k(v) = \frac{\exp\{\hat{a}_k(v)\}}{1 + \sum_{j=1}^{K-1} \exp\{\hat{a}_j(v)\}},$$

by δ -method we have

$$\begin{pmatrix} f\{\hat{\eta}_1(v)\} - f\{\eta_1(v)\} \\ f\{\hat{\eta}_2(v)\} - f\{\eta_2(v)\} \\ \cdots \\ f\{\hat{\eta}_{K-1}(v)\} - f\{\eta_{K-1}(v)\} \end{pmatrix} \\ = \mathbf{D}(v) \mathbf{A}(u) \hat{\Delta}_a(v) / g_0(v) + o_p\{(nh)^{-1/2}\} \\ = \mathbf{D}(v) n^{-1} \sum_{i=1}^n \frac{w_i K_h(V_i - v)}{g_0(v) G(T_i \wedge t_0)} \begin{pmatrix} Y_{i1} - \frac{\exp\{\bar{m}_1(V_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(V_i, v)\}} \\ \cdots \\ Y_{i(K-1)} - \frac{\exp\{\bar{m}_{K-1}(V_i, v)\}}{1 + \sum_{k=1}^{K-1} \exp\{\bar{m}_k(V_i, v)\}} \end{pmatrix} + o_p\{(nh)^{-1/2}\},$$

where $\mathbf{D}(v) = \text{diag}[f\{\eta_1(v)\}, \dots, f\{\eta_{K-1}(v)\}]$ and $\dot{f}(\cdot)$ is the derivative of $f(\cdot)$. Therefore by the central limit theorem

$$(nh)^{1/2} [f\{\hat{\eta}(v)\} - f\{\eta(v)\}] \rightarrow N\{0, \Sigma(v)\},$$

in distribution as $n \rightarrow \infty$.

Appendix B: The Justification of the Resampling Methods

To justify the resampling-based variance estimator, note that the variance estimator $\tilde{\sigma}_k^2(v)$ can be approximated by

$$\begin{aligned} & \frac{f^2\{\hat{\eta}_k(v)\}}{\hat{g}_0^2(v)} n^{-1} \sum_{i=1}^n E_{B_i} \left[\frac{w_i K_h(V_i^* - v)}{G^*(T_i \wedge t_0)} \left\{ Y_{ik} - \frac{e^{\hat{a}_k(v) + \hat{b}_k(V_i^* - v)}}{1 + \sum_{k=1}^{K-1} e^{\hat{a}_k(v) + \hat{b}_k(V_i^* - v)}} \right\} B_i \right]^2 \\ &= \frac{f^2\{\hat{\eta}_k(v)\}}{\hat{g}_0^2(v)} n^{-1} \sum_{i=1}^n \left[\frac{w_i K_h(\hat{V}_i - v)}{\hat{G}(T_i \wedge t_0)} \left\{ Y_{ik} - \frac{e^{\hat{a}_k(v) + \hat{b}_k(\hat{V}_i - v)}}{1 + \sum_{k=1}^{K-1} e^{\hat{a}_k(v) + \hat{b}_k(\hat{V}_i - v)}} \right\} \right]^2 + o_p(1), \end{aligned}$$

which uniformly converges to σ_k^2 , the asymptotical variance of $n^{1/2}[f\{\hat{\eta}_k(v)\} - f\{\eta_k(v)\}]$, in probability as $n \rightarrow \infty$, where E_{B_i} is the expectation with respect to the random weights $\{B_1, \dots, B_n\}$, which are independent of the observed data. The first approximation follows from the fact that $|\hat{\beta}^* - \hat{\beta}| + \sup_t |G^*(t) - \hat{G}(t)|$ is in the order of $O_p(n^{-1/2})$ and similar arguments used to bound the difference between $\hat{Q}_k(\Delta_a, \Delta_b; v)$ and $Q_k(\Delta_a, \Delta_b; v)$.

To justify the proposed procedure for constructing the simultaneous confidence band of $\eta_k(v)$, $v \in \mathcal{I}$, first note that we have already established that uniformly in v ,

$$\begin{aligned} & (nh)^{1/2} [f\{\hat{\eta}_k(v)\} - f\{\eta_k(v)\}] \\ &= (nh)^{1/2} \frac{f\{\eta_k(v)\}}{g_0(v)} n^{-1} \sum_{i=1}^n \frac{w_i K_h(V_i - v)}{G(T_i \wedge t_0)} \left\{ Y_{i1} - \frac{e^{\bar{m}_k(V_i, v)}}{1 + \sum_{k=1}^{K-1} e^{\bar{m}_k(V_i, v)}} \right\} + o_p(n^{-\delta_0}) \\ &= (nh)^{-1/2} \sum_{i=1}^n K \left(\frac{V_i - v}{h} \right) \xi_{ki} + o_p(n^{-\delta_0}), \end{aligned}$$

for some $\delta_0 > 0$, where

$$\xi_{ki} = \frac{\dot{f}\{\eta_k(V_i)\} w_i}{g_0(V_i) G(T_i \wedge t_0)} \{Y_{ik} - \eta_k(V_i)\}.$$

Let

$$S = \sup_{\mathcal{I}} (nh)^{1/2} |f\{\hat{\eta}_k(v)\} - f\{\eta_k(v)\}| / \tilde{\sigma}_k(v)$$

and immediately we have

$$\mathcal{S} = \sup_{\mathcal{I}} \left| \frac{1}{(nh)^{1/2}\sigma_k(v)} \sum_{i=1}^n K\left(\frac{V_i - v}{h}\right) \xi_{ki} \right| + o_p(n^{-\delta_0}).$$

From the strong approximation theorem, one may construct a sequence of standard bivariate Wiener processes $W_n(x, y)$ such that

$$\mathcal{S} = \sup_{\mathcal{I}} \left| \frac{1}{(nh)^{1/2}\sigma_k(v)} \int K\left(\frac{x - v}{h}\right) y dW_n\{M(x, y)\} \right| + o_p(n^{-\delta_0})$$

in an appropriate probability space, where $M(x, y)$ is the Rosenblatt transformation such that $M(V_i, \xi_i)$ is uniformly distributed on the unit square. Furthermore, it follows from the similar arguments in Bickel and Rosenblatt (1973) that

$$\text{pr}\{-2 \log(h)\}^{1/2}(\mathcal{S} - d_h) \leq s = \exp(-2e^{-x}) + o(1),$$

as $n \rightarrow \infty$, where

$$d_h = \{-2 \log(h)\}^{1/2} + \frac{1}{\{-2 \log(h)\}^{1/2}} \log \left\{ \frac{\int \dot{K}(t)^2 dt}{4\pi \int K(t)^2 dt} \right\}.$$

Unlike the supremum value of tight processes, \mathcal{S} itself does not converge in distribution, since $d_h \rightarrow \infty$ as $n \rightarrow \infty$. In parallel arguments \mathcal{S}^* , the resampling counterpart of \mathcal{S} , is equivalent to

$$\begin{aligned} & \sup_{\mathcal{I}} \left| \frac{1}{(nh)^{1/2}\sigma_k(v)} \sum_{i=1}^n K\left(\frac{V_i - v}{h}\right) \hat{\xi}_{ki} B_i \right| + o_p(n^{-\delta_0}) \\ &= \sup_{\mathcal{I}} \left| \frac{1}{(nh)^{1/2}\sigma_k(v)} \int K\left(\frac{x - v}{h}\right) y dW_n^*\{M^*(x, y)\} \right| + o_p(n^{-\delta_0}) \end{aligned}$$

for some $\delta_0 > 0$ where

$$\hat{\xi}_{ki} = \frac{f\{\hat{\eta}_k(\hat{V}_i)\} w_i}{\hat{g}_0(\hat{V}_i) \hat{G}(T_i \wedge t_0)} \{Y_{ik} - \hat{\eta}_k(\hat{V}_i)\}$$

and $W_n^*\{M^*(x, y)\}$ is a sequence of mean zero Gaussian processes, whose covariance function is identical to that of $W_n\{M(x, y)\}$ conditional on the observed data. Let $\mathcal{T}^* = \{-2\log(h)\}^{1/2}(\mathcal{S}^* - d_h)$ and $\mathcal{T} = \{-2\log(h)\}^{1/2}(\mathcal{S} - d_h)$. It follows that

$$|\text{pr}_B(\mathcal{T}^* \leq s) - \text{pr}(\mathcal{T} \leq s)| = o_p(n^{-\delta_0}),$$

which implies that we can use the conditional distribution of \mathcal{S}^* to approximate that of \mathcal{S} , where pr_B is conditional on the observed data.



Table 1: Coding of the covariates for the prostate cancer data

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



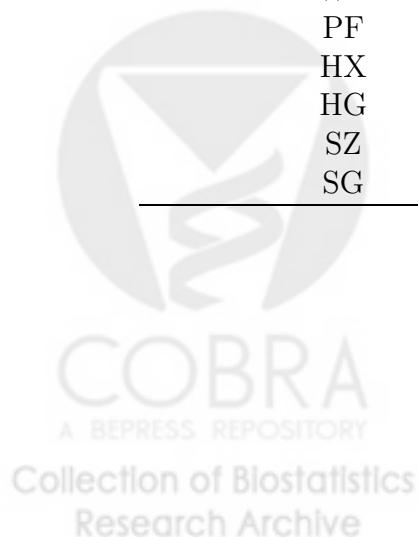
Table 2: Regression Coefficient Estimates for Model (2.2) with the data from the high dose groups

(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



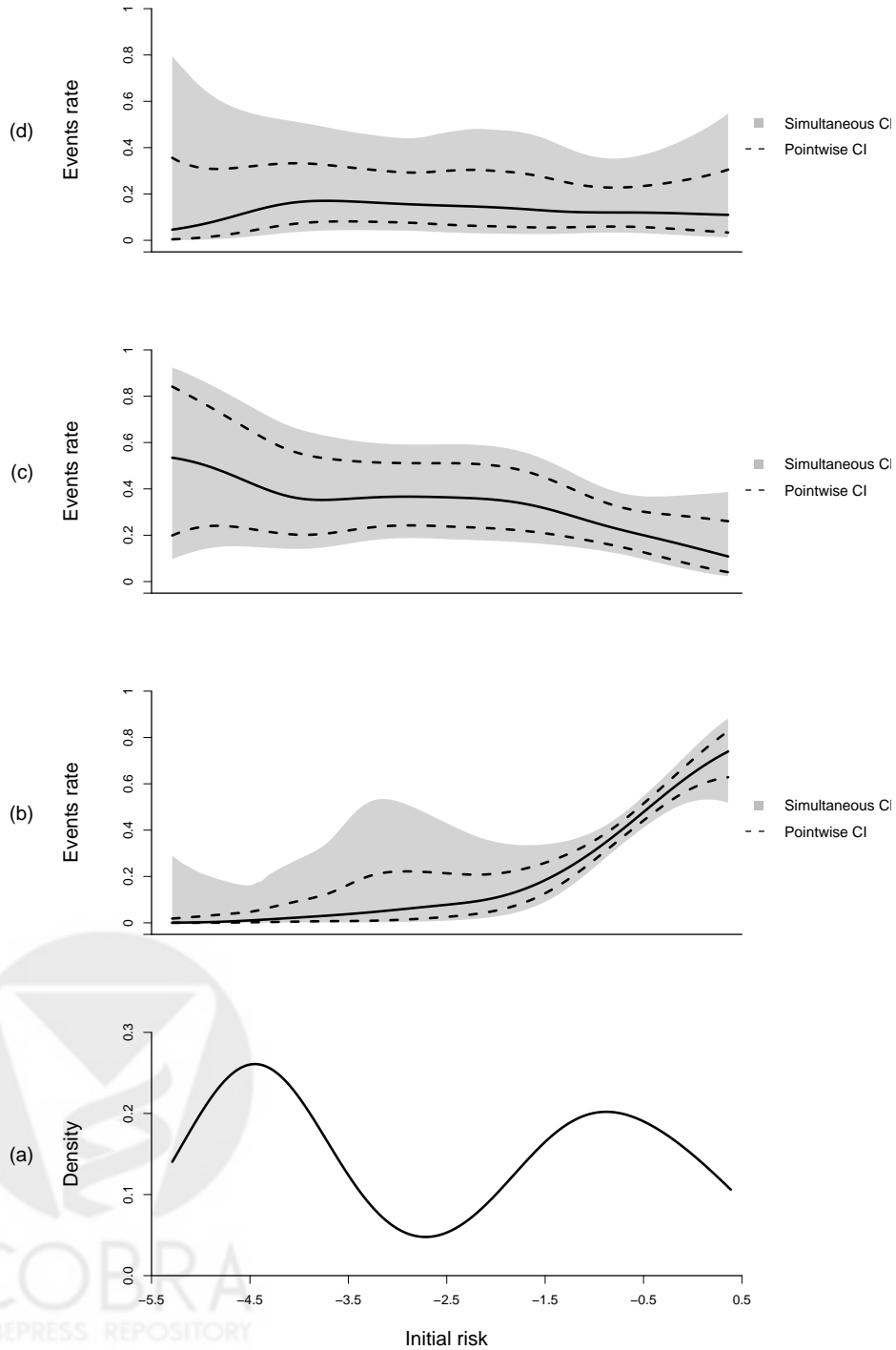


Figure 1: Consistent estimates (solid curve), pointwise 0.95 confidence intervals (enclosed by dotted curves) and simultaneous intervals (gray area) for various risks $\eta_k(v)$ at $t_0 = 5$ years: (a) The density function for the index score; (b) Inference for $\eta_1(v)$; (c) inference for $\eta_2(v)$;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

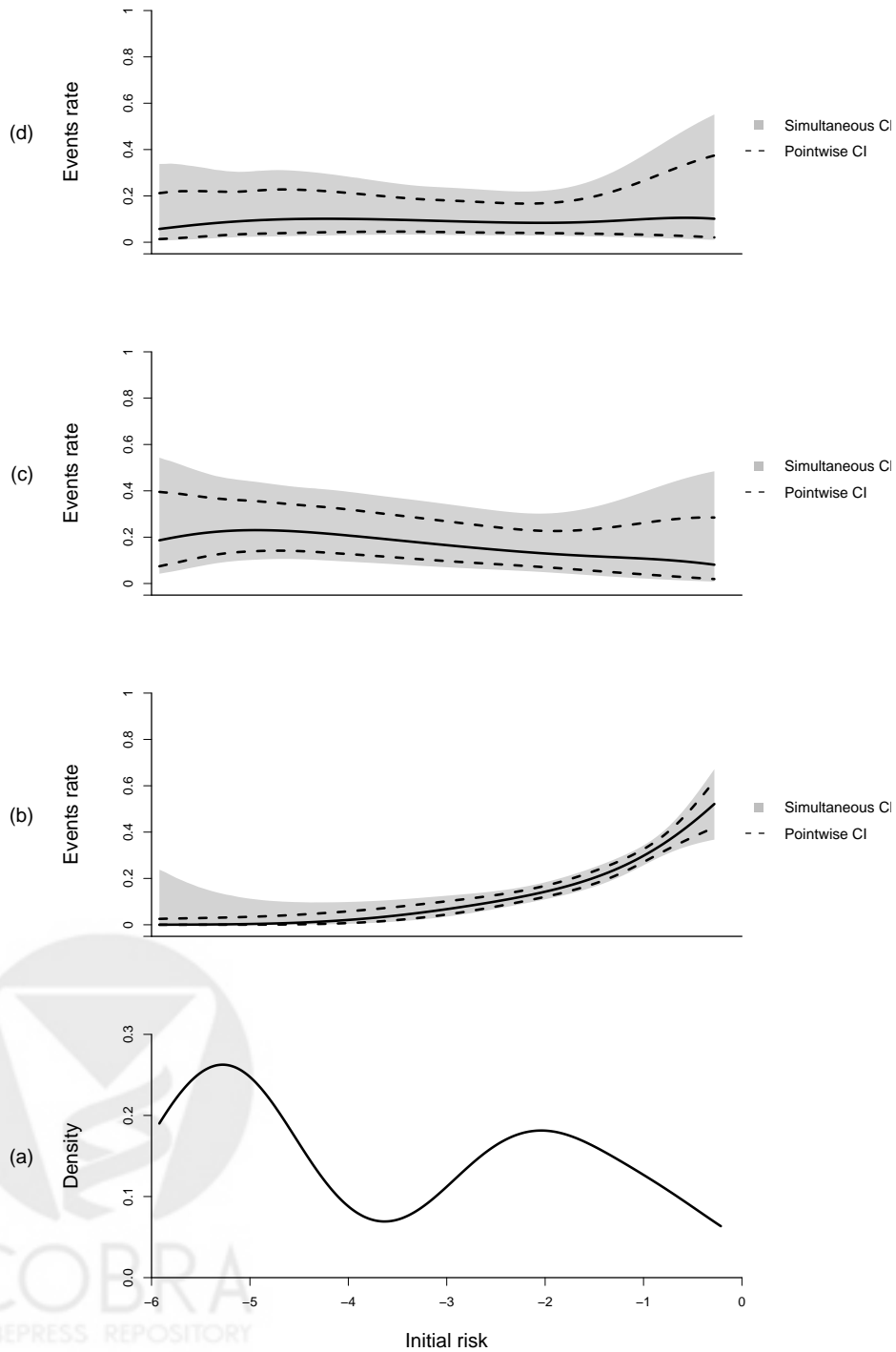


Figure 2: Consistent estimates (solid curve), pointwise 0.95 confidence intervals (enclosed by dotted curves) and simultaneous intervals (gray area) for various risks $\eta_k(v)$ at $t_0 = 2$ years: (a) The density function for the index score; (b) Inference for $\eta_1(v)$; (c) inference for $\eta_2(v)$;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