An improved fully nonparametric estimator of the marginal survival function based on case-control clustered data

David M. Zucker

Department of Statistics and Data Science,

The Hebrew University of Jerusalem

Mount Scopus, Jerusalem, Israel

email: david.zucker@mail.huji.ac.il

and

Malka Gorfine

Department of Statistics and Operations Research

Tel Aviv University, Ramat Gan, Israel

email: gorfinem@post.tau.ac.il

December 3, 2018

Abstract

A case-control family study is a study where individuals with a disease of interest (case probands) and individuals without the disease (control probands) are randomly sampled from a well-defined population. Possibly right-censored age at onset and disease status are observed for both probands and their relatives. Correlation among the outcomes within a family is induced by factors such as inherited genetic susceptibility, shared environment, and common behavior patterns. For this setting, we present a nonparametric estimator of the marginal survival function, based on local linear estimation of conditional survival functions. Asymptotic theory for the estimator is provided, and simulation results are presented showing that the method performs well. The method is illustrated on data from a prostate cancer study.

Keywords: case-control; family study; multivariate survival; nonparametric estimator; local linear

1 Introduction

Many epidemiological and medical studies focus on disease events that are rare, so that a random sample from the population provides very few observations where failure has occurred during the monitoring time. In such situations, it is a common practice to consider a case-control strategy. Separate samples of n_1 individuals in whom the event has already occurred (case probands) and n_0 individuals in whom the event has not yet occurred (control probands) are obtained. Age at onset or age at censoring and disease status of each proband and of one or more of his/her relatives are recorded. For example, in a prostate cancer study, case probands are men diagnosed with prostate cancer, control probands are men without prostate cancer, and each proband is interviewed to obtain detailed disease history information of his relatives. The goals of such case-control family studies are to evaluate the effect of genetic and environmental factors on disease risk, and to estimate the distribution of the age at onset. The analysis of such data is complicated by the case-control selection scheme used to ascertain the families and by the within-family dependence. It is of interest to estimate the marginal survival function under dependent failure times of family members by a fully nonparametric estimator which avoids specific assumptions about the form of the distribution or the dependence structure among failure times with a family.

The estimator that first comes to mind is the Kaplan-Meier survival curve estimator based on the survival data on the relatives. This estimator, however, is biased because it does not take the case-control sampling and the within-family dependence into account. Gorfine et al. (2017) demonstrated the serious bias that can arise with this naive Kaplan-Meier estimator.

Accordingly, Gorfine et al. proposed a nonparametric estimator for this problem, based on a kernel smoothing approach. They presented a simulation study showing that their estimator performs well in terms of bias. The estimator of Gorfine et al. is based on the median of random variables with a complicated dependence structure. Consequently, the asymptotic properties of the estimator were not derived. Also, their use of the median causes some efficiency loss. In addition, their bandwidth selection procedure was not specifically targeted to the estimand of

interest.

In the present paper we develop a new nonparametric estimator which performs well in terms of bias and much better in terms of variance than the estimator of Gorfine et al. In some scenarios the new estimator also outperforms Gorfine et al. in terms of bias. We provide asymptotic theory for the estimator. We also present a bandwidth selection procedure that is specifically targeted to the estimand of interest.

2 Preliminaries

The setup is as in Gorfine et al. (2017). We denote the maximum observation time among the probands by τ_0 and the maximum observation time among the relatives by τ. We write $n = n_0 + n_1$. We let J_i denote the number of relatives in family i and we let J denote the maximum number of relatives for a given proband. We view J_i as a random variable. We will write some of the formulas as if each proband has exactly J relatives, with the extra relatives taken to be censored at time 0. For family i, $i = 1, \ldots, n$, let T_{Pi} denote the failure time of the proband and $T_{Rij}, j = 1, \ldots, J_i$ the failure times of the relatives. Let C_{Pi} and C_{Rij} denote the corresponding censoring times, $X_{Pi} = \min(T_{Pi}, C_{Pi})$ and $X_{Rij} = \min(T_{Rij}, C_{Pij})$ the corresponding observed times, and $\delta_{Pi} = I(T_{Pi} \leq C_{Pi})$ and $\delta_{Rij} = I(T_{Rij} \leq C_{Rij})$ the corresponding event indicators. We assume that the censoring is independent of the survival times. The sample includes n_1 case probands, with each case proband frequency matched with a control probands, so that the total number of control probands is $n_0 = an_1$. Thus, the data consist of n_1 independent and identically distributed matched sets comprising one case family and a control families, and the observed data on family i consists of $(X_{Pi}, \delta_{Pi}, X_{Ri1}, \ldots, X_{RiJ_i}, \delta_{Ri1}, \ldots, \delta_{RiJ_i})$. We seek to estimate the marginal survival function

$$S(t) = \text{pr}(T_P > t) = \text{pr}(T_R > t)$$

which we are assuming is the same for probands and relatives, a common assumption in case-control family studies (Shih & Chatterjee, 2002; Chatterjee et al., 2006). We further assume

that the bivariate survival function for the proband and a given relative is the same for all relatives. The marginal survival distribution is assumed to be continuous with density $f(t)$. In the case where T_P and T_R have different marginal distributions, our procedure yields a consistent estimate of the marginal distribution of T_P.

Denote $S_0(u|t) = \text{pr}(T_R > u|T_P > t)$ and $S_1(u|t) = \text{pr}(T_R > u|T_P = t)$. As in Gorfine et al., we have $\text{pr}(T_R > u|X_P = t, \delta_P = 0) = S_0(u|t)$ and $\text{pr}(T_R > u|X_P = t, \delta_P = 0) = S_1(u|t)$.

We now develop an expression for $S(t)$ in terms of $S_0(u|t)$ and $S_1(u|t)$. Let $\lambda(t)$ and $\Lambda(t)$ denote the hazard and cumulative hazard functions corresponding to $S(t)$, and define $S_q^*(u|t) = (\partial/\partial t)S_q(u|t)$, $q = 0, 1$. Also define $\Lambda_q(u|t) = -\log S_q(u|t)$ and $\Lambda_q^*(u|t) = (\partial/\partial t)\Lambda_q(u|t)$, $q = 0, 1$. We can write $S_0^*(u|t) = -S_0(u|t)\Lambda_0^*(u|t)$. We then have the following:

$$
\begin{aligned}
&\text{pr}(T_P > t, T_R > u) = \int_t^\infty \text{pr}(T_R > u|T_P = x)f(x)dx \\
&\Rightarrow S(t)S_0(u|t) = \int_t^\infty S_1(u|x)f(x)dx \\
&\Rightarrow \frac{\partial}{\partial t}[S(t)S_0(u|t)] = -S_1(u|t)f(t) \\
&\Rightarrow -f(t)S_0(u|t) + S(t)S_0^*(u|t) = -S_1(u|t)f(t) \\
&\Rightarrow -\lambda(t)S_0(u|t) + S_0^*(u|t) = -S_1(u|t)\lambda(t) \\
&\Rightarrow \lambda(t)(S_0(u|t) - S_1(u|t)) = S_0^*(u|t) = -S_0(u|t)\Lambda_0^*(u|t) \\
&\Rightarrow \lambda(t)(S_0(u|t) - S_1(u|t))^2 = -S_0(u|t)(S_0(u|t) - S_1(u|t))\Lambda_0^*(u|t) \qquad (1)
\end{aligned}
$$

Now define

$$\psi(u,t) = \left[\int_0^\tau (S_0(v|t) - S_1(v|t))^2\, dv\right]^{-1} (S_0(u|t) - S_1(u|t))S_0(u|t) \qquad (2)$$

Note that the bracketed integral is nonzero provided that for every t there exists a set of u values of positive measure for which $S_0(u|t) \neq S_1(u|t)$. Integrating both sides of (1) and rearranging gives

$$\lambda(t) = -\int_0^\tau \psi(u,t)\Lambda_0^*(u|t)\, du \qquad (3)$$

$$\Lambda(t) = -\int_0^t \int_0^\tau \psi(u,s)\Lambda_0^*(u|s)\, du\, ds \qquad (4)$$

We use (4) to construct our estimator.

3 Estimation Procedure

In Gorfine et al., $S_0(u|t)$ and $S_1(u|t)$ were estimated using a generalized version of the kernel-smoothed Kaplan-Meier estimator proposed in an unpublished 1981 University of California at Berkeley technical report by R. Beran and examined in Dabrowska (1987), and the resulting estimators were used to construct an estimator of $S(t)$. Here, in light of (4), we work not only with $S_0(u|t)$ and $S_1(u|t)$ but also the derivative $\Lambda_0^*(u|t)$. Accordingly, we take a local linear estimation approach. Choose a symmetric kernel function K and a bandwidth h. Let $N_{Rij}(t) = \delta_{Rij} I(T_{Rij} \leq t)$ and $Y_{Rij}(t) = I(T_{Rij} \geq t)$, and write $Q_i = \delta_{Pi}$. Let $\chi(q_1, q_2) = I(q_1 = q_2)$. Define (with $q = 0, 1$)

$$\lambda_q(u|t) = \frac{\partial}{\partial u}\Lambda_q(u|t), \quad N_{Ri\bullet}(v) = \sum_{j=1}^{J} N_{Rij}(v), \quad Y_{Ri\bullet}(v) = \sum_{j=1}^{J} Y_{Rij}(v)$$

$$dM_{Rij}(v) = dN_{Rij}(v) - Y_{Rij}(v)\lambda_{Q_i}(v|X_{Pi})dv, \quad M_{Ri\bullet}(v) = \sum_{j=1}^{J} M_{Rij}(v)$$

$$\mathcal{Y}_q(s,v) = \frac{1}{n_q h}\sum_{i=1}^{n} \chi(Q_i, q) Y_{Ri\bullet}(v) K\left(\frac{X_{Pi} - s}{h}\right)$$

$$\bar{X}_{Pq}(s,v) = \mathcal{Y}_q(s,v)^{-1}\left[\frac{1}{n_q h}\sum_{i=1}^{n} \chi(Q_i, q) Y_{Ri\bullet}(v) K\left(\frac{X_{Pi} - s}{h}\right) X_{Pi}\right]$$

$$C_q(s,v) = \frac{1}{n_q h}\sum_{i=1}^{n} \chi(Q_i, q) Y_{Ri\bullet}(v) K\left(\frac{X_{Pi} - s}{h}\right)\left(X_{Pi} - \bar{X}_{Pq}(s,v)\right)^2$$

We can write

$$\frac{dN_{Ri\bullet}(v)}{Y_{Ri\bullet}(v)} = \Lambda_{Q_i}(dv|X_{Pi}) + \frac{dM_{Ri\bullet}(v)}{Y_{Ri\bullet}(v)}$$

with $E[dM_{Ri\bullet}(v)/Y_{Ri\bullet}(v)] = 0$. A first-order Taylor approximation gives the local linear representation

$$\frac{dN_{Ri\bullet}(v)}{Y_{Ri\bullet}(v)} \approx \Lambda_{Q_i}(dv|s) + \Lambda_{Q_i}^*(dv|s)(X_{Pi} - s) + \frac{dM_{Ri\bullet}(v)}{Y_{Ri\bullet}(v)}$$

We now carry out weighted linear least squares with response variable $dN_{Ri\bullet}(v)/Y_{Ri\bullet}(v)$, explanatory variable $X_{Pi} - s$, and weights $\chi(Q_i, q)Y_{Ri\bullet}(v)K((X_{Pi} - s)/h)$. This leads to the local

linear estimators

$$\widehat{\Lambda}_q^*(dv|s) = C_q(s,v)^{-1}\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\right]$$
$$\widehat{\Lambda}_q(dv|s) = \frac{(n_q h)^{-1}\sum_{i=1}^n \chi(Q_i,q)K((X_{Pi}-s)/h)dN_{Ri\bullet}(v)}{\mathcal{Y}_q(s,v)} - \widehat{\Lambda}_q^*(dv|s)(\bar{X}_{Pq}(s,v)-s)$$

that is, for a given u,

$$\widehat{\Lambda}_q^*(u|s) = \int_0^u \frac{1}{C_q(s,v)}\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\right] \quad (5)$$
$$\widehat{\Lambda}_q(u|s) = \int_0^u \frac{(n_q h)^{-1}\sum_i \chi(Q_i,q)K((X_{Pi}-t)/h)dN_{Ri\bullet}(v)}{\mathcal{Y}_q(s,v)}$$
$$-\int_0^u \frac{\bar{X}_{Pq}(s,v)-s}{C_q(s,v)}\left[\frac{1}{n_q h}\sum_i \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\right] \quad (6)$$

with the second equation leading to $\widehat{S}_q(u|s) = \exp(-\widehat{\Lambda}_q(u|s))$. We can now substitute $\widehat{S}_0(u|s)$, $\widehat{S}_0(u|s)$, and $\widehat{\Lambda}_0^*(u|s)$ into (3) and (4) to obtain estimators $\widehat{\lambda}(t)$ and $\widehat{\Lambda}(t)$ for $\lambda(t)$ and $\Lambda(t)$. We then take $\widehat{S}(t) = \exp(-\widehat{\Lambda}(t))$. When we want to emphasize the dependence on the bandwidth h, we will write $\widehat{S}(t;h)$.

4 Asymptotic Theory

We can write $\widehat{\Lambda}(t) - \Lambda(t) = \mathcal{A}(t) + \mathcal{A}'(t) + \mathcal{A}''(t)$ with

$$\mathcal{A}(t) = \int_0^t\int_0^\tau \psi(u,s)[\widehat{\Lambda}_0^*(u|s) - \Lambda_0^*(u|s)]\,du\,ds$$
$$\mathcal{A}'(t) = \int_0^t\int_0^\tau [\widehat{\psi}(u,s) - \psi(u,s))]\Lambda_0^*(u|s)\,du\,ds$$
$$\mathcal{A}''(t) = \int_0^t\int_0^\tau [\widehat{\psi}(u,s) - \psi(u,s)][\widehat{\Lambda}_0^*(u|s) - \Lambda_0^*(u|s)]\,du\,ds$$

We will provide a detailed asymptotic analysis of $\mathcal{A}(t)$. Similar arguments can be used to show that $\mathcal{A}'(t)$ converges in probability to zero at faster rate than $\mathcal{A}(t)$, and $\mathcal{A}''(t)$ is negligible in comparison with the other two terms.

In this section and in the appendix, we will write $m = n_0$ and assume that the indices have been arranged so that the control probands appear first in the list of probands, meaning that a sum over probands 1 to m is a sum over the control probands.

Defining

$$\Gamma(s,v) = h^{-2}C_0(s,v) = \frac{1}{mh}\sum_{i=1}^{m} Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)\left(\frac{X_{Pi}-\bar{X}_{P0}(s,v)}{h}\right)^2$$

we have

$$\mathcal{A}(t) = -\frac{1}{mh^3}\sum_{i=1}^{m}\int_0^t\int_0^\tau\int_0^u \psi(u,s)\Gamma(s,v)^{-1}K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))$$
$$(dN_{Ri\bullet}(v) - Y_{Ri\bullet}(v)(X_{Pi}-\bar{X}_{P0}(s,v))\lambda_0^*(v|s)dv)\,du\,ds$$

Now, the process M_{Rij} is a martingale with respect to the filtration $\mathcal{F}_{ijv} = \sigma(X_{Pi}, \{N_{Rij}(d), Y_{Rij}(d), d \in [0,v]\})$. Accordingly, for any process $\mathcal{P}(v)$ that is predictable with respect to $\mathcal{F}_{ijv}$, the process

$$\int_0^u \mathcal{P}(v)dM_{Rij}(v)$$

is a mean-zero martingale. It follows, even though the process $M_{Ri\bullet}$ does not have any martingale properties, that for any function $P(v,x)$ we have

$$E\left[\int_0^u P(v,X_{Pi})dM_{Ri\bullet}(v)\right] = 0 \tag{7}$$

We now write $\mathcal{A}(t) = -(\mathcal{A}_1(t) + \mathcal{A}_2(t))$, where

$$\mathcal{A}_1(t) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^t\int_0^\tau\int_0^u \psi(u,s)\Gamma(s,v)^{-1}Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))$$
$$(\lambda_0(v|X_{Pi}) - (X_{Pi}-\bar{X}_{P0}(s,v))\lambda_0^*(v|s))dv\ du\,ds$$
$$\mathcal{A}_2(t) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^t\int_0^\tau\int_0^u \psi(u,s)\Gamma(s,v)^{-1}K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))dM_{Ri\bullet}(v)\,du\,ds$$

In the appendix it is shown, under the conditions listed in the appended supplementary document that $(mh)^{1/2}\mathcal{A}_1(t) \to 0$ in probability and that

$$\mathcal{A}_2(t) = \frac{1}{m}\sum_{i=1}^{m}\zeta_i(t) + o_p((mh)^{-1/2})$$

with

$$\zeta_i(t) = \frac{1}{h}\int_0^t \int_0^\tau \int_0^u \psi(u,s)\gamma(s,v)^{-1}\left(\frac{X_{Pi}-s}{h}\right)\left[h^{-1}K\left(\frac{X_{Pi}-s}{h}\right)\right] dM_{Ri\bullet}(v)\, du\, ds$$

where γ is the limiting value of Γ. Define $\sigma_\zeta^2(t) = h\,\mathrm{var}(\zeta_i(t))$. We then obtain the following theorem.

Theorem 1. *For each $t \in [0, \tau_0]$, $\sigma_\zeta^2(t)$ converges to a limit $\kappa(t)$ and $(mh)^{1/2}\left(\widehat{\Lambda}(t) - \Lambda(t)\right)$ converges in distribution to $N(0, \kappa(t))$.*

The proof of this theorem appears in the appendix. Details are given in the appended supplementary document.

5 Practical Implementation Details

In preliminary work, we found that the performance of the estimator of $\widehat{S}(t)$ can be improved dramatically by introducing a time transformation that makes the proband observation times approximately uniformly distributed. Along the lines of Doksum et al. (2017), we propose transforming according to an estimate of the distribution function of the proband observation times, which leads to a modified form of nearest neighbor regression. In the appended supplementary document, we show that the consistency and asymptotic normality is maintained under the time transformation if a smooth estimate of the distribution function is used. We believe that this result holds as well when the empirical distribution function is used. In our numerical work, we used the empirical distribution function.

In the context of family survival data, it is usually reasonable to assume that $\mathrm{pr}(T_P > t, T_R > u) \geq \mathrm{pr}(T_P > t)\mathrm{pr}(T_R > u) = S(t)S(u)$ for all t and u, i.e., $S_0(u|t) \geq S(u)$ for all t and u. This condition implies that

$$\mathrm{pr}(T_R > u | T_P \leq C_P) \leq S(u) \leq \mathrm{pr}(T_R > u | T_P > C_P) \tag{8}$$

If we let $\widehat{S}_{KM,case}(t)$ and $\widehat{S}_{KM,control}(t)$ denote the Kaplan-Meier survival curve estimator based on the case relatives' survival data and the control relatives' survival data, respectively, the foregoing inequalities motivate modifying the estimator to the estimator $\tilde{S}(t)$ resulting from replacing $\widehat{S}(t)$ with $\widehat{S}_{KM,case}(t)$ if $\widehat{S}(t) \leq \widehat{S}_{KM,case}(t)$ and by $\widehat{S}_{KM,control}(t)$ if $\widehat{S}(t) \geq \widehat{S}_{KM,control}(t)$. We implemented this modification in our numerical work. The modification comes into play mainly when the event rate is extremely low or extremely high. Given the consistency of $\widehat{S}(t)$, if the inequalities in (8) are strict, then for large sample sizes the modification no longer comes into play. The inequalities in (8) are strict if the following mild condition holds: for each u there exists a set $\mathcal{T}(u)$ such that $P(C_P \in \mathcal{T}(u)) > 0$ and

$$\inf_{t \in \mathcal{T}(u)} P(T_P > t, T_R > u) - P(T_P > t)P(T_R > u) > 0$$

For bandwidth selection, we propose a bootstrap procedure. Let $\widehat{S}^C(u)$ denote the Kaplan-Meier estimate of the survival function of the time to censoring among the relatives (which is the same for case relatives and control relatives). In each bootstrap replication $b = 1, \ldots, B$, for each family i we generate J_i event times for proband i's relatives according to the survival function $\widehat{S}_{Q_i}(u|T_{Pi})$ and J_i censoring times for relatives according to the survival function $\widehat{S}^C(u)$. We then run our estimation procedure for a given h on the resulting data, obtaining the estimate $\widehat{S}(u; h, b)$. Denote

$$\bar{S}(t;h) = \frac{1}{B}\sum_{b=1}^{B} \widehat{S}(t;h,b), \quad V(t;h) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{S}(t;h,b) - \bar{S}(t;h))^2$$

$$\text{MSE}_{est}(t,h) = (\bar{S}(t;h) - \widehat{S}(t;h))^2 + V(t;h), \quad \text{IMSE}_{est}(h) = \int_0^{\tau} \text{MSE}_{est}(t,h)dt$$

We evaluate $\text{IMSE}_{est}(h)$ over a grid of h values in the range $(0, 1]$ and choose the h values with the minimum $\text{IMSE}_{est}(h)$.

To construct confidence intervals in finite samples with bandwidth selection, we use the percentile bootstrap method.

6 Simulation Study

We carried out a simulation study to evaluate the finite sample properties of the proposed estimator. Data were generated under frailty models in which the within-family dependence is expressed in terms of a shared frailty variate W_i, conditional on which the failure times of the family members are independent with hazard function $\lambda(t|W_i) = W_i\lambda_0(t)$. We manipulated five design factors, as follows: (1) frailty distribution: gamma or positive stable, (2) cumulative end-of-study event rate: high (60%) or low (15%), (3) number of case probands: 500 or 1000 (with 1:1 matching of control probands to case probands), (4) number of relatives per family: 1 or 4, and (5) strength of within-family dependence: low (Kendall tau of 1/3 between the failure of times of two members of the same family) and moderate (Kendall tau of 1/2). We took $\lambda_0(t) = \nu(\mu t)^{p-1}$ with $p = 4.6$, $\mu = 0.01$, and ν chosen so as to obtain the desired cumulative end-of-study event rate. The end of study age was taken to be 110 years. The overall censoring rate, including both interim and end-of-study censoring, was about 60% in the high event rate case and 90% in the low event rate case. The number of case probands was taken to be 500, with 1:1 matching of control probands to case probands. The data generation was carried out in the same manner as in Gorfine et al. We carried out 1024 simulation replications for each of the 32 combinations of the design. For each replication, we carried out 30 inner replications for the bootstrap bandwidth selection procedure and 100 outer replications for the percentile bootstrap confidence interval procedure. The initial bandwidth was 0.5 and the bandwidth search was done in two stages. In the first stage, we searched over $[0.1, 1]$ in steps of 0.1 and identified the h value h_1 with the lowest $\text{IMSE}_{est}(h)$. In the second stage, we searched over $h_1 - 0.05, h_1$, and $\min(h_1 + 0.05, 1)$ and chose the h value with the lowest $\text{IMSE}_{est}(h)$ to be the final h value. The kernel used was the triweight kernel $K(u) = (35/32)I(|u| \leq 1)(1 - u^2)^3$.

The results for 500 case probands are summarized in Fig. 1 and 2. The left two columns of each figure show the true survival curve, along with Gorfine et al.'s estimator and the new estimator. The finite-sample bias of the new estimator tends to be smaller, and in some settings, such as the gamma frailty model with very low event rates, its finite-sample bias is dramatically

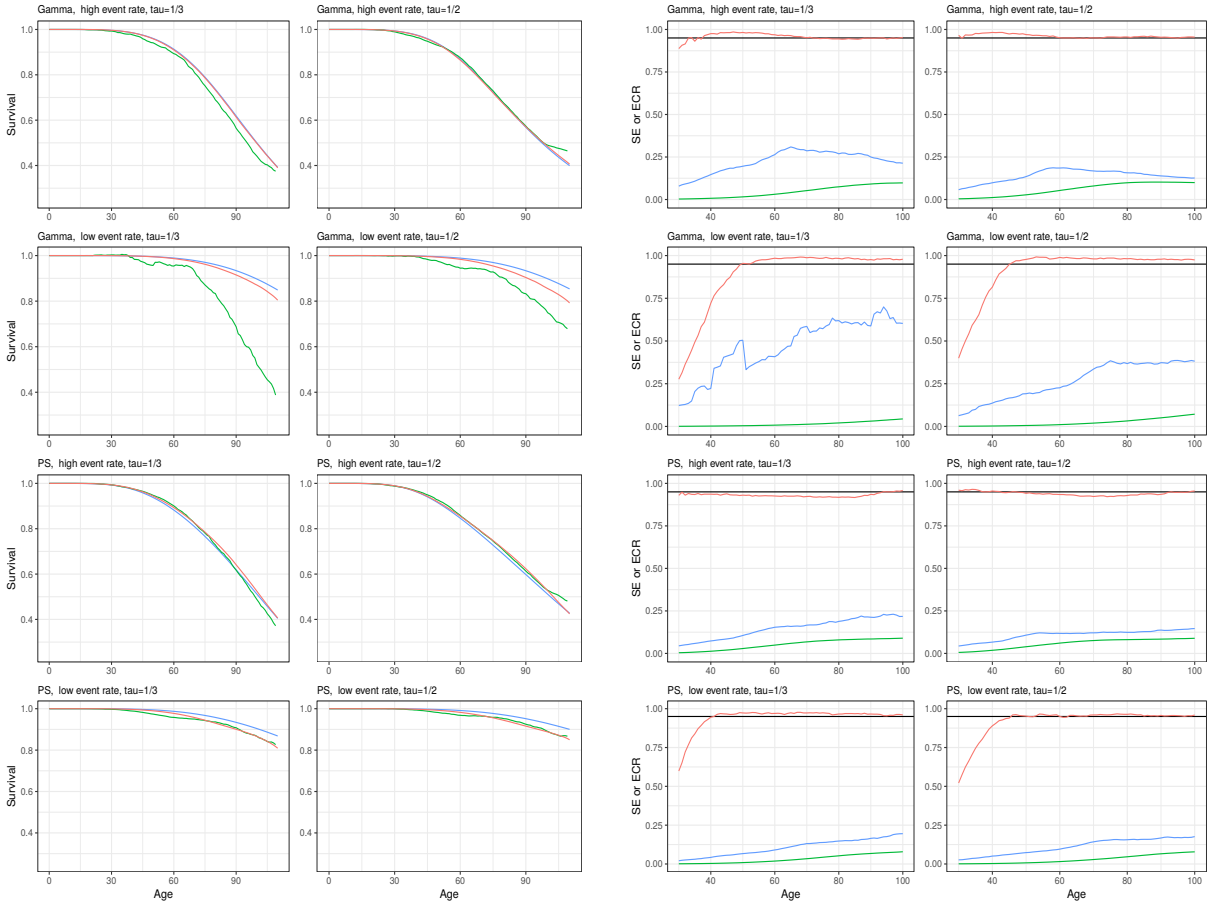

Figure 1: Simulation results, one relative for each proband: Left two columns include the true survival curve (blue); Gorfine et al. estimator (green); and the proposed estimators (red). Right two columns present the empirical standard errors of Gorfine et al. (blue) and the proposed estimator (green); and point-wise precentile-bootstrap 95% confidence interval coverage rates of the proposed estimator. The black horizontal line at 0.95 serves as a reference.

smaller. The right two columns of each figure summarize the point-wise 95% coverage rates of the percentile-bootstrap confidence interval of the proposed estimator, along with the standard errors of the Gorfine et al. estimator and the proposed estimators. Clearly, the proposed estimator substantially outperforms the old estimator in terms of efficiency. In general, the coverage rates are reasonably close to 95%, except at very early ages with small number of observed events. Similar results were obtained with 1000 case probands.

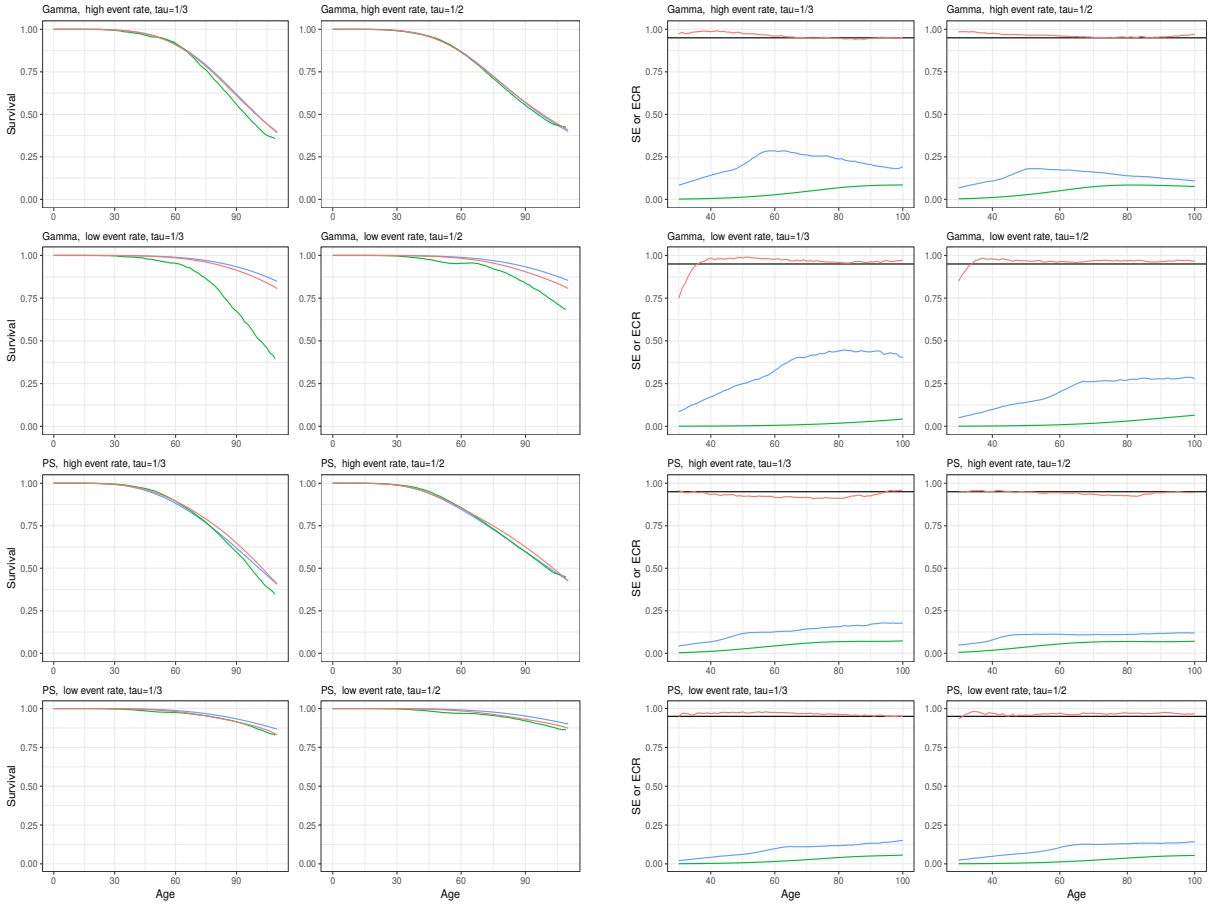

Figure 2: Simulation results, four relatives for each proband: Left two columns include the true survival curve (blue); Gorfine et al. estimator (green); and the proposed estimators (red). Right two columns present the empirical standard errors of Gorfine et al. (blue) and the proposed estimator (green); and point-wise precentile-bootstrap 95% confidence interval coverage rates of the proposed estimator. The black horizontal line at 0.95 serves as a reference.

7 Example

In this section we illustrate our method by re-analysing the data presented in Gorfine et al. (2017), population-based case-control family study of early onset prostate cancer (Stanford et al., 1999). Briefly, case participants were identified from the Seattle-Puget Sound Surveillance, Epidemiology, and End Results (SEER) cancer registry. Cases were those with age at diagnosis between 40 and 64 years. Controls were identified by use of random-digit dialing and they were frequency matched to case participants by age. The information collected on the relatives is the age at diagnosis for prostate cancer if the relative had prostate cancer or age at the last observation if the relative did not have prostate cancer. Here we use the information about age at onset or age at censoring and disease status that was observed for the probands and their fathers, brothers, and uncles. The following analysis is based on 730 prostate-cancer case probands, 693 control probands, and a total of 7316 relatives. Out of the 3793 case-probands relatives, 211 had prostate cancer, and out of the 3523 control-probands relatives, 102 had prostate cancer. The age range of the relatives with prostate cancer was 40–93. Given that frequency matching was used rather than exact age matching, and that the number of relatives per proband varied across the probands, we carried out the time transformation based on the empirical distribution of the proband observation times across all 7316 relatives in the data set. For bandwidth selection we used the same two-stage procedure as in the simulations.

Figure 3 and Table 1 present the estimates of prostate-cancer marginal survival function using the naive Kaplan-Meier estimator based on the relatives' data, Gorfine et al.'s estimator with nearest-neighbor smoothing and the median operator, the SEER survival curve based on the SEER Cancer Statistics Review 1975–2012, and the proposed estimator. In this dataset, Gorfine et al.'s estimator is closer to the SEER survival curve, but with very large point-wise standard errors compared to the proposed estimator. The standard errors reported in Table 1 are much larger than those reported in Gorfine et al. due to an error in the bootstrap code applied back then.

Table 1: Prostate cancer case-control family data

t	The proposed Estimator	Gorfine et al.	Naive KM	SEER
50	0.9997 (0.0007)	0.9918 (0.0311)	0.9991 (0.0003)	0.9958
52	0.9997 (0.0010)	0.9801 (0.0340)	0.9986 (0.0005)	0.9930
54	0.9993 (0.0012)	0.9784 (0.0413)	0.9981 (0.0006)	0.9902
56	0.9990 (0.0023)	0.9784 (0.0451)	0.9963 (0.0008)	0.9843
58	0.9910 (0.0037)	0.9784 (0.0451)	0.9945 (0.0010)	0.9783
60	0.9934 (0.0054)	0.9784 (0.0461)	0.9881 (0.0015)	0.9703
62	0.9908 (0.0063)	0.9678 (0.0501)	0.9848 (0.0017)	0.9603
64	0.9895 (0.0085)	0.9423 (0.0577)	0.9813 (0.0019)	0.9504

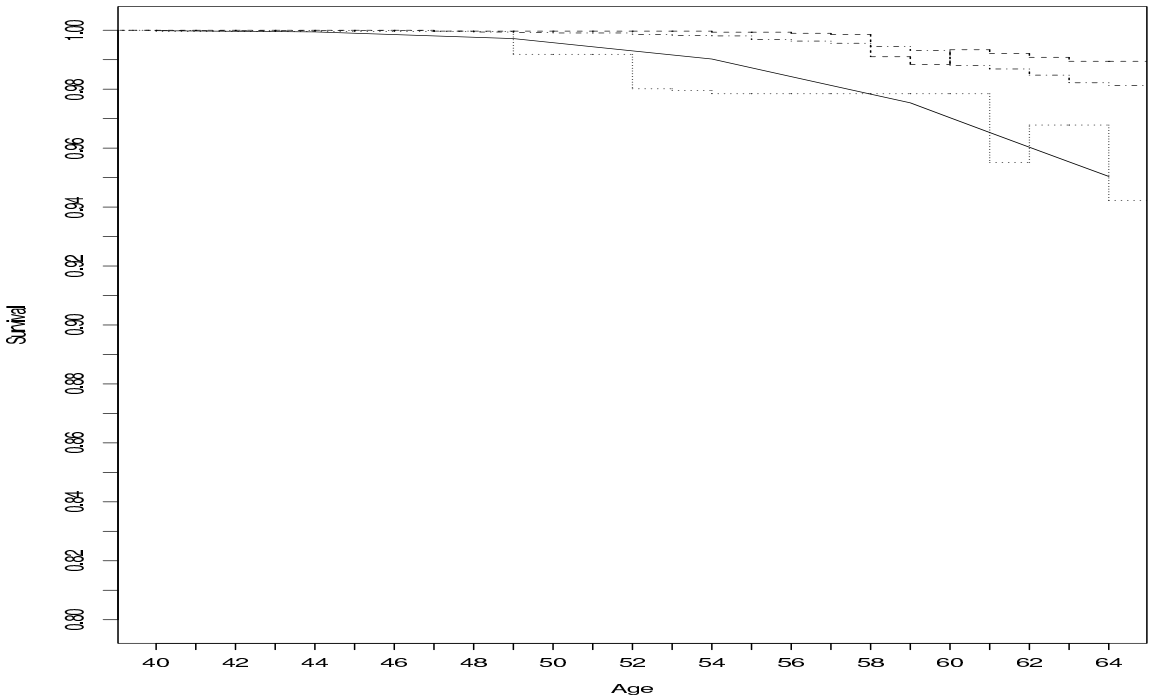

Figure 3: Prostate cancer case-control family data: the naive Kaplan-Meier estimator based on the relatives' data (dot-dashed line), Gorfine et al.'s estimator (dotted line), the SEER survival curve (solid line), and the proposed estimator (dashed line).

Acknowledgements

The authors would like to thank Dr. Stanford for her generosity in sharing the prostate cancer dataset that was used for illustrating the method. Malka Gorfine gratefully acknowledges support from the U.S.-Israel Binational Science Foundation in carrying out this work.

Supplementary material

We append a document with details of the proof of the asymptotic properties of the estimator. R code used to carry out the simulations and R code for applying the method to a data set may be found at `https://github.com/david-zucker/marginal-survival/`.

Appendix: Asymptotic Theory

A. *Preliminaries*

We first present some definitions. Let $g(x)$ denote the density of X_{Pi}. Define $y(s,v) = E[Y_{Ri\bullet}(v)|X_{Pi} = s, \delta_{Pi} = 0]$ and $\varphi(s,v) = g(s)y(s,v)$. In addition, define

$$\mathcal{I} = [h, \tau_0 - h], \quad \mathcal{U} = [0,h) \cup (\tau_0 - h, \tau_0], \quad Z_k(r) = r^k K(r), \quad \mu_k(\omega) = \int_{-1}^{\omega} Z_k(r)dr$$

$$A_k(s,v,h) = \frac{1}{mh}\sum_{i=1}^{m} Y_{Ri\bullet}(v) Z_k\left(\frac{X_{Pi} - s}{h}\right)$$

$$\eta_k(s,h) = \int_{-s/h}^{(\tau_0 - s)/h} Z_k(r)dr = \begin{cases} (-1)^k \mu_k(s/h) & s \in [0,h] \\ \mu_k(1) & s \in \mathcal{I} \\ \mu_k((\tau_0 - s)/h) & s \in [\tau_0 - h, \tau_0] \end{cases}$$

$$a_k(s,v,h) = \eta_k(s,h)\varphi(s,v)$$

Note that, by symmetry of K, $\mu_k(1) = 0$ for all odd k.

We now present two lemmas, whose proofs appear in the online Supplemental Materials. The notations O and o, and similarly O_p and o_p, should be understood as being uniform in the relevant arguments.

Lemma 1. *For k even we have*

$$\sup_{s,v} |A_k(s,v,h) - a_k(s,v,h)| = \begin{cases} O_p(m^{-(1-\nu)/2}(\log m)^{1/2}) & s \in \mathcal{I} \\ O_p(m^{-\nu}) & s \in \mathcal{U} \end{cases}$$

and for k odd we have

$$\sup_{s,v} |A_k(s,v,h)| = O_p(h) \textit{ for } s \in \mathcal{I}, \quad \sup_{s,v} |A_k(s,v,h) - a_k(s,v,h)| = O_p(h) \textit{ for } s \in \mathcal{U}$$

Lemma 2. *For $s \in \mathcal{I}$ we have*

$$\bar{X}_{P0}(s,v) - s = O_p(h^2), \quad \Gamma(s,v) = a_2(s,v,h) + O_p(m^{-(1-\nu)/2}(\log m)^{1/2})$$

and for $s \in \mathcal{U}$ we have

$$\bar{X}_{P0}(s,v) - s = \left(\frac{a_1(s,v,h)}{a_0(s,v,h)}\right) h + O_p(h^2), \quad \Gamma(s,v) = a_2(s,v,h) + O_p(h)$$

B. *Analysis of* $\mathcal{A}_1(t)$

We can write $\mathcal{A}_1(t)$ as

$$\mathcal{A}_1(t) = \int_0^t \int_0^\tau \int_0^u \psi(u|s)\Gamma(s,v)^{-1}\mathcal{S}(s,v)\,dv\,du\,ds$$

with

$$\mathcal{S}(s,v) = \frac{1}{mh^3}\sum_{i=1}^m Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))\left[\lambda_0(v|X_{Pi}) - (X_{Pi}-\bar{X}_{P0}(s,v))\lambda_0^*(v|s)\right]$$

By Taylor expansion, we can write

$$\lambda_0(v|X_{Pi}) = \lambda_0(v|s) + \lambda_0^*(v|s)(X_{Pi}-s) + \tfrac{1}{2}\lambda_0^{**}(v|s)(X_{Pi}-s)^2 + \mathcal{R}(s,v,X_{Pi})$$

in which $|\mathcal{R}(s,v,x)| \leq \mathcal{R}^*|s-x|^3$, with $\mathcal{R}^* = \sup_{s,v}|\lambda_0^{***}(v|s)|/6$, where $\lambda_0^{**}(v|s)$ and $\lambda_0^{***}(v|s)$ denote, respectively, the second and third partial derivatives of $\lambda_0(v|s)$ with respect to s We then have

$$\begin{aligned}&\lambda_0(v|X_{Pi}) - (X_{Pi}-\bar{X}_{P0}(s,v))\lambda_0^*(v|s)\\ &\quad= \left[\lambda_0(v|s) + \lambda^*(v|s)(\bar{X}_{P0}(s,v)-s)\right] + \tfrac{1}{2}\lambda_0^{**}(v|s)(X_{Pi}-s)^2 + \mathcal{R}(s,v,X_{Pi})\end{aligned}$$

The term in square brackets does not depend on i. Since

$$\sum_{i=1}^m Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v)) = 0$$

we get

$$\mathcal{S}(s,v) = \frac{1}{mh^3}\sum_{i=1}^m Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))\left[\tfrac{1}{2}\lambda_0^{**}(v|s)(X_{Pi}-s)^2 + \mathcal{R}(s,v,X_{Pi})\right]$$

We can then write $\mathcal{S}(s,v) = \mathcal{S}_1(s,v) + \mathcal{S}_2(s,v)$, where

$$\begin{aligned}\mathcal{S}_1(s,v) &= \tfrac{1}{2}\lambda_0^{**}(v|s)\left[A_3(s,v)h - (\bar{X}_{P0}(s,v)-s)A_2(s,v)\right]\\ \mathcal{S}_2(s,v) &= \frac{1}{mh^3}\sum_{i=1}^m Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))\mathcal{R}(s,v,X_{Pi})\end{aligned}$$

We have

$$\mathcal{S}_2(s,v) \leq h^2A_4(s,v) + \mathcal{R}^*h|\bar{X}_{P0}(s,v)-s|\left[\frac{1}{mh}\sum_{i=1}^m Y_{Ri\bullet}(v)K\left(\frac{X_{Pi}-s}{h}\right)\left|\frac{X_{Pi}-s}{h}\right|^3\right]$$

We now consider separately the case of $s \in \mathcal{I}$ and $s \in \mathcal{U}$. For $s \in \mathcal{I}$, the results of Lemmas 1 and 2 imply that $\mathcal{S}_1(s,v) = O_p(h^2)$ and $\mathcal{S}_2(s,v) = O_p(h^2)$, so that $\mathcal{S}(s,v) = O_p(h^2)$ and

$$\int_{[0,t]\cap\mathcal{I}} \int_0^\tau \int_0^u \psi(u|s)\Gamma(s,v)^{-1}\mathcal{S}(s,v)dv\,du\,ds = O(h^2)$$

For $s \in \mathcal{U}$, the results of Lemmas A1 and A2 imply that $\mathcal{S}_1(s,v) = O_p(h)$ and $\mathcal{S}_2(s,v) = O_p(h)$, so that $\mathcal{S}(s,v) = O_p(h)$ and

$$\int_{[0,t]\cap\mathcal{U}} \int_0^\tau \int_0^u \psi(u|s)\Gamma(s,v)^{-1}\mathcal{S}(s,v)dv\,du\,ds = O(h^2)$$

(recalling that the length of $\mathcal{U}$ is $2h$). We thus obtain $\mathcal{A}_1 = O_p(h^2)$, so that $(mh)^{1/2}\mathcal{A}_1(t) = o_p(1)$, since $\nu > 1/4$.

C. *Analysis of* $\mathcal{A}_2(t)$

Let $\rho(s,u,v) = \psi(u|s)/\varphi(s,v)$ and define

$$H(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \frac{\rho(s,u,v)}{\eta_2(s,h)} K\left(\frac{\xi - s}{h}\right)\left(\frac{\xi - s}{h}\right) ds$$
$$H_1(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \psi(u|s)(\Gamma(s,v)^{-1} - a_2(s,v,h)^{-1}) K\left(\frac{\xi - s}{h}\right)\left(\frac{\xi - s}{h}\right) ds$$
$$H_2(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \psi(u|s)\Gamma(s,v)^{-1} K\left(\frac{\xi - s}{h}\right)\left(\frac{s - \bar{X}_{P0}(s,v)}{h}\right) ds$$

Define further

$$\zeta_i(t) = \int_0^\tau \int_0^u H(u,v,t,X_{Pi})dM_{Ri\bullet}(v)du$$
$$\Delta_{1i}(t) = \int_0^\tau \int_0^u H_1(u,v,t,X_{Pi})dM_{Ri\bullet}(v)du$$
$$\Delta_{2i}(t) = \int_0^\tau \int_0^u H_2(u,v,t,X_{Pi})dM_{Ri\bullet}(v)du$$

We can then write $\mathcal{A}_2(t) = \mathcal{B}(t) + \mathcal{B}_1(t) + \mathcal{B}_2(t)$, where

$$\mathcal{B}(t) = \frac{1}{m}\sum_{i=1}^m \zeta_i(t), \quad \mathcal{B}_1(t) = \frac{1}{m}\sum_{i=1}^m \Delta_{1i}(t), \quad \mathcal{B}_2(t) = \frac{1}{m}\sum_{i=1}^m \Delta_{2i}(t)$$

Our claim is that $(mh)^{1/2}\mathcal{B}(t)$ is asymptotically mean-zero normal, and that $(mh)^{1/2}\mathcal{B}_1(t)$ and $(mh)^{1/2}\mathcal{B}_2(t)$ are both $o_p(1)$.

By (7), $E[\zeta_i(t)] = 0$. Thus, $\mathcal{B}$ is the sum of i.i.d. mean-zero terms. Accordingly, to show that $(mh)^{1/2}\,\mathcal{B}(t)$ is asymptotically mean-zero normal we need to show that $h\,\mathrm{var}(\zeta_i(t))$ converges to a limit $\kappa(t)$ and that $\zeta_i(t)$ satisfies the Lindeberg condition

$$s_m(t)^{-2} \sum_{i=1}^{m} E\left[\zeta_i(t)^2 I\left(\left|\frac{\zeta_i(t)}{s_m(t)}\right| > \epsilon\right)\right] \to 0 \quad \text{for all } \epsilon > 0 \tag{9}$$

where

$$s_m^2(t) = \mathrm{var}\left(\sum_{i=1}^{m} \zeta_i(t)\right) = m\left[\frac{\kappa(t)}{h} + O(1)\right].$$

The appended supplementary document provides a proof of the above two assertions, along with a proof that $(mh)^{1/2}\mathcal{B}_1(t)$ is $o_p(1)$. The proof of the corresponding result for $\mathcal{B}_2(t)$ is similar.

References

CHATTERJEE, N., KALAYLIOGLU, Z., SHIH, J. H. & GAIL, M. (2006). Casecontrol and case-only designs with genotype and family history data: estimating relative risk, residual familial aggregation, and cumulative risk. *Biometrics* **62**, 36–48.

DABROWSKA, D. M. (1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics* **14**, 181–197.

DOKSUM, K. A., JIANG, J., SUN, B. & WANG, S. (2017). Nearest neighbor estimates of regression. *Computational Statistics and Data Analysis* **110**, 64–74.

GORFINE, M., BORDO, N. & HSU, L. (2017). A fully nonparametric estimator of the marginal survival function based on casecontrol clustered age-at-onset data. *Biostatistics* **18**, 76–90.

SHIH, J. H. & CHATTERJEE, N. (2002). Analysis of survival data from casecontrol family studies. *Biometrics* **58**, 502–509.

STANFORD, J. L., WICKLUND, K. G., MCKNIGHT, B., DALING, J. R. & BRAWER, M. K. (1999). Vasectomy and risk of prostate cancer. *Cancer Epidemiology Biomarkers and Prevention* **8**, 881–886.

Supplemental Web Materials

An improved fully nonparametric estimator of the marginal survival function based on case-control clustered data

David M. Zucker

Department of Statistics and Data Science,

The Hebrew University of Jerusalem,

Mount Scopus, Jerusalem, Israel

email: david.zucker@mail.huji.ac.il

and

Malka Gorfine

Department of Statistics and Operations Research

Tel Aviv University, Ramat Gan, Israel

email: gorfinem@post.tau.ac.il

December 3, 2018

A. Introduction and Technical Assumptions for Asymptotic Results

This document provides details of the proof of the asymptotic properties of the estimator. Notation given in the main paper (including the Appendix) will be used throughout without repeating the definitions; additional notation will be defined as needed. In the development below, the notations O and o, and similarly O_p and o_p, should be understood as being uniform in the relevant arguments.

Below are the technical conditions assumed in deriving the asymptotic results.

1. The kernel K is symmetric, equal to zero outside of $[-1, 1]$, and equal to a polynomial inside $[-1, 1]$. In addition, K is twice differentiable with bounded derivatives over the entire real line, including the points -1 and 1.

2. The bandwidth $h = h_m$ is given by $h_m = \alpha_m m^{-\nu}$, where $1/4 < \nu < 1$ and $\alpha_m \to \alpha > 0$.

3. We have $g_{min} = \inf_{x\in[0,\tau_0]} g(x) > 0$ and $y_{min} = \inf_{s\in[0,\tau_0]} y(s, \tau) > 0$.

4. The first and second partial derivatives $\dot{\varphi}(s, v)$ and $\ddot{\varphi}(s, v)$ of $\varphi(s, v)$ with respect to s exist and are bounded uniformly over s and v. Note that Assumptions 3 and 4 imply that $\varphi_{min} = \inf_{s\in[0,\tau_0]} \inf_{v\in[0,\tau]} \varphi(s, v) > 0$.

5. The first and second partial derivatives of $\psi(u|s)$ with respect to s exist and are bounded uniformly over s and v.

6. The first three partial derivatives of $\lambda_0(v|s)$ with respect to s exist and are bounded uniformly over s and v.

B. Proofs of Lemmas A1 and A2

We begin with an expanded statement of Lemma A1, continue with the proof of this lemma, and then present Lemma A2 and its proof.

Lemma A1: For k even we have

$$E[A_k(s, v, h)] = \begin{cases} a_k(s, v, h) + O(h^2) & s \in \mathcal{I} \\ a_k(s, v, h) + O(h) & s \in \mathcal{U} \end{cases}$$

and for k odd we have

$$E[A_k(s, v, h)] = \begin{cases} O(h) & s \in \mathcal{I} \\ a_k(s, v, h) + O(h) & s \in \mathcal{U} \end{cases}$$

In addition, for any k,

$$\sup_{s,v} |A_k(s,v,h) - E[A_k(s,v,h)]| = O_p\left((mh)^{-1/2}(\log m)^{1/2}\right)$$

In general, we have

$$\sup_{s,v} |A_k(s,v,h) - a_k(s,v,h)| \leq \sup_{s,v} |A_k(s,v,h) - E[A_k(s,v,h)]| + \sup_{s,v} |E[A_k(s,v,h)] - a_k(s,v,h)|$$

When k is even, the first term dominates for $s \in I$ while the second term dominates for $s \in \mathcal{U}$, so that we obtain

$$\sup_{s,v} |A_k(s,v,h) - a_k(s,v,h)| = \begin{cases} O_p(m^{-(1-\nu)/2}\,(\log m)^{1/2}) & s \in \mathcal{I} \\ O_p(m^{-\nu}) & s \in \mathcal{U} \end{cases}$$

When k is odd, we get

$$\begin{aligned} &\sup_{s,v} |A_k(s,v,h)| = O_p(h) && s \in \mathcal{I} \\ &\sup_{s,v} |A_k(s,v,h) - a_k(s,v,h)| = O_p(h) && s \in \mathcal{U} \end{aligned}$$

Proof: The analysis of $E[A_k(s,v)]$ involves a combination of a conditioning argument with the usual change of variable + Taylor expansion argument. We have

$$\begin{aligned} E[A_k(s,v,h)] &= E\left[\frac{1}{h}\left(\frac{X_{Pi}-s}{h}\right)^k K\left(\frac{X_{Pi}-s}{h}\right) Y_{Ri\bullet}(v)\right] \\ &= E\left[\frac{1}{h}\left(\frac{X_{Pi}-s}{h}\right)^k K\left(\frac{X_{Pi}-s}{h}\right) E[Y_{Ri\bullet}(v)|X_{Pi},\delta_{Pi}=0]\right] \\ &= \frac{1}{h}\int_0^{\tau_0}\left(\frac{x-s}{h}\right)^k K\left(\frac{x-s}{h}\right) y(x,v)g(x)dx \\ &= \int_{-s/h}^{(\tau_0-s)/h} r^k K(r)\varphi(s+hr,v)dr \end{aligned}$$

By Taylor's theorem, we have

$$\varphi(s+hr,v) = \varphi(s,v) + \dot{\varphi}(s,v)(hr) + \tfrac{1}{2}R(s+hr,v)(hr)^2$$

where $|R(s+hr,v)| \leq \sup_{s,v} |\ddot{\varphi}(s,v)| < \infty$. So we get

$$\begin{aligned} E[A_k(s,v,h)] = \varphi(s,v)&\int_{-s/h}^{(\tau_0-s)/h} r^k K(r)dr \\ &+ h\dot{\varphi}(s,v)\int_{-s/h}^{(\tau_0-s)/h} r^{k+1}K(r)dr \\ &+ \tfrac{1}{2}h^2\int_{-s/h}^{(\tau_0-s)/h} r^{k+2}K(r)R(s+hr,v)dr \end{aligned}$$

and the claimed result follows.

We now turn to the analysis of $\sup_{s,v} |A_k(s,v,h) - E[A_k(s,v,h)]|$. We use Corollary 2.2 of Giné and Guillou (2002), a result concerning empirical processes. For $\bar{x} = (x_1, \ldots, x_J) \in \mathbb{R}^J$, define $L_v(\bar{x}) = \sum_{j=1}^J I(x_j \geq v)$ and

$$\Upsilon_{s,v,h}(x_0, \bar{x}) = Z_k\left(\frac{x_0 - s}{h}\right) L_v(\bar{x})$$

We can then write

$$A_k(s,v,h) = \frac{1}{mh} \sum_{i=1}^m \Upsilon_{s,v,h}(X_{Pi}, X_{R1}, \ldots, X_{RJ})$$

Since K is assumed polynomial on $[-1,1]$, the function $Z_k(r)$ satisfies Giné and Guillou's Condition (K_1). Hence, by the arguments in Giné and Guillou, the class

$$\left\{ Z_k\left(\frac{\cdot - s}{h}\right) : s \in \mathbb{R}, h > 0 \right\}$$

is a bounded, measurable VapnikChervonenkis (VC) class of functions on $\mathbb{R}$. Any set of the form $L_v(\bar{x}) = j$ can be expressed as the result of Boolean operations on half-spaces, and hence the class of sets $\{\{\bar{x} : L_v(\bar{x}) = j\}, v \in \mathbb{R}, j \in \{0, \ldots, J\}\}$ is a VC class (see Dudley, 1999, p. 141) (this is well known). Further, any set of the form $\{(x_0, \bar{x}) : \Upsilon_{s,v,h}(x_0, \bar{x}) \leq b\}$ with $b < 0$ can be expressed as

$$\bigcup_{j=1}^J \left(\{x_0 : Z_k\left(\frac{\cdot - s}{h}\right) \leq b/j\} \times \{\bar{x} : L_v(\bar{x}) = j\} \right)$$

and any set of this form with $b \geq 0$ can be expressed as

$$\left[\bigcup_{j=1}^J \left(\{x_0 : Z_k\left(\frac{\cdot - s}{h}\right) \leq b/j\} \times \{\bar{x} : L_v(\bar{x}) = j\} \right) \right] \cup (\mathbb{R} \times \{\bar{x} : L_v(\bar{x}) = 0\})$$

Recalling that the Cartesian product preserves the VC property, we can conclude that the class of functions $\Upsilon^* = \{\Upsilon_{s,v,h} : s \in [0, \tau_0], v \in [0, \tau], h > 0\}$ is a bounded VC class. Moreover, since the map $(s, v, h, x_0, \bar{x}) \mapsto \Upsilon_{s,v,h}(x_0, \bar{x})$ is jointly measurable, the class Υ^* is measurable (see Giné and Guillou, bottom of p. 911 to top of p. 912). This allows us to apply Giné and Guillou's Corollary 2.2.

We have $\sup |\Upsilon_{s,v,h}(x_0, \bar{x})| \leq U$ with $U = J \sup_r |r|^k K(r)$. Also, a simple standard calculation shows that $\mathrm{Var}(\Upsilon_{s,v,h}(X_{Pi}, X_{R1}, \ldots, X_{RJ})) \leq Rh$ for a constant R. Writing $\sigma^2 = Rh$ and letting $\mathfrak{C}$ denote the constant C in Giné and Guillou Eqn. (2.6), we find, after some simple algebra, that for m sufficiently large

$$\mathfrak{C}\sqrt{m}\sigma\sqrt{\log \frac{U}{\sigma}} \leq \rho\sqrt{mh \log m}$$

with $\rho = \mathfrak{C}\sqrt{2R\nu}$. Thus, writing $\mathcal{E}_{s,v,h} = E[\Upsilon_{s,v,h}(X_{Pi}, X_{R1}, \ldots, X_{RJ})]$ and applying Giné and Guillou's Corollary 2.2, for m sufficiently large we have

$$\begin{aligned}
&\Pr\left(\left(\frac{mh}{\log m}\right)^{1/2} \sup_{s,v} |A_k(s,v) - E[A_k(s,v)]| > \rho\right) \\
&\quad = \Pr\left(\left|\sum_{i=1}^{m} \Upsilon_{s,v,h}(X_{Pi}, X_{R1}, \ldots, X_{RJ}) - \mathcal{E}_{s,v,h}\right| > \rho\sqrt{mh\log m}\right) \\
&\quad \leq \Pr\left(\left|\sum_{i=1}^{m} \Upsilon_{s,v,h}(X_{Pi}, X_{R1}, \ldots, X_{RJ}) - \mathcal{E}_{s,v,h}\right| > \mathfrak{C}\sqrt{m}\sigma\sqrt{\log\frac{U}{\sigma}}\right) \\
&\quad \leq \mathfrak{L}_1 \exp\left(-\mathfrak{L}_2\frac{U}{\sigma}\right) \\
&\quad = \mathfrak{L}_1 \exp(-\mathfrak{L}_2[\log U/R - \log\alpha_m + \nu\log m]) \to 0
\end{aligned}$$

where $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are universal constants. This proves that

$$\sup_{s,v} |A_k(s,v,h) - E[A_k(s,v,h)]| = O_p\left((mh)^{-1/2}(\log m)^{1/2}\right)$$

Lemma A2: For $s \in \mathcal{I}$ we have

$$\begin{aligned}
&\bar{X}_{P0}(s,v) - s = O_p(h^2) \\
&\Gamma(s,v) = a_2(s,v,h) + O_p(m^{-(1-\nu)/2}(\log m)^{1/2})
\end{aligned}$$

and for $s \in \mathcal{U}$ we have

$$\begin{aligned}
&\bar{X}_{P0}(s,v) - s = \left(\frac{a_1(s,v,h)}{a_0(s,v,h)}\right)h + O_p(h^2) \\
&\Gamma(s,v) = a_2(s,v,h) + O_p(h)
\end{aligned}$$

Proof: Simple algebra yields

$$\begin{aligned}
\bar{X}_{P0}(s,v) - s &= A_0(s,v)^{-1}A_1(s,v)h \\
\Gamma(s,v) &= A_2(s,v) - A_0(s,v)\left(\frac{\bar{X}_{P0}(s,v) - s}{h}\right)^2 \\
&= A_2(s,v) - A_0(s,v)^{-1}(A_1(s,v))^2
\end{aligned}$$

The result now follows immediately from Lemma 1.

C. Analysis of $\mathcal{A}_2(t)$

As stated in the Appendix of the main paper, we let $\rho(s,u,v) = \psi(u|s)/\varphi(s,v)$ and define

$$H(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \frac{\rho(s,u,v)}{\eta_2(s,h)} K\left(\frac{\xi - s}{h}\right)\left(\frac{\xi - s}{h}\right) ds$$
$$H_1(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \psi(u|s)(\Gamma(s,v)^{-1} - a_2(s,v,h)^{-1}) K\left(\frac{\xi - s}{h}\right)\left(\frac{\xi - s}{h}\right) ds$$
$$H_2(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \psi(u|s)\Gamma(s,v)^{-1} K\left(\frac{\xi - s}{h}\right)\left(\frac{s - \bar{X}_{P0}(s,v)}{h}\right) ds$$

In addition, we define

$$\zeta_i(t) = \int_0^\tau \int_0^u H(u,v,t,X_{Pi}) dM_{Ri\bullet}(v) du$$
$$\Delta_{1i}(t) = \int_0^\tau \int_0^u H_1(u,v,t,X_{Pi}) dM_{Ri\bullet}(v) du$$
$$\Delta_{2i}(t) = \int_0^\tau \int_0^u H_2(u,v,t,X_{Pi}) dM_{Ri\bullet}(v) du$$

We can then write $\mathcal{A}_2(t) = \mathcal{B}(t) + \mathcal{B}_1(t) + \mathcal{B}_2(t)$, where

$$\mathcal{B}(t) = \frac{1}{m}\sum_{i=1}^m \zeta_i(t), \quad \mathcal{B}_1(t) = \frac{1}{m}\sum_{i=1}^m \Delta_{1i}(t), \quad \mathcal{B}_2(t) = \frac{1}{m}\sum_{i=1}^m \Delta_{2i}(t)$$

Our claim is that $\sqrt{mh}\,\mathcal{B}(t)$ is asymptotically mean-zero normal, and that $\sqrt{mh}\,\mathcal{B}_1(t)$ and $\sqrt{mh}\,\mathcal{B}_2(t)$ are both $o_p(1)$.

As noted in the Appendix of the main paper, $\mathcal{B}(t)$ is the sum of i.i.d. mean-zero terms. Accordingly, to show that $\sqrt{mh}\,\mathcal{B}(t)$ is asymptotically mean-zero normal we need to show that $h\mathrm{Var}(\zeta_i(t))$ converges to a limit $\kappa(t)$ and that $\zeta_i(t)$ satisfies the Lindeberg condition

$$s_m(t)^{-2}\sum_{i=1}^m E\left[\zeta_i(t)^2 I\left(\left|\frac{\zeta_i(t)}{s_m(t)}\right| > \epsilon\right)\right] \to 0 \quad \forall \epsilon > 0 \tag{1}$$

where

$$s_m(t)^2 = \mathrm{Var}\left(\sum_{i=1}^m \zeta_i(t)\right) = m\left[\frac{\kappa(t)}{h} + O(1)\right]$$

1. Analysis of $\mathrm{Var}(\zeta_i(t))$

We can write $H(u,v,t,\xi)$ as

$$H(u,v,t,\xi) = \frac{1}{h}\int_{-1}^1 rK(r) I\left(r \in \left[\frac{\xi - t}{h}, \frac{\xi}{h}\right]\right) \frac{\rho(\xi - hr, u, v)}{\eta_2(\xi - hr, h)} dr$$

The relevant range of ξ is $[0, \tau_0]$. The analysis of $H(u,v,t,\xi)$ divides into several cases. To ease the presentation, we assume that $t < \tau_0$. A similar development can be given for $t = \tau_0$.

Case 1, $\xi = \omega h$ with $\omega \in [0,1]$: In this we case we have $H(u,v,t,\xi) = -h^{-1}\mathcal{P}_1(\omega)\rho(0,u,v) + O(1)$, where

$$\mathcal{P}_1(\omega) = \int_{-\omega}^{1-\omega} \frac{rK(r)}{\mu_2(\omega+r)}\,dr - \frac{\mu_1(-1+\omega)}{\mu_2(1)}$$

Case 2, $\xi = (1+\omega)h$ with $\omega \in [0,1]$: In this we case we have $H(u,v,t,\xi) = -h^{-1}\mathcal{P}_2(\omega)\rho(0,u,v) + O(1)$, where

$$\mathcal{P}_2(\omega) = \int_{-1}^{-\omega} \frac{rK(r)}{\mu_2(1+\omega+r)}\,dr - \frac{\mu_1(\omega)}{\mu_2(1)}$$

Case 3, $\xi \in [2h, t-h]$: In this case the indicator equals 1 and $\xi - hr \in \mathcal{I}$ for all $r \in [-1,1]$, and hence, recalling that $\mu_1(1) = 0$, we get $H(u,v,t,\xi) = -\dot{\rho}(\xi,u,v) + O(h)$, where $\dot{\rho}(s,u,v)$ is the partial derivative of $\rho(s,u,v)$ with respect to s.

Case 4, $\xi = t + \omega h$ with $\omega \in [-1,1]$: In this case, $H(u,v,t,\xi) = -h^{-1}\mu_1(-\omega)\rho(t,u,v)/\mu_2(1) + O(1)$.

Case 5, $\xi > t+h$: In this case the indicator equals 0 for all $r \in [-1,1]$ and so $H(u,v,t,\xi) = 0$.

Define

$$\mathcal{V}(\xi,t) = E\left[\int_0^\tau \left(\int_0^u H(u,v,t,\xi)dM_{i\bullet}(v)du\right)^2 \middle| X_{Pi} = \xi, \delta_{Pi} = 0\right]$$
$$\mathcal{V}^*(\xi,\xi') = E\left[\left(\int_0^\tau \int_0^u \rho(\xi',u,v)dM_{i\bullet}(v)du\right)^2 \middle| X_{Pi} = \xi, \delta_{Pi} = 0\right]$$
$$\dot{\mathcal{V}}^*(\xi) = E\left[\left(\int_0^\tau \int_0^u \dot{\rho}(\xi,u,v)dM_{i\bullet}(v)du\right)^2 \middle| X_{Pi} = \xi, \delta_{Pi} = 0\right]$$

We then have

$$\mathcal{V}(\xi,t) = \begin{cases} h^{-2}\mathcal{V}^*(\xi,0)\mathcal{P}_1(\xi/h)^2 + O(h^{-1}) & \xi \in [0,h] \\ h^{-2}\mathcal{V}^*(\xi,0)\mathcal{P}_2(\xi/h-1)^2 + O(h^{-1}) & \xi \in [h,2h] \\ \dot{\mathcal{V}}^*(\xi) + O(h) & \xi \in [2h, t-h] \\ h^{-2}\mathcal{V}^*(\xi,t)(\mu_1(-(\xi-t)/h)/\mu_2(1))^2 + O(h^{-1}) & \xi \in [t-h, t+h] \\ 0 & \xi > t+h \end{cases}$$

Accordingly,

$$\mathrm{Var}(\zeta_i(t)) = E[\zeta_i(t)^2] = \int_0^{\tau_0} g(\xi)\mathcal{V}(\xi)d\xi = \mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3 + \mathcal{C}_4$$

where

$$\mathcal{C}_1 = \int_0^h g(\xi)\mathcal{V}(\xi,0,t)d\xi = h\int_0^1 g(\omega h)\mathcal{V}(\omega h,0,t)d\omega = h^{-1}g(0)\mathcal{V}^*(0,0)\int_0^1 \mathcal{P}_1(\omega)^2 d\omega + O(1)$$
$$\mathcal{C}_2 = \int_h^{2h} g(\xi)\mathcal{V}(\xi,t)d\xi = h\int_1^2 g(\omega h)\mathcal{V}(\omega h,0,t)d\omega = h^{-1}g(0)\mathcal{V}^*(0,0)\int_0^1 \mathcal{P}_2(\omega)^2 d\omega + O(1)$$
$$\mathcal{C}_3 = \int_{2h}^{t-h} g(\xi)\mathcal{V}(\xi,t)d\xi = \int_{2h}^{t-h} g(\xi)\dot{\mathcal{V}}^*(\xi)d\xi + O(h)$$
$$\mathcal{C}_4 = \int_{t-h}^{t+h} g(\xi)\mathcal{V}(\xi,t)d\xi = h^{-1}g(t)\mathcal{V}^*(t,t)\mu_2(1)^{-2}\int_{-1}^1 \mu_1(\omega)^2 d\omega + O(1)$$

In other words,

$$\mathrm{Var}(\zeta_i(t)) = \frac{\kappa(t)}{h} + O(1)$$

where

$$\kappa(t) = g(0)\mathcal{V}^*(0,0)\left[\int_0^1 \mathcal{P}_1(\omega)^2 d\omega + \int_0^1 \mathcal{P}_2(\omega)^2 d\omega\right] + g(t)\mathcal{V}^*(t,t)\mu_2(1)^{-2}\int_{-1}^1 \mu_1(\omega)^2 d\omega$$

2. <u>Proof of Lindeberg Condition</u>

We have

$$\begin{aligned}
|\zeta_i(t)| &= \left|\int_0^\tau \int_0^u H(u,v,t,X_{Pi})dM_{Ri\bullet}(v)du\right| \\
&\le \int_0^\tau \int_0^u |H(u,v,t,X_{Pi})|dN_{Ri\bullet}(v)du + \int_0^\tau \int_0^u |H(u,v,t,X_{Pi})|Y_{Ri\bullet}(v)\lambda(v)dvdu \\
&\le (1+\lambda_{max})J\tau \sup_{u,v,t,\xi} |H(u,v,t,\xi)| \\
&\le [(1+\lambda_{max})J\tau]\left[\int_{-1}^1 |r|K(r)dr\right]\mu_2(0)^{-1}\sup_{s,u,v}|\rho(s,u,v)|h^{-1} = \mathcal{M}h^{-1}
\end{aligned}$$

with $\mathcal{M}$ defined in the obvious manner. Thus,

$$\left|\frac{\zeta_i(t)}{s_m(t)}\right| \le \frac{\mathcal{M}}{[(\kappa(t)+O(h))mh]^{1/2}} \to 0$$

Thus, the Lindeberg condition (1) is satisfied.

3. <u>Analysis of $\mathcal{B}_1(t)$ and $\mathcal{B}_2(t)$</u>

We show here that $\sqrt{mh}\ \mathcal{B}_1(t) = o_p(1)$; the argument for $\sqrt{mh}\ \mathcal{B}_2(t)$ is similar. For simplicity of exposition, we present the proof for the case $J = 2$. Define the filtration $\mathcal{F}_v = \sigma(\mathcal{F}_0, \{N_{Rij}(d), Y_{Rij}(d);\ i = 1,\ldots,m; j = 1,2;\ d \in [0,v]\})$. We can write

$$\begin{aligned}
&\lim_{d\downarrow 0}\Pr(N_{Ri1}(t+d) - N_{Ri1}(t) = 1|\mathcal{F}_{v-}) \\
&\quad = Y_{Ri1}(v)\left[Y_{Ri2}(v)\lambda_0(v|v) + (1-Y_{Ri2}(v))(1-N_{Ri2}(v))\lambda_0(v|X_{Ri2}) + N_{Ri2}(v)\lambda_1(v|X_{Ri2})\right]
\end{aligned}$$

A similar equality holds for $\lim_{d\downarrow 0} \Pr(N_{Ri2}(t+d) - N_{Ri2}(t) = 1|\mathcal{F}_{v-})$. So if we define

$$\tilde{\lambda}_i(v) = Y_{Ri1}(v)\left[Y_{Ri2}(v)\lambda_0(v|v) + (1-Y_{Ri2}(v))(1-N_{Ri2}(v))\lambda_0(v|X_{Ri2}) + N_{Ri2}(v)\lambda_1(v|X_{Ri2})\right]$$
$$+ Y_{Ri2}(v)\left[Y_{Ri1}(v)\lambda_0(v|v) + (1-Y_{Ri1}(v))(1-N_{Ri1}(v))\lambda_0(v|X_{Ri1}) + N_{Ri1}(v)\lambda_1(v|X_{Ri1})\right]$$

then the process $\tilde{M}_{Ri\bullet}(v)$ defined by $d\tilde{M}_{Ri\bullet}(v) = \sum_{i=1}^m (dN_{Ri\bullet}(v) - \tilde{\lambda}_i(v)dv)$ is a martingale with respect to the filtration $\mathcal{F}_v$.

Define

$$\mathcal{B}_1(t,u) = \frac{1}{m}\sum_{i=1}^m \Delta_{1i}(t,u)$$

with

$$\Delta_{1i}(t,u) = \int_0^u H_1(u,v,t,X_{Pi})dM_{Ri\bullet}(v)$$

We can write $\sqrt{mh}\ \mathcal{B}_1(t,u) = \mathcal{B}_1^*(t,u,u) + \mathcal{B}_1^{**}(t,u)$, where

$$\mathcal{B}_1^*(t,u,w) = \sqrt{mh}\left[\frac{1}{m}\sum_{i=1}^m \int_0^u H_1(w,v,t,X_{Pi})d\tilde{M}_{Ri\bullet}(v)\right]$$
$$\mathcal{B}_1^{**}(t,u) = \sqrt{mh}\left[\frac{1}{m}\sum_{i=1}^m \int_0^u H_1(w,v,t,X_{Pi})(dM_{Ri\bullet}(v) - d\tilde{M}_{Ri\bullet}(v))\right]$$
$$= \sqrt{mh}\left[\frac{1}{m}\sum_{i=1}^m \int_0^u H_1(w,v,t,X_{Pi})(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))dv\right]$$

Note that $E[\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi})] = 0$.

Now, since $H_1(u,v,t,X_{Pi})$, viewed as a process in v, is predictable with respect to $\mathcal{F}_v$, the process $\mathcal{B}_1^*(t,u,w)$ viewed as a process in u, is a martingale with respect to $\mathcal{F}_v$, with predictable variation process given by

$$\langle \mathcal{B}_1^*(t,\cdot,w), \mathcal{B}_1^*(t,\cdot,w)\rangle(u) = h\int_0^u \left[\frac{1}{m}\sum_{i=1}^m H_1(w,v,t,X_{Pi})^2\tilde{\lambda}_i(v)\right]dv \tag{2}$$

We can write $H_1(w,v,t,\xi) = H_{1a}(w,v,t,\xi) + H_{1b}(w,v,t,\xi)$ with

$$H_{1a}(w,v,t,\xi) = \frac{1}{h^2}\int_{[0,t]\cap\mathcal{I}} \psi(w|s)(\Gamma(s,v)^{-1} - a_2(s,v,h)^{-1})K\left(\frac{\xi-s}{h}\right)\left(\frac{\xi-s}{h}\right)ds$$
$$H_{1b}(w,v,t,\xi) = \frac{1}{h^2}\int_{[0,t]\cap\mathcal{I}} \psi(w|s)(\Gamma(s,v)^{-1} - a_2(s,v,h)^{-1})K\left(\frac{\xi-s}{h}\right)\left(\frac{\xi-s}{h}\right)ds$$

Now,

$$H_{1a}(w,v,t,\xi) \le \sup_{s\in\mathcal{I},v\in[0,\tau]} |\Gamma(s,v)^{-1} - a_2(s,v)^{-1}| \sup_{w,s} |\psi(w|s)|\mathfrak{A}(\xi)$$

with

$$\mathfrak{A}(\xi) = \frac{1}{h^2}\int_0^t K\left(\frac{\xi - s}{h}\right)\left|\frac{\xi - s}{h}\right| ds$$

Defining

$$\mathfrak{B}(r) = \int_{-1}^{r} |r'|K(r')dr'$$

we have $\mathfrak{A}(\xi) \leq \mathfrak{A}'(\xi)$ with

$$\mathfrak{A}'(\xi) = h^{-1}\mathfrak{B}\left(\frac{t-\xi}{h}\right)$$

and

$$E[\mathfrak{A}'(X_{Pi})] = \int_0^{\tau_0} h^{-1}\mathfrak{B}\left(\frac{t-\xi}{h}\right) g(\xi)d\xi \leq \int_{-t/h}^{(\tau_0 - t)/h} \mathfrak{B}(\xi')g(t + h\xi')d\xi' \leq \mathfrak{B}(1)$$

Thus $\mathfrak{A}(X_{Pi}) = O_p(1)$. Hence, using the result of Lemma A2, we obtain

$$|H_{1a}(w, v, t, X_{Pi})| = O_p(m^{-(1-\nu)/2}(\log m)^{1/2}) = O_p(1)$$

since $\nu < 1$. Similarly, again using the result of Lemma A2, we find that $|H_{1b}(w, v, t, X_{Pi})| = O_p(1)$. Accordingly, the term in brackets in (2) is $O_p(1)$. It follows from Lenglart's inequality (see, e.g., Andersen and Gill, 1982, Thm. I.1(b)) that for any given w (and in particular for $w = u$), $\sup_{u\in[0,\tau]} |\mathcal{B}_1^*(t, u, w)| = O_p(\sqrt{h})$.

In regard to $\mathcal{B}_1^{**}(t, u)$, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\mathcal{B}_1^{**}(t, u)| = \sqrt{mh}\Bigg|\int_0^u \int_0^t \psi(u|s)(\Gamma(s, v)^{-1} - a_2(s, v, h)^{-1}) \\
\left[\frac{1}{mh^2}\sum_{i=1}^m K\left(\frac{X_{Pi} - s}{h}\right)\left(\frac{X_{Pi} - s}{h}\right)(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))\right] ds\, dv\Bigg| \\
\leq \sqrt{mh}\sqrt{\mathcal{Q}_1\mathcal{Q}_2}
\end{aligned}$$

with

$$\begin{aligned}
\mathcal{Q}_1 &= \int_0^u \int_0^t \psi(u|s)(\Gamma(s, v)^{-1} - a_2(s, v, h)^{-1})^2 ds\, dv \\
\mathcal{Q}_2 &= \int_0^u \int_0^t \psi(u|s)\left[\frac{1}{mh^2}\sum_{i=1}^m K\left(\frac{X_{Pi} - s}{h}\right)\left(\frac{X_{Pi} - s}{h}\right)(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))\right]^2 ds\, dv
\end{aligned}$$

Now,

$$
\begin{aligned}
E[\mathcal{Q}_2] &= \int_0^u \int_0^t \psi(u|s) E\left[\left\{\frac{1}{mh^2}\sum_{i=1}^m K\left(\frac{X_{Pi}-s}{h}\right)\left(\frac{X_{Pi}-s}{h}\right)(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))\right\}^2\right] ds\, dv \\
&= (mh^2)^{-1}\int_0^u \int_0^t \psi(u|s) E\left[\frac{1}{h^2}K^2\left(\frac{X_{Pi}-s}{h}\right)\left(\frac{X_{Pi}-s}{h}\right)^2(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))^2\right] ds\, dv \\
&= (mh^2)^{-1}\int_0^u \int_0^t \int_0^{\tau_0} h^{-2}\psi(u|s)K^2\left(\frac{\xi-s}{h}\right)\left(\frac{\xi-s}{h}\right)^2(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|\xi))^2 g(\xi) d\xi\, ds\, dv \\
&= O((mh^2)^{-1})
\end{aligned}
$$

so that $\mathcal{Q}_2 = O_p((mh^2)^{-1})$. In addition, by Lemma A2, $\mathcal{Q}_1 = O_p(m^{-(1-\nu)}(\log m))$. We thus find that $\mathcal{B}_1^{**}(t,u) = O_p(m^{-(1-2\nu)/2}(\log m)^{1/2}) = o_p(1)$.

D. Time Transformation According to the EDF of the Proband Observation Times

In this section, we sketch the proof that the consistency and asymptotic normality of our estimator is maintained under a time transformation based on an estimate G_m of the cumulative distribution function G of the proband observation times. In this proof we need to assume that G has four bounded derivatives and that a smooth estimate of G is used. We conjecture that the result holds without these conditions (in our simulations, we obtained good results taking G_m to be a linearly interpolated version of the empirical CDF), but we do not have a proof.

We take

$$
G_m(t) = \frac{1}{m}\sum_{i=1}^m \mathcal{K}\left(\frac{t - X_{Pi}}{b_m}\right)
$$

where

$$
\mathcal{K}(a) = \int_{-\infty}^a K(c)dc
$$

with b_m chosen as described in Schuster (1969) so that $G_m(t)$ and its first three derivatives converge uniformly to G and its first three derivatives. We define $D = G^{-1}$ and $D_m = G_m^{-1}$.

For any thrice differentiable inverse CDF function $\mathcal{D}$ and any $\bar{s} \in [0,1]$, define

$$\Lambda_{\mathcal{D}}(\bar{s}) = \Lambda(\mathcal{D}(\bar{s}))$$

$$S_{0,\mathcal{D}}(u|\bar{s}) = S_0(u|\mathcal{D}(\bar{s}))$$

$$\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \Lambda_0(u|\mathcal{D}(\bar{s}))$$

$$\Lambda^*_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial \bar{s}}\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \Lambda^*_0(u|\mathcal{D}(\bar{s}))\mathcal{D}'(\bar{s})$$

$$\lambda_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial u}\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \lambda_0(u|\mathcal{D}(\bar{s}))$$

$$\lambda^*_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial \bar{s}}\lambda_{0,\mathcal{D}}(u|\bar{s})$$

$$\lambda^{**}_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial^2}{\partial \bar{s}^2}\lambda_{0,\mathcal{D}}(u|\bar{s})$$

$$\lambda^{***}_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial^3}{\partial \bar{s}^3}\lambda_{0,\mathcal{D}}(u|\bar{s})$$

$$\psi_{\mathcal{D}}(u,\bar{s}) = \psi(u,\mathcal{D}(\bar{s}))$$

Also, for any distribution function $\mathcal{G}$, define

$$A_k(\bar{s},v,h,\mathcal{G}) = \frac{1}{mh}\sum_{i=1}^{m} Y_{Ri\bullet}(v) Z_k\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right)$$

$$\mathcal{M}(\bar{s},v,\mathcal{G}) = A_1(\bar{s},v,h,\mathcal{G})/A_0(\bar{s},v,h,\mathcal{G})$$

$$\Gamma(\bar{s},v,\mathcal{G}) = A_2(\bar{s},v,h,\mathcal{G}) - A_0(\bar{s},v,h,\mathcal{G})^{-1}(A_1(\bar{s},v,h,\mathcal{G}))^2$$

For any given nonrandom $\mathcal{G}$, the analogues of Lemmas A1 and A2 hold for the above quantities. Our estimator $\hat{\Lambda}^*_{0,D}(u|\bar{s})$ of $\Lambda^*_{0,D}(u|\bar{s})$ is

$$\hat{\Lambda}^*_{0,D}(u|\bar{s}) = \int_0^u \Gamma(\bar{s},v,G_m)^{-1}\left[\frac{1}{mh^3}\sum_{i=1}^{m} K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)(G_m(X_{Pi}) - \mathcal{M}(\bar{s},v,G_m))dN_{Ri\bullet}(v)\right]$$

our estimator of $\Lambda_D(\bar{s})$ is

$$\hat{\Lambda}_D(\bar{t}) = -\int_0^{\bar{t}}\int_0^{\tau} \hat{\psi}_D(u,\bar{s})\hat{\Lambda}^*_{0,D}(u|\bar{s})\,du\,d\bar{s}$$

and our estimator of $\Lambda(t)$ is $\hat{\Lambda}(t) = \hat{\Lambda}_D(G_m(t))$.

In the analysis of $\hat{\Lambda}_D(\bar{t}) - \Lambda_D(\bar{t})$, the analogues of $\mathcal{A}_1(t)$ and $\mathcal{A}_2(t)$ are $\mathcal{A}_1(\bar{t},G_m)$ and $\mathcal{A}_2(\bar{t},G_m)$,

where

$$\mathcal{A}_1(\bar{t},\mathcal{G}) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,\mathcal{G})^{-1}Y_{Ri\bullet}(v)K\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right)(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))$$

$$(\lambda_{0,\mathcal{G}^{-1}}(v|\mathcal{G}(X_{Pi}))-(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))\lambda^*_{0,D}(v|\bar{s}))dv\ du\,d\bar{s}$$

$$\mathcal{A}_2(\bar{t},\mathcal{G}) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,\mathcal{G})^{-1}K\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right)(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))dM_{Ri\bullet}(v)\,du\,d\bar{s}$$

We can write $\mathcal{A}_1(\bar{t},\mathcal{G}) = \mathcal{A}_{1a}(\bar{t},\mathcal{G}) - \mathcal{A}_{1b}(\bar{t},\mathcal{G})$, where

$$\mathcal{A}_{1a}(\bar{t},\mathcal{G}) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,\mathcal{G})^{-1}Y_{Ri\bullet}(v)K\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right)(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))$$

$$(\lambda_{0,\mathcal{G}^{-1}}(v|\mathcal{G}(X_{Pi}))-(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))\lambda^*_{0,\mathcal{G}^{-1}}(v|\bar{s}))dv\ du\,d\bar{s}$$

$$\mathcal{A}_{1b}(\bar{t},\mathcal{G}) = \int_0^{\bar{t}}\int_0^{\tau}\psi_D(u,\bar{s})(\Lambda^*_{0,\mathcal{G}^{-1}}(u|\bar{s})-\Lambda^*_{0,D}(u|\bar{s}))du\,d\bar{s}$$

By the same Taylor expansion argument as in the Appendix of the main paper, we can write

$$\mathcal{A}_{1a}(\bar{t},\mathcal{G}) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,\mathcal{G})^{-1}Y_{Ri\bullet}(v)K\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right)$$

$$(\mathcal{G}(X_{Pi})-\mathcal{M}(\bar{s},v,\mathcal{G}))\left[\tfrac{1}{2}\lambda^{**}_{0,\mathcal{G}^{-1}}(v|\bar{s})(\mathcal{G}(X_{Pi})-\bar{s})^2+\mathcal{R}(\bar{s},v,\mathcal{G}(X_{Pi}))\right]$$

where $|\mathcal{R}(\bar{s},v,\bar{x})| \le \mathcal{R}^*_m|\bar{s}-\bar{x}|^3$ with $\mathcal{R}^*_m = O_p(1)$. We have $\sqrt{mh}\,\mathcal{A}_{1a}(\bar{t},G)$ by the same argument as in the Appendix of the main paper, which leaves us to deal with $\mathcal{A}_{1a}(\bar{t},G_m)-\mathcal{A}_{1a}(\bar{t},G)$ and $\mathcal{A}_{1b}(\bar{t},G_m)$, The quantity $\mathcal{A}_{1a}(\bar{t},G_m)-\mathcal{A}_{1a}(\bar{t},G)$ can be broken up into a series of various terms. A typical term is

$$\Psi = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,G)^{-1}Y_{Ri\bullet}(v)\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]$$

$$(G(X_{Pi})-\mathcal{M}(\bar{s},v,G))\left[\tfrac{1}{2}\lambda^{**}_{0,D}(v|\bar{s})(G(X_{Pi})-\bar{s})^2+\mathcal{R}(\bar{s},v,G(X_{Pi}))\right]dv\ du\,d\bar{s}$$

We can write $\Psi = \Psi_1 - \Psi_2$, where

$$\Psi_1 = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,G)^{-1}Y_{Ri\bullet}(v)\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]$$

$$(G(X_{Pi})-\bar{s})\left[\tfrac{1}{2}\lambda^{**}_{0,D}(v|\bar{s})(G(X_{Pi})-\bar{s})^2+\mathcal{R}(\bar{s},v,G(X_{Pi}))\right]dv\ du\,d\bar{s}$$

$$\Psi_2 = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})\Gamma(\bar{s},v,G)^{-1}Y_{Ri\bullet}(v)\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]$$

$$(\mathcal{M}(\bar{s},v,G)-\bar{s})\left[\tfrac{1}{2}\lambda^{**}_{0,D}(v|\bar{s})(G(X_{Pi})-\bar{s})^2+\mathcal{R}(\bar{s},v,G(X_{Pi}))\right]dv\ du\,d\bar{s}$$

Write $\Delta_i = (G_m(X_{Pi}) - G(X_{Pi}))/h$ and $\Delta = \|G_m - G\|_\infty/h$. We have

$$
\begin{aligned}
|\Psi_1| &\le O_p(1)\left[\frac{1}{m}\sum_{i=1}^{m}\int_0^{\bar{t}}\left|K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right) - K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right| \right.\\
&\qquad\qquad \left.\left\{\left|\frac{G(X_{Pi})-\bar{s}}{h}\right|^3 + \left|\frac{G(X_{Pi})-\bar{s}}{h}\right|^4\right\} d\bar{s}\right]\\
&= O_p(1)\left[\frac{h}{m}\sum_{i=1}^{m}\int_{-G(X_{Pi})/h}^{(\bar{t}-G(X_{Pi}))/h}\left|K\left(r + \frac{G_m(X_{Pi})-G(X_{Pi})}{h}\right) - K(r)\right|(|r|^3+|r|^4)dr\right]\\
&\le O_p(1)\left[\frac{h}{m}\sum_{i=1}^{m}\int_{-1-|\Delta|}^{1+|\Delta|}(|r|^3+|r|^4)dr|K(r+\Delta_i)-K(r)|dr\right]\\
&\le O_p(1)\left[\frac{h}{m}\sum_{i=1}^{m}|\Delta_i|\int_{-1-|\Delta|}^{1+|\Delta|}(|r|^3+|r|^4)dr\right]\\
&\le O_p(1)[h\Delta(1+\Delta)^4]
\end{aligned}
$$

Recalling that $\|G_m - G\|_\infty = O_p(m^{-1/2})$, we obtain $|\Psi_1| = O_p(m^{-1/2})$. Thus $\sqrt{mh}\,\Psi_1 = o_p(1)$. Similarly, using Lemma A2, $\sqrt{mh}\,\Psi_2 = o_p(1)$.

Next, regarding $\mathcal{A}_{1b}(\bar{t}, G_m)$, we can write

$$
\begin{aligned}
\Lambda^*_{0,D_m}(u|\bar{s}) - \Lambda^*_{0,D}(u|\bar{s}) &= \Lambda^*_0(u|D_m(\bar{s}))D'_m(\bar{s}) - \Lambda^*_0(u|D(\bar{s}))D'(\bar{s})\\
&= (\Lambda^*_0(u|D_m(\bar{s})) - \Lambda^*_0(u|D(\bar{s})))D'_m(\bar{s}) + \Lambda^*_0(u|D(\bar{s}))(D'_m(\bar{s}) - D'(\bar{s}))\\
&= \Lambda^{**}_{0,D}(u|s^*)(D_m(\bar{s}) - D(\bar{s})) + \Lambda^*_0(u|D(\bar{s}))(D'_m(\bar{s}) - D'(\bar{s}))
\end{aligned}
$$

with s^* between $D_m(\bar{s})$ and $D(\bar{s})$. The first of the above two terms is $O_p(m^{-1/2})$. Regarding the second term, using integration by parts we can write

$$
\begin{aligned}
&\int_0^{\bar{t}}\psi_D(u,\bar{s})\Lambda^*_0(u|D(\bar{s}))(D'_m(\bar{s}) - D'(\bar{s}))d\bar{s}\\
&\quad = [\psi_D(u,\bar{s})\Lambda^*_0(u|D(\bar{s}))(D_m(\bar{s}) - D(\bar{s}))]_0^{\bar{t}} - \int_0^{\bar{t}}\left[\frac{\partial}{\partial\bar{s}}\psi_D(u,\bar{s})\Lambda^*_0(u|D(\bar{s}))\right](D_m(\bar{s}) - D(\bar{s}))d\bar{s}\\
&\quad = O_p(m^{-1/2})
\end{aligned}
$$

Thus $\sqrt{mh}\,\mathcal{A}_{1b}(\bar{t}, G_m) = o_p(1)$.

We turn now to $\mathcal{A}_2(\bar{t}, G_m)$. By the same arguments as before we find that $\sqrt{mh}\,\mathcal{A}_2(\bar{t}, G)$ converges in distribution to a mean-zero normal distribution. This leaves us to deal with $\mathcal{A}_2(\bar{t}, G_m) - \mathcal{A}_2(\bar{t}, G)$. This quantity can be broken up into a series of terms, a typical one of which is

$$\int_0^{\tau}\Omega(\bar{t}, u)du$$

with

$$\Omega(\bar{t},u) = \frac{1}{mh^3}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})a_2(\bar{s},v,G)^{-1}\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]$$

$$(G(X_{Pi})-\bar{s})\,dM_{Ri\bullet}(v)\,du\,d\bar{s}$$

$$= \frac{1}{mh^2}\sum_{i=1}^{m}\int_0^{\bar{t}}\int_0^{\tau}\int_0^{u}\psi_D(u,\bar{s})a_2(\bar{s},v,G)^{-1}\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]$$

$$\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)dM_{Ri\bullet}(v)\,du\,d\bar{s}$$

This term can be dealt using an argument similar to that used for $\mathcal{B}_1(t,u)$. We can write $\Omega(u) = \Omega^*(\bar{t},u,w)+\Omega^{**}(\bar{t},u)$ with

$$\Omega^*(\bar{t},u,w) = \frac{1}{m}\sum_{i=1}^{m}\int_0^{u}\tilde{H}(w,v,\bar{t},X_{Pi})\tilde{M}_{Ri}(v)$$

$$\Omega^{**}(\bar{t},u) = \frac{1}{m}\sum_{i=1}^{m}\int_0^{u}\tilde{H}(w,v,\bar{t},X_{Pi})(\tilde{\lambda}_i(v)-Y_{Ri\bullet}(v)\lambda_{0,D}(v|G(X_{Pi})))dv$$

$$\tilde{H}(w,v,t,X_{Pi}) = \frac{1}{h^2}\int_0^{\bar{t}}\psi_D(u,\bar{s})a_2(\bar{s},v,G)^{-1}$$

$$\left[K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right]\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)d\bar{s}$$

We have

$$|\tilde{H}(w,v,\bar{t},X_{Pi})| = \tilde{H}_a(w,v,\bar{t},X_{Pi})+\tilde{H}_b(w,v,\bar{t},X_{Pi})$$

with

$$\tilde{H}_a(w,v,\bar{t},X_{Pi}) = \frac{1}{h^3}\int_0^{\bar{t}}|\psi_D(w,\bar{s})|a_2(\bar{s},v,G)^{-1}$$

$$\left|K'\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)\right|\left|\frac{G(X_{Pi})-\bar{s}}{h}\right||G_m(X_{Pi})-G(X_{Pi})|\,d\bar{s}$$

$$\tilde{H}_b(w,v,t,X_{Pi}) = \frac{1}{h^4}\int_0^{\bar{t}}|\psi_D(w,\bar{s})|a_2(\bar{s},v,G)^{-1}$$

$$|K''(\upsilon_i)|\left|\frac{G(X_{Pi})-\bar{s}}{h}\right|(G_m(X_{Pi})-G(X_{Pi}))^2$$

$$I\left(\min\left\{\left|\frac{G(X_{Pi})-\bar{s}}{h}\right|,\left|\frac{G_m(X_{Pi})-\bar{s}}{h}\right|\right\}\leq 1\right)d\bar{s}$$

where υ_i is a value between $(G(X_{Pi})-\bar{s})/h$ and $(G_m(X_{Pi})-\bar{s})/h$. Now,

$$\tilde{H}_a(w,v,\bar{t},X_{Pi}) \leq O_p(1)h^{-1}\|G_m-G\|_\infty\tilde{\mathfrak{A}}(G(X_{Pi}))$$

with

$$\tilde{\mathfrak{A}}(\xi) = \frac{1}{h^2}\int_0^{t}K'\left(\frac{\xi-s}{h}\right)\left|\frac{\xi-s}{h}\right|ds$$

By the same argument as used before for $\mathfrak{A}(X_{Pi})$, we have $\tilde{\mathfrak{A}}(G(X_{Pi})) = O_p(1)$. Thus,

$$\tilde{H}_a(w, v, \bar{t}, X_{Pi}) \leq O_p(1)h^{-1}\|G_m - G\|_\infty = O_p(1)$$

By a similar argument,

$$\tilde{H}_b(w, v, \bar{t}, X_{Pi}) \leq O_p(1)h^{-2}\|G_m - G\|_\infty^2 = O_p(1)$$

Hence, by the same argument as used before for $\mathcal{B}_1^*(t, u, u)$, we find that $\sqrt{mh}\ \Omega^*(\bar{t}, u) = o_p(1)$. Finally, by the same argument as used for $\mathcal{B}_1^{**}(t, u)$, we obtain $\sqrt{mh}\ \Omega^{**}(\bar{t}, u) = o_p(1)$.

Finally, we have

$$\begin{aligned}\hat{\Lambda}(t) - \Lambda(t) &= \hat{\Lambda}_D(G_m(t)) - \Lambda_D(G(t)) \\ &= (\hat{\Lambda}_D(G(t)) - \Lambda_D(G(t))) + (\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t)))\end{aligned}$$

We have just shown that $\sqrt{mh}\ (\hat{\Lambda}_D(G(t)) - \Lambda_D(G(t)))$ converges in distribution to a mean-zero normal distribution. We now show that $\sqrt{mh}\ (\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t))) = o_p(1)$. We have

$$|\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t))| \leq \|G_m - G\|_\infty \sup_{\bar{s}\in[0,1]} \left| \int_0^\tau \hat{\psi}_D(u, \bar{s})\hat{\Lambda}_{0,D}^*(u|\bar{s})\, du \right|$$

we know that $\|G_m - G\|_\infty = O_p(m^{-1/2})$, and we have

$$\sup_{\bar{s}\in[0,1]} \left| \int_0^\tau \hat{\psi}_D(u, \bar{s})\hat{\Lambda}_{0,D}^*(u|\bar{s})\, du \right| = O_p(1)$$

so we get $\sqrt{mh}\ (\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t))) = o_p(1)$ as desired.

Additional References

Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study, *Annals of Statistics* **10**: 1100–1120.

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge.

Giné, E. and Guillou, A. (2002). Rates of strong consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **38**: 907-922.

Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Annals of Mathematical Statistics* **40**: 1187–1195.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.