Electronic Journal of Statistics
Vol. 13 (2019) 5415–5453
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1651

Fully nonparametric estimation 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
91905 Jerusalem, Israel
e-mail: david.zucker@mail.huji.ac.il

and

Malka Gorfine

Department of Statistics and Operations Research
Tel Aviv University
Ramat Gan, Israel
e-mail: gorfinem@post.tau.ac.il

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, making this paper the first to present for this data setting a fully nonparametric estimator with proven consistency. Simulation results are presented showing that the method performs well. The method is illustrated on data from a prostate cancer study.

MSC 2010 subject classifications: Primary 62N02; secondary 62G20.
Keywords and phrases: Case-control; family study; multivariate survival data; nonparametric estimator; local linear.

Received May 2019.

Contents

1. Introduction

In epidemiology, a key quantity of interest is the cumulative incidence of a given disease in a given population, i.e., the distribution of the time T to disease onset for a random member of the target population. The most direct way to obtain an estimate of this quantity is by taking a random sample from the target population and following them up for occurrences of the disease. For rare diseases, however, this approach is impractical because it would take an enormous sample size to accrue a sufficient number of disease events.

A common powerful alternative avenue for estimating the population cumulative incidence is the case-control family study, in which the cases are oversampled. Specifically, 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. We record age at disease onset or age at censoring and disease status of each proband and of one or more members of a designated set of the proband's relatives, which we call the individual's key relatives. For example, in Stanford et al.'s [12] study of prostate cancer, case probands were men diagnosed with prostate cancer, control probands were men without prostate cancer, and each proband was interviewed to obtain detailed disease history information on his key relatives, which included the proband's father, brothers, and uncles. The goals of a case-control family study are to (1) assess familial aggregation of the disease, (2) assess the effect of genetic and environmental factors on disease risk, and (3) to estimate the population cumulative incidence of the disease. This paper focuses on the last of these three goals. We will work in terms of the survival function $S(t) = P(T > t)$, which we will refer to as the marginal survival function.

The analysis of case-control family data is complicated by two factors: the case-control selection scheme used to ascertain the families (i.e., cases are oversampled) and by the within-family dependence (i.e., family members are not independent given the observed covariates). A review of semiparametric models and estimation procedures based on copula or frailty approaches for estimating the effect of genetic and environmental factors on disease risk can be found in Gorfine et al. [9].

It is of interest to provide a fully nonparametric estimator of the marginal survival function that takes into account the dependence of the failure times of family members and the case-control sampling design. By fully nonparametric estimation, we mean estimation that avoids specific assumptions about the form of the distribution or the dependence structure among failure times within a family. The estimator that first comes to mind is the Kaplan-Meier survival curve estimator based on the survival data of 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. [9] demonstrated that the bias can be either upward or downward, depending on the dependence structure, and the coverage rates of 95% confidence intervals can be as low as zero.

So far, Gorfine et al. [9] is the only published work proposing a fully nonparametric estimator for this problem. Their estimator is based on a kernel smoothing approach. Through an extensive simulation study they showed that their estimator performs well in terms of bias. The estimator of [9] is based on the median of random variables with a complicated dependence structure. Consequently, the asymptotic properties of the estimator were not derived. In addition, their bandwidth selection procedure was not specifically targeted to the estimand of interest.

In the present paper we approach the problem with a different strategy, with the following novel components: (i) A new identity (Equation (2.4) below) is derived that links the marginal survival function to the conditional survival function for a relative given the status of the proband as case or control. Using this identity, we develop a new estimator for the problem. (ii) The asymptotic properties of the proposed estimator is established. This is the first paper to present for this data setting a fully nonparametric estimator with proven consistency. (iii) The new estimator works with means rather than medians and performs well in terms of bias and much better in terms of variance than the estimator of [9]. In some scenarios the new estimator also outperforms the estimator of [9] in terms of bias. (iv) In contrast with [9], we also present a bandwidth selection procedure that is specifically targeted to the estimand of interest. (v) Although it is traditionally assumed in case-control family studies that the marginal survival distribution is the same for a given individual and all his/her key relatives, see, e.g., [3, 9, 11], and this assumption is often reasonable, our proposed procedure provides a consistent estimator of the marginal survival function without requiring this assumption. This is a novel result for the case-control family design. R code for carrying out the simulations reported in this paper and for applying the method to a data set is available at the following `Github` site:

https://github.com/david-zucker/marginal-survival

2. Preliminaries

We begin by presenting some general definitions. We then describe the data we will work with and note relevant connections to the general definitions. Finally, we present the key identity that serves as the basis for our estimation procedure.

As above, we let T denote the time to disease onset for a randomly selected member of the population and let T' be the time to onset of a randomly selected key relative. In some cases the set of key relatives consists of only one relative, for example the individual's father, while in other cases it consists of several relatives, for example the individual's brothers. It is commonly assumed that T and T' are identically distributed, see, e.g., [3, 9, 11]. This assumption is often reasonable, but for our main result we do not require it. In fact, theoretically we could replace T' with any variable W collected on the relatives that satisfies $P(T > t, W > u) \geq P(T > t)P(W > u)$ for all t and u, with strict inequality for every t on a set of u values of positive measure. Our goal is to estimate the marginal survival function $S(t) = P(T > t)$. The marginal survival distribution is assumed to be continuous with density $f(t) = -(d/dt)S(t)$. We define $S_0(u|t) = P(T' > u|T > t)$ and $S_1(u|t) = P(T' > u|T = t)$. We assume here that the joint distribution of (T, T') is the same for all relatives. This assumption can be relaxed by introducing different types of proband-relative relationships.

We now describe the data we will work with. The setup is as in Gorfine et al. [9]. We have a sample of n_1 individuals with the disease, which we refer to as the case probands. Each case proband is frequency matched by age with a disease-free control probands. The total number of control probands is $n_0 = an_1$, and the total number of probands in the study is $n = n_0+n_1$. Proband i's observation time X_{Pi} is the age at disease onset for a case proband and the age at sampling for a control proband. We define δ_{Pi} to be 1 if proband i is a case proband and 0 if proband i is a control proband. For proband i we have data on J_i relatives. We view J_i as a random variable, and let J denote the maximum number of relatives for a given proband. For each relative ij we determine whether the relative has the disease ($\delta_{Rij} = 1$) or not ($\delta_{Rij} = 0$). If the relative has the disease, we ascertain the age of disease onset. If the relative does not have the disease, he/she is regarded as censored at the latest age for which his/her disease status was known. The censoring is assumed independent of the survival time. For each relative ij, we denote by X_{Rij} the relative's age at the time of disease onset or censoring. 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 denote the maximum observation time among the probands by τ_0 and the maximum observation time among the relatives by τ. 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. As in Gorfine et al. [9], the conditional survival function for a relative in family i given the information on the corresponding proband is $S_0(\cdot|X_{Pi})$ if the proband is a control and $S_1(\cdot|X_{Pi})$ if the proband is a case.

We now develop a new representation 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 $S^*_q(u|t) = -S_q(u|t)\Lambda^*_0(u|t)$. We then have

$$P(T > t, T' > u) = \int_t^\infty P(T' > u|T = x)f(x)dx$$

which yields the following:

$$\begin{aligned}
&S(t)S_0(u|t) = \int_t^\infty S_1(u|x)f(x)dx \\
&\Rightarrow \frac{\partial}{\partial t}[S(t)S_0(u|t)] = -S_1(u|t)f(t) \\
&\Rightarrow -f(t)S_0(u|t) + S(t)S^*_0(u|t) = -S_1(u|t)f(t) \\
&\Rightarrow -\lambda(t)S_0(u|t) + S^*_0(u|t) = -S_1(u|t)\lambda(t) \\
&\Rightarrow \lambda(t)(S_0(u|t) - S_1(u|t)) = S^*_0(u|t) = -S_0(u|t)\Lambda^*_0(u|t) \\
&\Rightarrow \lambda(t)(S_0(u|t) - S_1(u|t))^2 = -S_0(u|t)(S_0(u|t) - S_1(u|t))\Lambda^*_0(u|t) \qquad (2.1)
\end{aligned}$$

Now define

$$\psi(u, t) = \left[\int_0^\tau (S_0(v|t) - S_1(v|t))^2\, dv\right]^{-1} (S_0(u|t) - S_1(u|t))S_0(u|t) \qquad (2.2)$$

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 (2.1) and rearranging gives

$$\lambda(t) = -\int_0^\tau \psi(u, t)\Lambda^*_0(u|t)\, du \qquad (2.3)$$

$$\Lambda(t) = -\int_0^t \int_0^\tau \psi(u, s)\Lambda^*_0(u|s)\, du\, ds \qquad (2.4)$$

We use the key identity (2.4) 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 by Beran [2] and examined in Dabrowska [4], and the resulting estimators were used to construct an estimator of $S(t)$. Here, in light of (2.4), 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 $N_{Rij}(t) = \delta_{Rij} I(X_{Rij} \leq t)$ and $Y_{Rij}(t) = I(X_{Rij} \geq t)$, and write $Q_i = \delta_{Pi}$. Let $\chi(q_1, q_2) = I(q_1 = q_2)$. Define (with $q = 0, 1$)

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

$$dM_{Rij}(v) = dN_{Rij}(v) - Y_{Rij}(v)\lambda_{Q_i}(v|X_{Pi})dv$$

$$M_{Ri\bullet}(v) = \sum_{j=1}^{J} M_{Rij}(v)$$

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

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

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

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 fitting where the response variable is $dN_{Ri\bullet}(v)/Y_{Ri\bullet}(v)$, the explanatory variable is $X_{Pi} - s$, and the weights are $\chi(Q_i, q)Y_{Ri\bullet}(v)K((X_{Pi} - s)/h)$. This leads to the local linear estimators

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

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

that is, for a given u,

$$\widehat{\Lambda}^*_q(u|s) = \int_0^u \frac{1}{C_q(s, v)}\left[\frac{1}{n_q h}\sum_{i=1}^{n} \chi(Q_i, q) K\left(\frac{X_{Pi} - s}{h}\right)\right.$$

$$(X_{Pi} - \bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\Bigg] \tag{3.1}$$

$$\begin{aligned}\widehat{\Lambda}_q(u|s) = &\int_0^u \mathcal{Y}_q(s,v)^{-1}\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K((X_{Pi}-t)/h)dN_{Ri\bullet}(v)\right]\\ &-\int_0^u \frac{\bar{X}_{Pq}(s,v)-s}{C_q(s,v)}\Bigg[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)\\ &\qquad (X_{Pi}-\bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\Bigg]\end{aligned} \tag{3.2}$$

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 (2.3) and (2.4) 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

$$\begin{aligned}\mathcal{A}(t) &= \int_0^t \int_0^\tau \psi(u,s)[\widehat{\Lambda}_0^*(u|s) - \Lambda_0^*(u|s)]\,du\,ds\\ \mathcal{A}^*(t) &= \int_0^t \int_0^\tau [\widehat{\psi}(u,s) - \psi(u,s))]\Lambda_0^*(u|s)\,du\,ds\\ \mathcal{A}^{**}(t) &= \int_0^t \int_0^\tau [\widehat{\psi}(u,s) - \psi(u,s)][\widehat{\Lambda}_0^*(u|s) - \Lambda_0^*(u|s)]\,du\,ds\end{aligned}$$

We will provide a detailed asymptotic analysis of $\mathcal{A}(t)$ and an outline of the proof that $\mathcal{A}^*(t)$ converges in probability to zero at a faster rate than $\mathcal{A}(t)$. A similar argument can be used to show that $\mathcal{A}^{**}(t)$ is negligible in comparison with the other two terms. The result that $\mathcal{A}^*(t)$ converges in probability to zero at a faster rate than $\mathcal{A}(t)$ can be explained by the fact that in nonparametric regression, the error in estimating the derivative of the regression function converges to zero at a slower rate than the error in estimating the regression function itself; see Fan and Gijbels [7] (Theorem 3.1, page 62). Details regarding $\mathcal{A}^*(t)$ can be found in Section 8.4.

In this section, Section 8, and 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. Recall that $n_0 = an_1$.

Define $\Gamma(s,v) = h^{-2}C_0(s,v)$. We then have

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

and thus

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

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

$$E\left[\int_0^u P(v,X_{Pi})dM_{Ri\bullet}(v)\right] = 0 \tag{4.1}$$

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 Section 8 we establish the asymptotic properties of our estimator. We list below the assumptions we make in obtaining this result.

Assumptions

1. Relative ij's event time and censoring time are conditionally independent given X_{Pi} and δ_{Pi}.
2. 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.
3. The bandwidth $h = h_m$ is given by $h_m = \alpha_m m^{-\nu}$, where $1/4 < \nu < 1/3$ and $\alpha_m \to \alpha > 0$ as $m \to \infty$.
4. The random variable X_{Pi} has a density $g(x)$ satisfying
$$g_{min} = \inf_{x\in[0,\tau_0]} g(x) \ > \ 0.$$
5. Defining $y_q(s,v) = E[Y_{Ri\bullet}(v)|X_{Pi}=s, \delta_{Pi}=q]$, we have
$$y_q^{(min)} = \inf_{s\in[0,\tau_0]} y_q(s,\tau) \ > \ 0.$$

6. Defining $\varphi(s,v) = g(s)y_0(s,v)$, 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 4 and 5 imply that $\varphi_{min} = \inf_{s\in[0,\tau_0],v\in[0,\tau]} \varphi(s,v) > 0$.
7. The first and second partial derivatives of $\psi(v|s)$ with respect to s exist and are bounded uniformly over s and v.
8. The first three partial derivatives of $\lambda_0(v|s)$ with respect to s exist and are bounded uniformly over s and v.

Assumption 1 is an independent censoring assumption which is standard in survival analysis. Assumption 2 is a condition that can always be arranged to be satisfied, since the kernel K is chosen by the user and there are many kernel functions that satisfy the requisite conditions. Assumption 3 is a type of assumption always made in kernel estimation, and ensures that the bandwidth goes to 0 fast enough, as the sample size increases, to make the bias term ($\mathcal{A}_1(t)$ in our case) converge to 0 but not so fast that the random error term ($\mathcal{A}_2(t)$) gets out of control. Assumption 4 amounts to saying that the survival time and censoring time have a continuous distribution on $[0,\tau_0]$ and that the proband observation times adequately cover the entire interval $[0,\tau_0]$, which is a reasonable assumption in most applications. No assumption is made on the censoring times of the relatives. Assumption 5 is a type of assumption commonly made in survival analysis, and ensures that the fraction of relatives at risk remains positive throughout the interval $[0,\tau]$.

Under the above assumptions, we show that $(mh)^{1/2}\mathcal{A}_1(t) \to 0$ in probability as $m \to \infty$ and that

$$\mathcal{A}_2(t) = \mathcal{B}(t) + o_p((mh)^{-1/2}) \tag{4.2}$$

where

$$\mathcal{B}(t) = \frac{1}{m}\sum_{i=1}^{m} \zeta_i(t)$$

with

$$\zeta_i(t) = \frac{1}{h}\int_0^t \int_0^\tau \int_0^u \psi(u,s)\gamma(s,v)^{-1}\left(\frac{X_{Pi}-s}{h}\right) \left[h^{-1}K\left(\frac{X_{Pi}-s}{h}\right)\right] dM_{Ri\bullet}(v)\,du\,ds \tag{4.3}$$

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

The proof of the result for $\widehat{\Lambda}(t)$ is given in Section 8. Here we give a brief sketch. A basic background result is a result on the behavior of

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

where $Z_k(r) = r^k K(r)$. The behavior of $E[A_k(s,v,h)]$ is established using the usual change of variable + Taylor expansion argument used in the analysis of kernel estimators. The behavior of $\sup_{s,v} |A_k(s,v,h) - E[A_k(s,v,h)]|$ is established using Corollary 2.2 of Giné and Guillou [8], a result concerning empirical processes involving kernel terms. Given this background result, we proceed to the analysis of $\mathcal{A}_1$ and $\mathcal{A}_2$.

The main steps in the proof of convergence in probability of $(mh)^{1/2}\mathcal{A}_1(t)$ to 0 is a second-order Taylor expansion of $\lambda_0(v|X_{Pi})$ around s ((8.1) below), the resulting representation of $\lambda_0(v|X_{Pi}) - (X_{Pi} - \bar{X}_{P0}(s,v))\lambda_0^*(v|s)$ in (8.2) below, and the observation that the first-order term in the representation (8.2) drops out because of the identity (8.3). The results on $A_k(s,v,h)$ are used to deal with the remaining terms.

Regarding (4.2), Equation (4.1) implies that $E[\zeta_i(t)] = 0$. Thus, $\mathcal{B}(t)$ 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 an appropriate Lindeberg condition. The remainder term in (4.2) is dealt with by a somewhat lengthy technical argument, in which a key role is played by the martingale result in Theorem I.1 of Anderson and Gill [1].

The result for $\widehat{S}(t)$ follows immediately from the result for $\widehat{\Lambda}(t)$ and the delta method.

5. Practical implementation details

In general, kernel estimators suffer from areas where the variable being smoothed over has low density, and indeed, we encountered this problem in preliminary work, particularly with survival distributions with low event rate in the initial part of the time line. 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. [5], 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 appendix, 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 $P(T > t, T' > u) \geq P(T > t)P(T' > u) = S(t)S(u)$ for all t and u, i.e., $S_0(u|t) \geq S(u)$ for all t and u. This condition implies that for any t

$$P(T' > u | T \leq t) \leq S(u) \leq P(T' > u | T > t) \tag{5.1}$$

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 (which happens in some applications) or extremely high (which rarely happens in practice). Given the consistency of $\widehat{S}(t)$, if the inequalities in (5.1) are strict, then for large sample sizes the modification no longer comes into play. The inequalities in (5.1) are strict if the following mild condition holds: for each u there exists a set $\mathcal{T}(u)$ such that $P(X_{Pi} \in \mathcal{T}(u)) > 0$ and

$$\inf_{t \in \mathcal{T}(u)} P(T > t, T' > u) - S(t)S(u) > 0$$

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 $\widehat{S}_{Q_i}(u|X_{Pi})$ 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

$$\begin{aligned}
\bar{S}(t; h) &= \frac{1}{B} \sum_{b=1}^{B} \widehat{S}(t; h, b) \\
V(t; h) &= \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{S}(t; h, b) - \bar{S}(t; h))^2 \\
\text{MSE}_{est}(t, h) &= (\bar{S}(t; h) - \widehat{S}(t; h))^2 + V(t; h) \\
\text{IMSE}_{est}(h) &= \int_0^\tau \text{MSE}_{est}(t, h) dt
\end{aligned}$$

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 evaluated 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 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. These choices yielded good performance in our simulations; in applying the method to a single dataset the user may wish to use larger values.

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 Figures 1–4. Figures 1 and 3 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 smaller. Figures 2 and 4 summarize the pointwise 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 a small number of observed events. Similar results were obtained with 1000 case probands.

7. Example

In this section we illustrate our method by re-analyzing the data presented in Gorfine et al. from a population-based case-control family study of early onset prostate cancer reported by Stanford et al. [12]. Briefly, case participants were identified from the Seattle-Puget Sound Surveillance, Epidemiology, and

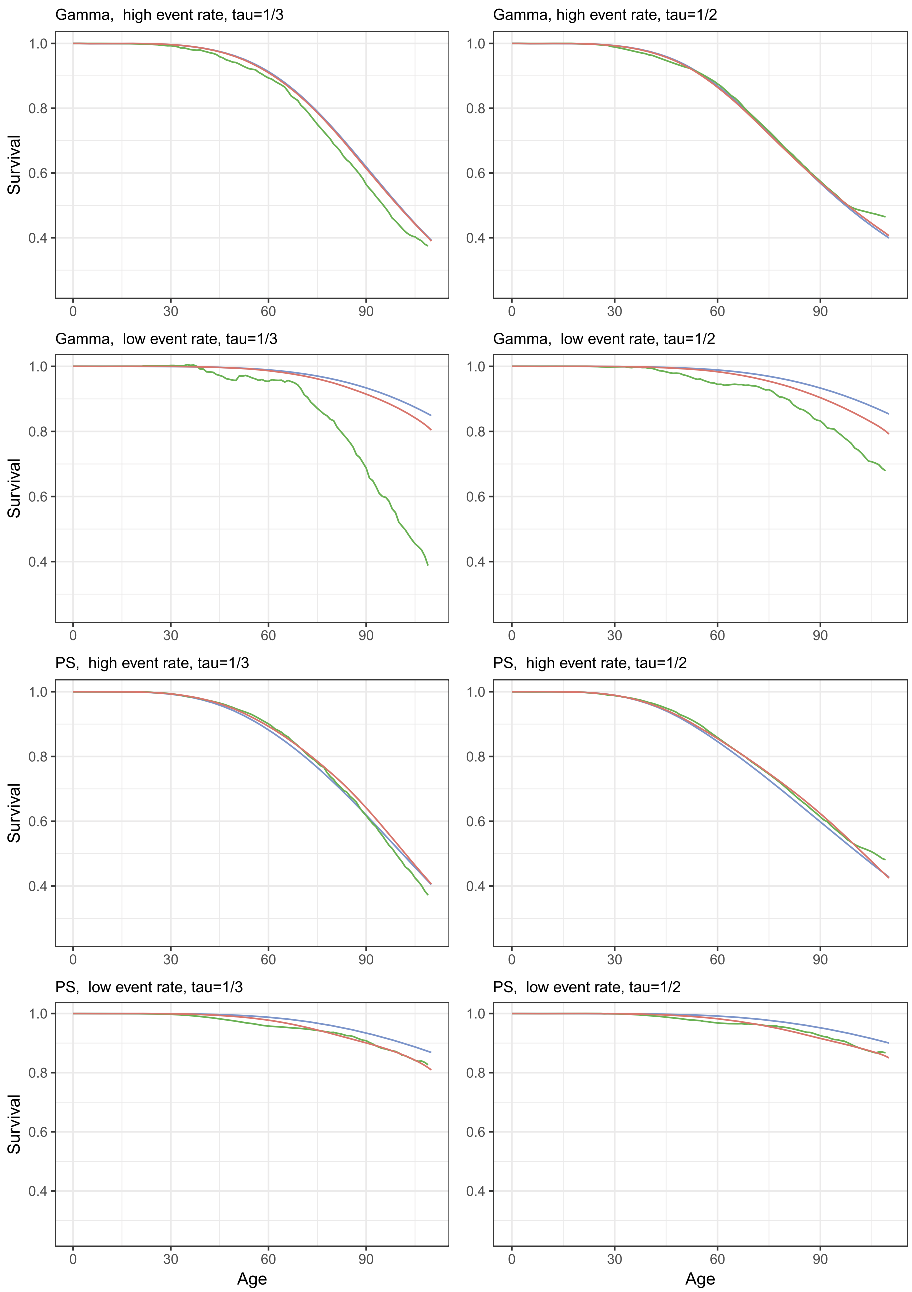

FIG 1. *Simulation results, one relative for each proband: the true survival curve (blue); Gorfine et al. estimator (green); and the proposed estimators (red).*

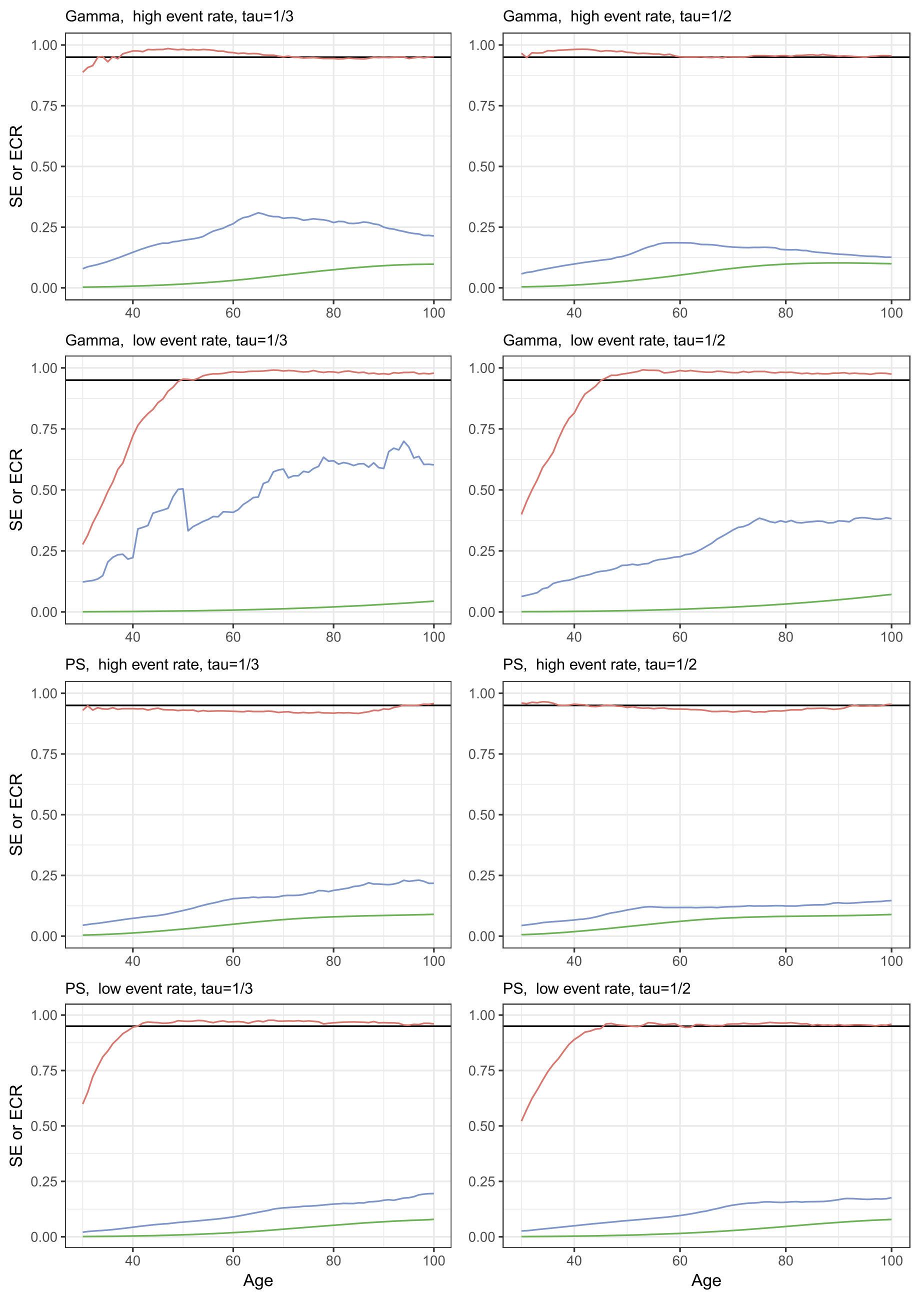

Fig 2. *Simulation results, one relative for each proband: the empirical standard errors of Gorfine et al. (blue) and the proposed estimator (green); and point-wise precentile-bootstrap 95% confidence interval coverage rates of the proposed estimator. The black horizontal line at 0.95 serves as a reference.*

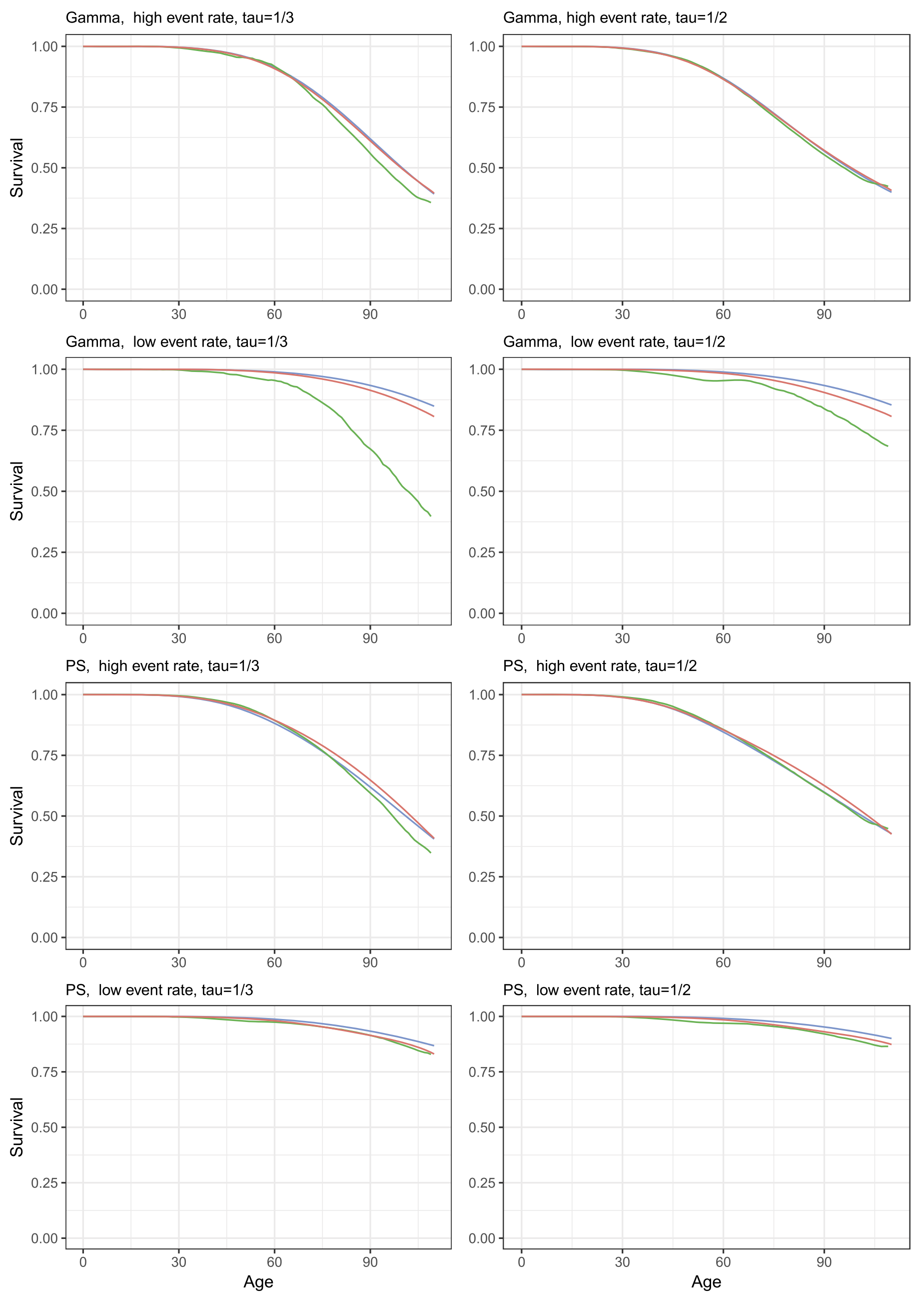

FIG 3. *Simulation results, four relatives for each proband: the true survival curve (blue); Gorfine et al. estimator (green); and the proposed estimators (red).*

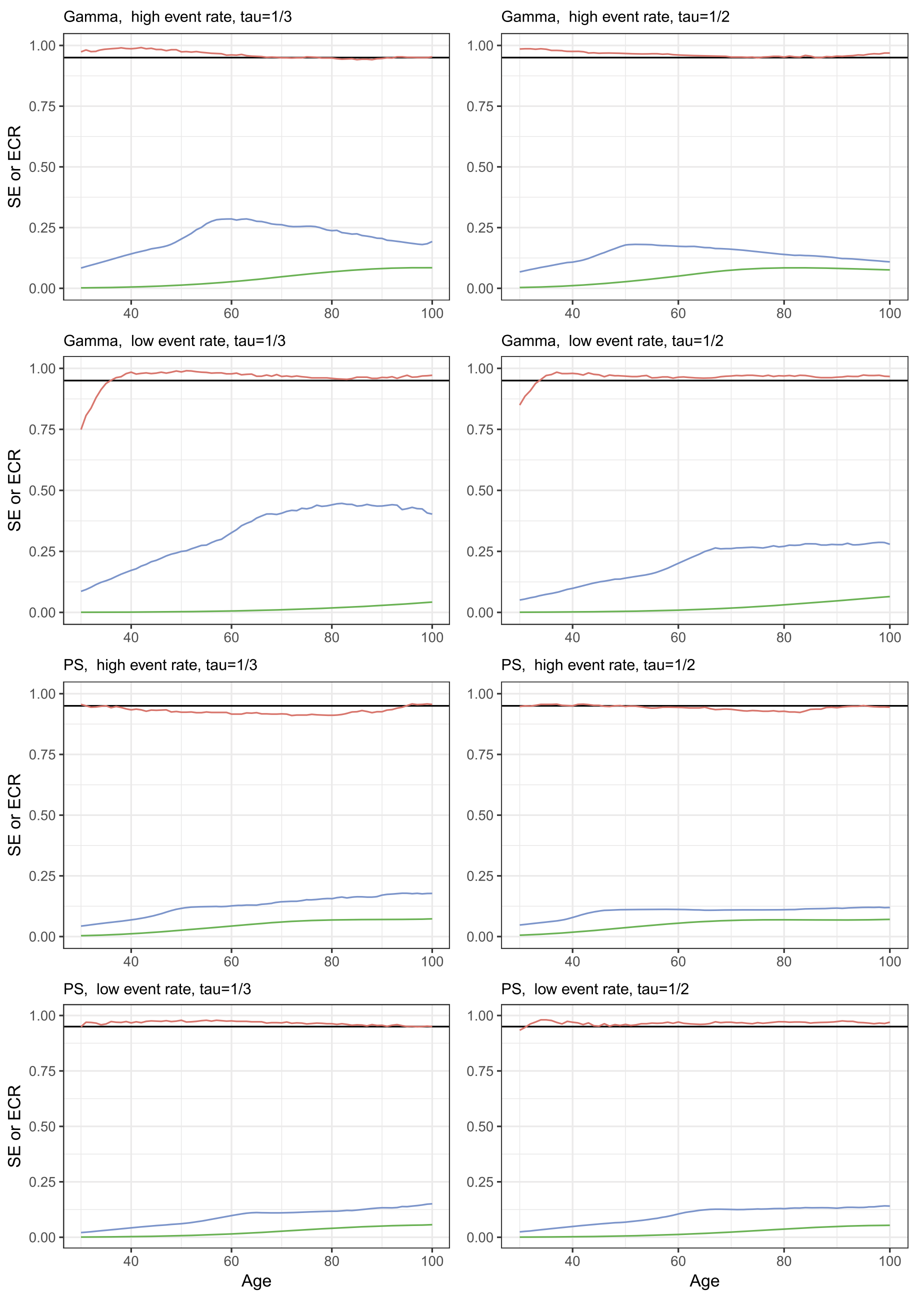

Fig 4. *Simulation results, four relatives for each proband: the empirical standard errors of Gorfine et al. (blue) and the proposed estimator (green); and point-wise precentile-bootstrap 95% confidence interval coverage rates of the proposed estimator. The black horizontal line at 0.95 serves as a reference.*

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

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 pointwise 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.

8. Proofs

We present here the proofs of our asymptotic distribution results. 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.

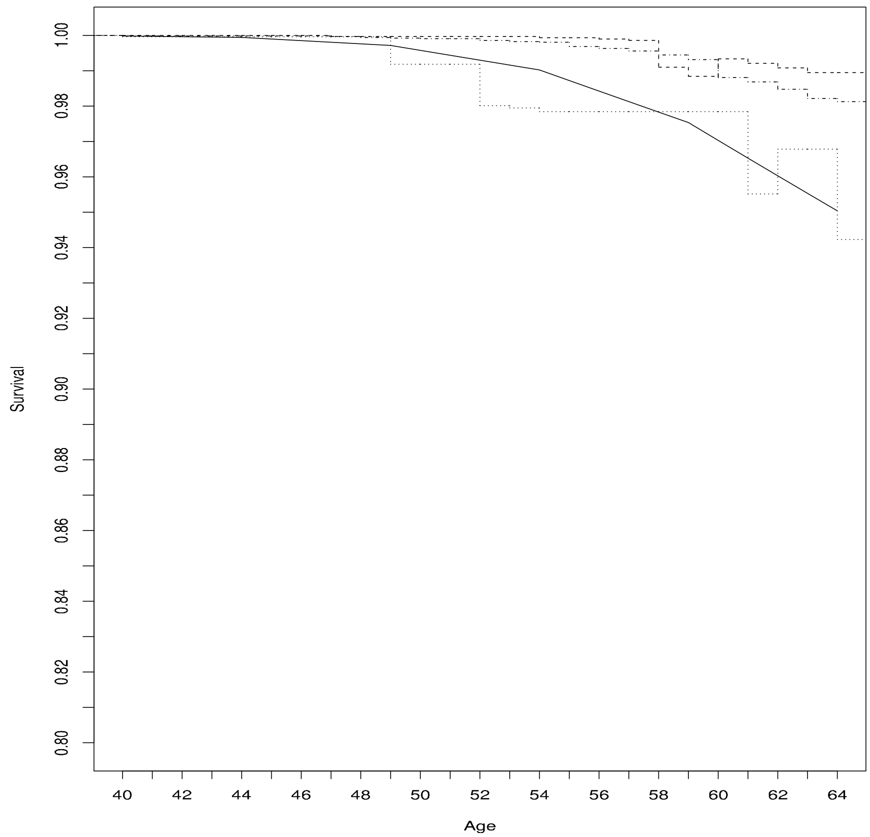

FIG 5. *Prostate cancer case-control family data: the naive Kaplan-Meier estimator based on the relatives' data (dot-dashed line), Gorfine et al.'s estimator (dotted line), the SEER survival curve (solid line), and the proposed estimator (dashed line).*

8.1. Preliminaries

Define

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

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

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

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

We now present two lemmas.

Lemma 8.1. *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

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

Proof. The proof of this lemma is given in the appendix. □

Lemma 8.2. *We have*

$$\sup_{s\in\mathcal{I},v\in[0,\tau]} |\bar{X}_{P0}(s,v) - s| = O_p(h^2)$$
$$\sup_{s\in\mathcal{I},v\in[0,\tau]} |\Gamma(s,v) - a_2(s,v,h)| = O_p(m^{-(1-\nu)/2} (\log m)^{1/2})$$
$$\sup_{s\in\mathcal{U},v\in[0,\tau]} \left| \bar{X}_{P0}(s,v) - s - \left(\frac{a_1(s,v,h)}{a_0(s,v,h)} \right) h \right| = O_p(h^2)$$
$$\sup_{s\in\mathcal{U},v\in[0,\tau]} |\Gamma(s,v) - a_2(s,v,h)| = O_p(h)$$

Proof. Simple algebra yields

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

The result now follows immediately from Lemma 8.1. □

8.2. 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

$$\lambda_0(v|X_{Pi}) = \lambda_0(v|s) + \lambda_0^*(v|s)(X_{Pi}-s) \\ + \tfrac{1}{2}\lambda_0^{**}(v|s)(X_{Pi}-s)^2 + \mathcal{R}(s,v,X_{Pi}) \tag{8.1}$$

in which $|\mathcal{R}(s,v,x)| \le \mathcal{R}^*|s-x|^3$, 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

$$\lambda_0(v|X_{Pi}) - (X_{Pi} - \bar{X}_{P0}(s,v))\lambda_0^*(v|s) \\ = \left[\lambda_0(v|s) + \lambda^*(v|s)(\bar{X}_{P0}(s,v)-s)\right] \\ + \tfrac{1}{2}\lambda_0^{**}(v|s)(X_{Pi}-s)^2 + \mathcal{R}(s,v,X_{Pi}) \tag{8.2}$$

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

$$\sum_{i=1}^m Y_{Ri\bullet}(v) K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi} - \bar{X}_{P0}(s,v)) = 0 \tag{8.3}$$

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

$$\mathcal{S}_1(s,v) = \tfrac{1}{2}\lambda_0^{**}(v|s)\left[A_3(s,v)h - (\bar{X}_{P0}(s,v)-s)A_2(s,v)\right]$$

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

We have

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

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 1 and 2 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 $\nu > 1/4$.

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

We begin with some notation. Let $\rho(s,u,v) = \psi(u|s)/\varphi(s,v)$ and define

$$H(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \frac{\rho(s,u,v)}{\eta_2(s,h)} K\left(\frac{\xi-s}{h}\right)\left(\frac{\xi-s}{h}\right) ds$$

$$H_1(u,v,t,\xi) = \frac{1}{h^2}\int_0^t \psi(u|s)(\Gamma(s,v)^{-1} - a_2(s,v,h)^{-1}) \\ K\left(\frac{\xi-s}{h}\right)\left(\frac{\xi-s}{h}\right) ds$$

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

The quantity $\zeta_i(t)$ in (4.3) can be written as

$$\zeta_i(t) = \int_0^\tau \int_0^u H(u,v,t,X_{Pi}) dM_{Ri\bullet}(v) du$$

In addition, define

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

The remainder term in (4.2) can then be written as $\mathcal{B}_1(t) + \mathcal{B}_2(t)$, where

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

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 $o_p(1)$.

<u>Analysis of $\mathrm{Var}(\zeta_i(t))$</u>

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

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

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

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

Proof of Lindeberg Condition

Define

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

We need to show that

$$s_m(t)^{-2}\sum_{i=1}^m E\left[\zeta_i(t)^2 I\left(\left|\frac{\zeta_i(t)}{s_m(t)}\right| > \epsilon\right)\right] \to 0 \quad \text{for all } \epsilon > 0 \tag{8.4}$$

We have

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

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 (8.4) is satisfied.

Analysis of $\mathcal{B}_1(t)$ and $\mathcal{B}_2(t)$

We show here that $(mh)^{1/2}\ \mathcal{B}_1(t) = o_p(1)$; the argument for $(mh)^{1/2}\ \mathcal{B}_2(t)$ is similar. Define $\mathcal{F}_0 = \sigma(X_{Pi}, \delta_{Pi}, J_i; i = 1, \ldots, m)$. 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

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

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

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

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

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

with

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

We can write $(mh)^{1/2}\ \mathcal{B}_1(t, u) = \mathcal{B}_1^*(t, u, u) + \mathcal{B}_1^{**}(t, u)$, where

$$
\begin{aligned}
\mathcal{B}_1^*(t, u, w) &= (mh)^{1/2}\left[\frac{1}{m}\sum_{i=1}^n \int_0^u H_1(w, v, t, X_{Pi})d\tilde{M}_{Ri\bullet}(v)\right] \\
\mathcal{B}_1^{**}(t, u) &= (mh)^{1/2}\left[\frac{1}{m}\sum_{i=1}^n \int_0^u H_1(w, v, t, X_{Pi})(dM_{Ri\bullet}(v) - d\tilde{M}_{Ri\bullet}(v))\right] \\
&= (mh)^{1/2}\left[\frac{1}{m}\sum_{i=1}^n \int_0^u H_1(w, v, t, X_{Pi})(\tilde{\lambda}_i(v) - Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi}))dv\right]
\end{aligned}
$$

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 $\mathcal{F}_u$, with predictable variation process given by

$$\langle \mathcal{B}_1^*(t,\cdot,w), \mathcal{B}_1^*(t,\cdot,w)\rangle(u) = h\int_0^u \left[\frac{1}{m}\sum_{i=1}^n H_1(w,v,t,X_{Pi})^2\tilde{\lambda}_i(v)\right] dv \tag{8.5}$$

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

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

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

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 $\mathfrak{A}(\xi) \le \mathfrak{A}'(\xi)$ with

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

and

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

Thus $\mathfrak{A}(X_{Pi}) = O_p(1)$. Hence, using the result of Lemma 2, 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 2, we find that $|H_{1b}(w,v,t,X_{Pi})| = O_p(1)$. Accordingly, the term in brackets in (8.5) is $O_p(1)$. It follows from Lenglart's inequality (see, e.g., Andersen and Gill [1], 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

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

with

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

Now,

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

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

8.4. Analysis of $\mathcal{A}^*(t)$

The term $\mathcal{A}^*(t)$ is

$$\mathcal{A}^*(t) = \int_0^t \int_0^\tau [\widehat{\psi}(u,s) - \psi(u,s))]\Lambda_0^*(u|s)\, du\, ds$$

where

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

Defining

$$\mathcal{Z}(t) = \int_0^\tau (S_0(v|t) - S_1(v|t))^2\, dv$$

we can write

$$\psi(u,t) = \mathcal{Z}(t)^{-1}(S_0(u|t) - S_1(u|t))S_0(u|t)$$

The estimate of $S_q(u|s)$ (for $q = 0, 1$) is given by $\widehat{S}_q(u|s) = \exp(-\widehat{\Lambda}_q(u|s))$ with

$$\begin{aligned}\widehat{\Lambda}_q(u|s) = \int_0^u \mathcal{Y}_q(s,v)^{-1} &\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K((X_{Pi}-s)/h)dN_{Ri\bullet}(v)\right] \\ &- \int_0^u \frac{\bar{X}_{Pq}(s,v) - s}{C_q(s,v)}\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)\right. \\ &\left.(X_{Pi} - \bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\right] \qquad (8.6)\end{aligned}$$

Let $\widehat{\mathcal{Z}}(t)$ denote the estimate of $\mathcal{Z}(t)$ obtained by plugging in these estimates.

We can write

$$\begin{aligned}\widehat{\psi}(u,t) - \psi(u,t) &= \widehat{\mathcal{Z}}(t)^{-1}(\widehat{S}_0(u|t) - \widehat{S}_1(u|t))\widehat{S}_0(u|t) \\ &\quad - \mathcal{Z}(t)^{-1}(S_0(u|t) - S_1(u|t))S_0(u|t) \\ &= -(S_0(u|t) - S_1(u|t))S_0(u|t)\mathcal{Z}(t)^{-2}(\widehat{\mathcal{Z}}(t) - \mathcal{Z}(t)) \\ &\quad + \mathcal{Z}(t)^{-1}(2S_0(u|t) - S_1(u|t))(\widehat{S}_0(u|t) - S_0(u|t)) \\ &\quad - \mathcal{Z}(t)^{-1}S_0(u|t)(\widehat{S}_1(u|t) - S_1(u|t)) \\ &\quad + \text{higher order terms} \qquad (8.7) \\ &= -(S_0(u|t) - S_1(u|t))S_0(u|t)\mathcal{Z}(t)^{-2}(\widehat{\mathcal{Z}}(t) - \mathcal{Z}(t)) \\ &\quad - \mathcal{Z}(t)^{-1}(2S_0(u|t) - S_1(u|t))S_0(u|t)(\widehat{\Lambda}_0(u|t) - \Lambda_0(u|t)) \\ &\quad + \mathcal{Z}(t)^{-1}S_0(u|t)S_1(u|t)(\widehat{\Lambda}_1(u|t) - \Lambda_1(u|t)) \\ &\quad + \text{higher order terms} \qquad (8.8)\end{aligned}$$

with

$$\widehat{\mathcal{Z}}(t) - \mathcal{Z}(t) = -2\int_0^\tau (S_0(u|t) - S_1(u|t)S_0(u|t)(\widehat{\Lambda}_0(u|t) - \Lambda_0(u|t))du$$

$$+ 2\int_0^\tau (S_0(u|t) - S_1(u|t)S_1(u|t)(\widehat{\Lambda}_1(u|t) - \Lambda_1(u|t))du$$
$$+ \text{higher order terms} \qquad (8.9)$$

In the above, we have used the Taylor expansion

$$\exp(-\widehat{\Lambda}_q(u|t)) - \exp(-\Lambda_q(u|t)) = -\exp(-\Lambda_q(u|t))(\widehat{\Lambda}_q(u|t) - \Lambda_q(u|t)) + O((\widehat{\Lambda}_q(u|t) - \Lambda_q(u|t))^2)$$

and the higher order terms in (8.8) and (8.9) are terms involving the quantity $(\widehat{\Lambda}_q(u|t) - \Lambda_q(u|t))^2$, which are negligible in comparison with the terms involving the quantity $\widehat{\Lambda}_q(u|t) - \Lambda_q(u|t)$.

Thus, the leading terms in $\mathcal{A}^*(t)$ are of the form

$$\begin{aligned}
\bar{\mathcal{A}}^*(t) &= \int_0^t \int_0^\tau \mathcal{W}(u,s)\,(\widehat{\Lambda}_q(u|t) - \Lambda_q(u|t))\,du\,ds \\
&= \int_0^t \int_0^\tau \mathcal{W}(u,s)\left\{\int_0^u \mathcal{Y}_q(s,v)^{-1}\right. \\
&\qquad \left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K((X_{Pi}-t)/h)dN_{Ri\bullet}(v)\right] \\
&\quad - \int_0^u \frac{\bar{X}_{Pq}(s,v)-s}{C_q(s,v)}\left[\frac{1}{n_q h}\sum_{i=1}^n \chi(Q_i,q)K\left(\frac{X_{Pi}-s}{h}\right)\right. \\
&\qquad \left.\left.(X_{Pi}-\bar{X}_{Pq}(s,v))dN_{Ri\bullet}(v)\right]\right\}du\,ds \\
&\quad - \int_0^t \int_0^\tau \mathcal{W}(u,s)\int_0^u \lambda_q(v)\,dv\,du\,ds
\end{aligned}$$

Since in our previous asymptotic analysis we focused on the case of $q = 0$, using notation pertaining to that case, we will continue to do so here; exactly the same analysis applies to the case of $q = 1$.

Based on (8.6), we can express $\bar{\mathcal{A}}^*(t)$ as the sum of the following four terms

$$\begin{aligned}
\bar{\mathcal{A}}_1^*(t) &= \int_0^t \int_0^\tau \mathcal{W}(u,s)\int_0^u \mathcal{Y}_0(s,v)^{-1} \\
&\qquad \left[\frac{1}{mh}\sum_{i=1}^n K((X_{Pi}-s)/h)Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi})\right]dv\,du\,ds \\
&\quad - \int_0^t \int_0^\tau \mathcal{W}(u,s)\int_0^u \lambda_0(v)\,dv\,du\,ds \\
\bar{\mathcal{A}}_2^*(t) &= \int_0^t \int_0^\tau \mathcal{W}(u,s)\int_0^u \mathcal{Y}_0(s,v)^{-1} \\
&\qquad \left[\frac{1}{mh}\sum_{i=1}^n K((X_{Pi}-s)/h)dM_{Ri\bullet}(v)\right]du\,ds
\end{aligned}$$

$$\bar{\mathcal{A}}_3^*(t) = h^{-2}\int_0^t\int_0^\tau \mathcal{W}(u,s)\int_0^u (\bar{X}_{P0}(s,v)-s)\Gamma(s,v)^{-1}$$
$$\left[\frac{1}{mh}\sum_{i=1}^n K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi})\right] dv\,du\,ds$$
$$\bar{\mathcal{A}}_4^*(t) = h^{-2}\int_0^t\int_0^\tau \mathcal{W}(u,s)\int_0^u (\bar{X}_{P0}(s,v)-s)\Gamma(s,v)^{-1}$$
$$\left[\frac{1}{mh}\sum_{i=1}^n K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-\bar{X}_{P0}(s,v))dM_{Ri\bullet}(v)\right] du\,ds$$

The terms $\bar{\mathcal{A}}_1^*(t)$ and $\bar{\mathcal{A}}_3^*(t)$ can be shown to be $O_p(h^2)$ using a Taylor expansion argument similar to that used to analyze $\mathcal{A}_1(t)$. The term $\bar{\mathcal{A}}_2^*(t)$ can be written as

$$\bar{\mathcal{A}}_2^*(t) = \frac{1}{m}\sum_{i=1}^n\int_{-1}^1\int_0^\tau \mathcal{W}(u,X_{Pi}-hr)I\left(r\in\left[\frac{X_{Pi}-t}{h},\frac{X_{Pi}}{h}\right]\right)K(r)$$
$$\int_0^u \frac{Y_{Ri\bullet}(v)\lambda_0(v|X_{Pi})dM_{Ri\bullet}(v)}{\mathcal{Y}_0(X_{Pi}-hr,v)}\,du\,dr$$
$$= \bar{\mathcal{A}}_{21}^*(t) + \bar{\mathcal{A}}_{22}^*(t)$$

with

$$\bar{\mathcal{A}}_{21}^*(t) = \frac{1}{m}\sum_{i=1}^n\int_{-1}^1\int_0^\tau \mathcal{W}(u,X_{Pi}-hr)I\left(r\in\left[\frac{X_{Pi}-t}{h},\frac{X_{Pi}}{h}\right]\right)K(r)$$
$$\int_0^u y_0(X_{Pi}-hr,v)^{-1}dM_{Ri\bullet}(v)\,du\,dr$$
$$\bar{\mathcal{A}}_{22}^*(t) = \frac{1}{m}\sum_{i=1}^n\int_{-1}^1\int_0^\tau \mathcal{W}(u,X_{Pi}-hr)I\left(r\in\left[\frac{X_{Pi}-t}{h},\frac{X_{Pi}}{h}\right]\right)K(r)$$
$$\int_0^u [\mathcal{Y}_0(X_{Pi}-hr,v)^{-1}-y_0(X_{Pi}-hr,v)^{-1}]dM_{Ri\bullet}(v)\,du\,dr$$

Now, $\bar{\mathcal{A}}_{21}^*(t)$ is the mean of i.i.d. bounded mean-zero random variables and thus is $O_p(m^{-1/2})$ by the central limit theorem. Using arguments along the lines used for $\mathcal{B}_1(t)$, $\bar{\mathcal{A}}_{22}^*(t)$ can be shown to be negligible compared with $\bar{\mathcal{A}}_{21}^*(t)$. Next, we can write $\bar{\mathcal{A}}_4^*(t) = \bar{\mathcal{A}}_{41}^*(t) + \bar{\mathcal{A}}_{42}^*(t)$ with

$$\bar{\mathcal{A}}_{41}^*(t) = h^{-2}\int_0^t\int_0^\tau \mathcal{W}(u,s)\int_0^u (\bar{X}_{P0}(s,v)-s)\Gamma(s,v)^{-1}$$
$$\left[\frac{1}{mh}\sum_{i=1}^n K\left(\frac{X_{Pi}-s}{h}\right)(X_{Pi}-s)dM_{Ri\bullet}(v)\right] du\,ds$$
$$= h^{-1}\int_0^t\int_0^\tau \mathcal{W}(u,s)\int_0^u (\bar{X}_{P0}(s,v)-s)\Gamma(s,v)^{-1}$$

$$\left[\frac{1}{mh}\sum_{i=1}^{n} K\left(\frac{X_{Pi}-s}{h}\right)\left(\frac{X_{Pi}-s}{h}\right) dM_{Ri\bullet}(v)\right] du\, ds$$

$$= \frac{1}{m}\sum_{i=1}^{n}\int_{-1}^{1}\int_{0}^{\tau} \mathcal{W}(u, X_{Pi}-hr) I\left(r \in \left[\frac{X_{Pi}-t}{h}, \frac{X_{Pi}}{h}\right]\right) rK(r)$$

$$\int_0^u h^{-1}(\bar{X}_{P0}(X_{Pi}-hr, v) - (X_{Pi}-hr))\Gamma(X_{Pi}-hr), v)^{-1}$$

$$dM_{Ri\bullet}(v)\, du\, dr$$

and

$$\bar{\mathcal{A}}^*_{42}(t) = h^{-2}\int_0^t\int_0^\tau \mathcal{W}(u,s)\int_0^u (\bar{X}_{P0}(s,v)-s)^2\Gamma(s,v)^{-1}$$

$$\left[\frac{1}{mh}\sum_{i=1}^{n} K\left(\frac{X_{Pi}-s}{h}\right) dM_{Ri\bullet}(v)\right] du\, ds$$

$$= \frac{1}{m}\sum_{i=1}^{n}\int_{-1}^{1}\int_{0}^{\tau} \mathcal{W}(u, X_{Pi}-hr) I\left(r \in \left[\frac{X_{Pi}-t}{h}, \frac{X_{Pi}}{h}\right]\right) K(r)$$

$$\int_0^u h^{-2}(\bar{X}_{P0}(X_{Pi}-hr, v) - (X_{Pi}-hr))^2$$

$$\Gamma(X_{Pi}-hr, v)^{-1} dM_{Ri\bullet}(v)\, du\, dr$$

Using arguments along the lines used for $\mathcal{B}_1(t)$, these two terms can be shown to be negligible compared with $\bar{\mathcal{A}}^*_{21}(t)$.

Appendix A: Appendix

A.1. Proof of Lemma 8.1

We begin with an expanded statement of the lemma and then proceed with the proof.

Lemma 8.1. *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

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

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

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

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 $|R(s+hr,v)| \le \sup_{s,v} |\ddot\varphi(s,v)| < \infty$. 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 [8], a result concerning empirical processes involving kernel terms. For $\bar{x} = (x_1, \ldots, x_J) \in \mathbb{R}^J$, define

$$L_v(\bar{x}) = \sum_{j=1}^{J} I(x_j \geq v)$$
$$\Upsilon_{s,v,h}(x_0, \bar{x}) = Z_k\left(\frac{x_0 - s}{h}\right) L_v(\bar{x})$$

We can then write

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

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 Vapnik-Chervonenkis (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 [6], 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

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

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

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

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

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

A.2. Extension to time transformation 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

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

where

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

with b_m chosen as described in Schuster [10] 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

$$\begin{aligned}
&\Lambda_{\mathcal{D}}(\bar{s}) = \Lambda(\mathcal{D}(\bar{s})) \\
&S_{0,\mathcal{D}}(u|\bar{s}) = S_0(u|\mathcal{D}(\bar{s})) \\
&\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \Lambda_0(u|\mathcal{D}(\bar{s})) \\
&\Lambda^*_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial \bar{s}}\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \Lambda^*_0(u|\mathcal{D}(\bar{s}))\mathcal{D}'(\bar{s})
\end{aligned}$$

$$\begin{aligned}
&\lambda_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial u}\Lambda_{0,\mathcal{D}}(u|\bar{s}) = \lambda_0(u|\mathcal{D}(\bar{s})) \\
&\lambda^*_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial}{\partial \bar{s}}\lambda_{0,\mathcal{D}}(u|\bar{s}) \\
&\lambda^{**}_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial^2}{\partial \bar{s}^2}\lambda_{0,\mathcal{D}}(u|\bar{s}) \\
&\lambda^{***}_{0,\mathcal{D}}(u|\bar{s}) = \frac{\partial^3}{\partial \bar{s}^3}\lambda_{0,\mathcal{D}}(u|\bar{s}) \\
&\psi_{\mathcal{D}}(u,\bar{s}) = \psi(u,\mathcal{D}(\bar{s}))
\end{aligned}$$

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

$$\begin{aligned}
A_k(\bar{s},v,h,\mathcal{G}) &= \frac{1}{mh}\sum_{i=1}^m Y_{Ri\bullet}(v) Z_k\left(\frac{\mathcal{G}(X_{Pi})-\bar{s}}{h}\right) \\
\mathcal{M}(\bar{s},v,\mathcal{G}) &= A_1(\bar{s},v,h,\mathcal{G})/A_0(\bar{s},v,h,\mathcal{G}) \\
\Gamma(\bar{s},v,\mathcal{G}) &= A_2(\bar{s},v,h,\mathcal{G}) - A_0(\bar{s},v,h,\mathcal{G})^{-1}(A_1(\bar{s},v,h,\mathcal{G}))^2
\end{aligned}$$

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

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

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 Section 8 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 $(mh)^{1/2}\,\mathcal{A}_{1a}(\bar{t},G)$ by the same argument as in Section 8 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

$$|\Psi_1|\le O_p(1)\left[\frac{1}{m}\sum_{i=1}^m\int_0^{\bar{t}}\left|K\left(\frac{G_m(X_{Pi})-\bar{s}}{h}\right)-K\left(\frac{G(X_{Pi})-\bar{s}}{h}\right)\right|\right.$$

$$\left.\left\{\left|\frac{G(X_{Pi})-\bar{s}}{h}\right|^3+\left|\frac{G(X_{Pi})-\bar{s}}{h}\right|^4\right\}d\bar{s}\right]$$

$$=O_p(1)\left[\frac{h}{m}\sum_{i=1}^m\int_{-G(X_{Pi})/h}^{(\bar{t}-G(X_{Pi}))/h}\left|K\left(r+\frac{G_m(X_{Pi})-G(X_{Pi})}{h}\right)-K(r)\right|\right.$$

$$\left.(|r|^3+|r|^4)\,dr\right]$$

$$\le O_p(1)\left[\frac{h}{m}\sum_{i=1}^m\int_{-1-|\Delta|}^{1+|\Delta|}(|r|^3+|r|^4)dr|K(r+\Delta_i)-K(r)|dr\right]$$

$$\le O_p(1)\left[\frac{h}{m}\sum_{i=1}^m|\Delta_i|\int_{-1-|\Delta|}^{1+|\Delta|}(|r|^3+|r|^4)dr\right]$$

$$\le O_p(1)[h\Delta(1+\Delta)^4]$$

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

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

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

$$=(\Lambda^*_0(u|D_m(\bar{s}))-\Lambda^*_0(u|D(\bar{s})))D'_m(\bar{s})+\Lambda^*_0(u|D(\bar{s}))(D'_m(\bar{s})-D'(\bar{s}))$$

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

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

$$\int_0^{\bar{t}}\psi_D(u,\bar{s})\Lambda^*_0(u|D(\bar{s}))(D'_m(\bar{s})-D'(\bar{s}))d\bar{s}$$

$$= [\psi_D(u,\bar{s})\Lambda_0^*(u|D(\bar{s}))(D_m(\bar{s}) - D(\bar{s}))]_0^{\bar{t}}$$
$$- \int_0^{\bar{t}} \left[\frac{\partial}{\partial \bar{s}}\psi_D(u,\bar{s})\Lambda_0^*(u|D(\bar{s}))\right](D_m(\bar{s}) - D(\bar{s}))d\bar{s}$$
$$= O_p(m^{-1/2})$$

Thus $(mh)^{1/2}\,\mathcal{A}_{1b}(\bar{t}, G_m) = o_p(1)$.

We turn now to $\mathcal{A}_2(\bar{t}, G_m)$. By the same arguments as before we find that $(mh)^{1/2}\,\mathcal{A}_2(\bar{t}, G)$ 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

$$\tilde{H}_a(w, v, \bar{t}, X_{Pi}) = \frac{1}{h^3}\int_0^{\bar{t}} |\psi_D(w, \bar{s})| a_2(\bar{s}, v, G)^{-1}$$
$$\left|K'\left(\frac{G_m(X_{Pi}) - \bar{s}}{h}\right)\right| \left|\frac{G(X_{Pi}) - \bar{s}}{h}\right| |G_m(X_{Pi}) - G(X_{Pi})| \, d\bar{s}$$
$$\tilde{H}_b(w, v, t, X_{Pi}) = \frac{1}{h^4}\int_0^{\bar{t}} |\psi_D(w, \bar{s})| a_2(\bar{s}, v, G)^{-1}$$
$$|K''(\upsilon_i)| \left|\frac{G(X_{Pi}) - \bar{s}}{h}\right| (G_m(X_{Pi}) - G(X_{Pi}))^2$$
$$I\left(\min\left\{\left|\frac{G(X_{Pi}) - \bar{s}}{h}\right|, \left|\frac{G_m(X_{Pi}) - \bar{s}}{h}\right|\right\} \le 1\right) d\bar{s}$$

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

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

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,

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

By a similar argument,

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

Hence, by the same argument as used before for $\mathcal{B}_1^*(t, u, u)$, we find that $(mh)^{1/2}\ \Omega^*(\bar{t}, u) = o_p(1)$. Finally, by the same argument as used for $\mathcal{B}_1^{**}(t, u)$, we obtain $(mh)^{1/2}\ \Omega^{**}(\bar{t}, u) = o_p(1)$.

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 $(mh)^{1/2}\ (\hat{\Lambda}_D(G(t)) - \Lambda_D(G(t)))$ converges in distribution to a mean-zero normal distribution. We now show that $(mh)^{1/2}\ (\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t))) = o_p(1)$. We have

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

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 $(mh)^{1/2}\ (\hat{\Lambda}_D(G_m(t)) - \hat{\Lambda}_D(G(t))) = o_p(1)$ as desired.

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.

References

[1] 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. MR0673646

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, University of California at Berkeley.

[3] Chatterjee, N., Kalaylioglu, Z., Shih, J. H., and Gail, M. (2006). Case-control and case-only designs with genotype and family history data: estimating relative risk, residual familial aggregation and cumulative risk. *Biometrics* **62** 36–48. MR2226554

[4] Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics* **14** 181–197. MR0932943

[5] Doksum, K. A., Jiang, J., Sun, B. and Wang, S. (2017). Nearest neighbor estimates of regression. *Computational Statistics and Data Analysis* **110** 64–74. MR3612608

[6] Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press, Cambridge. MR1720712

[7] Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London. MR1383587

[8] 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. MR1955344

[9] Gorfine, M., Bordo, N., and Hsu, L. (2017). A fully nonparametric estimator of the marginal survival function based on case-control clustered age-at-onset data. *Biostatistics* **18** 76–90. MR3612275

[10] Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Annals of Mathematical Statistics* **40** 1187–1195. MR0247723

[11] Shih, J. H., and Chatterjee, N. (2002). Analysis of survival data from case-control family studies. *Biometrics* **58** 502–509. MR1925547

[12] Stanford, J. L., Wicklund, K. G., McKnight, B., Daling, J. R., and Brawer, M. K. (1999). Vasectomy and risk of prostate cancer. *Cancer Epidemiology Biomarkers and Prevention* **8** 881–886.

[13] van der Vaart, A (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. MR1652247