Asymptotics and Additional Simulation Results for "On Estimation of the Hazard Function from Population-based Case-Control Studies"

Li Hsu, Malka Gorfine and David Zucker

A. Proof of Martingale Representation of $\widehat{\Lambda}(\cdot, \theta)$

We begin by defining some basic quantities. Let

$$\begin{aligned}
Q_{ij}(t,\Lambda) &= I(X_{i0}<t)Y_{ij}(t)\mathcal{R}(S(X_{i0}),S(t-),\delta_{i0};\theta_j)S(t-),\\
Q^{(1)}_{ij}(t,\Lambda) &= I(X_{i0}<t)Y_{ij}(t)\frac{\partial}{\partial u}\mathcal{R}(u,S(t-),\delta_{i0};\theta_j)|_{u=S(X_{i0})}S(t-)S(X_{i0}),\\
Q^{(2)}_{ij}(t,\Lambda) &= I(X_{i0}<t)Y_{ij}(t)\frac{\partial}{\partial u}\mathcal{R}(S(X_{i0}),u,\delta_{i0};\theta_j)|_{u=S(t-)}S(t-)^2\\
&\quad +I(X_{i0}<t)Y_{ij}(t)\mathcal{R}(S(X_{i0}),S(t-),\delta_{i0};\theta_j)S(t-),
\end{aligned}$$

where $S(t)=\exp\{-\Lambda(t)\}$. Note that the Q's are also functions of θ; but for simplicity of presentation, we suppress θ whenever there is no confusion. Furthermore, let $Q(t,\Lambda)=n^{-1}\sum_{i=1}^n\sum_{j=1}^M Q_{ij}(t,\Lambda)$, $Q^{(1)}(t,\Lambda)=n^{-1}\sum_{i=1}^n\sum_{j=1}^M Q^{(1)}_{ij}(t,\Lambda)$, $Q^{(2)}(t,\Lambda)=n^{-1}\sum_{i=1}^n\sum_{j=1}^M Q^{(2)}_{ij}(t,\Lambda)$ and $\bar{N}(t)=n^{-1}\sum_{i=1}^n\sum_{j=1}^M\int_0^t I(X_{i0}<s)N_{ij}(ds)$.

In what follows, for a given random variable ξ_{ij} depending only on the data on individual ij, we denote $E^{\#}\{\xi_{ij}\}=(1-\alpha)E\{\xi_{ij}|\delta_{i0}=0\}+\alpha E\{\xi_{ij}|\delta_{i0}=1\}$. Due to condition C4 and the strong law of large numbers, the above functions converge uniformly in t and θ to $q(t)=E^{\#}\{Q_{ij}(t,\Lambda)\}$, $q^{(1)}(t)=E^{\#}\{Q^{(1)}_{ij}(t,\Lambda)\}$, $q^{(2)}(t)=E^{\#}\{Q^{(2)}_{ij}(t,\Lambda)\}$ and $n(t)=E^{\#}\{\int_0^t I(X_{i0}<s)N_{ij}(ds)\}$, respectively. Let $M_{ij}(dt)=N_{ij}(dt)-Y_{ij}(t)\mathcal{R}(S(X_{i0}),$ $S(t-),\delta_{i0};\theta_j)\ S(t-)\Lambda(dt,\theta)$, where $\Lambda(t,\theta)=\int_0^t E^{\#}\{\mathcal{R}(S(X_{i0}),S(t-),\delta_{i0};\theta_{0j})S(t-)\}/$

$E^{\#}\{\mathcal{R}(S(X_{i0}), S(t-), \delta_{i0}; \theta_j) S(t-)\}\Lambda(ds)$ and $\theta_0 = (\theta_{01}, \ldots, \theta_{0m})$ is the true value of $\theta = (\theta_1, \ldots, \theta_m)$. Clearly, $\Lambda(t, \theta_0) = \Lambda(t)$. Now we can write the first stage estimator as

$$
\begin{aligned}
&\widetilde{\Lambda}(t,\theta) - \Lambda(t,\theta) \\
&= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^M I(X_{i0} < s) M_{ij}(ds)}{Q(s,\Lambda)} + \int_0^t \left\{ \frac{1}{Q(s,\widetilde{\Lambda})} - \frac{1}{Q(s,\Lambda)} \right\} \bar{N}(ds) \\
&\approx \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^M I(X_{i0} < s) M_{ij}(ds)}{Q(s,\Lambda)} \\
&\quad + \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^M Q_{ij}^{(1)}(s,\Lambda)\{\widetilde{\Lambda}(X_{i0},\theta) - \Lambda(X_{i0},\theta)\}}{Q(s,\Lambda)^2} \bar{N}(ds) \\
&\quad + \int_0^t \frac{Q^{(2)}(s,\Lambda)}{Q(s,\Lambda)^2} \{\widetilde{\Lambda}(s-,\theta) - \Lambda(s-,\theta)\} \bar{N}(ds).
\end{aligned}
$$

The second term of the above equation can be written, by interchanging the order of integration, as

$$
\begin{aligned}
&\int_0^t \sum_{i=1}^n \sum_{j=1}^M \frac{n^{-1} Q_{ij}^{(1)}(s,\Lambda)}{Q(s,\Lambda)^2} \left[\int_0^s \{\widetilde{\Lambda}(u-,\theta) - \Lambda(u-)\} \tilde{N}_{i0}(du) \right] \bar{N}(ds) \\
&= \int_0^t \{\widetilde{\Lambda}(s-,\theta) - \Lambda(s-)\} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^M \mathcal{Y}_{ij}(s,t) \tilde{N}_{i0}(ds),
\end{aligned}
$$

where $\mathcal{Y}_{ij}(s,t) = \int_s^t Q_{ij}^{(1)}(u,\Lambda)/Q(u,\Lambda)^2 \bar{N}(du)$ and $\tilde{N}_{i0}(t) = I(X_{i0} < t)$. An argument similar to that of Yang and Prentice (1999) and Zucker (2005) yields the following representation

$$
\widetilde{\Lambda}(t,\theta) - \Lambda(t,\theta) = \widehat{p}(t) \int_0^t \frac{1}{\widehat{p}(s) Q(s,\Lambda)} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^M I(X_{i0} < s) M_{ij}(ds), \qquad (1)
$$

where

$$
\widehat{p}(t) = \prod_{s<t} \left[1 + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \mathcal{Y}_{ij}(s,t) \tilde{N}_{i0}(ds) + \frac{Q^{(2)}(s,\Lambda)}{Q(s,\Lambda)^2} \bar{N}(ds) \right],
$$

and $\widehat{p}(\cdot)$ is a product integral of empirical processes. By the Glivenko-Cantelli theorem (van der Vaart and Wellner, 1996) and the continuous mapping theorem, $\widehat{p}(t)$ converges to a limit $p(t)$ uniformly in t as $n \to \infty$. By Lemma A.1 of Spiekerman and Lin (1998), we have that $\|n^{1/2}\{\widehat{p}(t)\int_0^t\{\widehat{p}(s-)Q(s,\Lambda)\}^{-1}n^{-1}\sum_{i=1}^n\sum_{j=1}^M I(X_{i0} < s)M_{ij}(ds) - p(t)\int_0^t\{p(s-)q(s,\Lambda)\}^{-1}n^{-1}\sum_{i=1}^n\sum_{j=1}^M I(X_{i0} < s)M_{ij}(ds)\}\|$ converges to 0 in probability. Since $\sum_{i=1}^n M_{ij}(s)$ is square-integrable zero mean martingale with respect to the filtration of the jth relative of all families, by standard martingale theory, we have that $\|p(t)\int_0^t\{p(s-)q(s,\Lambda)\}^{-1}n^{-1}\sum_{i=1}^n\sum_{j=1}^M I(X_{i0} < s)M_{ij}(ds)\}\|$ converges to 0 in probability and that $n^{-1/2}p(t)\int_0^t\{p(s-)q(s,\Lambda)\}^{-1}\sum_{i=1}^n\sum_{j=1}^M I(X_{i0} < s)M_{ij}(ds)\}$ converges to a zero mean Gaussian process.

For the second-stage estimator, we similarly define the following parallel notation to the Q functions,

$$
\begin{aligned}
H_{ij}(t,\Lambda) &= Y_{ij}(t)\mathcal{R}(S(X_{i0}),S(t-),\delta_{i0};\theta_j)S(t-) \\
H_{ij}^{(1)}(t,\Lambda) &= Y_{ij}(t)\frac{\partial}{\partial u}\mathcal{R}(u,S(t-),\delta_{i0};\theta_j)|_{u=S(X_{i0})}S(t-)S(X_{i0}) \\
H_{ij}^{(2)}(t,\Lambda) &= Y_{ij}(t)\frac{\partial}{\partial u}\mathcal{R}(S(X_{i0}),u,\delta_{i0};\theta_j)|_{u=S(t-)}S(t-)^2 \\
&\quad +Y_{ij}(t)\mathcal{R}(S(X_{i0}),S(t-),\delta_{i0};\theta_j)S(t-).
\end{aligned}
$$

Similar to the Q functions, we also suppress the dependence of the H's on θ for simplicity of presentation. Let $H(t,\Lambda) = n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}(t,\Lambda)$, $H^{(1)}(t,\Lambda) = n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(1)}(t,\Lambda)$ and $H^{(2)}(t,\Lambda) = n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(2)}(t,\Lambda)$ with the respective limits denoted by $h(t,\Lambda)$, $h^{(1)}(t,\Lambda)$ and $h^{(2)}(t,\Lambda)$. Also let $N(t) = n^{-1}\sum_{i=1}^n\sum_{j=1}^M N_{ij}(t)$. The second-stage estimator can then be approximated by

$$
\begin{aligned}
&\widehat{\Lambda}(t,\theta)-\Lambda(t,\theta)\\
\approx\ & \int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M M_{ij}(ds)}{H(s,\Lambda)}\\
&+\int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(1)}(s,\Lambda)I(X_{i0}\geq s)\{\widetilde{\Lambda}(X_{i0},\theta)-\Lambda(X_{i0},\theta)\}}{H(s,\Lambda)^2}N(ds)\\
&+\int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(1)}(s,\Lambda)I(X_{i0}< s)\{\widehat{\Lambda}(X_{i0},\theta)-\Lambda(X_{i0},\theta)\}}{H(s,\Lambda)^2}N(ds)\\
&+\int_0^t \frac{H^{(2)}(s,\Lambda)}{H(s,\Lambda)^2}\{\widehat{\Lambda}(s-,\theta)-\Lambda(s-,\theta)\}N(ds).
\end{aligned}
$$

Combining with the first-stage estimator (1) and using a similar argument as used before, we have

$$
\begin{aligned}
&\widehat{\Lambda}(t,\theta)-\Lambda(t,\theta)\\
\approx\ & \int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M M_{ij}(ds)}{H(s,\Lambda)}+\int_0^\tau \frac{\widehat{\mathcal{A}}(t,u)}{p(u)q(u,\Lambda)}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^M I(X_{i0}<u)M_{ij}(du)\\
&+\int_0^t\{\widehat{\Lambda}(s-,\theta)-\Lambda(s-,\theta)\}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^M\mathcal{B}_{ij}(s,t)\tilde{N}_{i0}(ds)\\
&+\int_0^t\{\widehat{\Lambda}(s-,\theta)-\Lambda(s-,\theta)\}\frac{H^{(2)}(s,\Lambda)}{H(s,\Lambda)^2}N(ds)
\end{aligned}
$$

where $\mathcal{B}_{ij}(s,t)=\int_s^t Q_{ij}^{(1)}(u,\Lambda)/H(u,\Lambda)^2N(du)$ and

$$
\widehat{\mathcal{A}}(t,u)=\int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(1)}(s,\Lambda)I(X_{i0}\geq s\vee u)p(X_{i0})}{H(s,\Lambda)^2}N(ds).
$$

Let

$$
\widehat{p}_1(t)=\prod_{s<t}\left[1+\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^M\mathcal{B}_{ij}(s,t)\tilde{N}_{i0}(ds)+\frac{H^{(2)}(s,\Lambda)}{H(s,\Lambda)^2}N(ds)\right],
$$

and

$$
\widehat{\mathcal{A}}_1(t,u)=\int_0^t \frac{n^{-1}\sum_{i=1}^n\sum_{j=1}^M H_{ij}^{(1)}(s,\Lambda)I(X_{i0}\geq s\vee u)p(X_{i0})}{\widehat{p}_1(s)H(s,\Lambda)^2}N(ds).
$$

Then the second stage estimator can be written as

$$\begin{aligned}\widehat{\Lambda}(t,\theta)-\Lambda(t,\theta) &= \widehat{p}_1(t)\int_0^t \frac{1}{\widehat{p}_1(s)H(s,\Lambda)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M} M_{ij}(ds)\\ &\quad +\widehat{p}_1(t)\int_0^\tau \frac{\widehat{\mathcal{A}}_1(t,u)}{p(u)q(u,\Lambda)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M} I(X_{i0}<u)M_{ij}(du).\end{aligned}$$

By similar argument to that used for $\widetilde{\Lambda}(t,\theta)$, we can show that

$$n^{1/2}\{\widehat{\Lambda}(t,\theta)-\Lambda(t,\theta)\} = \mathcal{M}_1(t)+\mathcal{M}_2(t)+o_p(1) \tag{2}$$

with

$$\begin{aligned}\mathcal{M}_1(t) &= n^{-\frac{1}{2}}p_1(t)\int_0^t \frac{1}{p_1(s)h(s,\Lambda)}\sum_{i=1}^{n}\sum_{j=1}^{M} M_{ij}(ds)\\ \mathcal{M}_2(t) &= n^{-\frac{1}{2}}p_1(t)\int_0^\tau \frac{\mathcal{A}_1(t,u)}{p(u)q(u,\Lambda)}\sum_{i=1}^{n}\sum_{j=1}^{M} I(X_{i0}<u)M_{ij}(du),\end{aligned}$$

where $\mathcal{A}$ is the limit of $\widehat{\mathcal{A}}$. Hence, $n^{1/2}\{\widehat{\Lambda}(t,\theta)-\Lambda(t,\theta)\}$ can be represented by a sum of i.i.d. martingales over n individuals.

B. Proof of Asymptotic Theory for $\int_0^t \{\tilde{\phi}(u) - \phi(u, \bar{\beta})\} \lambda(u) du$

The parameter of interest is the cumulative baseline hazard function $\Lambda_0(t, \bar{\beta})$, which can be obtained by

$$\Lambda_0(t, \bar{\beta}) = \int_0^t \phi(u, \bar{\beta}) \Lambda(du)$$

where

$$\phi(u, \bar{\beta}) = E\{\exp(-\bar{\beta}^T Z_i) | \delta_{i0} = 1, X_{i0} = u\},$$

with $\bar{\beta}$ being the limiting value of $\hat{\beta}$, as described in the main paper. We define the following notation: g is the density of $(X_{i0}|\delta_{i0} = 1)$, G is the corresponding cumulative distribution function, and τ is the maximum follow-up time. For given function φ, define $\|\varphi\|_\infty = \sup_{t \in [0,\tau]} |\varphi(t)|$. Additionally, we let $\mathcal{M}$ denote a generic constant which may change from use to use but will always be independent of any varying quantities.

Let $K(\cdot)$ be a symmetric kernel function and $K_h(x) = K(x/h)/h$. Write

$$w_i(x) = \frac{\delta_{i0}}{n_1} K_h(x - X_{i0})$$

where n_1 is the number of cases and $h > 0$. Note that the definition of $w_i(x)$ here is slightly different from that in the main text of the paper; we have inserted the factor $1/n_1$, mainly for the ease of presentation in the following proof. The quantity h will depend on n, but we will suppress this dependence from the notation. We assume that $h \sim n^{-\nu}$ with $\nu \in (\frac{1}{4}, \frac{1}{2})$. Let $Y_i^* = \exp(-\widehat{\beta}^T Z_i)$. The local linear estimator for $\phi(x)$ is

$$\widehat{\phi}(x) = \sum_{i=1}^n c_i(x) Y_i^*,$$

where

$$c_i(x) = \frac{w_i(x)}{\sum_{j=1}^n w_j(x)} - \frac{\bar{U}_w(x) w_i(x)(U_i(x) - \bar{U}_w(x))}{\sum_{j=1}^n w_j(x)(U_j(x) - \bar{U}_w(x))^2}$$

with $U_i(x) = x - X_{i0}$ and $\bar{U}_w(x) = \sum_{i=1}^n w_i(x)U_i(x)/\sum_{i=1}^n w_i(x)$.

The resulting estimator of $\Lambda_0(t)$ is

$$\hat{\Lambda}_0(t) = \int_0^t \hat{\phi}(u)\hat{\Lambda}(du)$$

Define $\widetilde{\phi}(x)$ to be equal to the expression for $\hat{\phi}(x)$ with Y_i^* replaced by $Y_i = \exp(\bar{\beta}^T Z_i)$. We can write

$$\begin{aligned}\hat{\Lambda}_0(t) - \Lambda_0(t,\bar{\beta}) &= \int_0^t \phi(u,\bar{\beta})\{\hat{\Lambda}(du) - \Lambda(du)\} + \int_0^t \{\tilde{\phi}(u) - \phi(u,\bar{\beta})\}\lambda(u)du \\ &\quad + \int_0^t \{\hat{\phi}(u) - \tilde{\phi}(u)\}\lambda(u)du + \int_0^t \{\hat{\phi}(u) - \phi(u,\bar{\beta})\}\{\hat{\Lambda}(du) - \Lambda(du)\}\end{aligned} \tag{3}$$

In the appendix of the paper we show that the first and third terms are asymptotically equivalent to the sum of i.i.d. variables and the fourth term is asymptotically negligible. In what follows we will focus on the second term. Conditions C6 to C7 in the Appendix of the paper are used in the development for the asymptotic theory for $\widehat{\phi}(t)$.

For ease of presentation we use $\sum_i^*$ to denote summation over the set of i with $\delta_{i0} = 1$, because $\widetilde{\phi}$ only involves cases. Let $\mathcal{I}$ denote the interval $[h, \tau - h]$ and $\mathcal{B}$ denote the set $[0, h) \cup (\tau - h, \tau]$. Define

$$\widehat{g}(x) = \sum_i{}^* w_i(x) = \frac{1}{n_1}\sum_i{}^* K_h(x - X_{i0})$$

which is the standard kernel density estimate of $g(x)$ with kernel K. Write $\bar{g}(x) = E[\hat{g}(x)]$.

Now define

$$g^*(x) = g(x)\int_{-x/h}^{(\tau-x)/h} K(v)dv$$

i.e., $g^*(x) = g(x)$ for $x \in \mathcal{I}$ and $g^*(x) = \mu(c)g(x)$ for $x = ch$ or $x = \tau - ch$ with $c \in [0,1]$, where

$$\mu(c) = \int_{-1}^{c} K(u)du$$

From standard kernel estimation theory, the following results are known (it is here that we use the differentiability assumption on g stated in Condition C7):

$$\sup_{x \in \mathcal{I}} |\bar{g}(x) - g^*(x)| = O(h^2) = O(n_1^{-2\nu}) \tag{4}$$

$$\sup_{x \in \mathcal{B}} |\bar{g}(x) - g^*(x)| = O(h) = O(n_1^{-\nu}) \tag{5}$$

$$\sup_{x \in [0,\tau]} |\hat{g}(x) - \bar{g}(x)| = O_p((\sqrt{n_1}\, h)^{-1}) = O_p\left(n_1^{-\frac{1}{2}+\nu}\right) \tag{6}$$

The first two of the above results are proven by Taylor expansion. The third result is proved by a combination of Donsker's theorem for the empirical cdf and an integration by parts argument; see Schuster (1969) and the proof of Claim 4 below. As a consequence of the above, along with the assumption that $\nu \in (\frac{1}{4}, \frac{1}{2})$, we have

$$\sup_{x \in [0,\tau]} |\hat{g}(x) - g^*(x)| = O_p\left(n_1^{-\frac{1}{2}+\nu}\right) \tag{7}$$

Next, define $\Delta_i = Y_i - \phi(X_{i0}, \bar{\beta})$. Note that Y_i and Δ_i are bounded, since X_{i0} and Z_i are bounded. For future reference, we denote by Δ_{max} an upper bound on Δ_i. We can write $\tilde{\phi}(x) - \phi(x, \bar{\beta})$ as

$$\tilde{\phi}(x) - \phi(x, \bar{\beta}) = \frac{D(x)}{g^*(x)} - \frac{D(x)(\hat{g}(x) - g^*(x))}{\hat{g}(x)g^*(x)} + \frac{A(x)}{\hat{g}(x)} - \bar{U}_w(x)\,\frac{B(x)}{C(x)} + \bar{U}_w(x)^2\,\frac{D(x)}{C(x)} - \bar{U}_w(x)\,\frac{H(x)}{C(x)} \tag{8}$$

where

$$
\begin{aligned}
A(x) &= \sum_i{}^{*} w_i(x)\{\phi(X_{i0}, \bar{\beta}) - [\phi(x, \bar{\beta}) + \phi'(x, \bar{\beta})(x - X_{i0})]\} \\
B(x) &= \sum_i{}^{*} w_i(x)(U_i(x) - \bar{U}_w(x))\{\phi(X_{i0}, \bar{\beta}) - [\phi(x, \bar{\beta}) + \phi'(x, \bar{\beta})(x - X_{i0})]\} \\
&= \sum_i{}^{*} w_i(x)U_i(x)\{\phi(X_{i0}, \bar{\beta}) - [\phi(x, \bar{\beta}) + \phi'(x, \bar{\beta})(x - X_{i0})]\} - \bar{U}_w(x)A(x) \\
C(x) &= \sum_i{}^{*} w_i(x)(U_i(x) - \bar{U}_w(x))^2 \\
D(x) &= \sum_i{}^{*} w_i(x)\Delta_i \\
H(x) &= \sum_i{}^{*} w_i(x)U_i(x)\Delta_i
\end{aligned}
$$

Denote

$$
\begin{aligned}
\mathbb{H}_{n_1}(t) &= \sqrt{n_1}\int_0^t (\tilde{\phi}(x) - \phi(x))\lambda(x)dx \\
\mathbb{D}_{n_1}(t) &= \sqrt{n_1}\int_0^t \left(\frac{D(x)}{g^*(x)}\right)\lambda(x)dx \\
\mathcal{E}_1(t) &= \sqrt{n_1}\int_0^t \frac{D(x)(\hat{g}(x) - g^*(x))}{\hat{g}(x)g^*(x)}\lambda(x)dx \\
\mathcal{E}_2(t) &= \sqrt{n_1}\int_0^t \left(\frac{A(x)}{\hat{g}(x)}\right)\lambda(x)dx \\
\mathcal{E}_3(t) &= \sqrt{n_1}\int_0^t \bar{U}_w(x)\left(\frac{B(x)}{C(x)}\right)\lambda(x)dx \\
\mathcal{E}_4(t) &= \sqrt{n_1}\int_0^t \bar{U}_w(x)^2\left(\frac{D(x)}{C(x)}\right)\lambda(x)dx \\
\mathcal{E}_5(t) &= \sqrt{n_1}\int_0^t \bar{U}_w(x)\left(\frac{G(x)}{C(x)}\right)\lambda(x)dx
\end{aligned}
$$

We can then write

$$
\mathbb{H}_{n_1}(t) = \mathbb{D}_{n_1}(t) - \mathcal{E}_1 + \mathcal{E}_2 - \mathcal{E}_3 + \mathcal{E}_4 - \mathcal{E}_5
$$

In addition, defining

$$\begin{aligned}
\Omega(t,x_0) &= \int_0^t \frac{\lambda(x)}{g^*(x)} K_h(x-x_0)dx \\
\Omega^\circ(t,x_0) &= \left(\frac{\lambda(x_0)}{g(x_0)}\right) I(x_0 \le t) \\
\mathbb{D}^\circ_{n_1}(t) &= \frac{1}{\sqrt{n_1}} \sum_i{}^{*} \Omega^\circ(t, X_{i0})\Delta_i \\
\mathcal{E}_6(t) &= \frac{1}{\sqrt{n_1}} \sum_i{}^{*} \Upsilon(t, X_{i0})\Delta_i
\end{aligned}$$

with $\Upsilon(t,x) = \Omega(t,x) - \Omega^\circ(t,x)$, we can write

$$\mathbb{D}_{n_1}(t) = \mathbb{D}^\circ_{n_1}(t) + \mathcal{E}_6(t) \tag{9}$$

Our main claim is that $\mathbb{H}_{n_1}(t)$ is asymptotically equivalent to $\mathbb{D}^\circ_{n_1}(t)$ and that $\mathbb{D}^\circ_{n_1}(t)$ converges in distribution to a mean-zero Gaussian process. We will prove this by showing that $\mathbb{D}^\circ_{n_1}(t)$ converges in distribution to a mean-zero Gaussian process and that the terms $\mathcal{E}_1, \ldots, \mathcal{E}_6$ converge in probability to zero uniformly in t. Since the supremum of $\mathcal{E}_6(t)$ over an interval of t values may not be measurable, we will use outer probability and expectation when dealing with such quantities.

Claim 1. a. The process $\mathbb{D}^\circ_{n_1}(t)$ converges in distribution in $C[0,\tau]$ to $\mathbb{B}(V(t))$, where $\mathbb{B}$ is a Brownian motion process and

$$V(t) = \int_0^t \left(\frac{\lambda(x)}{g(x)}\right)^2 \sigma^2(x) g(x) dx,$$

where $\sigma^2(x) = \mathrm{Var}(Y_i | X_{i0} = x, \delta_{i0} = 1)$.

b. $\mathcal{E}_6(t)$ converges uniformly in outer probability to zero.

Proof: We take up the two parts of the claim in turn.

a. Write $Q^*(t, x_0, \Delta) = (\lambda(x_0)/g(x_0))\Delta I(x_0 \leq t)$ and $Q_i^\circ(t) = Q^*(t, X_{i0}, \Delta_i) = \Omega^\circ(t, X_{i0})\Delta_i$. Since $E(\Delta_i | X_{i0}, \delta_{i0} = 1) = 0$, we have $E[Q_i^\circ(t)] = 0$. A simple computation yields $\mathrm{Var}(Q_i^\circ(t)) = V(t)$. Since X_{i0} and Δ_i are bounded, λ is bounded, and g is bounded below, the classical central limit theorem implies that for any fixed t, $\mathbb{D}_{n_1}^\circ(t)$ converges in distribution to $N(0, V(t))$. Similarly, the finite-dimensional distributions of $\mathbb{D}_{n_1}^\circ(t)$ converge to those of $\mathbb{B}(V(t))$. Since $I(\cdot \leq t)$ as t ranges over $[0, \tau]$ is a Donsker class, so is $Q^*(t, \cdot, \cdot)$, and we thus obtain the claimed result.

b. We will write $\mathcal{E}_{6n}(t)$ to emphasize the dependence of $\mathcal{E}_6(t)$ on n. Write $Q_i(t) = \Upsilon(t, X_{i0})\Delta_i$. As before, $E[Q_i(t)] = 0$. We have

$$
\begin{aligned}
\mathrm{Var}(Q_i(t)) &= E\{Q_i(t)^2\} \\
&= E\{(\Omega(t, X_{i0}) - \Omega^\circ(t, X_{i0}))^2 \sigma^2(X_{i0})\} \\
&= E\{\Omega(t, X_{i0})^2 \sigma^2(X_{i0})\} - 2E\{\Omega(t, X_{i0})\Omega^\circ(t, X_{i0})\sigma^2(X_{i0})\} + E\{\Omega^\circ(t, X_{i0})^2 \sigma^2(X_{i0})\} \\
&= \int_0^\tau \left\{ \frac{1}{h} \int_0^t \frac{\lambda(x)}{g^*(x)} K\left(\frac{x-y}{h}\right) dx \right\}^2 \sigma^2(y) g(y) dy \\
&\quad - 2\int_0^t \left\{ \frac{1}{h} \int_0^t \frac{\lambda(x)}{g^*(x)} K\left(\frac{x-y}{h}\right) dx \right\} \frac{\lambda(y)}{g(y)} \sigma^2(y) g(y) dy \\
&\quad + \int_0^t \left(\frac{\lambda(y)}{g(y)}\right)^2 \sigma^2(y) g(y) dy \\
&= \int_0^\tau \left\{ \int_{-1}^1 \left(\frac{\lambda(y - hv)}{f^*(y - hv)}\right) K(v) I(y - hv \in [0, t]) dv \right\}^2 \sigma^2(y) g(y)\, dy \\
&\quad - 2\int_0^t \left\{ \int_{-1}^1 \left(\frac{\lambda(y - hv)}{g^*(y - hv)}\right) K(v) I(y - hv \in [0, t]) dv \right\} \lambda(y) \sigma^2(y)\, dy \\
&\quad + \int_0^t \left(\frac{\lambda(y)^2}{g(y)}\right) \sigma^2(y) dy
\end{aligned}
$$

Noting that as $n \to \infty$ we have $h \to 0$ and $I(y - hv \in [0, t]) \to I(y \in [0, t])$ for all y in

$[0,t]$ except for the boundary points 0 and t, we find that as $n \to \infty$

$$\int_0^\tau \left\{ \int_{-1}^1 \left(\frac{\lambda(y-hv)}{g^*(y-hv)} \right) K(v) I(y-hv \in [0,t]) dv \right\}^2 \sigma^2(y) g(y)\, dy \to \int_0^t \left(\frac{\lambda(y)^2}{g(y)} \right) \sigma^2(y) dy$$
$$\int_0^t \left\{ \int_{-1}^1 \left(\frac{\lambda(y-hv)}{g^*(y-hv)} \right) K(v) I(y-hv \in [0,t]) dv \right\} \lambda(y) \sigma^2(y)\, dy \to \int_0^t \left(\frac{\lambda(y)^2}{g(y)} \right) \sigma^2(y) dy$$

and thus $\mathrm{Var}(Q_i(t)) \to 0$. This implies that $\mathcal{E}_{6n}(t)$ converges pointwise in probability to zero

We are now going to prove a tightness condition: that for every positive ϵ and η there exist a positive number δ and an integer n_0 such that $\Pr^*(\sup_{|s-t|\le\delta} |\mathcal{E}_{6n}(t) - \mathcal{E}_{6n}(s)| \ge \epsilon) \le \eta$, for all $n \ge n_0$, where $\Pr^*$ denotes outer probability. This condition in conjunction with the pointwise convergence in probability implies the uniform convergence of $\mathcal{E}_6(t)$ in outer probability, since the pointwise convergence implies that the supremum of $|\mathcal{E}_6(t)|$ over any finite set of t values converges to zero. By the argument in the proof of Billingsley (1968, Thm. 8.3), it suffices to show the following: for every positive ϵ and η, there exist a number $\gamma \in (0,1)$ and an integer n_0 such that

$$\gamma^{-1} \Pr^*(\sup_{s \in [t, t+\gamma]} |\mathcal{E}_{6n}(s)| > \epsilon) \le \eta, \quad t \in [0,\tau],\, n \ge n_0 \tag{10}$$

Define

$$\Upsilon(t,s,y) = \Upsilon(s,y) - \Upsilon(t,y) = \int_t^s a(x) K_h(y-x) dx - a^\circ(y) I(t < y \le s)$$

with $a(x) = \lambda(x)/g^*(x)$ and $a^\circ(x) = \lambda(x)/g(x)$. Note that $|\Upsilon(t,s,y)| \le \|a\|_\infty + \|a^\circ\|_\infty$. Note also that, since $K(u) = 0$ when $|u| > 1$, $\Upsilon(t,s,y) = 0$ for $y \notin [t-h, s+h]$. For the remainder of the proof, these are the only properties of $\Upsilon(t,s,y)$ that we will use; in particular, we will not be using the fact that when h is small $\Upsilon(t,y)$ is small for most values of t and y, since there are some values of t and y for which this is not the case.

Since $\Upsilon(t,s,y)=0$ for $y \notin [t-h, s+h]$, then using the representation (9) leads to

$$\begin{aligned}
\mathcal{E}_{6n}(s) - \mathcal{E}_{6n}(t) &= \frac{1}{\sqrt{n}} \sum_i{}^* \Upsilon(t,s,X_{i0})\Delta_i \\
&= \frac{1}{\sqrt{n}} \sum_i{}^* \Upsilon(t,s,X_{i0}) I(X_{i0} \in [t-h, s+h])\Delta_i \\
&= \sum_{i=N(t-h)+1}^{N(s+h)} \Upsilon(t,s,X_{\zeta(i)0})\Delta_{\zeta(i)}
\end{aligned}$$

For the moment, let us condition on $\mathcal{X}$. Write $N(t,s,h) = N(s+h) - N(t-h)$. From the preceding displayed equation, we find that

$$\sup_{s \in [t, t+\gamma]} |\mathcal{E}_6(s) - \mathcal{E}_6(t)| = n^{-1/2} \max_{1 \le j \le N(t,t+\gamma,h)} |\mathbb{S}_j|$$

with

$$\mathbb{S}_j = \sum_{i=N(t-h)+1}^{N(t-h)+j} \Upsilon(t,s,X_{\zeta(i)0})\Delta_{\zeta(i)}$$

Now,

$$E[\mathbb{S}_j^4|\mathcal{X}] = \sum_{i_1}\sum_{i_2}\sum_{i_3}\sum_{i_4} \left\{ \left(\prod_{r=1}^{4} \Upsilon(t,s,X_{(i_r)0}) \right) E\left[\prod_{r=1}^{4} \Delta_{\zeta(i_r)} \middle| \mathcal{X} \right] \right\}$$

where each of the summations above ranges from $N(t-h)+1$ to $N(t-h)+j$.

Next, for a given x, define $\mathcal{Q}(x,\xi)$ to be the quantile function of the conditional distribution of Δ given $X_{i0} = x$, i.e., the (generalized) inverse of the distribution function $F_{\Delta|X=x}$ of Δ_i given $X_{i0} = x$. We can then represent Δ_i as $\Delta_i = \mathcal{Q}(X_{i0}, \xi_i)$, where ξ_i is a $U(0,1)$ random variable independent of X_{i0}. Since $E[\Delta_i|X_{i0}] = 0$, it follows that for any $U(0,1)$ random variable ξ° and any fixed x, we have $E[Q(x,\xi^\circ)] = 0$. Since Δ_i is bounded, $\mathcal{Q}$ is a bounded function.

We can now write

$$E[\mathbb{S}_j^4|\mathcal{X}] = \sum_{i_1}\sum_{i_2}\sum_{i_3}\sum_{i_4} \left\{ \left(\prod_{r=1}^{4} \Upsilon(t,s,X_{(i_r)0}) \right) E\left[\prod_{r=1}^{4} \mathcal{Q}(X_{\zeta(i_r)}, \xi_{\zeta(i_r)}) \middle| \mathcal{X} \right] \right\}$$

Note that all the terms inside the quadruple summation are bounded. Since (by the above construction) the ξ_i's all independent of each other and of $\mathcal{X}$, and $E[\mathcal{Q}(X_{\zeta(i)0}, \xi_{\zeta(i)})|\mathcal{X}] = 0$, it follows that all the terms in the quadruple summation above drop out except in the following cases: (a) $i_1 = i_2 = i_3 = i_4$, (b) $i_1 = i_2, i_3 = i_4$, and $i_1 \neq i_3$, (c) $i_1 = i_3, i_2 = i_4$, and $i_1 \neq i_2$, (d) $i_1 = i_4, i_2 = i_3$, and $i_1 \neq i_2$. Among these four cases there are a total of $3j^2 - 2j$ terms. So we have $E[\mathbb{S}_j^4|\mathcal{X}] \leq \mathcal{M}j^2$ for every j. It follows from Theorem B of Serfling (1970) that $E[(\max_{1\leq j\leq N(t,t+\gamma,h)} |\mathbb{S}_j|)^4|\mathcal{X}] \leq \mathcal{M}(N(t+\gamma+h) - N(t-h))^2$. (Because of the boundedness of the random variables we are dealing with, the quantities q_n in Serfling's Eqn. (3.6) can be replaced with fixed constants.) Thus (with E^* denoting outer expectation)

$$E^*[(\sup_{s\in[t,t+\gamma]} |\mathcal{E}_6(s) - \mathcal{E}_6(t)|)^4 \mid \mathcal{X}] \leq \mathcal{M} \left(\frac{N(t+\gamma+h) - N(t-h)}{n} \right)^2$$

and so

$$\begin{aligned} E^*[(\sup_{s\in[t,t+\gamma]} |\mathcal{E}_6(s) - \mathcal{E}_6(t)|)^4] &\leq \mathcal{M}E \left[\left(\frac{N(t+\gamma+h) - N(t-h)}{n} \right)^2 \right] \\ &\leq \mathcal{M} \left((\gamma+2h)^2 + \frac{(\gamma+2h)}{n} \right) \\ &\leq \mathcal{M}^* \left(\gamma^2 + h^2 + \frac{(\gamma+2h)}{n} \right) \end{aligned}$$

where $\mathcal{M}^*$ is a fixed constant. Accordingly, for any $\epsilon > 0$, Markov's inequality yields

$$\Pr{}^*(\sup_{s\in[t,t+\gamma]} |\mathcal{E}_6(s) - \mathcal{E}_6(t)| > \epsilon) \leq (\mathcal{M}^*/\epsilon^4) \left(\gamma^2 + h^2 + \frac{\gamma+2h}{n} \right)$$

Now let η be given. Take γ small enough so that $(\mathcal{M}^*/\epsilon^4)\gamma \leq \frac{1}{2}\eta$. With γ thus chosen, find n_0 large enough such that

$$\left(\frac{\mathcal{M}^*}{\epsilon^4\gamma} \right) \left(h_{n_1}^2 + \frac{\gamma + 2h_{n_1}}{n_1} \right) \leq \tfrac{1}{2}\eta$$

for $n \geq n_0$. We then obtain (10), and so the proof is complete.

Claim 2. The term $\mathcal{E}_1(t)$ converges in probability to zero uniformly in t.

Proof: Define the event

$$\mathcal{G} = \{\inf_x |\hat{g}(x)| \geq \tfrac{1}{4} g_{min} \text{ and } \sup_x |\hat{g}(x) - g^*(x)| \leq 1\}$$

Since $\hat{g}$ converges uniformly in probability to g^*, $\Pr(\mathcal{G}^c) \to 0$. Thus, defining $\mathcal{E}_1^*(t) = \mathcal{E}_1(t)\mathrm{I}(\mathcal{G})$ where $\mathrm{I}(\cdot)$ is an indicator function, it suffices to prove that $\sup_t |\mathcal{E}_1^*(t)|$ converges in probability to zero. We have

$$\mathcal{E}_1^*(t) = \left\{ \sqrt{n_1} \int_0^t \frac{D(x)(\hat{g}(x) - g^*(x))}{\hat{g}(x) g^*(x)} \lambda(x) dx \right\} \mathrm{I}(\mathcal{G}) = \frac{1}{\sqrt{n_1}} \sum_i{}^{*} \alpha_1(t, X_{i0}) \Delta_i$$

with

$$\alpha_1(t, y) = \left\{ \frac{1}{h} \int_0^t K\left(\frac{x - y}{h} \right) \frac{\hat{g}(x) - g^*(x)}{\hat{g}(x)} \frac{\lambda(x)}{g^*(x)} dx \right\} \mathrm{I}(\mathcal{G})$$

Note that

$$|\alpha_1(t, y)| \leq \mathcal{M}\mathrm{I}(\mathcal{G}) \sup_x |\hat{g}(x) - g^*(x)| \Omega(t, y) \leq \mathcal{M}\mathrm{I}(\mathcal{G}) \sup_x |\hat{g}(x) - g^*(x)|$$

By the same argument as used to prove tightness in the proof of Claim 1, we obtain

$$E[(\sup_t |\mathcal{E}_1^*(t)|)^4 | \mathcal{X}] \leq \mathcal{M}\mathrm{I}(\mathcal{G}) \sup_x |\hat{g}(x) - g^*(x)|$$

so that

$$E[(\sup_t |\mathcal{E}_1^*(t)|)^4] \leq \mathcal{M} E[\sup_x |\hat{g}(x) - g^*(x)| \mathrm{I}(\mathcal{G})]$$

Since $\sup_x |\hat{g}(x) - g^*(x)| \mathrm{I}(\mathcal{G})$ is uniformly bounded and converges in probability to zero, it follows (see Van der Vaart, 2000, Section 2.5) that $E[(\sup_t |\mathcal{E}_1^*(t)|)^4] \to 0$, which yields the desired result.

Claim 3. *The term $\mathcal{E}_2(t)$ converges in probability to zero uniformly in t.*

Proof: Recall the expression for $\mathcal{E}_2(t)$:

$$\begin{aligned}\mathcal{E}_2(t) &= \sqrt{n_1}\int_0^t \left(\frac{A(x)}{\hat{g}(x)}\right)\lambda(x)dx \\ &= \sqrt{n_1}\int_0^t \left(\frac{\lambda(x)}{\hat{g}(x)}\right)\left[\sum_i{}^{*} w_i(x)\{\phi(X_{i0},\bar{\beta}) - [\phi(x,\bar{\beta}) + \phi'(x,\bar{\beta})(x - X_{i0})]\}\right]dx\end{aligned}$$

We can write

$$\begin{aligned}\mathcal{E}_2(t) = &\frac{1}{\sqrt{n_1}}\sum_i{}^{*}\left[\frac{1}{h}\int_0^t K\left(\frac{x - X_{i0}}{h}\right)\left(\frac{\lambda(x)}{\hat{g}(x)}\right)\{\phi(X_{i0},\bar{\beta}) - [\phi(x,\bar{\beta}) + \phi'(x,\bar{\beta})(x - X_{i0})]\}dx\right] \\ = &\frac{1}{\sqrt{n_1}}\sum_i{}^{*}\int_{-1}^1 I(X_{i0} + vh \in [0,t])\left(\frac{\lambda(X_{i0} + vh)}{\hat{g}(X_{i0} + vh)}\right) \\ &\{\phi(X_{i0},\bar{\beta}) - [\phi((X_{i0} + vh,\bar{\beta}) + \phi'((X_{i0} + vh,\bar{\beta})vh\}K(v)dv\end{aligned}$$

Now,

$$|\{\phi(X_{i0},\bar{\beta}) - [\phi((X_{i0} + vh,\bar{\beta}) + \phi'((X_{i0} + vh,\bar{\beta})vh\}| \le \tfrac{1}{2}\|\phi''\|_\infty v^2h^2 = \mathcal{M}n^{-2\nu}v^2$$

In addition, from (7) plus the assumption that $g(x)$ is bounded below, we have

$$\sup_x \hat{g}(x)^{-1} \le \{\inf_x g^*(x) + O_p(n^{-\frac{1}{2}+\nu})\}^{-1} = O_p(1)$$

So the integrand in the second line above is bounded by $\mathcal{M}n^{-2\nu}$. This gives $\sup_t|\mathcal{E}_2(t)| \le \mathcal{M}n^{-2\nu+\frac{1}{2}}$, which tends to 0 since $\nu > \frac{1}{4}$.

Claim 4. We have

$$\sup_{x\in[0,\tau]}\left|\bar{U}_w(x) - \left(\frac{\int_{(x-\tau)/h}^{x/h} vK(v)dv}{\int_{(x-\tau)/h}^{x/h} K(v)dv}\right)h\right| = O_p(n_1^{-\frac{1}{2}})$$

In particular, since K is symmetric, so that $\int_{-1}^1 vK(v)dv = 0$, we have $\sup_{x\in\mathcal{I}}|\bar{U}_w(x)| = O_p(n_1^{-\frac{1}{2}})$.

Proof: We have $\bar{U}_w(x) = Q(x)/\hat{g}(x)$, where

$$Q(x) = \frac{1}{n_1 h} \sum_i{}^{*} K\left(\frac{x - X_{i0}}{h}\right)(x - X_{i0})$$

Let $\tilde{K}(v) = vK(v)$ and denote by $\mathbb{G}_n(y)$ the empirical distribution function of the X_{i0}'s.

We then have

$$\begin{aligned}
|Q(x) - E[Q(x)]| &= \left|\frac{1}{h}\int K\left(\frac{x-y}{h}\right)(x-y)\{d\mathbb{G}_n(y) - dG(y)\}\right| \\
&= \left|\int \tilde{K}\left(\frac{x-y}{h}\right)\{d\mathbb{G}_n(y) - dG(y)\}\right| \\
&\leq \|\mathbb{G}_n - G\|_\infty \left[\frac{1}{h}\int \left|\tilde{K}'\left(\frac{y-x}{h}\right)\right| dy\right] \\
&\leq \mathcal{M}\|\mathbb{G}_n - G\|_\infty = \mathcal{M} n_1^{-\frac{1}{2}} \qquad (11)
\end{aligned}$$

This result, along with (7), implies that

$$\sup_{x\in[0,\tau]} |Q(x) - E[Q(x)]|/\hat{g}(x) = O_p(n^{-\frac{1}{2}}).$$

Next,

$$\begin{aligned}
E[Q(x)] &= \frac{1}{h}\int_0^\tau K\left(\frac{x-y}{h}\right)(x-y)g(y)dy \\
&= h\int_{(x-\tau)/h}^{x/h} vK(v)g(x-hv)dv \\
&= hg(x)\int_{(x-\tau)/h}^{x/h} vK(v)dv + h\int_{(x-\tau)/h}^{x/h} K(v)[g(x-hv) - g(x)]dv
\end{aligned}$$

Now, $|g(x-hv) - g(x)| \leq \|g'\|_\infty hv$. Recall that $h \sim n^{-\nu}$ with $\nu > \frac{1}{4}$, we have

$$E[Q(x)] = hg(x)\int_{(x-\tau)/h}^{x/h} vK(v)dv + o(n_1^{-\frac{1}{2}})$$

This result along with (7) lead to the desired conclusion.

Claim 5. The quantity

$$B_1(x) = \sum_i^* w_i(x) U_i(x) \{\phi(X_{i0,\bar{\beta}}) - [\phi(x, \bar{\beta}) + \phi'(x, \bar{\beta})(x - X_{i0})]$$

satisfies $\sup_{x \in [0,\tau]} |B_1(x)| = O_p(n_1^{-3\nu})$.

Proof: This result is proved using arguments similar to those used to prove (4)-(6) and Claim 4.

Claim 6. $\inf_{x \in [0,\tau]} C(x) \geq \mathcal{M} n_1^{-2\nu}$.

Proof: We have

$$C(x) = \sum_i^* w_i(x) U_i^2(x) - \hat{g}(x) \bar{U}_w(x)^2$$

Denote the first term by $\tilde{C}(x)$. By an argument similar to that used in the proof of Claim 4, we have $\sup_x |\tilde{C}(x) - E[\tilde{C}(x)]| = O_p(n_1^{-(\frac{1}{2}+\nu)})$. In addition, we have

$$\begin{aligned}
E[\tilde{C}(x)] &= \frac{1}{h} \int_0^\tau K\left(\frac{x-y}{h}\right)(x-y)^2 g(y) dy \\
&= h^2 \int_{(x-\tau)/h}^{x/h} v^2 K(v) g(x - hv) dv \\
&= \left[g(x) \int_{(x-\tau)/h}^{x/h} v^2 K(v) dv \right] h^2 + O(h^3)
\end{aligned}$$

For $x \in \mathcal{I}$, we get

$$E[\tilde{C}(x)] = \left[g(x) \int_{-1}^{1} v^2 K(v) dv \right] h^2 + O(h^3)$$

and Claim 4 implies that $\sup_{x \in \mathcal{I}} \bar{U}_w(x)^2 = O_p(n^{-1})$. This yields the desired conclusion for $x \in \mathcal{I}$. For $x \in \mathcal{B}$, the above developments in conjunction with Claim 4 and (7) lead to

$$C(x) = \left[g(x) \int_{(x-\tau)/h}^{x/h} v^2 K(v) dv - g^*(x) \left(\frac{\int_{(x-\tau)/h}^{x/h} v K(v) dv}{\int_{(x-\tau)/h}^{x/h} K(v) dv} \right)^2 \right] h^2 + o_p(h^2)$$

with the o_p uniform in x. Since, by definition,

$$g^*(x) = g(x) \int_{(x-\tau)/h}^{x/h} K(v)dv$$

we get

$$C(x) = \omega(x)g(x)h^2 + o_p(h^2)$$

with

$$\omega(x) = \frac{\int_{(x-\tau)/h}^{x/h} v^2 K(v)dv}{\int_{(x-\tau)/h}^{x/h} K(v)dv} - \left(\frac{\int_{(x-\tau)/h}^{x/h} vK(v)dv}{\int_{(x-\tau)/h}^{x/h} K(v)dv} \right)^2$$

If we consider $x = ch$ for $c \in [0,1]$, we have (for $h \leq \frac{1}{2}\tau$) $\omega(x) = \mathrm{Var}(\kappa|\kappa > -c)$, where κ is a random variable with density K. This quantity is strictly positive for every $c \in [0,1]$, since for every c the region $\{v > -c\}$ has positive mass under the density K. It is clear that $\mathrm{Var}(\kappa|\kappa > -c)$ is continuous in c, and so its minimum over $[0,1]$ is attained for some $c^* \in [0,1]$, and, as just argued, $\mathrm{Var}(\kappa|\kappa > -c^*)$ is strictly positive. The same argument can be made for $x = \tau - ch, c \in [0,1]$. We thus find that $\min_{x\in[0,h]\cup[\tau-h,\tau]} \omega(x)$ is strictly positive. From this result, along with the fact that $g(x) \geq g_{min}$, the desired conclusion follows.

Claim 7. The term $\mathcal{E}_3(t)$ converges in probability to zero uniformly in t.

Proof: Let us write $\mathcal{E}_3(t) = \mathcal{E}_{31}(t) - \mathcal{E}_{32}(t)$ with

$$\mathcal{E}_{31}(t) = \sqrt{n_1} \int_0^t \bar{U}_w(x) \left(\frac{B_1(x)}{C(x)} \right) \lambda(x)dx$$
$$\mathcal{E}_{32}(t) = \sqrt{n_1} \int_0^t \bar{U}_w(x)^2 \left(\frac{A(x)}{C(x)} \right) \lambda(x)dx$$

Regarding $\mathcal{E}_{31}(t)$ we have

$$
\begin{aligned}
|\mathcal{E}_{31}(t)| &= \left| \frac{1}{\sqrt{n_1}} \sum_i^* \int_0^t \left[\frac{1}{h} K\left(\frac{x - X_{i0}}{h} \right) (x - X_{i0}) \right] \left(\frac{B_1(x)}{C(x)} \right) \lambda(x) dx \right| \\
&= \left| \frac{h}{\sqrt{n_1}} \sum_i^* \int_{-1}^1 \left[I(X_{i0} + vh \in [0,t]) \left(\frac{B_1(X_{i0} + hv)}{C(X_{i0} + hv)} \right) \lambda(X_{i0} + hv) \right] vK(v) dv \right| \\
&\leq \mathcal{M} h \sqrt{n_1} \left[\frac{\sup_{x \in [0,\tau]} |B_1(x)|}{\inf_{x \in [0,\tau]} C(x)} \right] = O_p(n_1^{-2\nu + \frac{1}{2}}) = o_p(1)
\end{aligned}
$$

using the results of Claims 5 and 6 and the fact that $\nu > \frac{1}{4}$. Regarding $\mathcal{E}_{32}(t)$, using Claims 4 and 6 we have

$$
|\mathcal{E}_{32}(t)| \leq \left[\frac{\sup_{x \in [0,\tau]} \bar{U}_w(x)^2}{\inf_{x \in [0,\tau]} C(x)} \right] \left[\sqrt{n_1} \int_0^t A(x) \lambda(x) dx \right] = O_p(1) \left[\sqrt{n_1} \int_0^t A(x) \lambda(x) dx \right]
$$

and the bracketed term is $o_p(1)$ by the argument used in the proof of Claim 3.

Claim 8. The term $\mathcal{E}_4(t)$ converges in probability to zero uniformly in t.

Proof: Let us write $\mathcal{E}_4(t) = \mathcal{E}_{41}(t) + \mathcal{E}_{42}(t)$ with

$$
\begin{aligned}
\mathcal{E}_{41}(t) &= \sqrt{n_1} \int_{[0,t] \cap \mathcal{I}} \bar{U}_w(x)^2 \left(\frac{D(x)}{C(x)} \right) \lambda(x) dx = \frac{1}{\sqrt{n_1}} \sum_i^* \alpha_2(t, X_{i0}) \Delta_i \\
\mathcal{E}_{42}(t) &= \sqrt{n_1} \int_{[0,t] \cap \mathcal{B}} \bar{U}_w(x)^2 \left(\frac{D(x)}{C(x)} \right) \lambda(x) dx = \frac{1}{\sqrt{n_1}} \sum_i^* \alpha_3(t, X_{i0}) \Delta_i
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_2(t, y) &= \frac{1}{h} \int_{[0,t] \cap \mathcal{I}} K\left(\frac{x - y}{h} \right) \frac{\bar{U}_w(s)^2}{C(x)} \lambda(x) dx \leq \mathcal{M} \left[\frac{\sup_{x \in \mathcal{I}} \bar{U}_w(x)^2}{\inf_{x \in \mathcal{I}} C(x)} \right] \\
\alpha_3(t, y) &= \frac{1}{h} \int_{[0,t] \cap \mathcal{B}} K\left(\frac{x - y}{h} \right) \frac{\bar{U}_w(s)^2}{C(x)} \lambda(x) dx \leq \mathcal{M} \left[\frac{\sup_{x \in \mathcal{B}} \bar{U}_w(x)^2}{\inf_{x \in \mathcal{B}} C(x)} \right] I(y \in [0, 2h] \cup [\tau - 2h, \tau])
\end{aligned}
$$

By the argument used to prove tightness in the proof of Claim 1, we obtain

$$
\begin{aligned}
E[\sup_{t \in [0,\tau]} \mathcal{E}_{41}(t)^4 | \mathcal{X}] &\leq \mathcal{M} \left[\frac{\sup_{x \in \mathcal{I}} \bar{U}_w(x)^2}{\inf_{x \in \mathcal{I}} C(x)} \right] \\
E[\sup_{t \in [0,\tau]} \mathcal{E}_{42}(t)^4 | \mathcal{X}] &\leq \mathcal{M} \left[\frac{\sup_{x \in \mathcal{B}} \bar{U}_w(x)^2}{\inf_{x \in \mathcal{B}} C(x)} \right] \left(\frac{N(2h) + (N(\tau) - N(\tau - 2h))}{n} \right)^2
\end{aligned}
$$

By Claims 4 and 6, we have

$$\frac{\sup_{x\in\mathcal{I}}\bar{U}_w(x)^2}{\inf_{x\in\mathcal{I}}C(x)} = O_p(n_1^{-1+2\nu}), \quad \frac{\sup_{x\in\mathcal{B}}\bar{U}_w(x)^2}{\inf_{x\in\mathcal{B}}C(x)} = O_p(1)$$

We can now use arguments similar to those in the proof of Claim 2 to show that $\sup_{t\in[0,\tau]}|\mathcal{E}_{41}(t)| \xrightarrow{P} 0$ and $\sup_{t\in[0,\tau]}|\mathcal{E}_{42}(t)| \xrightarrow{P} 0$.

Claim 9. The term $\mathcal{E}_5(t)$ converges in probability to zero uniformly in t.

Proof: The proof is the same as that of Claim 8, except that we replace $\alpha_2(t,y)$ by

$$\tilde{\alpha}_2(t,y) = \frac{1}{h}\int_{[0,t]\cap\mathcal{I}}(x-y)K\left(\frac{x-y}{h}\right)\frac{\bar{U}_w(s)}{C(x)}\lambda(x)dx \le \mathcal{M}\left[\frac{h\sup_{x\in\mathcal{I}}\bar{U}_w(x)}{\inf_{x\in\mathcal{I}}C(x)}\right]$$

and similarly with $\alpha_3(t,y)$.

Claim 10. $\sup_{x\in[0,\tau]}|\tilde{\phi}(x) - \phi(x,\bar{\beta})| = O_p(n_1^{-\frac{1}{2}+\nu})$

Proof: By integration by parts arguments similar to the one used to prove (11), we can show that

$$\sup_x |D(x)| = O_p(n_1^{-\frac{1}{2}+\nu})$$
$$\sup_x |A(x)| = O_p(n_1^{-\frac{1}{2}})$$
$$\sup_x |H(x)| = O_p(n_1^{-\frac{1}{2}})$$

Using these results and the results of the preceding claims, we can easily verify that each of the terms in (8) converges uniformly in probability to 0.

Finally, we present a result that is used in the Appendix of the paper to deal with the first term of (3). The proof is identical to the proof that $\tilde{\phi}(x)$ converges uniformly in x in probability to $\phi(x,\bar{\beta})$.

Claim 11. The term

$$\sum_{i=1}^{n} c_i(x) Z_i Y_i \tag{12}$$

converges uniformly in x in probability to $\xi(x) = E[Z_i Y_i | X_{i0} = x]$.

C. Proof of term (IV), $\int_0^t \{\widehat{\phi}(u) - \phi(u, \bar{\beta})\}\{\widehat{\Lambda}(du, \widehat{\theta}) - \Lambda(du)\}$, is asymptotically negligible

In this section, we provide the detailed proof of term (IV), $\int_0^t \{\widehat{\phi}(u) - \phi(u, \bar{\beta})\}\{\widehat{\Lambda}(du, \widehat{\theta}) - \Lambda(du)\}$, is $o_p(n_1^{-1/2})$, and thus asymptotically negligible. We cannot use an integration by parts argument because we have not examined the asymptotic properties of the derivative $\widetilde{\phi}'(t)$, so we take a different approach. Equation (2) gives a representation of $n^{1/2}\{\widehat{\Lambda}(t, \theta) - \Lambda(t, \theta)\}$ as the sum of two martingale terms $\mathcal{M}_1(t)$ and $\mathcal{M}_2(t)$. We will work with the term in (IV) arising from $\mathcal{M}_1(t)$; the argument for the term arising from $\mathcal{M}_2(t)$ is similar.

We have

$$n^{-1/2}\mathcal{M}_1(t) = \frac{p_1(t)}{n} \int_0^t \rho(s) \sum_{i=1}^{n} \sum_{j=1}^{M} M_{ij}(ds)$$

where $\rho(s) = [p_1(s)h(s, \Lambda)]^{-1}$. From the definition of $\hat{p}_1$ we can see that $p_1(s)$ is bounded below, and Condition C1 implies that $h(s, \Lambda)$ is bounded below. Thus $\rho(s)$ is bounded above by some quantity $\rho_{\max}$. Now, an alternate martingale representation of $dN_{ij}(s)$ is given by

$$dN_{ij}(s) = Y_{ij}(s)\lambda(t|X_{i0}, \delta_{i0}, Z_{i0})dt + d\widetilde{M}_{ij}(s)$$

with $\widetilde{M}_{ij}$ being a martingale with respect to the filtration $\mathcal{F}_t^* = \sigma(X_{i0}, \delta_{i0}, Z_{i0}, Y_{ij}(s), N_{ij}(s), s \leq t, j = 1, \ldots, m, i = 1, \ldots, n)$. Letting $g(s|X_{i0}, \delta_{i0}, Z_{i0}) = \lambda(s|X_{i0}, \delta_{i0}, Z_{i0}) - \lambda(s|X_{i0}, \delta_{i0})$, we can write

$$dM_{ij}(s) = Y_{ij}(s)g(s|X_{i0}, \delta_{i0}, Z_{i0})ds + d\widetilde{M}_{ij}(s).$$

By the innovation theorem, $E\{Y_{ij}(s)g(s|X_{i0},\delta_{i0},Z_{i0})\}=0$. Noting that

$$n^{-1/2}\mathcal{M}_1(du)=n^{-1}p_1'(u)\int_0^u \rho(s)\sum_{i=1}^{n}\sum_{j=1}^{m}M_{ij}(ds)du+n^{-1}p_1(u)\rho(u)\sum_{i=1}^{n}\sum_{j=1}^{M}M_{ij}(du),$$

we see that (IV) can be decomposed into the following terms:

$$\begin{aligned}
\Gamma_1(t) &= \int_0^t (\tilde{\phi}(u)-\phi(u))p_1'(u)\int_0^u \rho(s)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{M}Y_{ij}(s)g(s|X_{i0},\delta_{i0},Z_{i0})ds\right\}du\\
\Gamma_2(t) &= \int_0^t p_1(u)\rho(u)(\tilde{\phi}(u)-\phi(u))\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{M}Y_{ij}(u)g(u|X_{i0},\delta_{i0},Z_{i0})\right\}du\\
\Gamma_3(t) &= \int_0^t (\tilde{\phi}(u)-\phi(u))p_1'(u)\int_0^u \rho(s)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{M}d\widetilde{M}_{ij}(s)\right\}dsdu\\
\Gamma_4(t) &= \int_0^t p_1(u)\rho(u)(\tilde{\phi}(u)-\phi(u))\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{M}d\widetilde{M}_{ij}(u)\right\}
\end{aligned}$$

Consider first Γ_1. Let $\mathcal{G}(s)$ denote the term enclosed in curly brackets. This term is $n^{1/2}$ times the sample average of the mean-zero r.v.'s $\sum_{j=1}^{M}Y_{ij}(s)g(s|X_{i0},\delta_{i0},Z_{i0})$. Conditions C2 and C4 imply that $g(s|X_{i0},\delta_{i0},Z_{i0})$ is continuous as a function of s, and so it follows from empirical process theory that $\mathcal{G}(s)$ converges to a Gaussian process, which implies that $\|\mathcal{G}\|=O_p(1)$. This fact, in conjunction with the fact that $\|\tilde{\phi}-\phi\|\to_p 0$ (Claim 10), easily yields the result that $\|\Gamma_1\|\to_p 0$. The term Γ_2 can be dealt with in a similar way. Now examining $\Gamma_4(t)$, then from the definition of $d\widetilde{M}_{ij}$ and the fact that $\tilde{\phi}$ is $\mathcal{F}_0^*$-measurable, it follows that $\Gamma_4(t)$ is a mean-zero martingale w.r.t. $\mathcal{F}_t^*$. The predictable variation process is given by

$$\begin{aligned}
\langle\Gamma_4,\Gamma_4\rangle(t) &= \int_0^t p_1(u)^2\rho(u)^2(\tilde{\phi}(u)-\phi(u))^2\left\{\frac{1}{n}\sum_i\sum_j Y_{ij}(s)\lambda(s|X_{i0},\delta_{i0},Z_{i0})\right\}ds\\
&\le \tau\rho_{\max}^2\|p_1\|\|\tilde{\phi}-\phi\|^2\left\|\frac{1}{n}\sum_i\sum_j Y_{ij}(\cdot)\lambda(\cdot|X_{i0},\delta_{i0},Z_{i0})\right\|.
\end{aligned}$$

By the uniform law of large numbers we have $\|n^{-1}\Sigma_i\Sigma_j Y_{ij}(\cdot)\lambda(\cdot|X_{i0},\delta_{i0},Z_{i0})\|=O_p(1)$ and we know already that $\|\tilde{\phi}-\phi\|\to_p 0$. So we get $\langle\Gamma_4,\Gamma_4\rangle(\tau)\to_p 0$. By Andersen and Gill (1982, Theorem I.1.b.) (corollary of Lenglart's inequality), this implies that $\|\Gamma_4\| \to_p 0$.

Regarding $\Gamma_3(t)$, we have

$$\|\Gamma_3\| \leq \tau \|p_1'\| \|\tilde{\phi} - \phi\|^2 \left\{ \sup_u \left| \int_0^u \rho(s) \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^M d\widetilde{M}_{ij}(s) \right\} ds \right| \right\}$$

Applying Andersen and Gill (1982, Theorem I.1.b) again, we find that the term in curly brackets on the right side of the inequality is $O_p(1)$, and so $\|\Gamma_3\| \to_p 0$.

D. Efficiency comparison of the one-stage and two-stage estimators of $\Lambda(\cdot)$

We simulated 1,000 datasets each with 3000 cases and controls, and one relative per case or control. The correlation of failure times of family members was generated according to the Clayton-Oakes model (i.e. the Gamma frailty distribution) with $\theta = 3.0$. Table S1 shows the comparison of the one-stage estimator (3) and two-stage estimator (4) of $\Lambda(t)$ at selected ages $t = 40, 50, 60, 70$, and 80. Both estimators show little bias; however, the two-stage estimator has considerably smaller standard errors at all ages than the one-stage estimator, suggesting the second stage can improve the efficiency considerably.

Table S1: Comparison of the one-stage and two-stage estimators of $\Lambda(t)$

		One-Stage		Two-Stage	
	True	Est	SE	Est	SE
40	0.016	0.016	0.009	0.015	0.001
50	0.043	0.042	0.012	0.042	0.003
60	0.097	0.094	0.019	0.096	0.006
70	0.195	0.187	0.030	0.190	0.012
80	0.353	0.339	0.046	0.345	0.024

E. Misspecification Effect of the Copula Function

We conducted a simulation study when the true copula distribution is inverse Gaussian and positive stable and the fitting model is the Clayton-Oakes copula model (i.e. Gamma frailty distribution) (Table S2 and S3). For each copula distribution, we considered two situations, roughly corresponding to moderate and strong dependences. This was achieved by adjusting the parameters in the inverse Gaussian and positive stable distributions such that when it was fit by using the Clayton-Oakes model, the dependence parameter was around 1 or 3 roughly. The dependence parameter in the Clayton-Oakes model also possesses a desirable interpretation of constant cross ratio between paired failure times (Oakes, 1989).

For each scenario we simulated 1,000 datasets each with 3000 cases and controls, and one relative for each case or control. As a comparison, we also included the results from the Nelson-Aalen estimator using the relatives' only data without accounting for the case-control sampling and the dependence between relatives and probands. It can be seen that the naive Nelson-Aalen estimator is biased upward and as expected, the bias becomes more substantial as the dependence is stronger. In contrast, the proposed estimator is generally unbiased when the dependence is moderate. The bias increases as the dependence becomes stronger; however, the bias is mostly less than 5% for both inverse Gaussian and Positive Stable distributions. The bias becomes more noticeable when at older age (e.g., 70 and 80 years old), probably because the dependence is overestimated under the misspecified Clayton-Oakes model at late age. However, even in this situation, the bias is 10-15% or less within true values.

The coverage probabilities of 95% confidence intervals for the proposed estimator are much better than the naive Nelson-Aalen estimator under all scenarios considered at almost all ages. Here we define the confidence interval as providing successful coverage if it overlaps with a window of $\pm 5\%$ around the true values. The exception is when the age is 80 years old and the true frailty distribution is positive stable with strong dependence. This is because the positive stable induces a steep decreasing dependence function over time, and when the dependence is strong, the strength of dependence decreases rapidly and at age 80 it is nearly independent. As such, the naive Nelson-Aalen estimator that assumes independence performs well, whereas the proposed estimator underestimates the hazard function.

Table S2: Summary of estimates for Λ when the true distribution is inverse Gaussian

	Moderate Dependence							Strong Dependence						
		Naive			Proposed				Naive			Proposed		
	True	Est	SE	CP*	Est	SE	CP	True	Est	SE	CP	Est	SE	CP
θ					0.622	0.077						3.348	0.316	
$\Lambda(40)$	0.051	0.056	0.004	88.6	0.051	0.003	97.7	0.045	0.075	0.004	0	0.045	0.003	97.7
$\Lambda(50)$	0.135	0.147	0.006	82.3	0.135	0.007	99.5	0.107	0.171	0.007	0	0.103	0.007	97.5
$\Lambda(60)$	0.289	0.316	0.011	80.1	0.290	0.012	99.9	0.199	0.307	0.010	0	0.187	0.011	93.6
$\Lambda(70)$	0.532	0.577	0.019	82.3	0.530	0.021	99.7	0.323	0.476	0.016	0	0.29	0.018	83.7
$\Lambda(80)$	0.870	0.935	0.033	90.0	0.858	0.036	99.2	0.475	0.671	0.026	0	0.409	0.027	63.4

* CP: Coverage probabilities of 95% confidence intervals overlapping a window of $\pm 5\%$ around the true values.

Table S3: Summary when the true distribution is Positive Stable

	Moderate Dependence							Strong Dependence						
		Naive			Proposed				Naive			Proposed		
	True	Est	SE	CP	Est	SE	CP	True	Est	SE	CP	Est	SE	CP
theta					0.964	0.098						2.552	0.204	
Lambda(40)	0.093	0.110	0.005	35.6	0.102	0.005	88.4	0.145	0.181	0.007	0.9	0.154	0.008	95.7
Lambda(50)	0.209	0.235	0.008	56.3	0.218	0.009	97.7	0.28	0.332	0.01	2.6	0.282	0.013	99.1
Lambda(60)	0.404	0.439	0.013	79.8	0.406	0.015	99.6	0.478	0.548	0.015	15.1	0.455	0.020	96.8
Lambda(70)	0.704	0.747	0.022	94.1	0.686	0.026	99.3	0.749	0.836	0.025	48.2	0.672	0.032	74.1
Lambda(80)	1.137	1.186	0.043	97.1	1.072	0.048	93.9	1.107	1.209	0.041	79.4	0.936	0.049	36.2

* CP: Coverage probabilities of 95% confidence intervals overlapping a window of $\pm 5\%$ around the true values.

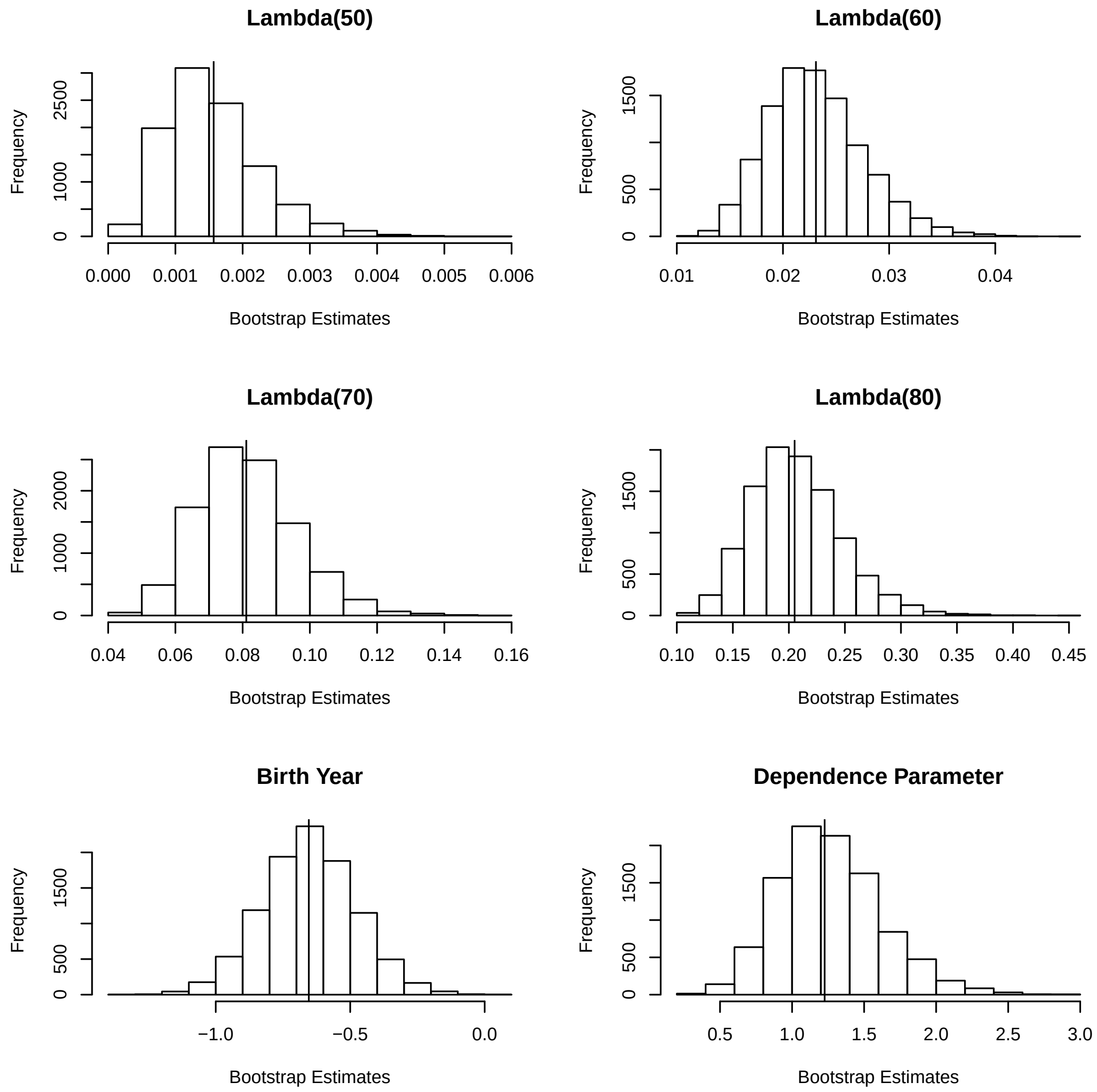
F. Exploration of Bootstrap-based SE Estimates

For the real data analysis of the prostate cancer study, we increased the number of the weighted bootstrap samples to 10,000 to evaluate the stability of the standard error estimates. Figure S1 shows the histogram of bootstrap estimates of hazard function at selected age 50, 60, 70, and 80 years old, as well as bootstrap estimates of birth cohort and the dependence parameter. The distributions of these estimates are roughly normal with slight skewness to right for $\Lambda(50)$, showing that bootstrap-based standard deviation estimator is likely to be a good approximation to true SE. We also evaluated the stability of bootstrap SE estimates over 500, 1000, 2000, and 4000 (non-overlapping) bootstrap samples (Table S4). It can be seen that there is very little variation across the number of bootstrap samples.

Table S4: Standard deviations of bootstrap estimates of hazard function at selected ages, birth cohort, and dependence parameter

	# of Bootstrap Samples			
	500	1000	2000	4000
$\Lambda(50)$	0.001	0.001	0.001	0.001
$\Lambda(60)$	0.005	0.005	0.004	0.005
$\Lambda(70)$	0.015	0.015	0.014	0.015
$\Lambda(80)$	0.041	0.040	0.040	0.041
Birth Cohort	0.183	0.176	0.172	0.172
Dependence	0.358	0.359	0.354	0.357

Figure S1: Histograms of bootstrap estimates of $\widehat{\Lambda}(t)$ at $t = 50, 60, 70$, and 80, and parameters for birth cohort and dependence.



References

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Stat.* **40**, 1187-1195.

Serfling, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Stat.* **41**, 1227-1234.

Van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press.