Supplementary Materials to
"Efficient Estimation of Semiparametric Transformation Models for the Cumulative Incidence of Competing Risks"

Lu Mao and D. Y. Lin†

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, USA

S.1 Recursive formula for computing the profile likelihood

We take the derivative of the log-likelihood in (4) with respect to d_{kj} to obtain

$$\frac{\partial l_n(\boldsymbol{\beta}, \boldsymbol{\Lambda})}{\partial d_{kj}} = d_{kj}^{-1} + \sum_{l=j}^{m_k} e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_{(kl)}} H_k'\big(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_{(kl)}} \Lambda_{(kl)}\big) \\ - \sum_{t_l \geq t_{kj}} \sum_{t_l \leq \widetilde{T}_i < t_{l+1}} I(\widetilde{D}_i = 0) \frac{\exp\Big\{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i + H_k'\big(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i} \Lambda_{kl}\big)\Big\}}{\sum_{k=1}^K \exp\big\{ - G_k(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i} \Lambda_{kl})\big\} - K + 1}. \qquad \text{(S.1)}$$

Setting (S.1) to 0 for d_{kj} and $d_{k,j+1}$, we obtain

$$d_{k,j+1}^{-1} = d_{kj}^{-1} + e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_{(kj)}} H_k'\big(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_{(kj)}} \Lambda_{(kj)}\big) \\ - \sum_{t_{kj} \leq t_l < t_{k,j+1}} \sum_{t_l \leq \widetilde{T}_i < t_{l+1}} I(\widetilde{D}_i = 0) \frac{\exp\Big\{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i + H_k'\big(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i} \Lambda_{kl}\big)\Big\}}{\sum_{k=1}^K \exp\big\{ - G_k(e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}_i} \Lambda_{kl})\big\} - K + 1}. \qquad \text{(S.2)}$$

Because the second term on the right side of (S.2) involves only the d_l's with $t_l < t_{k,j+1}$, this equation indeed defines a recursive formula starting with $d_{k1}, k = 1, \cdots, K$.

S.2. Proof of Theorem 1

Let $\mathbb{P}_n$ denote the empirical measure and P the underlying probability measure. Write $N_k(t) = I(\widetilde{T} \leq t, \xi\widetilde{D} = k)$ $(k = 0, 1, \cdots, K)$ and $\widetilde{N}(t) = I(\widetilde{T} \leq t, \xi = 0)$. The proof consists of three major steps.

†*Address for correspondence:* D. Y. Lin, Department of Biostatistics, CB #7420, University of North Carolina, Chapel Hill NC 27599-7420, USA.
E-mail: lin@bios.unc.edu

Step 1. We show that for large n, the NPMLE exists, or equivalently, $\widehat{\Lambda}_k(\tau) < \infty$. The log-likelihood function is

$$
\begin{aligned}
l_n(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = {}& \mathbb{P}_n \sum_{k=1}^{K} \int_0^\tau \left[H_k \left\{ \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} + \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(t) + \log \Lambda_k\{t\} \right] dN_k(t) \\
& + \mathbb{P}_n \int_0^\tau \log \left(\sum_{k=1}^{K} \exp \left[H_k \left\{ \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} + \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(t) \right] \Lambda_k\{t\} \right) d\widetilde{N}(t) \\
& + \mathbb{P}_n \int_0^\tau \log S(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) dN_0(t). \qquad \text{(S.3)}
\end{aligned}
$$

Conditions (C2) and (C3) imply that, for large n with probability one, there exists a subject with $\widetilde{T} = \tau$ and $\widetilde{D} = 0$. For this subject, if $\Lambda_k(\tau) = \infty$ for some k, then $S(\widetilde{T}; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \le 0$ and the corresponding term in (S.3) is $-\infty$. Thus, $\widehat{\Lambda}_k(\tau) < \infty$.

Step 2. Let $\mathcal{Z}$ be the support of $\boldsymbol{Z}$ equipped with the uniform norm. We show that for every $\boldsymbol{z} \in \mathcal{Z}$, with probability one,

$$
\liminf_n S(\tau; \boldsymbol{z}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) > 0. \qquad \text{(S.4)}
$$

Note that (S.4) implies that $\limsup_n \widehat{\Lambda}_k(\tau) < \infty$ almost surely for each k.

Suppose that $\liminf_n S(\tau; \boldsymbol{z}_0, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \le 0$ for some $\boldsymbol{z}_0$. By the continuity of S in $\boldsymbol{z}$, there exists a small neighborhood $\mathcal{Z}_0 \subset \mathcal{Z}$ of $\boldsymbol{z}_0$ with $\Pr(\boldsymbol{z} \in \mathcal{Z}_0) > 0$ such that

$$
\liminf_n \sup_{\boldsymbol{z} \in \mathcal{Z}_0} S(\tau; \boldsymbol{z}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \le 0.
$$

We take a subsequence, still indexed by n, such that $\lim_n \sup_{\boldsymbol{z} \in \mathcal{Z}_0} S(\tau; \boldsymbol{z}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \le 0$. Choose $M > 0$ such that $\sup_{t \in [0,\tau]} |\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{z}(t)| \le M$ for every $\boldsymbol{\beta}_k$ and $\boldsymbol{z} \in \mathcal{Z}$, and choose $\epsilon_0 \in (0, K^{-1})$. Define $\overline{\Lambda}_k(t) = (\widehat{\Lambda}_k(t) \wedge \widetilde{M}_k) \vee \widetilde{M}_k/2$, where $\widetilde{M}_k = e^{-M} G_k^{-1}\{-\log(1 - K^{-1} + \epsilon_0)\} > 0$. Since $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Lambda}}$ are the NPMLEs,

$$
l_n(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \ge l_n(\widehat{\boldsymbol{\beta}}, \overline{\boldsymbol{\Lambda}}), \qquad \text{(S.5)}
$$

where $\overline{\boldsymbol{\Lambda}} = (\overline{\Lambda}_1. \cdots, \overline{\Lambda}_K)$. To derive a contradiction, we will show that the left side of (S.5) goes to $-\infty$ and the right side is bounded away from $-\infty$. To this end, we will use the following inequalities

$$
K^{-1} \sum_{k=1}^{K} \log a_k \le \log \left(\sum_{k=1}^{K} a_k \right) \le \sum_{k=1}^{K} \log a_k + \log K,
$$

where $a_1, \cdots, a_K$ are any positive constants. Clearly,

$$
\begin{aligned}
l_n(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \le {}& \sum_{k=1}^{K} \left[\log \left\{ \widehat{\Lambda}_k(\tau) \sup_{y \le \widehat{\Lambda}_k(\tau) e^M} G_k'(y) \right\} + M \right] \mathbb{P}_n \left\{ N_k(\tau) + \widetilde{N}(\tau) \right\} + \log K \mathbb{P}_n \widetilde{N}(\tau) \\
& + \log \sup_{\boldsymbol{z} \in \mathcal{Z}_0} S(\tau; \boldsymbol{z}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \mathbb{P}_n(\widetilde{T} = \tau, \widetilde{D} = 0, \boldsymbol{Z} \in \mathcal{Z}_0).
\end{aligned}
$$

By (C4), we can show that the right side goes to $-\infty$. On the other hand,

$$
\begin{aligned}
l_n(\widehat{\boldsymbol{\beta}}, \overline{\boldsymbol{\Lambda}}) \geq & \sum_{k=1}^{K} \left[\log \left\{ 2^{-1} \widetilde{M}_k \inf_{y \leq \widetilde{M}_k e^M} G_k'(y) \right\} - G_k(\widetilde{M}_k e^M) - M \right] \mathbb{P}_n \{ N_k(\tau) + K^{-1} \widetilde{N}(\tau) \} \\
& + \log \epsilon_0 \mathbb{P}_n(\widetilde{T} = \tau, \widetilde{D} = 0),
\end{aligned}
$$

which is bounded away from $-\infty$. Thus, we obtain a contradiction, so (S.4) holds. It then follows from Helly's selection lemma that, along a subsequence, $\widehat{\Lambda}_k(t) \to \Lambda_k^*(t)$ weakly for some increasing function $\Lambda_k^*(t)$ and $\widehat{\boldsymbol{\beta}} \to \boldsymbol{\beta}^*$ for some vector $\boldsymbol{\beta}^*$.

Step 3. We show that $\Lambda_k^* = \Lambda_{k0}$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. Define $\Lambda_k^\epsilon(t) = \int_0^t \{1 + \epsilon h_k(s)\} d\Lambda_k(s)$ as a path through Λ_k indexed by ϵ, where $h_k \in BV_1$. By differentiating the log-likelihood of Λ_k^ϵ for a single subject with respect to ϵ at 0, we obtain the score operator for Λ_k as

$$
\begin{aligned}
\dot{l}_{2k}(\boldsymbol{\beta}, \boldsymbol{\Lambda})[h_k] = & \int_0^\tau \left[H_k' \left\{ \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} h_k(s) d\Lambda_k(s) + h_k(t) \right] d\widetilde{N}_k(t; \boldsymbol{\beta}, \boldsymbol{\Lambda}) \\
& - \int_0^\tau \left\{ \widetilde{\Psi}_k(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \int_0^t h_k(s) e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} dN_0(t), \qquad \text{(S.6)}
\end{aligned}
$$

where $\widetilde{N}_k(t; \boldsymbol{\beta}, \boldsymbol{\Lambda}) = N_k(t) + w_k(\boldsymbol{\beta}, \boldsymbol{\Lambda}) \widetilde{N}(t)$, $w_k(\boldsymbol{\beta}, \boldsymbol{\Lambda}) = F_k'(\widetilde{T}; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) / \sum_{l=1}^K F_l'(\widetilde{T}; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda})$, and $\widetilde{\Psi}_k(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) = S^{-1}(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \exp \left[H_k \left\{ \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} \right]$.

By changing the order of integrations, we re-write (S.6) as

$$
\begin{aligned}
\dot{l}_{2k}(\boldsymbol{\beta}, \boldsymbol{\Lambda})[h_k] = & \int_0^\tau h_k(s) d\widetilde{N}_k(s; \boldsymbol{\beta}, \boldsymbol{\Lambda}) \\
& + \int_0^\tau h_k(s) e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} \int_s^\tau H_k' \left\{ \int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} d\Lambda_k(s) \right\} d\widetilde{N}_k(s; \boldsymbol{\beta}, \boldsymbol{\Lambda}) d\Lambda_k(s) \\
& - \int_0^\tau h_k(s) e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(s)} \int_s^\tau \widetilde{\Psi}_k(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) dN_0(t) d\Lambda_k(s).
\end{aligned}
$$

By definition of the NPMLEs, $\mathbb{P}_n \dot{l}_{2k}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})[h_k] = 0$ for all $h_k \in BV_1$. We take $h_k(\cdot) = I(\cdot \leq t)$ to obtain

$$
\widehat{\Lambda}_k(t) = \int_0^t \frac{\mathbb{P}_n d\widetilde{N}_k(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})}{\phi_k(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})},
$$

where

$$
\begin{aligned}
\phi_k(t; \boldsymbol{\beta}, \boldsymbol{\Lambda}) = & \mathbb{P}_n e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(t)} \int_t^\tau \widetilde{\Psi}_k(s; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) dN_0(s) \\
& - \mathbb{P}_n e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(t)} \int_t^\tau H_k' \left\{ \int_0^s e^{\boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{Z}(u)} d\Lambda_k(u) \right\} d\widetilde{N}_k(s; \boldsymbol{\beta}, \boldsymbol{\Lambda}).
\end{aligned}
$$

By Step 2 and the continuity of $S(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda})$ in $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$, and with $\boldsymbol{\Lambda}$ equipped with the weak topology, there exist a neighborhood of $\boldsymbol{\beta}^*$, denoted by $\mathcal{B}$, and a neighborhood of Λ_k^*, denoted by $\mathcal{A}_k$, such that $S(t; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda})$ is uniformly bounded away from zero. Therefore, $\{\widetilde{\Psi}_k(\cdot; \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) : \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{\beta} \in$

$\mathcal{B}, \Lambda_k \in \mathcal{A}_k, k = 1, \cdots, K\}$ is a class of functions on $[0, \tau]$ that are uniformly bounded and with total variation uniformly bounded and is thus Donsker (van der Vaart and Wellner (1996), chapter 2.10). By the permanence of the Donsker property and the uniform law of large numbers,

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{A}_k,k=1,\cdots,K} |\phi_k(t;\boldsymbol{\beta},\boldsymbol{\Lambda}) - \phi_k^*(t;\boldsymbol{\beta},\boldsymbol{\Lambda})| \to 0, \tag{S.7}$$

where

$$\phi_k^*(t;\boldsymbol{\beta},\boldsymbol{\Lambda}) = Pe^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(t)}\int_t^\tau \widetilde{\Psi}_k(s;\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\Lambda})dN_0(s)$$
$$- Pe^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(t)}\int_t^\tau H_k'\left\{\int_0^s e^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(u)}d\Lambda_k(u)\right\}d\widetilde{N}_k(s;\boldsymbol{\beta},\boldsymbol{\Lambda}).$$

By (S.7) and the continuity of ϕ_k^* in $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$, we have $\phi_k(\cdot;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) \to \phi_k^*(\cdot;\boldsymbol{\beta}^*,\boldsymbol{\Lambda}^*)$ uniformly. We can show that for large n, $\phi_k(t;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$ is uniformly bounded away from 0. Furthermore, $\mathbb{P}_n d\widetilde{N}_k(\cdot;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) = \mathbb{P}_n d\widetilde{N}_k(\cdot;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) + o_p(1)$ uniformly by similar Donsker property arguments.

We define

$$\widetilde{\Lambda}_k(t) = \int_0^t \frac{\mathbb{P}_n d\widetilde{N}_k(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)}{\phi_k(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)}.$$

It follows from the uniform law of large numbers and Condition (C5) that uniformly

$$\widetilde{\Lambda}_k(t) \to \int_0^t \frac{Pd\widetilde{N}_k(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)}{\phi_k^*(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)} = \Lambda_{k0}(t).$$

By the lower bound on $\phi_k(s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$, $\widehat{\Lambda}_k$ is absolutely continuous with respect to $\widetilde{\Lambda}_k$. Furthermore, $d\widehat{\Lambda}_k/d\widetilde{\Lambda}_k$ converges uniformly to a bounded function $\eta(\cdot)$. Thus, $\Lambda_k^*(t) = \int_0^t \eta(s)d\Lambda_{k0}(s)$, and $\Lambda_k^*(t)$ is absolutely continuous with respect to the Lebesgue measure. We denote the derivatives of $\Lambda_k^*(t)$ and $\Lambda_{k0}(t)$ by $\lambda_k^*(t)$ and $\lambda_{k0}(t)$, respectively. By the uniform convergence of the log-likelihood to its expectation and that of the function $d\widehat{\Lambda}_k/d\widetilde{\Lambda}_k$ to λ_k^*/λ_{k0}, together with the Kullback-Leibler criterion,

$$\sum_{k=1}^K\int_0^\tau \log\left(\lambda_k^*(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k^*(s)\right\}+\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(t)\right]\right)dN_k(t)$$
$$+\int_0^\tau \log\left(\sum_{k=1}^K\lambda_k^*(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k^*(s)\right\}+\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(t)\right]\right)d\widetilde{N}(t)$$
$$+\int_0^\tau \log S(t;\boldsymbol{Z},\boldsymbol{\beta}^*,\boldsymbol{\Lambda}^*)dN_0(t)$$
$$=\sum_{k=1}^K\int_0^\tau \log\left(\lambda_{k0}(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_{k0}(s)\right\}+\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(t)\right]\right)dN_k(t)$$
$$+\int_0^\tau \log\left(\sum_{k=1}^K\lambda_{k0}(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_{k0}(s)\right\}+\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(t)\right]\right)d\widetilde{N}(t)$$
$$+\int_0^\tau \log S(t;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)dN_0(t)$$

almost surely. In the case that $\widetilde{T}=t$ and $\xi\widetilde{D}=k$,

$$\lambda_k^*(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k^*(s)\right\}+\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(t)\right]$$

$$=\lambda_{k0}(t)\exp\left[H_k\left\{\int_0^t e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_{k0}(s)\right\}+\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(t)\right].$$

We integrate both sides to obtain $e^{\boldsymbol{\beta}_k^{*\mathrm{T}}\boldsymbol{Z}(t)}\lambda_k^*(t)=e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(t)}\lambda_{k0}(t)$. It then follows from (C2) that $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$ and $\Lambda_k^*(t)=\Lambda_{k0}(t)$. Thus, with probability one, $\widehat{\boldsymbol{\beta}}\to\boldsymbol{\beta}_0$ and $\widehat{\Lambda}_k(t)\to\Lambda_{k0}(t)$ pointwise. The latter can be strengthened to uniform convergence since Λ_{k0} is continuous.

S.3. Proof of Theorem 2

We denote the empirical process by $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$. The score operator for Λ_k is given in (S.6). The score function for $\boldsymbol{\beta}$ is $\dot{\boldsymbol{l}}_1\equiv(\dot{\boldsymbol{l}}_{11}^{\mathrm{T}},\cdots,\dot{\boldsymbol{l}}_{1K}^{\mathrm{T}})^{\mathrm{T}}$, where

$$\dot{\boldsymbol{l}}_{1k}(\boldsymbol{\beta},\boldsymbol{\Lambda})=\int_0^\tau H_k'\left\{\int_0^t e^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k(s)\right\}\int_0^t\boldsymbol{Z}(s)e^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k(s)d\widetilde{N}_k(t;\boldsymbol{\beta},\boldsymbol{\Lambda})+\int_0^\tau\boldsymbol{Z}(t)d\widetilde{N}_k(t;\boldsymbol{\beta},\boldsymbol{\Lambda})$$
$$-\int_0^\tau\widetilde{\Psi}_k(t;\boldsymbol{Z},\boldsymbol{\beta},\boldsymbol{\Lambda})\int_0^t\boldsymbol{Z}(s)e^{\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_k(s)dN_0(t).$$

For $\delta>0$ sufficiently small, the class of functions

$$\left\{\dot{\boldsymbol{l}}_1(\boldsymbol{\beta},\boldsymbol{\Lambda}),\dot{l}_{2k}(\boldsymbol{\beta},\boldsymbol{\Lambda})[h_k]:||\boldsymbol{\beta}-\boldsymbol{\beta}_0||+\sum_{k=1}^K\sup_{t\in[0,\tau]}|\Lambda_k(t)-\Lambda_{k0}(t)|<\delta,h_k\in BV_1,k=1,\cdots,K\right\}$$

is Donsker. Thus, by the consistency of $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$, the continuity of the score functions in the parameters, and the dominated convergence theorem,

$$\mathbb{G}_n\left\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})+\sum_{k=1}^K\dot{l}_{2k}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})[h_k]\right\}=\mathbb{G}_n\left\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)+\sum_{k=1}^K\dot{l}_{2k}(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)[h_k]\right\}+o_p(1)\qquad(\mathrm{S.8})$$

uniformly in $(\boldsymbol{v},\boldsymbol{w})$. It remains to show that the map $W:l^\infty(\mathcal{V}\times\mathcal{W})\to l^\infty(\mathcal{V}\times\mathcal{W})$ given by

$$W(\boldsymbol{\beta},\boldsymbol{\Lambda})[\boldsymbol{v},\boldsymbol{w}]=P\left\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\boldsymbol{\beta},\boldsymbol{\Lambda})+\sum_{k=1}^K\dot{l}_{2k}(\boldsymbol{\beta},\boldsymbol{\Lambda})[h_k]\right\}$$

is Fréchet differentiable at $(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)$ with a derivative that is continuously invertible.

It is straightforward to show that

$$\left.\frac{\partial}{\partial\epsilon}\right|_{\epsilon=0}W\left(\boldsymbol{\beta}_0+\epsilon\widetilde{\boldsymbol{v}},\boldsymbol{\Lambda}_0+\epsilon\int\widetilde{\boldsymbol{w}}d\boldsymbol{\Lambda}_0\right)=\widetilde{\boldsymbol{v}}^{\mathrm{T}}\boldsymbol{B}_1[\boldsymbol{v},\boldsymbol{w}]+\sum_{k=1}^K\int B_{2k}[\boldsymbol{v},\boldsymbol{w}]\tilde{h}_k d\Lambda_{k0},\qquad(\mathrm{S.9})$$

where the operator $(\boldsymbol{B}_1, B_{21}, \ldots, B_{2K})[\boldsymbol{v}, \boldsymbol{w}]$ can be expressed as

$$-\begin{pmatrix} \boldsymbol{v} \\ \phi_1^*(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)h_1(t) \\ \vdots \\ \phi_K^*(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)h_K(t) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)\boldsymbol{v} + \sum_{k=1}^K \int h_k(t)\boldsymbol{\nu}_{1k}(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)d\Lambda_{k0}(t) + \boldsymbol{v} \\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\zeta}_{21}(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) + \sum_{j=1}^K \int h_j(s)\nu_{21j}(s,t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)d\Lambda_{j0}(s) \\ \vdots \\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\zeta}_{2K}(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) + \sum_{j=1}^K \int h_j(s)\nu_{2Kj}(s,t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)d\Lambda_{j0}(s) \end{pmatrix}, \tag{S.10}$$

$\boldsymbol{\zeta}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) = -P\{\dot{\boldsymbol{l}}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)\dot{\boldsymbol{l}}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)^{\mathrm{T}}\}$, and $\boldsymbol{\zeta}_{2k}, \boldsymbol{\nu}_{1k}$ and ν_{2kj} are certain functions. We show that the operator $\boldsymbol{B} \equiv (\boldsymbol{B}_1, B_{21}, \ldots, B_{2K})$ is invertible on its range.

In light of Theorem 1,

$$\phi_k^*(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) = \frac{Pd\widetilde{N}_k(t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)/dt}{\lambda_{k0}(t)} > 0$$

under Conditions (C1) and (C3). Thus, the first term in (S.10) is an invertible operator. Because the second term is a compact operator, it suffices to show that the operator $\boldsymbol{B}$ is one-to-one (Rudin (1973), pages 99-103). Suppose that for some $(\boldsymbol{v},\boldsymbol{w}) \in \mathcal{V}\times\mathcal{W}$, $\boldsymbol{B}(\boldsymbol{v},\boldsymbol{w}) = \boldsymbol{0}$. We then wish to show that $(\boldsymbol{v},\boldsymbol{w}) = \boldsymbol{0}$. By (S.9),

$$P\left(\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\boldsymbol{\beta},\boldsymbol{\Lambda}) + \sum_{k=1}^K \dot{l}_{2k}(\boldsymbol{\beta},\boldsymbol{\Lambda})[h_k]\right)^2 = -\left.\frac{\partial}{\partial\epsilon}\right|_{\epsilon=0} W\left(\boldsymbol{\beta}_0 + \epsilon\boldsymbol{v}, \boldsymbol{\Lambda}_0 + \epsilon\int \boldsymbol{w}d\boldsymbol{\Lambda}_0\right) = 0.$$

Thus, with probability one,

$$\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\boldsymbol{\beta},\boldsymbol{\Lambda}) + \sum_{k=1}^K \dot{l}_{2k}(\boldsymbol{\beta},\boldsymbol{\Lambda})[h_k] = 0. \tag{S.11}$$

Let $\boldsymbol{v} = (\boldsymbol{v}_1^{\mathrm{T}}, \cdots, \boldsymbol{v}_K^{\mathrm{T}})^{\mathrm{T}}$, and take $dN_k(t) = 1$. It follows from (S.11) that

$$h_k(t) + \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{Z}(t) = -\left[\int_0^t \left\{h_k(s) + \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{Z}(s)\right\} e^{\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_{k0}(s)\right] H_k'\left\{\int_0^t e^{\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{Z}(s)}d\Lambda_{k0}(s)\right\},$$

which is a homogeneous integral equation of $h_k(t) + \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{Z}(t)$ with 0 as the only solution. Thus, it follows from (C2) that $\boldsymbol{v}_k = \boldsymbol{0}$ and $h_k(\cdot) = 0$. Therefore, $\boldsymbol{B}$ is one-to-one and thus invertible. Consequently, the derivative of W is continuously invertible.

For $(\boldsymbol{v},\boldsymbol{w}) \in \mathcal{V}\times\mathcal{W}$, denote $(\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{w}}) = \boldsymbol{B}^{-1}(\boldsymbol{v},\boldsymbol{w})$. It follows from (S.8) that

$$\sqrt{n}\Big\{\boldsymbol{v}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sum_{k=1}^K \int_0^\tau h_k d(\widehat{\Lambda}_k - \Lambda_{k0})\Big\} = -\mathbb{G}_n\Big\{\widetilde{\boldsymbol{v}}^{\mathrm{T}}\dot{\boldsymbol{l}}_1(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) + \sum_{k=1}^K \dot{l}_{2k}(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)[\widetilde{h}_k]\Big\} + o_p(1) \tag{S.12}$$

uniformly in $(\boldsymbol{v},\boldsymbol{w})$. Thus, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}_0)$ is asymptotically Gaussian. Take $h_k = 0$ and $\boldsymbol{v}$ to be the unit coordinate vectors in (S.12) to find that the influence function of $\widehat{\boldsymbol{\beta}}$ lies in the tangent space, i.e., the closed linear span of the score functions. By the semiparametric efficiency theory (Bickel *et al.*, 1993), $\widehat{\boldsymbol{\beta}}$ is semiparametric efficient.

S.4 Profile likelihood and information matrix

We justify the use of the negative inverse of the second derivative of the profile log-likelihood to estimate the covariance matrix of $\widehat{\boldsymbol{\beta}}$ by verifying the conditions in Theorem 1 of Murphy and van der Vaart (2000). From the proof of Theorem 2, the invertibility of the whole information operator implies the invertibility of the information operator for $\boldsymbol{\Lambda}$. This ensures that there is a "least favorable direction" $\boldsymbol{h}_k$, which is a vector of dimension p with components in BV_1 such that the parametric model $\boldsymbol{\epsilon} \to \big(\boldsymbol{\epsilon}, (\Lambda_{1\boldsymbol{\epsilon}}, \cdots, \Lambda_{K\boldsymbol{\epsilon}})\big)$ with $d\Lambda_{k\boldsymbol{\epsilon}} = \big\{1 + (\boldsymbol{\epsilon} - \boldsymbol{\beta}_0)^{\mathrm{T}}\boldsymbol{h}_k\big\}d\Lambda_{k0}$ is a least favorable submodel. Given $\widetilde{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}_0$, let $\widehat{\boldsymbol{\Lambda}}_{\widetilde{\boldsymbol{\beta}}} \equiv \big(\widehat{\Lambda}_{1\widetilde{\boldsymbol{\beta}}}, \cdots, \widehat{\Lambda}_{K\widetilde{\boldsymbol{\beta}}}\big)$ denote the maximizer of $l_n(\widetilde{\boldsymbol{\beta}}, \cdot)$. We can show that

$$\sup_{t\in[0,\tau]} \sum_{k=1}^{K} \left|\widehat{\Lambda}_{k\widetilde{\boldsymbol{\beta}}}(t) - \Lambda_{k0}(t)\right| = O_p(||\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| + n^{-1/2})$$

by using the arguments in the proofs of Theorems 1 and 2 and the smoothness of the likelihood in $\boldsymbol{\beta}$. Then the no-bias condition follows from the smoothness property of the score functions in the parameters. Finally, we can verify the Donsker properties of the first and second derivatives of the least favorable submodel log-likelihood since functions of uniformly bounded variation are Donsker.

For the use of the inverse information matrix for $\boldsymbol{\beta}$ and d_{kj}'s as an estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Lambda}}$, the justification is similar to Theorem 3 of Zeng and Lin (2007). We note that the first and second derivatives of our log-likelihood are smooth on a neighborhood of $(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)$.

S.5. Technical details of model checking procedures

Using the (functional) delta method, we will show that $W_{kn}(\boldsymbol{x}, t)$ is asymptotically equivalent to

$$\begin{aligned}\widetilde{W}_{kn}(\boldsymbol{x}, t) = \mathbb{G}_n\Bigg\{ &\int_0^t f(\boldsymbol{x}, u; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)dM_k(u; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) \\ &+ (\dot{\boldsymbol{l}}_1, \dot{\boldsymbol{l}}_2)\boldsymbol{B}^{-1}P\Big(\boldsymbol{g}_{k11}(\boldsymbol{x}, t), \cdots, \boldsymbol{g}_{k1K}(\boldsymbol{x}, t), g_{k21}(\cdot, \boldsymbol{x}, t), \cdots, g_{k2K}(\cdot, \boldsymbol{x}, t)\Big)\Bigg\} \qquad \text{(S.13)}\end{aligned}$$

uniformly in $\boldsymbol{x}$ and t, where $\dot{\boldsymbol{l}}_1$ is the score function for $\boldsymbol{\beta}$, $\dot{\boldsymbol{l}}_2$ is the score operator for $\boldsymbol{\Lambda}$, and $\boldsymbol{B}$ is the information operator, all evaluated at $(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)$,

$$\begin{aligned}\boldsymbol{g}_{k1j}(\boldsymbol{x}, t) = -\int_0^t & f(\boldsymbol{x}, s; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)Y(s)\widetilde{\Psi}_j(s; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\int_0^s \boldsymbol{Z}(u)e^{\boldsymbol{\beta}_{j0}^{\mathrm{T}}\boldsymbol{Z}(u)}d\Lambda_{j0}(u) \\ &\times \Psi_k(s; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)d\Lambda_{k0}(s) - I(j = k)\int_0^t f(\boldsymbol{x}, s; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\Bigg[\boldsymbol{Z}(s) + \\ &\times H_k'\left\{\int_0^s e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(u)}d\Lambda_{k0}(u)\right\}\int_0^s \boldsymbol{Z}(u)e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}(u)}d\Lambda_{k0}(u)\Bigg]Y(s)\Psi_k(s; \boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)d\Lambda_{k0}(s),\end{aligned}$$

and

$$g_{k2j}(u,\boldsymbol{x},t) = -\int_0^t I(s \geq u) f(\boldsymbol{x},s;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) Y(s) \Psi_k(s;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) \exp\{\boldsymbol{\beta}_{k0}^{\mathrm{T}} \boldsymbol{Z}(u)\}$$
$$\times \left[\widetilde{\Psi}_j(s;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) + I(j=k) H_k' \left\{ \int_0^s e^{\boldsymbol{\beta}_{k0}^{\mathrm{T}} \boldsymbol{Z}(\tilde{u})} d\Lambda_{k0}(\tilde{u}) \right\} \right] d\Lambda_{k0}(s)$$
$$- I(j=k) I(u \leq t) f(\boldsymbol{x},u;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0) Y(u) \Psi_k(u;\boldsymbol{Z},\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0).$$

We replace the unknown quantities in $\widetilde{W}_{kn}$ by their empirical counterparts. Specifically, we estimate the functions $\boldsymbol{g}_{k1j}$ and g_{k2j} by replacing $(\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)$ with $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$. Denote the resulting expressions by $\widehat{\boldsymbol{g}}_{k1j}$ and $\widehat{g}_{k2j}$. Recall from Section 2.2 that $t_1,\cdots,t_m$ are the distinct failure times and $\delta_1,\cdots,\delta_m$ are the corresponding failure types. We treat the jump sizes of $\boldsymbol{\Lambda}$ at the t_j's as Euclidean parameters, and, along with $\boldsymbol{\beta}$, we calculate the score vector for the ith subject, denoted by $\widetilde{\boldsymbol{l}}_i$, and the information matrix $\mathcal{I}_n$. Let $\widetilde{\boldsymbol{g}}_{k1n}(\boldsymbol{x},t) = \mathbb{P}_n\big(\widehat{\boldsymbol{g}}_{k11}(\boldsymbol{x},t)^{\mathrm{T}},\cdots,\widehat{\boldsymbol{g}}_{k1K}(\boldsymbol{x},t)^{\mathrm{T}}\big)^{\mathrm{T}}$ and $\widetilde{\boldsymbol{g}}_{k2n}(\boldsymbol{x},t) = \mathbb{P}_n\big(\widehat{g}_{k2\delta_1}(t_1,\boldsymbol{x},t),\cdots,\widehat{g}_{k2\delta_m}(t_m,\boldsymbol{x},t)\big)^{\mathrm{T}}$. Also, let $\widehat{S}_i(\boldsymbol{x},t) = \widetilde{\boldsymbol{l}}_i^{\mathrm{T}} \mathcal{I}_n^{-1}\big(\widetilde{\boldsymbol{g}}_{k1n}(\boldsymbol{x},t)^{\mathrm{T}},\widetilde{\boldsymbol{g}}_{k2n}(\boldsymbol{x},t)^{\mathrm{T}}\big)^{\mathrm{T}}$. Then we obtain

$$\widehat{W}_{kn}(\boldsymbol{x},t) = n^{-1/2}\sum_{i=1}^n \left\{ \int_0^t f(\boldsymbol{x},u;\boldsymbol{Z}_i,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) dM_{ki}(u;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) + \widehat{S}_i(\boldsymbol{x},t) \right\} Q_i,$$

as given in Section 2.4.

Let $\mathcal{X}$ denote the space of $\boldsymbol{x}$, and consider W_{kn} and $\widehat{W}_{kn}$ as random elements in $l^\infty(\mathcal{X}\times[0,\tau])$. In addition, let BL_1 be the space of Lipschitz functions on $l^\infty(\mathcal{X}\times[0,\tau])$ that are uniformly bounded by 1 and with Lipschitz norm bounded by 1. It is convenient to metrize the laws on $l^\infty(\mathcal{X}\times[0,\tau])$ by $\rho(\mathbb{Z}_1,\mathbb{Z}_2) = \sup_{h\in BL_1}|Eh(\mathbb{Z}_1) - Eh(\mathbb{Z}_2)|$, where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are random elements in $l^\infty(\mathcal{X}\times[0,\tau])$ (van der Vaart and Wellner 1996). We impose the following regularity conditions on the function $f(\boldsymbol{x},t;\boldsymbol{\beta},\boldsymbol{\Lambda})$, whose dependence on $\boldsymbol{Z}$ is suppressed for notational simplicity.

(D1) For some $\delta > 0$, the class of functions

$$\left\{ f(\boldsymbol{x},t;\boldsymbol{\beta},\boldsymbol{\Lambda}) : \boldsymbol{x}\in\mathcal{X}, t\in[0,\tau], ||\boldsymbol{\beta}-\boldsymbol{\beta}_0|| + \sum_{k=1}^K \sup_{s\in[0,\tau]} |\Lambda_k(s) - \Lambda_{k0}(s)| < \delta \right\}$$

is a uniformly bounded P-Donsker class.

(D2) There exists a constant $M > 0$ such that, with probability one, the total variation of $f(\boldsymbol{x},\cdot;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)$ is bounded by M for all $\boldsymbol{x}\in\mathcal{X}$.

(D3) For all $(\boldsymbol{\beta},\boldsymbol{\Lambda})$ such that $\boldsymbol{\beta}\to\boldsymbol{\beta}_0$ and $\boldsymbol{\Lambda}\to\boldsymbol{\Lambda}_0$,

$$\sup_{\boldsymbol{x},t} E|f(\boldsymbol{x},t;\boldsymbol{\beta},\boldsymbol{\Lambda}) - f(\boldsymbol{x},t;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)| \to 0.$$

Remark S.1. Conditions (D1) and (D2) are satisfied by all processes considered in the main text. Condition (D3) is satisfied by W_{kc}, W_{kp}, and W_{ko} and is also satisfied by W_{kl} and W_{ktr} if there is at least one continuous covariate.

Theorem S.1. Under Conditions (C1)-(C4) and (D1)-(D3),

$$\sup_{h \in BL_1} |E_{\boldsymbol{Q}} h(\widehat{W}_{kn}) - Eh(W_{kn})| \longrightarrow 0$$

almost surely, where $E_{\boldsymbol{Q}}$ denotes expectation with respect to $\boldsymbol{Q}$.

Proof. Our main task is to show that

$$W_{kn} = \widetilde{W}_{kn} + o_p(1) \qquad \text{in } l^\infty(\mathcal{X} \times [0, \tau]). \tag{S.14}$$

Then the conditional distribution of $\widehat{W}_{kn}$ can be shown to be asymptotically the same as the distribution of $\widetilde{W}_{kn}$ by using the uniform central limit theorem (van der Vaart and Wellner 1996, Thm 2.11.1).

To show (S.14), we define

$$\Lambda_{kc}(t; \boldsymbol{\beta}, \boldsymbol{\Lambda}) = \int_0^t \Psi_k(s; \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) d\Lambda_k(s).$$

Then $M_k(t; \boldsymbol{\beta}, \boldsymbol{\Lambda}) = N_k(t) - \int_0^t Y(s) d\Lambda_{kc}(s; \boldsymbol{\beta}, \boldsymbol{\Lambda})$. Clearly,

$$\begin{aligned} W_{kn} &= \mathbb{G}_n \int_0^t f(\boldsymbol{x}, s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) dM_k(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) \\ &\quad + \sqrt{n} P\left[\int_0^t f(\boldsymbol{x}, s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) Y(s) d\{\Lambda_{kc}(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) - \Lambda_{kc}(s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}\right] \\ &\quad + \sqrt{n} P\left[\int_0^t \{f(\boldsymbol{x}, s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) - f(\boldsymbol{x}, s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})\} Y(s) d\{\Lambda_{kc}(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) - \Lambda_{kc}(s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}\right]. \end{aligned} \tag{S.15}$$

Because $\Lambda_{kc}(\cdot; \boldsymbol{\beta}, \boldsymbol{\Lambda})$ is a Hadamard differentiable function of $(\boldsymbol{\beta}, \boldsymbol{\Lambda})$, for almost every $\boldsymbol{Z}$, $\sqrt{n}\{\Lambda_{kc}(\cdot; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) - \Lambda_{kc}(\cdot; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)\}$ converges to a zero-mean Gaussian process on $[0, \tau]$. Then, by Conditions (D2) and (D3), the third term on the right side of (S.15) is $o_p(1)$. By the delta method and the linearization result on $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}})$ given in the proof of Theorem 2, the second term is asymptotically linear in the second term on the right side of (S.13). The proof is complete if we can show that

$$\mathbb{G}_n \int_0^t f(\boldsymbol{x}, s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) dM_k(s; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Lambda}}) = \mathbb{G}_n \int_0^t f(\boldsymbol{x}, s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) dM_k(s; \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0) + o_p(1)$$

uniformly in $\boldsymbol{x}$ and t. By Conditions (D1) and (D2), together with the permanence of the Donsker property, the class of functions

$$\left\{\int_0^t f(\boldsymbol{x}, s; \boldsymbol{\beta}, \boldsymbol{\Lambda}) dM_k(s; \boldsymbol{\beta}, \boldsymbol{\Lambda}) : \boldsymbol{x} \in \mathcal{X}, t \in [0, \tau], ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| + \sum_{k=1}^K \sup_{s \in [0,\tau]} |\Lambda_k(s) - \Lambda_{k0}(s)| < \delta\right\}$$

is Donsker. Thus, it suffices to show that

$$\sup_{\boldsymbol{x},t} P\left(\int_0^t f(\boldsymbol{x},s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})dM_k(s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) - \int_0^t f(\boldsymbol{x},s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)dM_k(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)\right)^2 \to 0 \tag{S.16}$$

in probability. Note that

$$\begin{aligned}
&P\left|\int_0^t f(\boldsymbol{x},s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})dM_k(s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) - \int_0^t f(\boldsymbol{x},s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)dM_k(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)\right| \\
\le&\int_0^t P|f(\boldsymbol{x},s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) - f(\boldsymbol{x},s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)||dM_k(s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})| \\
&+P\left|\int_0^t f(\boldsymbol{x},s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)d\{\Lambda_{kc}(s;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}}) - \Lambda_{kc}(s;\boldsymbol{\beta}_0,\boldsymbol{\Lambda}_0)\}\right|.
\end{aligned} \tag{S.17}$$

The second term on the right side of (S.17) is uniformly $o_p(1)$. For the first term, note that from Step 2 in the proof of Theorem 1, for every $\boldsymbol{Z}$, $\Psi_k(s;\boldsymbol{Z},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$ is uniformly bounded as $n\to\infty$. Thus, the total variation of $M_k(\cdot;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Lambda}})$ is uniformly bounded in n for every $\boldsymbol{Z}$. Therefore, Condition (D3) implies that the first term is uniformly $o_p(1)$. Hence, (S.16) follows from the uniform boundedness of f. This completes the proof of (S.14).

To show that the conditional distribution of $\widehat{W}_{kn}$ is asymptotically equivalent to the distribution of $\widetilde{W}_{kn}$, we appeal to Theorem 2.11.1 of van der Vaart and Wellner (1996). With $\widehat{W}_{kn}$ playing the role of $\sum_{i=1}^{m_n} Z_{ni}$ in that theorem, where the Q_i are the random quantities, it suffices to show that $\widehat{W}_{kn}$ converges to a Gaussian process indexed by $(\boldsymbol{x},t,\boldsymbol{\beta},\boldsymbol{\Lambda})$. The ρ function in that theorem can be chosen to be the sum of the distances of the four components of the index. Then, the first and second displays of that theorem follow from the continuity of the influence function of $\widetilde{W}_{kn}$ in the index. The entropy integral condition (2.11.2) and the (almost sure) convergence of the covariance function are direct consequences of $\widehat{W}_{kn}$ being a multiplier Gaussian process.

We state below the consistency results on the supremum tests. We omit the proofs, which can be obtained by extending the arguments of Chen et al. (2012). As in Chen et al. (2012), we assume that covariates are time-independent.

(a) *Omnibus tests.* The results in (i) and (ii) pertain to the goodness-of-fit tests for a particular risk and all risks, respectively.

  (i) The test $\sup_{\boldsymbol{z},t}|W_{ko}(\boldsymbol{z},t)|$ is consistent against any alternative hypothesis such that there do not exist $\boldsymbol{\beta}_k$ and $c>0$ such that $\Lambda_k'(0;\boldsymbol{Z}) = c\exp(\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z})$.

  (ii) The test $\max_{1\le k\le K}\sup_{\boldsymbol{z},t}|W_{ko}(\boldsymbol{z},t)|$ is consistent against any alternative hypothesis such that there do not exist $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$ such that $\Lambda_k(t;\boldsymbol{Z}) = G_k(\exp(\boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{Z})\,\Lambda_k(t))$ for all k, t, and $\boldsymbol{Z}$.

(b) *Functional form of covariates.* Assume that the components of $\boldsymbol{Z}$ are independent. The test $\sup_{z,t} |W_{kc}^{(j)}(z,t)|$ is consistent against any alternative such that $\Lambda_k(t;\boldsymbol{Z}) = G_k(\exp(\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}^{(-j)})$ $g(Z_j)\Lambda_{k0}(t))$ for some $\boldsymbol{\beta}_{k0}$ and Λ_{k0}, where $\boldsymbol{Z}^{(-j)}$ is the covariate vector with Z_j removed, and g is not an exponential function.

(c) *Link function.* Assume that for any $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, $E\{g(\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}))|\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\boldsymbol{Z})\} = c_0\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\boldsymbol{Z})$ for some $c_0 > 0$ implies that $g(x) = cx^a$ for some constants a and c. Then the test $\sup_{x,t}|W_{kl}(x,t)|$ is consistent against any alternative that $\Lambda_k(t;\boldsymbol{Z}) = G_k(g(\exp(\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z}))\ \Lambda_{k0}(t))$ for some $\boldsymbol{\beta}_{k0}$ and Λ_{k0}, where $g(x)$ is not a mononomial function in the form of cx^a.

(d) *Proportionality.* Assume that Z is binary and that $xG_k''(x)/G_k'(x) \neq -1$. Then the test $\sup_t |W_{kp}(t)|$ is consistent against any alternative such that $\Lambda_k(t;Z) = G_k(\exp(\beta_k(t)Z)\ \Lambda_k(t))$ with $\beta_k'(0) \neq 0$.

(e) *Transformation function.* Assume that for any $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, $E\{g(\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}))|\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\boldsymbol{Z})\} = \exp(\boldsymbol{\beta}_2^{\mathrm{T}}\boldsymbol{Z})$ implies that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Then the supremum test $\sup_{x,t}|W_{ktr}(x,t)|$ is consistent against any alternative such that $\Lambda_k(t;\boldsymbol{Z}) = G_{k0}(\exp(\boldsymbol{\beta}_{k0}^{\mathrm{T}}\boldsymbol{Z})\Lambda_{k0}(t))$ for some $\boldsymbol{\beta}_{k0}$, Λ_{k0}, and G_{k0}, where G_{k0} is different from the adopted transformation G_k.

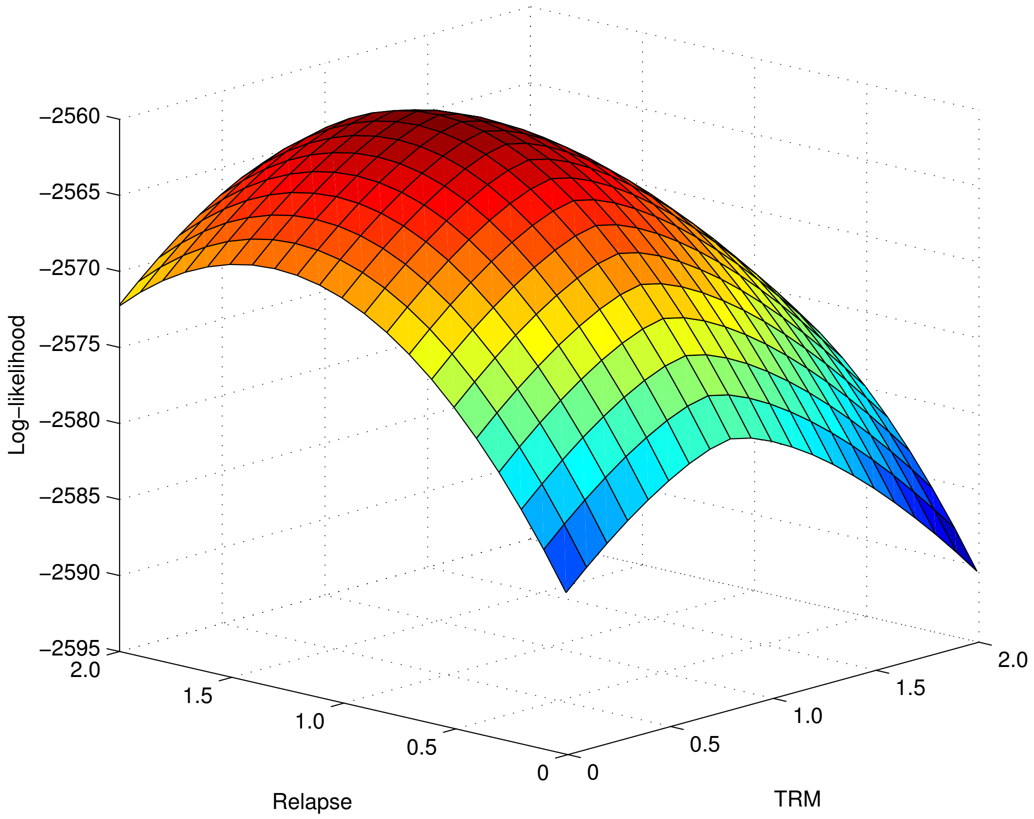

Fig. S.1. Log-likelihood surface for pairs of transformation functions $G_k(x) = r^{-1}\log(1+rx)$ $(k=1,2)$ in the analysis of the bone marrow transplantation data.

Table S.1. Simulation results on the NPMLE for the regression parameters under transformation models†

			β_{11}				β_{12}			
n	r	(β_{11}, β_{12})	Bias	SE	SEE	CP	Bias	SE	SEE	CP
100	0	(0, 0)	0.003	0.395	0.391	0.945	0.008	0.608	0.608	0.950
		(0, 0.5)	−0.008	0.398	0.393	0.945	−0.002	0.609	0.605	0.943
		(0.5, 0.5)	−0.001	0.401	0.397	0.945	−0.001	0.611	0.608	0.946
	0.5	(0, 0)	0.004	0.403	0.401	0.948	0.008	0.620	0.626	0.957
		(0, 0.5)	0.005	0.407	0.405	0.946	0.003	0.621	0.619	0.947
		(0.5, 0.5)	0.000	0.412	0.417	0.956	−0.001	0.623	0.623	0.953
	1	(0, 0)	0.005	0.411	0.408	0.945	−0.004	0.633	0.633	0.952
		(0, 0.5)	−0.007	0.414	0.408	0.944	−0.007	0.636	0.641	0.958
		(0.5, 0.5)	−0.007	0.417	0.418	0.951	0.004	0.639	0.633	0.942
200	0	(0, 0)	−0.005	0.269	0.272	0.955	−0.006	0.464	0.464	0.950
		(0, 0.5)	−0.001	0.273	0.278	0.960	−0.007	0.468	0.466	0.948
		(0.5, 0.5)	0.005	0.274	0.274	0.951	−0.001	0.469	0.467	0.946
	0.5	(0, 0)	−0.006	0.275	0.274	0.948	−0.005	0.473	0.470	0.946
		(0, 0.5)	−0.004	0.279	0.281	0.953	0.005	0.474	0.473	0.947
		(0.5, 0.5)	−0.006	0.281	0.281	0.952	0.005	0.476	0.470	0.943
	1	(0, 0)	−0.002	0.281	0.283	0.954	0.007	0.483	0.488	0.956
		(0, 0.5)	0.002	0.286	0.288	0.954	−0.004	0.487	0.484	0.946
		(0.5, 0.5)	−0.002	0.289	0.285	0.943	−0.003	0.489	0.494	0.959
500	0	(0, 0)	−0.003	0.163	0.165	0.953	0.002	0.285	0.283	0.947
		(0, 0.5)	−0.005	0.164	0.167	0.957	0.001	0.288	0.283	0.944
		(0.5, 0.5)	0.003	0.167	0.164	0.946	0.005	0.293	0.288	0.941
	0.5	(0, 0)	−0.004	0.165	0.163	0.945	−0.002	0.291	0.294	0.954
		(0, 0.5)	0.000	0.166	0.166	0.954	0.003	0.294	0.290	0.945
		(0.5, 0.5)	0.003	0.166	0.166	0.949	0.004	0.295	0.290	0.945
	1	(0, 0)	−0.002	0.169	0.169	0.951	−0.002	0.297	0.301	0.955
		(0, 0.5)	−0.002	0.173	0.174	0.953	0.000	0.297	0.303	0.955
		(0.5, 0.5)	0.001	0.178	0.179	0.954	−0.004	0.301	0.302	0.951

†See the note to Table 1.

Table S.2. Simulation results on the estimation of the cumulative hazard function under transformation models

n	r	t	$\Lambda_1(t)$	Bias	SE	SEE	CP
100	0	1	0.059	0.000	0.030	0.029	0.957
		2	0.081	0.001	0.028	0.028	0.948
	0.5	1	0.059	0.001	0.026	0.022	0.960
		2	0.081	0.000	0.030	0.030	0.952
	1	1	0.059	0.000	0.024	0.026	0.943
		2	0.081	0.000	0.033	0.030	0.951
200	0	1	0.059	−0.001	0.017	0.019	0.952
		2	0.081	0.000	0.022	0.021	0.950
	0.5	1	0.059	−0.002	0.017	0.017	0.950
		2	0.081	0.001	0.025	0.025	0.955
	1	1	0.059	0.001	0.018	0.018	0.949
		2	0.081	0.002	0.023	0.022	0.946
500	0	1	0.059	0.000	0.010	0.011	0.956
		2	0.081	0.000	0.012	0.012	0.944
	0.5	1	0.059	0.000	0.011	0.012	0.954
		2	0.081	−0.001	0.015	0.015	0.948
	1	1	0.059	−0.001	0.012	0.013	0.943
		2	0.081	0.000	0.013	0.013	0.954

†See the note to Table 1.

Table S.3. Comparison of the semiparametric and parametric MLEs of β_{11}†

		Semiparametric				Parametric			
n	r	Bias	SE	SEE	CP	Bias	SE	SEE	CP
100	0	0.004	0.391	0.388	0.943	−0.007	0.365	0.365	0.956
	0.5	0.004	0.400	0.399	0.953	−0.002	0.372	0.371	0.957
	1	0.006	0.412	0.406	0.945	0.008	0.387	0.385	0.945
200	0	−0.002	0.277	0.276	0.955	−0.007	0.261	0.262	0.950
	0.5	−0.003	0.296	0.292	0.946	−0.010	0.274	0.276	0.957
	1	−0.003	0.309	0.305	0.943	−0.002	0.287	0.288	0.944
500	0	0.004	0.169	0.173	0.959	0.006	0.164	0.159	0.947
	0.5	0.008	0.175	0.172	0.950	−0.005	0.160	0.160	0.958
	1	−0.004	0.173	0.180	0.940	−0.007	0.171	0.172	0.944

†See the note to Table 1.

Table S.4. Comparison of the semiparametric and parametric MLEs under mild mis-specification of the parametric distribution for Λ_k†

			Semiparametric				Parametric			
n	Parameter	Value	Bias	SE	SEE	CP	Bias	SE	SEE	CP
100	β_{11}	0.500	0.003	0.225	0.221	0.958	0.007	0.220	0.205	0.940
	$\Lambda_1(0.5)$	0.047	0.009	0.024	0.024	0.944	0.057	0.020	0.018	0.141
	$\Lambda_1(1.5)$	0.346	0.003	0.055	0.054	0.956	−0.106	0.048	0.037	0.279
200	β_{11}	0.500	0.006	0.146	0.147	0.956	0.001	0.135	0.131	0.958
	$\Lambda_1(0.5)$	0.047	−0.003	0.017	0.016	0.956	0.059	0.016	0.014	0.041
	$\Lambda_1(1.5)$	0.346	−0.008	0.037	0.038	0.944	−0.101	0.034	0.029	0.076
500	β_{11}	0.500	−0.001	0.095	0.096	0.947	0.000	0.092	0.087	0.959
	$\Lambda_1(0.5)$	0.047	0.005	0.013	0.013	0.947	0.058	0.012	0.009	$<.001$
	$\Lambda_1(1.5)$	0.346	−0.005	0.024	0.025	0.951	−0.105	0.020	0.016	$<.001$

†See the note to Table 1.

Table S.5. Comparison of the semiparametric and parametric MLEs of β_{11} under severe mis-specification of the parametric distribution for Λ_k†

	Semiparametric				Parametric			
n	Bias	SE	SEE	CP	Bias	SE	SEE	CP
100	0.008	0.228	0.232	0.952	−0.043	0.223	0.215	0.945
200	0.005	0.151	0.152	0.947	−0.041	0.143	0.136	0.937
500	−0.003	0.103	0.101	0.951	−0.039	0.098	0.083	0.930

†See the note to Table 1.

Table S.6. Simulation results on the regression parameters of one risk under mis-specification of the other risk†

		β_{11}				β_{12}			
n	(β_{11}, β_{12})	Bias	SE	SEE	CP	Bias	SE	SEE	CP
100	(0, 0)	0.001	0.384	0.384	0.936	0.002	0.606	0.606	0.949
	(0, 0.5)	0.010	0.373	0.371	0.963	0.004	0.667	0.669	0.952
	(0.5, 0.5)	0.010	0.432	0.434	0.961	0.000	0.638	0.640	0.955
200	(0, 0)	0.005	0.280	0.277	0.958	0.012	0.465	0.467	0.956
	(0, 0.5)	0.001	0.271	0.271	0.944	0.003	0.465	0.463	0.933
	(0.5, 0.5)	0.006	0.294	0.293	0.961	0.010	0.454	0.451	0.963
500	(0, 0)	0.009	0.171	0.170	0.939	0.005	0.281	0.284	0.935
	(0, 0.5)	0.006	0.170	0.169	0.948	0.011	0.280	0.278	0.955
	(0.5, 0.5)	0.002	0.173	0.175	0.954	0.002	0.266	0.264	0.951

†See the note to Table 1.

Table S.7. Proportional cause-specific hazards analysis of the bone marrow transplantation data

		Est	SE	p-value
TRM				
	Years 2001-2005	−0.461	0.133	<0.001
	Unrelated donor	0.743	0.128	<0.001
	Prior auto-HCT	−0.362	0.148	0.014
	TX > 24 months	0.316	0.133	0.017
Relapse				
	Years 2001-2005	0.348	0.134	0.009
	Unrelated donor	0.770	0.120	<0.001
	Prior auto-HCT	0.309	0.137	0.024
	TX > 24 months	0.425	0.122	<0.001

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Baltimore: Johns Hopkins University Press.

Chen, L., Lin, D. Y. and Zeng, D. (2012) Checking semiparametric transformation models with censored data. *Biostatistics*, **13**, 18-31.

Murphy, S. A. and van der Vaart, A. W. (2000) On profile likelihood. *J. Am. Statist. Ass.*, **95**, 449-465.

Rudin, W. (1973) *Functional Analysis*, New York: McGraw-Hill.

van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Zeng, D. and Lin, D. Y. (2007) Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Statist. Soc.* B, **69**, 507-564.