Supplementary Materials for Semiparametric Regression Analysis of Multiple Right- and Interval-Censored Events

Fei Gao, Donglin Zeng, David Couper, and D. Y. Lin

S. 1 PROOFS OF ASYMPTOTIC RESULTS

Let $\mathbb{P}_n$ denote the empirical measure for n independent subjects, $\mathbb{P}$ denote the true probability measure, and $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ denote the empirical process. The proofs of Theorems 1, 2, and 3 make use of five lemmas, which are stated and proved in Section S.4.

S.1.1 Proof of Theorem 1

We first show the existence of the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$. Let $\widetilde{M} = \sum_{k=1}^{K} \sup_{t \in \mathcal{U}_k} \sup_{\boldsymbol{X}_k(t), \boldsymbol{\beta}} |\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t)| + \sum_{k=K_1+1}^{K} |\gamma_k|$. For any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space, the integrand in the ith term of $l_n(\boldsymbol{\theta}, \mathcal{A})$ is bounded by

$$O(1) \prod_{k=K_1+1}^{K} \left[\left(\Lambda_k\{Y_{ik}\} e^{\widetilde{M}|\boldsymbol{b}_i|} \right)^{\Delta_{ik}} \left\{ 1 + \int_0^{Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\}^{-\Delta_{ik}} \right] \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}).$$

Thus, $l_n(\boldsymbol{\theta}, \mathcal{A})$ attains the maximum for finite values of Λ_k for $k = K_1 + 1, \ldots, K$, so the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ exists by allowing $\widehat{\Lambda}_k(\tau_k) = \infty$ for $k = 1, \ldots, K_1$.

We shall prove that $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) < \infty$ with probability 1 for any $\epsilon > 0$ and $k = 1, \ldots, K_1$ and that $\limsup_n \widehat{\Lambda}_k(\tau_k) < \infty$ with probability 1 for $k = K_1 + 1, \ldots, K$. By definition, $l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}, \mathcal{A}) \geq 0$ for any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space. We wish to show that if $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) = \infty$ for some $\epsilon > 0$ for $k = 1, \ldots, K_1$ or $\limsup_n \widehat{\Lambda}_k(\tau_k) = \infty$ for $k = K_1 + 1, \ldots, K$, then this difference must be negative, which is a contradiction. The key is to construct a suitable function in the parameter space that converges uniformly to $\mathcal{A}_0$.

For $k = 1, \ldots, K_1$, we define the step function $\widetilde{\Lambda}_k$ with $\widetilde{\Lambda}_k(t) = \Lambda_{k0}(t)$ for $t = t_{k1}, \ldots, t_{k,m_k}$ such that it converges uniformly to Λ_{k0}. For $k = K_1 + 1, \ldots, K$, we construct function $\widetilde{\Lambda}_k$ by imitating $\widehat{\Lambda}_k$. Specifically, by differentiating $l_n(\boldsymbol{\theta}, \mathcal{A})$ with respect to $\Lambda_k\{Y_{ik}\}$ and setting the derivative to 0, we find that $\widehat{\Lambda}_k$ satisfies the equation

$$\frac{\Delta_{ik}}{\widehat{\Lambda}_k\{Y_{ik}\}} = \sum_{j=1}^{n} \frac{\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right) J_{2k}\left(Y_{ik}, \boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\gamma}_k\right) \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right) \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}}, \tag{S.1}$$

where

$$J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) = \prod_{k=1}^{K_1} \left[\exp\left\{-\int_0^{L_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s)\right\} - \exp\left\{-\int_0^{R_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s)\right\}\right]$$

$$\times \prod_{K_1+1}^{K} \left[\left\{e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_k b_1 + b_2} \Lambda_k\{Y_k\}\right\}^{\Delta_k} \exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_k b_1 + b_2} d\Lambda_k(s)\right\}\right],$$

and

$$J_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) = I(Y_k \geq t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + \gamma_k b_1 + b_2}.$$

We replace $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathcal{A}}$ on the right side of equation (S.1) by $\boldsymbol{\theta}_0$ and $\mathcal{A}_0$, respectively, to obtain a similar function. We denote the solution as $\widetilde{\Lambda}_k$. By the Glivenko-Cantelli result in Lemma 1, $\widetilde{\Lambda}_k$ converges uniformly to Λ_{k0} in $\mathcal{U}_k$ for $k = K_1 + 1, \ldots, K$. We denote $\widetilde{\mathcal{A}} = (\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_K)$.

Clearly, $n^{-1}\left\{l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\right\} \geq 0$. Let $\delta_{ikm} = I(U_{ikm} < T_{ik} \leq U_{ik,m+1})$ for $i = 1, \ldots, n$, $k = 1, \ldots, K_1$, and $m = 0, \ldots, M_{ik}$, where $U_{ik,M_{ik}+1} = \infty$. By the fact that $e^{-|x|}(1+y) \leq 1 + e^x y \leq e^{|x|}(1+y)$, we obtain

$$\begin{aligned} 0 &\leq n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) \\ &\leq O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log\left(n \widehat{\Lambda}_k\{Y_{ik}\}\right) \\ &\quad + n^{-1} \sum_{i=1}^{n} \left[\log \int_{\boldsymbol{b}} \prod_{k=K_1+1}^{K} \left\{\frac{e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2}}}{1 + \int_0^{Y_{ik}} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_{ik}(t) + \gamma_k b_{i1} + b_{i2}} d\widehat{\Lambda}_k(t)}\right\}^{\Delta_{ik}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}\right] \end{aligned}$$

$$
\begin{aligned}
&\le \quad O(1)+n^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\log\left(n\widehat{\Lambda}_k\{Y_{ik}\}\right) \\
&\quad +n^{-1}\sum_{i=1}^{n}\left(\log\int_{\boldsymbol{b}}\prod_{k=K_1+1}^{K}\left[\frac{e^{\widetilde{M}\|\boldsymbol{b}\|}}{e^{-\widetilde{M}\|\boldsymbol{b}\|}\left\{1+\widehat{\Lambda}_k(Y_{ik})\right\}}\right]^{\Delta_{ik}}\phi\left(\boldsymbol{b};\widehat{\boldsymbol{\Sigma}}\right)d\boldsymbol{b}\right) \\
&\le \quad O(1)+n^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\log\left(n\widehat{\Lambda}_k\{Y_{ik}\}\right)-n^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\left[\Delta_{ik}\log\left\{1+\widehat{\Lambda}_k(Y_{ik})\right\}\right].
\end{aligned}
$$

We first show that $\limsup_n \widehat{\Lambda}_k(\tau_k) < \infty$ using the partitioning idea of Murphy (1994). Specifically, we construct a sequence $u_{k0}=\tau_k > u_{k1} > \cdots > u_{k,Q_k}=0$. Then,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\log\left(n\widehat{\Lambda}_k\{Y_{ik}\}\right)-n^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\left[\Delta_{ik}\log\left\{1+\widehat{\Lambda}_k(Y_{ik})\right\}\right] \\
\le \quad & O(1)+\sum_{k=K_1+1}^{K}\sum_{q=0}^{Q_k-1}n^{-1}\sum_{i=1}^{n}I(Y_{ik}\in[u_{k,q+1},u_{kq}))\log\left(n\widehat{\Lambda}_k\{Y_{ik}\}\right) \\
& -\sum_{k=K_1+1}^{K}n^{-1}\sum_{i=1}^{n}I(Y_{ik}=\tau_k)\Delta_{ik}\log\left\{1+\widehat{\Lambda}_k(\tau_k)\right\} \\
& -\sum_{k=K_1+1}^{K}\sum_{q=0}^{Q_k-1}n^{-1}\sum_{i=1}^{n}\Delta_{ik}I\left(Y_{ik}\in[u_{k,q+1},u_{kq})\right)\log\left\{1+\widehat{\Lambda}_k(u_{k,q+1})\right\},
\end{aligned}
$$

which is further bounded by

$$
\begin{aligned}
& -(2n)^{-1}\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\Delta_{ik}I(Y_{ik}=\tau_k)\log\left\{1+\widehat{\Lambda}_k(\tau_k)\right\} \\
& -\sum_{k=K_1+1}^{K}\left\{(2n)^{-1}\sum_{i=1}^{n}\Delta_{ik}I(Y_{ik}=\tau_k)-n^{-1}\sum_{i=1}^{n}\Delta_{ik}I\left(Y_{ik}\in[u_1,u_0)\right)\right\}\log\left\{1+\widehat{\Lambda}_k(\tau_k)\right\} \\
& -\sum_{k=K_1+1}^{K}\sum_{q=1}^{Q_k-1}\left\{n^{-1}\sum_{i=1}^{n}\Delta_{ik}I\left(Y_{ik}\in[u_{kq},u_{k,q-1})\right)-n^{-1}\sum_{i=1}^{n}\Delta_{ik}I\left(Y_{ik}\in[u_{k,q+1},u_{kq})\right)\right\} \\
& \times\log\left\{1+\widehat{\Lambda}_k(u_{kq})\right\}.
\end{aligned}
$$

Note that u_{kq} is chosen such that the coefficients in front of $\log\{1+\widehat{\Lambda}_k(u_{kq})\}$ are all negative when n is large enough. Thus, the corresponding terms cannot diverge to ∞. However, if

$\widehat{\Lambda}_k(\tau_k)$ diverges to ∞, then the first term diverges to $-\infty$. We conclude that there exists some $M^* < \infty$ such that $\max_{K_1+1 \le k \le K} \limsup_n \widehat{\Lambda}_k(\tau_k) \le M^*$ for $k = K_1+1, \ldots, K$.

We denote $\widetilde{\mathcal{A}}^* = (\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_{K_1}, \widehat{\Lambda}_{K_1+1}, \ldots, \widehat{\Lambda}_K)$. Then,

$$\begin{aligned}
0 &\le n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}^*) \\
&\le O(1) + n^{-1} \sum_{i=1}^n \left(\log \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \left[\exp\left\{ -e^{\widetilde{M}\|\boldsymbol{b}\|} \widehat{\Lambda}_k(U_{ik,M_{ik}}) \right\} \right]^{\delta_{ik,M_{ik}}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b} \right) \\
&\le O(1) + n^{-1} \sum_{i=1}^n \left(\log \int_{\|\boldsymbol{b}\| \le 1} \prod_{k=1}^{K_1} \left[\exp\left\{ -e^{\widetilde{M}\|\boldsymbol{b}\|} \widehat{\Lambda}_k(U_{i,M_{ik}}) \right\} \right]^{\delta_{i,M_{ik}}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b} \right) \\
&\quad + n^{-1} \sum_{i=1}^n \left\{ \log \int_{\|\boldsymbol{b}\| > 1} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b} \right\} \\
&\le O(1) - n^{-1} \sum_{i=1}^n \sum_{k=1}^{K_1} \delta_{ik,M_{ik}} e^{\widetilde{M}} \widehat{\Lambda}_k(U_{ik,M_{ik}}).
\end{aligned}$$

Therefore, for $k = 1, \ldots, K_1$, $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) < \infty$ with probability 1 for any $\epsilon > 0$. By choosing a sequence of ϵ decreasing to 0, it then follows from Helly's selection lemma that along a subsequence, $\widehat{\Lambda}_k \to \Lambda_{k*}$ pointwise on any interior set of $\mathcal{U}_k$ and $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_* \equiv (\boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)$. We denote $\mathcal{A}_* = (\Lambda_{1*}, \ldots, \Lambda_{K*})$.

We now show that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$. First, we consider the differentiability of Λ_{k*} for $k = K_1+1, \ldots, K$. By the definition of $\widetilde{\Lambda}_k$, $\widehat{\Lambda}_k(t)$ is absolutely continuous with respect to $\widetilde{\Lambda}_k(t)$, and

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\mathbb{P}_n \nu_k\left(s, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\left|\mathbb{P}_n \nu_k\left(s, \mathcal{O}; \widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)\right|} d\widetilde{\Lambda}_k(s), \tag{S.2}$$

where

$$\nu_k(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\int_{\boldsymbol{b}} J_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) J_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}.$$

To take limits on the two sides of equation (S.2), we first show that the denominator of the integrand is uniformly bounded away from zero. It follows from the Glivenko-Cantelli property in Lemma 1 that

$$\sup_{t\in\mathcal{U}_k}\left|\mathbb{P}_n\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0\right)-\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0\right)\right|\to_{a.s.}0$$

and

$$\sup_{t\in\mathcal{U}_k}\left|\mathbb{P}_n\nu_k\left(t,\mathcal{O};\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)\right|\to_{a.s.}0.$$

Note that for any $\epsilon>0$,

$$\limsup_n\widehat{\Lambda}_k(\tau_k)\geq\int_0^{\tau_k}\frac{\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0\right)}{\epsilon+\left|\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)\right|}d\Lambda_{k0}(s).$$

Let $\epsilon\to 0$. By the Monotone Convergence Theorem,

$$\int_0^{\tau_k}\frac{\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0\right)}{\left|\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)\right|}d\Lambda_{k0}(t)<\infty.$$

We claim that $\min_{t\in\mathcal{U}_k}|\mathbb{P}\nu_k(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*)|>0$. If this inequality does not hold, then there exists some $t_*\in\mathcal{U}_k$ such that $\mathbb{P}\nu_k(t_*,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*)=0$. The function $\mathbb{P}\nu_k(t_*,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*)$ is right-differentiable almost everywhere. Thus, there exists $\delta>0$ such that for $t\in(t_*,t_*+\delta)$,

$$\left|\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)\right|=\left|\mathbb{P}\nu_k\left(t,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)-\mathbb{P}\nu_k\left(t_*,\mathcal{O};\boldsymbol{\theta}_*,\mathcal{A}_*\right)\right|\leq O(1)|t-t_*|$$

almost everywhere. Hence,

$$\int_{t_*}^{t_*+\delta}\frac{1}{|t-t_*|}d\Lambda_{k0}(t)<\infty,$$

which is a contradiction. By taking the limits on both sides of (S.2), we conclude that $\Lambda_{k*}(t)$ is absolutely continuous with respect to $\Lambda_{k0}(t)$, so that $\Lambda_{k*}(t)$ is differentiable with respect to t. In addition, $d\widehat{\Lambda}_k(t)/d\widetilde{\Lambda}_k(t)$ converges to $d\Lambda_{k*}(t)/d\Lambda_{k0}(t)$ uniformly in t.

Define

$$m(\boldsymbol{\theta},\mathcal{A})=\log\left\{\frac{L(\boldsymbol{\theta},\mathcal{A})+L(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})}{2}\right\}$$

and

$$\mathcal{M}=\left\{m(\boldsymbol{\theta},\mathcal{A}):\boldsymbol{\theta}\in\Theta,\mathcal{A}\in\mathcal{D}_{1,\infty}\times\cdots\times\mathcal{D}_{K_1,\infty}\times\mathcal{D}_{K_1+1,M}\times\cdots\times\mathcal{D}_{K,M}\right\},$$

where $L(\boldsymbol{\theta}, \mathcal{A})$ is the objective function for a single subject, and $\mathcal{D}_{k,c} = \{\Lambda : \Lambda$ is increasing with $\Lambda(0) = 0, \Lambda(\tau_k) \leq c\}$. By the concavity of the log function,

$$\mathbb{P}_n m(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) \geq \frac{1}{2}\left\{\mathbb{P}_n \log L(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) + \mathbb{P}_n \log L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\right\} \geq \mathbb{P}_n l(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) = \mathbb{P}_n m(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}).$$

It follows from Lemma 1 that the class $\mathcal{M}$ is Glivenko-Cantelli. Thus,

$$\begin{aligned}
0 &\leq \mathbb{P}_n m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \mathbb{P}_n m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right) \\
&= \mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right)\right\} + o_P(1) \\
&= \mathbb{P}\log\left[\frac{L(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) + L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})}{2L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})}\right] + o_P(1) \\
&= \mathbb{P}\log\left\{\frac{1}{2} + \frac{\prod_{k=K_1+1}^{K}\widehat{\Lambda}_k\{Y_k\}^{\Delta_k}\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right)\psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}})d\boldsymbol{b}}{2\prod_{k=K_1+1}^{K}\widetilde{\Lambda}_k\{Y_k\}^{\Delta_k}\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \widetilde{\mathcal{A}}\right)\psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0)d\boldsymbol{b}}\right\} + o_P(1) \\
&\rightarrow \mathbb{P}\left[\log\left\{\frac{1}{2} + \frac{\prod_{k=K_1+1}^{K}\Lambda'_{k*}(Y_k)^{\Delta_k}\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*, \mathcal{A}_*\right)\psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*)d\boldsymbol{b}}{2\prod_{k=K_1+1}^{K}\Lambda'_{k0}(Y_k)^{\Delta_k}\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0\right)\psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0)d\boldsymbol{b}}\right\}\right],
\end{aligned}$$

such that the negative Kullback-Leibler information is positive. The identifiability result in Lemma 3 implies that $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_* = \boldsymbol{\gamma}_0$, $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_0$, and $\Lambda_{k*}(t_k) = \Lambda_{k0}(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$. We conclude that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \rightarrow 0$ and $|\widehat{\Lambda}_k(t_k) - \Lambda_{k0}(t_k)| \rightarrow 0$ for any $t_k \in \mathcal{U}_k$. Because $\mathcal{A}_0$ is continuous, $\widehat{\mathcal{A}}$ converges uniformly to $\mathcal{A}_0$ on $\prod_k \mathcal{U}_k$.

S.1.2 Proof of Theorem 2

Let

$$H_{1k}(t, u, v, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})Q_1(t, u, v, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}, \Lambda_k)\psi(\boldsymbol{b}; \boldsymbol{\Sigma})}{\int_{\boldsymbol{b}'} J_1(\boldsymbol{b}', \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})\psi(\boldsymbol{b}'; \boldsymbol{\Sigma})d\boldsymbol{b}'}$$

for $k = 1, \ldots, K_1$, and

$$H_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}, \gamma_k)\psi(\boldsymbol{b}; \boldsymbol{\Sigma})}{\int_{\boldsymbol{b}'} J_1(\boldsymbol{b}', \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})\psi(\boldsymbol{b}'; \boldsymbol{\Sigma})d\boldsymbol{b}'}$$

for $k = K_1+1, \ldots, K$, where

$$\begin{aligned}
& Q_1(t, u, v, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}, \Lambda_k) \\
= \;& e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(t)+b_1} \left[\frac{I(v \geq t)\exp\left\{-\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\}}{\exp\left\{-\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\} - \exp\left\{-\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\}} \right. \\
& \left. - \frac{I(u \geq t)\exp\left\{-\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\}}{\exp\left\{-\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\} - \exp\left\{-\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\}} \right],
\end{aligned}$$

and

$$Q_2(t, u, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}, \gamma_k) = -I(u \geq t) e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(t)+\gamma_k b_1 + b_2}.$$

Then, the score function for $\boldsymbol{\theta}$ is $\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \mathcal{A}) = (\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \mathcal{A})^{\mathrm{T}}, l_{\gamma_{K_1+1}}(\boldsymbol{\theta}, \mathcal{A}), \ldots, l_{\gamma_K}(\boldsymbol{\theta}, \mathcal{A}), l_{\sigma_1^2}(\boldsymbol{\theta}, \mathcal{A}), l_{\sigma_2^2}(\boldsymbol{\theta}, \mathcal{A}))^{\mathrm{T}}$, where

$$\begin{aligned}
\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \mathcal{A}) \;=\; & \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{1k}(t, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) d\boldsymbol{b} \boldsymbol{X}_k(t) d\Lambda_k(t), \\
& + \sum_{k=K_1+1}^{K} \left\{ \Delta_k \boldsymbol{X}_k(Y_k) + \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) d\boldsymbol{b} \boldsymbol{X}_k(t) d\Lambda_k(t) \right\},
\end{aligned}$$

$$l_{\gamma_k}(\boldsymbol{\theta}, \mathcal{A}) = \Delta_k \frac{\int_{\boldsymbol{b}} b_1 J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}} + \int_0^{\tau_k} \int_{\boldsymbol{b}} b_1 H_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) d\boldsymbol{b} d\Lambda_k(t),$$

$$l_{\sigma_j^2}(\boldsymbol{\theta}, \mathcal{A}) = \frac{\int J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \phi'_{\sigma_j^2}(b_j; \sigma_j^2) \phi(b_{3-j}; \sigma_{3-j}^2) d\boldsymbol{b}}{\int J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}$$

for $j = 1, 2$, and $\phi'_{\sigma_j^2}(b_j; \sigma_j^2)$ is the derivative of $\phi(b_j; \sigma_j^2)$ with respect to σ_j^2. The score operator for $\mathcal{A}$ along the submodel $d\mathcal{A}_{\epsilon,\boldsymbol{h}} = ((1+\epsilon h_1)d\Lambda_1, \ldots, (1+\epsilon h_K)d\Lambda_K)^{\mathrm{T}}$ for $\boldsymbol{h} = (h_1, \ldots, h_K)$ with $h_k \in L_2(\mu_k)$ for $k = 1, \ldots, K_1$ and $h_k \in BV_1(\mathcal{U}_k)$ for $k = K_1+1, \ldots, K$ is

$$l_{\mathcal{A}}(\boldsymbol{\theta}, \mathcal{A})(\boldsymbol{h}) \;=\; \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{1k}(t, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) d\boldsymbol{b} h_k(t) d\Lambda_k(t)$$

$$+\sum_{k=K_1+1}^{K}\left\{\Delta_k h_k(Y_k)+\int_0^{\tau_k}\int_{\boldsymbol{b}} H_{2k}(t,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}h_k(t)d\Lambda_k(t)\right\},$$

where $BV_1(\mathcal{B})$ denotes the set of functions on $\mathcal{B}$ with total variation bounded by 1.

Clearly,

$$\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})\right\}=-\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)\right\},$$

and

$$\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h})\right\}=-\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h})-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)(\boldsymbol{h})\right\}.$$

We apply the Taylor series expansions at $(\boldsymbol{\theta}_0,\mathcal{A}_0)$ to the right sides of the above two equations. In light of Lemma 5, the second-order terms are bounded by

$$\begin{aligned}
& O_P(1)\sqrt{n}E\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\left\{\widehat{\Lambda}_k(U_{km})-\Lambda_{k0}(U_{km})\right\}^2+\sum_{k=K_1+1}^{K}\left\{\widehat{\Lambda}_k(Y_k)-\Lambda_{k0}(Y_k)\right\}^2\right.\\
& \left.+\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right]\\
= & \sqrt{n}\left\{O_P(n^{-2/3})+O_P(1)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+O_P(1)\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+O_P(1)\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right\}\\
= & O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})\right\} = & -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}}-\mathcal{A}_0)\right\}\\
& +O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h})\right\} = & -\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h},\widehat{\mathcal{A}}-\mathcal{A}_0)\right\}\\
& +O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right),
\end{aligned}$$

where $\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the second derivative of $\boldsymbol{l}(\boldsymbol{\theta}, \mathcal{A})$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\boldsymbol{h})$ is the derivative of $\boldsymbol{l}_{\boldsymbol{\theta}}$ along the submodel $d\mathcal{A}_{\epsilon,h}$, $l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ with respect to $\boldsymbol{\theta}$, and $l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}, \widehat{\mathcal{A}} - \mathcal{A}_0)$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ along the submodel $d\mathcal{A}_0 + \epsilon d(\widehat{\mathcal{A}} - \mathcal{A}_0)$. All derivatives are evaluated at $(\boldsymbol{\theta}_0, \mathcal{A}_0)$.

If the least favorable direction exists, we denote it as $\boldsymbol{h}^* = (\boldsymbol{h}_1^*, \ldots, \boldsymbol{h}_K^*)$, where $\boldsymbol{h}_k^*$ $(k = 1, \ldots, K_1)$ is $(p + K_2 + 2)$-dimensional vector of functions in $L_2(\mu_k)$, and $\boldsymbol{h}_k^*$ $(k = K_1 + 1, \ldots, K)$ is $(p + K_2 + 2)$-dimensional vector of functions in $L_2(\mathcal{U}_k)$. We first show the existence of $\boldsymbol{h}^*$, which is the solution to $l_{\mathcal{A}}^* l_{\mathcal{A}}(\boldsymbol{h}^*) = l_{\mathcal{A}}^* \boldsymbol{l}_{\boldsymbol{\theta}}$ with $l_{\mathcal{A}}^*$ as the adjoint operator of $l_{\mathcal{A}}$. Let $\mathcal{Q} = \prod_{k=1}^{K_1} L_2(\mu_k) \times \prod_{k=K_1+1}^{K} L_2(\mathcal{U}_k)$. We equip $\mathcal{Q}$ with an inner product defined as

$$< \boldsymbol{h}^{(1)}, \boldsymbol{h}^{(2)} >= \sum_{k=1}^{K_1} \int_{\mathcal{U}_k} h_k^{(1)} h_k^{(2)} d\mu_k(t) + \sum_{k=K_1+1}^{K} \int_0^{\tau_k} h_k^{(1)} h_k^{(2)} d\Lambda_{k0}(t),$$

where $\boldsymbol{h}^{(1)} = (h_1^{(1)}, \ldots, h_K^{(1)})$ and $\boldsymbol{h}^{(2)} = (h_1^{(2)}, \ldots, h_K^{(2)})$. On the same space, we define

$$\begin{aligned} \|\boldsymbol{h}\| &= \mathbb{P}\left\{l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)(\boldsymbol{h})^2\right\}^{1/2} \\ &= \mathbb{P}\left(\left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{1k}(t, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b} h_k(t) d\Lambda_{k0}(t) \right.\right. \\ &\quad \left.\left. + \sum_{k=K_1+1}^{K} \left\{\Delta h_k(Y_k) + \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b} h_k(t) d\Lambda_{k0}(t)\right\}\right]^2\right)^{1/2} \end{aligned}$$

for $\boldsymbol{h} = (h_1, \ldots, h_K)$. It is easy to show that $\|\cdot\|$ is a seminorm on $\mathcal{Q}$. Furthermore, if $\|\boldsymbol{h}\| = 0$, then $\mathbb{P}\{l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h})^2\} = 0$. Thus, with probability 1, $l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}) = 0$. By the arguments in the proof of Lemma 5, $h_k(t_k) = 0$ for $t_k \in \mathcal{U}_k$ for $k = 1, \ldots, K$. Clearly, $\|\boldsymbol{h}\| \leq c < \boldsymbol{h}, \boldsymbol{h} >^{1/2}$ for some constant c by the Cauchy-Schwarz inequality. According to the bounded inverse theorem in Banach spaces, we have $< \boldsymbol{h}, \boldsymbol{h} >^{1/2} \leq \widetilde{c}\|\boldsymbol{h}\|$ for another constant $\widetilde{c}$. By the Lax-Milgram theorem (Zeidler, 1995), $\boldsymbol{h}^*$ exists and for any $t_k \in \mathcal{U}_k$,

$$\int_0^{\tau_k} \mathbb{P}\left\{\sum_{m=0}^{M_k} \delta_{km} \int_{\boldsymbol{b}} H_{1k}(t_k, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b} \int_{\boldsymbol{b}} H_{1k}(s, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b}\right\}$$

$$\times \boldsymbol{h}_k^*(s)d\Lambda_{k0}(s)$$

$$= \mathbb{P}\left\{\sum_{m=0}^{M_k} \delta_{km} \int_{\boldsymbol{b}} H_{1k}(t_k, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b} \boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\right\} \tag{S.3}$$

for $k = 1, \ldots, K_1$ and

$$\begin{aligned}
&\int_0^{\tau_k} \mathbb{P}\left(\left[I(t_k \leq C_k)\exp\{-\Lambda_k(t_k)\} + \int_{\boldsymbol{b}} H_{2k}(t_k, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)d\boldsymbol{b}\right] \int_{\boldsymbol{b}} H_{2k}(s, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)d\boldsymbol{b}\right) \\
&\times \boldsymbol{h}_k^*(s)d\Lambda_{k0}(s) + \mathbb{P}\left[I(t_k \leq C_k)\exp\{-\Lambda_k(t_k)\}\right] \boldsymbol{h}_k^*(t_k) \\
&= \mathbb{P}\left\{E\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) | T_k = t_k\} I(t_k \leq C_k)\exp\{-\Lambda_k(t_k)\} + \int_{\boldsymbol{b}} H_{2k}(t_k, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)d\boldsymbol{b}\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\right\}
\end{aligned}$$

for $k = K_1 + 1, \ldots, K$. We differentiate (S.3) with respect to t_k to obtain

$$q_{k1}(t_k)\boldsymbol{h}_k^*(t_k) + \sum_{k'=1}^{K} \int_{t_k}^{\tau_k} q_{k2}(s, t_k)\boldsymbol{h}_{k'}^*(s)ds + \int_0^{t_k} q_{k3}(s, t_k)\boldsymbol{h}_k^*(s)ds = \boldsymbol{q}_{k4}(t_k),$$

where $q_{k1}(t_k) > 0$ and q_{kj} $(k = 1, \ldots, K; j = 1, 2, 3)$ and $\boldsymbol{q}_{k4}$ $(k = 1, \ldots, K)$ are continuously differentiable functions. Thus, $\boldsymbol{h}^*$ can be expanded to be a continuously differentiable function in $[0, \tau_k]^K$ with bounded total variation. It then follows that

$$\begin{aligned}
&\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})\right\} - \mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\} \\
=& -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}} - \mathcal{A}_0)\right\} + \sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}^*, \widehat{\mathcal{A}} - \mathcal{A}_0)\right\} \\
&+ O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right) \\
=& \sqrt{n}\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) - l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\
&+ O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}$$

Using arguments in the proof of Lemma 2, we can show that $\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*)$ belongs to a Donsker class. It follows from Lemma 4 that the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible. Then, $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$, and

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(\mathbb{P}\left[\{\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*)\}^{\otimes 2}\right]\right)^{-1} \mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\} + o_P(1).$$

The influence function for $\widehat{\boldsymbol{\theta}}$ is the efficient influence function, such that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

S.1.3 Proof of Theorem 3

Let $\widehat{\mathcal{A}}^*$ be the estimator of $\mathcal{A}$ in the bootstrap sample. We denote $\widehat{\mathbb{P}}_n$ as the bootstrap empirical distribution and $\widehat{\mathbb{G}}_n = \sqrt{n}(\widehat{\mathbb{P}}_n - \mathbb{P}_n)$ as the bootstrap empirical process. Using arguments in the proof of Theorem 2, we can show that

$$
\begin{aligned}
\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)\right\} &= \sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}^*, \widehat{\mathcal{A}}^*\right)\right\} \\
&= \sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^*\right) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}\left(\widehat{\mathcal{A}} - \widehat{\mathcal{A}}^*\right)\right\} \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\mathbb{G}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)(\boldsymbol{h})\right\} &= \sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)(\boldsymbol{h}) - l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}^*, \widehat{\mathcal{A}}^*\right)(\boldsymbol{h})\right\} \\
&= \sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^*\right) + l_{\mathcal{A}\mathcal{A}}\left(\boldsymbol{h}, \widehat{\mathcal{A}} - \widehat{\mathcal{A}}^*\right)\right\} \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)\right\} - \widehat{\mathbb{G}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)(\boldsymbol{h}^*)\right\} \\
= \;& \sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^*\right) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}\left(\widehat{\mathcal{A}} - \widehat{\mathcal{A}}^*\right)\right\} - \sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^*\right) + l_{\mathcal{A}\mathcal{A}}\left(\boldsymbol{h}^*, \widehat{\mathcal{A}} - \widehat{\mathcal{A}}^*\right)\right\} \\
& + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right) \\
= \;& \sqrt{n}\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)(\boldsymbol{h}^*)\right\}^{\otimes 2}\right]\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^*\right) \\
& + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}
$$

By the arguments in the proof of Theorem 2,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^{*}\right)=\left(\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)\left(\boldsymbol{h}^{*}\right)\right\}^{\otimes 2}\right]\right)^{-1}\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}^{*})\right\}+o_P(1)$$
$$=\left(\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)\left(\boldsymbol{h}^{*}\right)\right\}^{\otimes 2}\right]\right)^{-1}\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}^{*})\right\}+o_P(1),$$

where the last equality follows from Theorem 3.6.1 of van der Vaart and Wellner (1996). Therefore, $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^{*})$ converges weakly to a zero-mean normal random vector, and $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^{*})$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ have the same asymptotic distribution.

S. 2 SIMULATION STUDIES UNDER MISSPECIFIED MODELS

To investigate the performance of the proposed dynamic prediction methods under misspecified models, we conducted a series of simulation studies by generating the event times from the proportional odds models with random effects. The cumulative hazard functions take the form of

$$\Lambda_k(t;\boldsymbol{X}_k,b_1)=\log\left\{1+\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}\lambda_k(s)ds\right\}$$

for $k=1,2$ and

$$\Lambda_k(t;\boldsymbol{X}_k,b_1,b_2)=\log\left\{1+\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k b_1+b_2}\lambda_k(s)ds\right\}$$

For $k=3,4,5$. We set the parameters to be the same as in the simulation studies with a terminal event. We fit models (1) and (2) and estimated the cumulative incidence functions of events 3 and 4 given that at the first monitoring time $t_0=1$, event 2 has occurred but events 1, 3, and 4 have not. The results are shown in Figure S.3.

S. 3 ARBITRARY COMBINATION OF INTERVAL- AND RIGHT-CENSORED EVENTS

We consider a more general setting where both the symptomatic and asymptomatic events can be right- or interval-censored. Let η_k $(k=1,\ldots,K)$ denote, by the values 1 versus 0, whether T_k is interval- or right-censored. If $\eta_k = 1$, let (L_k, R_k) denote the interval that brackets T_k; otherwise, let (Y_k, Δ_k) denote the right-censored observation. For a random sample of n subjects, the data consist of $\{\mathcal{O}_i : i = 1, \ldots, n\}$, where

$$\mathcal{O}_i = \{\eta_k, \eta_k L_{ik}, \eta_k R_{ik}, (1-\eta_k) Y_{ik}, (1-\eta_k)\Delta_{ik}, \boldsymbol{X}_{ik}(\cdot) : k = 1, \ldots, K\}.$$

The likelihood function concerning the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \Lambda_1, \ldots, \Lambda_K)$ is then given by

$$\begin{aligned}
&\prod_{i=1}^{n} \int_{\boldsymbol{b}_i} \prod_{k=1}^{K_1} \Bigg(\left[\exp\left\{ - \int_0^{L_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} - \exp\left\{ - \int_0^{R_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} \right]^{\eta_k} \\
&\qquad \times \left[\left\{ e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + b_{i1}} \lambda_k(Y_{ik}) \right\}^{\Delta_{ik}} \exp\left\{ - \int_0^{Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} \right]^{1-\eta_k} \Bigg) \\
&\times \prod_{k=K_1+1}^{K} \Bigg(\left[\exp\left\{ - \int_0^{L_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\} - \exp\left\{ - \int_0^{R_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\} \right]^{\eta_k} \\
&\left[\left\{ e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2}} \lambda_k(Y_{ik}) \right\}^{\Delta_{ik}} \exp\left\{ - \int_0^{Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\} \right]^{1-\eta_k} \Bigg) \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}) d\boldsymbol{b}_i.
\end{aligned}$$

We adopt the nonparametric maximum likelihood estimation approach. For $k = 1, \ldots, K$ with $\eta_k = 1$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{k,m_k} < \infty$ be the ordered sequence of all L_{ik} and R_{ik} with $R_{ik} < \infty$. For $k = 1, \ldots, K$ with $\eta_k = 0$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{k,m_k} < \infty$ be the ordered sequence of all Y_{ik} with $\Delta_{ik} = 1$. The estimator for Λ_k $(k = 1, \ldots, K)$ is a step function that jumps only at $t_{k1}, \ldots, t_{k,m_k}$ with respective jump sizes $\lambda_{k1}, \ldots, \lambda_{k,m_k}$.

For $k = 1, \ldots, K$ with $\eta_k = 1$, we let $R_{ik}^* = I(R_{ik} = \infty) L_{ik} + I(R_{ik} < \infty) R_{ik}$ and introduce independent Poisson random variables W_{ikl} $(l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*)$ with means $\lambda_{kl} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1})$ if $k \leq K_1$ or $\lambda_{kl} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + \gamma_k b_{i1} + b_{i2})$ otherwise. Let $A_{ik} = \sum_{t_{kl} \leq L_{ik}} W_{ikl}$ and $B_{ik} = I(R_{ik} < \infty) \sum_{L_{ik} < t_{kl} \leq R_{ik}} W_{ikl}$. The objective function in the nonparametric maximum likelihood estimation can be viewed as the observed-data likelihood for $\{A_{ik} = 0, B_{ik} > 0 : i = 1, \ldots, n; k = 1, \ldots, K, \eta_k = 1\} \cup \{Y_{ik}, \Delta_{ik} : i = 1, \ldots, n; k =$

$1,\dots,K,\eta_k=0\}$ with $(W_{ikl},\boldsymbol{b}_i)$ $(i=1,\dots,n;k=1,\dots,K,\eta_k=1;l=1,\dots,m_k,t_{kl}\le R^*_{ik})$ as latent variables.

We propose an EM algorithm with W_{ikl} and $\boldsymbol{b}_i$ as missing data. In the M-step, we update $\boldsymbol{\beta}$ by solving the equation

$$\sum_{i=1}^{n}\left[\sum_{k=1}^{K_1}\left\{\eta_k\sum_{l=1}^{m_k}\widehat{E}\left(W_{ikl}\right)I(t_{kl}\le R^*_{ik})\left[\boldsymbol{X}_{ikl}-\frac{\sum_{j=1}^{n}\boldsymbol{X}_{jkl}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+b_{j1}\right)\right\}}{\sum_{j=1}^{n}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+b_{j1}\right)\right\}}\right]\right.\right.$$
$$\left.+(1-\eta_k)\Delta_{ik}\left(\boldsymbol{X}_{ik}(Y_{ik})-\frac{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\boldsymbol{X}_{jk}(Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+b_{j1})\right\}\right]}{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+b_{j1}\right\}\right]}\right)\right\}$$
$$+\sum_{k=K_1+1}^{K}\left\{\eta_k\sum_{l=1}^{m_k}\widehat{E}\left(W_{ikl}\right)I(t_{kl}\le R^*_{ik})\left[\boldsymbol{X}_{ikl}-\frac{\sum_{j=1}^{n}\boldsymbol{X}_{jkl}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+\gamma_k b_{j1}+b_{j2}\right)\right\}}{\sum_{j=1}^{n}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+\gamma_k b_{j1}+b_{j2}\right)\right\}}\right]\right.$$
$$\left.\left.+(1-\eta_k)\Delta_{ik}\left(\boldsymbol{X}_{ik}(Y_{ik})-\frac{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\boldsymbol{X}_{jk}(Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2})\right\}\right]}{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2}\right\}\right]}\right)\right\}\right]=\boldsymbol{0}.$$

We update γ_k by solving the equation

$$\sum_{i=1}^{n}\left[\left\{\eta_k\sum_{l=1}^{m_k}I(t_{kl}\le R^*_{ik})\widehat{E}\left(W_{ikl}b_{i1}\right)+(1-\eta_k)\Delta_{ik}\widehat{E}\left(b_{i1}\right)\right\}\right.$$
$$-\sum_{l=1}^{m_k}\left\{\eta_k I(t_{kl}\le R^*_{ik})\widehat{E}\left(W_{ikl}\right)+(1-\eta_k)\Delta_{ik}I(Y_{ik}=t_{kl})\right\}$$
$$\left.\times\frac{\sum_{j=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{jk})+(1-\eta_k)I(t_{kl}\le Y_{jk})\right\}\widehat{E}\left\{b_{j1}\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+\gamma_k b_{j1}+b_{j2}\right)\right\}}{\sum_{j=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{jk})+(1-\eta_k)I(t_{kl}\le Y_{jk})\right\}\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+\gamma_k b_{j1}+b_{j2}\right)\right\}}\right]=0.$$

We update λ_{kl} by

$$\lambda_{kl}=\frac{\sum_{i=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{ik})\widehat{E}\left(W_{ikl}\right)+(1-\eta_k)\Delta_{ik}I(Y_{ik}=t_{kl})\right\}}{\sum_{i=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{ik})+(1-\eta_k)I(t_{kl}\le Y_{ik})\right\}\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+b_{i1}\right)\right\}}$$

for $k=1,\dots,K_1$ and $l=1,\dots,m_k$ and

$$\lambda_{kl}=\frac{\sum_{i=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{ik})\widehat{E}\left(W_{ikl}\right)+(1-\eta_k)\Delta_{ik}I(Y_{ik}=t_{kl})\right\}}{\sum_{i=1}^{n}\left\{\eta_k I(t_{kl}\le R^*_{ik})+(1-\eta_k)I(t_{kl}\le Y_{ik})\right\}\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+\gamma_k b_{i1}+b_{i2}\right)\right\}}$$

for $k = K_1 + 1, \ldots, K$ and $l = 1, \ldots, m_k$. Finally, we update σ_j^2 by $\sigma_j^2 = \sum_{i=1}^n \widehat{E}(b_{ij}^2)/n$ for $j = 1, 2$. In the E-step, we evaluate the conditional expectation of W_{ikl} ($k = 1, \ldots, K, \eta_k = 1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) and the other terms of $\boldsymbol{b}_i$ given the observed data $\widetilde{\mathcal{O}}_i$ for $i = 1, \ldots, n$. Specifically, the conditional expectation of W_{ikl} ($k = 1, \ldots, K_1, \eta_k = 1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) given $\widetilde{\mathcal{O}}_i$ and $\boldsymbol{b}_i$ is

$$I(L_{ik} < t_{kl} \leq R_{ik} < \infty) \frac{\lambda_{kl} \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}\right)}{1 - \exp\left(-\sum_{L_{ik} < t_{kl'} \leq R_{ik}} \lambda_{kl'} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl'} + b_{i1}}\right)}$$

and the conditional expectation of W_{ikl} ($k = K_1 + 1, \ldots, K, \eta_k = 1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) given $\widetilde{\mathcal{O}}_i$ and $\boldsymbol{b}_i$ is

$$I(L_{ik} < t_{kl} \leq R_{ik} < \infty) \frac{\lambda_{kl} \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + \gamma_k b_{i1} + b_{i2}\right)}{1 - \exp\left(-\sum_{L_{ik} < t_{kl'} \leq R_{ik}} \lambda_{kl'} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl'} + \gamma_k b_{i1} + b_{i2}}\right)}.$$

We iterate between the E-step and M-step until convergence.

S. 4 SOME USEFUL LEMMAS

Lemma 1. Under Conditions 1–5, the classes of functions

$$\widetilde{\mathcal{H}}_1 \equiv \left\{ \int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b} : \boldsymbol{\theta} \in \Theta, \mathcal{A} \in \mathcal{D}_1 \right\}$$

and

$$\widetilde{\mathcal{H}}_{2k} \equiv \left\{ \int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) J_{2k}(t, b, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b} : \boldsymbol{\theta} \in \Theta, t \in \mathcal{U}_k, \mathcal{A} \in \mathcal{D}_1 \right\}$$

for $k = K_1 + 1, \ldots, K$ are $\mathbb{P}$-Glivenko-Cantelli, where $\mathcal{D}_1 = \mathcal{D}_{1,\infty} \times \cdots \times \mathcal{D}_{K_1,\infty} \times \mathcal{D}_{K_1+1,M} \times \cdots \times \mathcal{D}_{K,M}$, and M is a finite constant.

Proof. Define

$$W_k(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}, \Lambda_k) = \frac{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s)}{\Lambda_k(\tau_k)}$$

for $k=1,\ldots,K_1$, where $\boldsymbol{\beta}\in\mathcal{B}$ and $\Lambda_k\in\mathcal{D}_{k,\infty}$. The class of functions $\{e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}:\boldsymbol{\beta}\in\mathcal{B}\}$, with $\boldsymbol{X}$ and b_1 as random variables, is a VC class with VC-index V. Thus, the class $\mathcal{W}_k\equiv\{W_k(t,\boldsymbol{X},b_1;\boldsymbol{\beta},\Lambda_k):\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{D}_{k,\infty}\}$ is a convex hull of the VC-class with the $L_2(\mathbb{P})$-bracketing number given by $O\{\exp(\epsilon^{-2V/(V+2)})\}$ (van der Vaart and Wellner, 1996, pp. 142–145).

For any $(\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)})$ in $\mathcal{B}\times\mathcal{D}_{k,\infty}$, $t_k\in\mathcal{U}_k$, and any positive constant M, if $\Lambda_k^{(1)}(\tau_k)>M$ and $\Lambda_k^{(2)}(\tau_k)>M$, then

$$\begin{aligned}
&\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(2)}(s)\right\}\right|\\
\leq\;& 2\exp\left(-Me^{-\widetilde{M}-|b_1|}\right).
\end{aligned}$$

If $\Lambda_k^{(1)}(\tau_k)\leq M$ and $\Lambda_k^{(2)}(\tau_k)\leq M$, then

$$\begin{aligned}
&\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(2)}(s)\right\}\right|\\
\leq\;& \sup_{\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{D}_{k,\infty},\Lambda_k(\tau_k)\leq M}\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k(s)\right\}\right|\\
&\times\left\{\left|W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)-W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)}\right)\right|M\right.\\
&\left.+W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\left|\Lambda_k^{(1)}(\tau_k)-\Lambda_k^{(2)}(\tau_k)\right|\right\}\\
\leq\;& \left|W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)-W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)}\right)\right|M+e^{\widetilde{M}+|b_1|}\left|\Lambda_k^{(1)}(\tau_k)-\Lambda_k^{(2)}(\tau_k)\right|.
\end{aligned}$$

In the remaining scenario, we assume, without loss of generality, that $\Lambda_k^{(1)}(\tau_k)\leq M$ and $\Lambda_k^{(2)}(\tau_k)>M$. Then,

$$\begin{aligned}
&\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(2)}(s)\right\}\right|\\
\leq\;& \sup_{\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{D}_{k,\infty},\Lambda_k(\tau_k)\leq M}\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k(s)\right\}\right|\\
&\times\left[\left|\exp\left\{-\Lambda_k^{(1)}(\tau_k)W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\}-\exp\left\{-MW_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\}\right|\right.\\
&\left.+\left|\exp\left\{-MW_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\}-\exp\left\{-\Lambda_k^{(2)}(\tau_k)W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)}\right)\right\}\right|\right]
\end{aligned}$$

$$\leq \left(e^{\widetilde{M}+|b_1|}|\Lambda_k^{(1)}(\tau_k)-M|\right)+2\exp\left(-Me^{-\widetilde{M}-|b_1|}\right).$$

Because there exist M/ϵ ϵ-brackets to cover $[0,M]$, the above results imply that there exist $O\left\{\exp\left(\epsilon^{-2V/(V+2)}\right)\right\}\times M/\epsilon$ brackets $(\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)})$ such that

$$\begin{aligned}
&\left|\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(2)}(s)\right\}\right|\\
\leq\ &\left(M+e^{\widetilde{M}+|b_1|}\right)\epsilon+2\exp\left(-e^{-\widetilde{M}-|b_1|}M\right).
\end{aligned}$$

Therefore, there exist $O\left\{\exp\left(\epsilon^{-2V/(V+2)}\right)/\epsilon\right\}$ ϵ-brackets to cover $\{\exp\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k(s)\}:\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{D}_{k,\infty}\}$ in $L_2(\mathbb{P})$.

For any $(\boldsymbol{\beta}^{(1)},\boldsymbol{\gamma}^{(1)},\Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)},\boldsymbol{\gamma}^{(2)},\Lambda_k^{(2)})$ in $\mathcal{B}\times\mathcal{G}\times\mathcal{D}_{k,M}$ for $k=K_1+1,\ldots,K$,

$$\begin{aligned}
&\left|\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(2)}b_1+b_2}d\Lambda_k^{(2)}(s)\right\}\right|\\
\leq\ &\sup_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\gamma}\in\mathcal{G},\Lambda_k\in\mathcal{D}_{k,M}}\left|\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k b_1+b_2}d\Lambda_k(s)\right\}\right|\\
&\times\left|\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2}d\Lambda_k^{(1)}(s)-\int_0^{Y_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(2)}b_1+b_2}d\Lambda_k^{(2)}(s)\right|\\
\leq\ &\left\{C^*e^{\widetilde{M}\|\boldsymbol{b}\|}\left(\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\|+\left|\gamma_k^{(1)}-\gamma_k^{(2)}\right|\right)+\left|\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2}d\left(\Lambda_k^{(1)}-\Lambda_k^{(2)}\right)(s)\right|\right\}\\
\leq\ &e^{\widetilde{M}\|\boldsymbol{b}\|}\left\{C^*\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\|+C^*\left|\gamma_k^{(1)}-\gamma_k^{(2)}\right|+\left|\Lambda_k^{(1)}(Y_k)-\Lambda_k^{(2)}(Y_k)\right|+\int_0^{\tau_k}\left|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)\right|ds\right\},
\end{aligned}$$

where the last inequality follows from integration by parts. By Theorem 2.7.5 of van der Vaart and Wellner (1996), the bracketing number of $\mathcal{B}\times\mathcal{G}\times\mathcal{D}_{k,M}$ is of order $O\{\exp(\epsilon^{-1})\}$. Thus, the bracketing number of $\widetilde{\mathcal{H}}_1$ is of order $O\{\exp(\epsilon^{-2V/(V+2)}+\epsilon^{-1})\epsilon^{-1}\}$. Therefore, the class $\widetilde{\mathcal{H}}_1$ is Glivenko-Cantelli. Because $I(Y_k\geq t)$ is Glivenko-Cantelli, $\widetilde{\mathcal{H}}_{2k}$ is Glivenko-Cantelli by the preservation of the Glivenko-Cantelli property under the product.

Lemma 2. Under Conditions 1−5, the classes of functions

$$\mathcal{H}_1\equiv\left\{\int_{\boldsymbol{b}} J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}:\boldsymbol{\theta}\in\Theta,\mathcal{A}\in\mathcal{D}_2\right\}$$

and

$$\mathcal{H}_{2k} \equiv \left\{ \int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) J_{2k}(t, b, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b} : \boldsymbol{\theta} \in \Theta, t \in \mathcal{U}_k, \mathcal{A} \in \mathcal{D}_2 \right\}$$

for $k = K_1 + 1, \ldots, K$ are uniformly Donsker, where $\mathcal{D}_2 = \mathcal{D}_{1,M} \times \cdots \times \mathcal{D}_{K,M}$, and M is a finite constant.

Proof. As in the proof of Lemma 1, for any $(\boldsymbol{\beta}^{(1)}, \Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)}, \Lambda_k^{(2)})$ in $\mathcal{B} \times \mathcal{D}_{k,M}$ and $t_k \in \mathcal{U}_k$, we have

$$\begin{aligned} & \left| \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(1)}(s) \right\} - \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(2)}(s) \right\} \right| \\ \leq \; & e^{\widetilde{M} + |b_1|} \left\{ C^* \left\| \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)} \right\| + \left| \Lambda_k^{(1)}(t) - \Lambda_k^{(2)}(t) \right| + \int_0^{\tau_k} \left| \Lambda_k^{(1)}(s) - \Lambda_k^{(2)}(s) \right| ds \right\}. \end{aligned}$$

Thus,

$$\begin{aligned} & \left| J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)}, \mathcal{A}^{(1)}) - J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)}, \mathcal{A}^{(2)}) \right| \\ \leq \; & \widetilde{C} e^{2\widetilde{M}\|\boldsymbol{b}\|} \Bigg\{ \left\| \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)} \right\| + \left\| \boldsymbol{\gamma}^{(1)} - \boldsymbol{\gamma}^{(2)} \right\| + \sum_{k=1}^{K} \int_0^{\tau_k} \left| \Lambda_k^{(1)}(s) - \Lambda_k^{(2)}(s) \right| ds \\ & + \sum_{k=1}^{K_1} \left\{ \left| \Lambda_k^{(1)}(L_k) - \Lambda_k^{(2)}(L_k) \right| + \left| \Lambda_k^{(1)}(R_k) - \Lambda_k^{(2)}(R_k) \right| \right\} + \sum_{k=K_1+1}^{K} \left| \Lambda_k^{(1)}(Y_k) - \Lambda_k^{(2)}(Y_k) \right| \Bigg\}, \end{aligned}$$

where $\widetilde{C}$ is a constant. By the arguments in the proof of Lemma 1, the bracketing numbers of $\mathcal{H}_1$ and $\mathcal{H}_{2k}$ are of the order $O\{\exp(\epsilon^{-1})\}$. Thus, $\mathcal{H}_1$ and $\mathcal{H}_{2k}$ are uniformly Donsker.

Lemma 3. Under Conditions 1–5, if there exist $\boldsymbol{\beta}_* \in \mathcal{B}$, $\boldsymbol{\gamma}_* \in \mathcal{G}$, $\boldsymbol{\Sigma}_* \in \mathcal{S}$, and strictly increasing and continuously differentiable $\Lambda_{k*}(t)$ for $k = 1, \ldots, K$ and $t \in \mathcal{U}_k$ with $\Lambda_{k*}(0) = 0$ such that

$$\begin{aligned} & \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s) \right\} - \exp\left\{ -\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s) \right\} \right] \right) \\ & \times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k*} b_1 + b_2} \Lambda_{k*}(Y_k) \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k*} b_1 + b_2} d\Lambda_{k*}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b} \end{aligned}$$

$$= \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ - \int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} - \exp\left\{ - \int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} \right] \right)$$
$$\times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k0} b_1 + b_2} \Lambda_{k0}(Y_k) \right\}^{\Delta_k} \exp\left\{ - \int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \tag{S.4}$$

with probability 1, then $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_* = \boldsymbol{\gamma}_0$, $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_0$, and $\Lambda_{k*}(t) = \Lambda_{k0}(t)$ for $k = 1, \ldots, K$ and $t \in \mathcal{U}_k$.

Proof. For any $k \in \{1, \ldots, K_1\}$ and $m \in \{0, \ldots, M_k\}$, we set $\delta_{km'} = 1$ in equation (S.4) for $m' = m, \ldots, M_k$ and take the sum of the resulting equations to obtain

$$\begin{aligned}
&\int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \exp\left\{ - \int_0^{U_{km}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s) \right\} \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k*} b_1 + b_2} \Lambda_{k*}(Y_k) \right\}^{\Delta_k} \right. \\
&\left. \exp\left\{ - \int_0^{Y_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k*} b_1 + b_2} d\Lambda_{k*}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b} \\
= \ &\int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \exp\left\{ - \int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k0} b_1 + b_2} \Lambda_{k0}(Y_k) \right\}^{\Delta_k} \right. \\
&\left. \exp\left\{ - \int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}.
\end{aligned}$$

Because m is arbitrary, we can replace U_{km} in the above equation by any $t_k \in \mathcal{U}_k$. For $k = K_1 + 1, \ldots, K$, we set $\Delta_k = 1$ and integrate Y_k from 0 to $t_k \in \mathcal{U}_k$ to obtain

$$\begin{aligned}
&\int_{\boldsymbol{b}} \exp\left\{ - \sum_{k=1}^{K_1} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s) - \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k*} b_1 + b_2} d\Lambda_{k*}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b} \\
= &\int_{\boldsymbol{b}} \exp\left\{ - \sum_{k=1}^{K_1} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) - \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}.
\end{aligned} \tag{S.5}$$

For any $k = 1, \ldots, K_1$, we set $t_{k'} = 0$ for $k' \neq k$ in (S.5) to obtain

$$\int_{b_1} \exp\left\{ -e^{b_1} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s) \right\} \phi(b_1; \sigma_{1*}^2) db_1$$

$$= \int_{b_1} \exp\left\{-e^{b_1}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\}\phi(b_1;\sigma_{10}^2)db_1.$$

By the arguments in the proof of Theorem 1 of Elbers and Ridder (1982), we find $\sigma_{1*}^2 = \sigma_{10}^2$ and

$$\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k*}(s) = \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k0}(s). \tag{S.6}$$

We differentiate both sides with respect to t_k and take the logarithm to obtain

$$\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}_k(t_k) + \log\lambda_{k*}(t_k) = \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(t_k) + \log\lambda_{k0}(t_k) \tag{S.7}$$

for $t_k \in \mathcal{U}_k$ and $k = 1, \ldots, K_1$. For $k = K_1+1, \ldots, K$, we set $t_{k'} = 0$ for $k' \notin \{1, k\}$ in (S.5) to obtain

$$\int_{\boldsymbol{b}} \exp\left\{-e^{b_1}\int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_1(s)} d\Lambda_{10}(s) - e^{\gamma_{k*}b_1+b_2}\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\}\phi(b_1;\sigma_{10}^2)\phi(b_2;\sigma_{2*}^2)d\boldsymbol{b}$$
$$= \int_{\boldsymbol{b}} \exp\left\{-e^{b_1}\int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_1(s)} d\Lambda_{10}(s) - e^{\gamma_{k0}b_1+b_2}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\}\phi(b_1;\sigma_{10}^2)\phi(b_2;\sigma_{20}^2)d\boldsymbol{b}.$$

We let $b_{3k*} = \gamma_{k*}b_1 + b_2$ and $b_{3k0} = \gamma_{k0}b_1 + b_2$ to obtain

$$\begin{aligned}
&\int_{b_1} \exp\left\{-e^{b_1}\int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_1(s)} d\Lambda_{10}(s)\right\} \\
&\times \left[\int_{b_{3k*}} \exp\left\{-e^{b_{3k*}}\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\}\phi(b_{3k*}-\gamma_{k*}b_1;\sigma_{2*}^2)db_{3k*}\right]db_1 \\
= &\int_{b_1} \exp\left\{-e^{b_1}\int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_1(s)} d\Lambda_{10}(s)\right\} \\
&\times \left[\int_{b_{3k0}} \exp\left\{-e^{b_{3k0}}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\}\phi(b_{3k0}-\gamma_{k0}b_1;\sigma_{20}^2)db_{3k0}\right]db_1.
\end{aligned}$$

We apply the inverse Laplace transform to both sides to obtain

$$\begin{aligned}
&\int_{b_{3k*}} \exp\left\{-e^{b_{3k*}}\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\}\phi(b_{3k*}-\gamma_{k*}b_1;\sigma_{2*}^2)db_{3k*} \\
= &\int_{b_{3k0}} \exp\left\{-e^{b_{3k0}}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\}\phi(b_{3k0}-\gamma_{k0}b_1;\sigma_{20}^2)db_{3k0}
\end{aligned}$$

for any b_1. By the arguments in the proof of Theorem 1 of Elbers and Ridder (1982), we find $\sigma_{2*}^2 = \sigma_{20}^2$, $\gamma_{k*} = \gamma_{k0}$, and

$$\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s) = \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s) \tag{S.8}$$

for $k = K_1+1, \ldots, K$. We differentiate both sides with respect to t_k and take the logarithm to obtain

$$\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k*}(t_k) = \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k0}(t_k) \tag{S.9}$$

for $t_k \in \mathcal{U}_k$ and $k = K_1 + 1, \ldots, K$. By Condition 5, (S.7), and (S.9), $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$ and $\lambda_{k*}(t_k) = \lambda_{k0}(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$. We let $\boldsymbol{X}_k(t) = 0$ by redefining $\boldsymbol{X}_k(t)$ to center at a deterministic function in the support of $\boldsymbol{X}_k(t)$ in (S.6) and (S.8) to obtain $\Lambda_{k*}(t_k) = \Lambda_{k0}(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$.

Lemma 4. Under Conditions 1–5, the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible.

Proof. If the matrix is singular, then there exists a vector $\boldsymbol{v} \equiv (\boldsymbol{v}_1, \boldsymbol{v}_2, v_3, v_4)^{\mathrm{T}}$ with $\boldsymbol{v}_1 \in \mathbb{R}^p$, $\boldsymbol{v}_2 \equiv (v_{2,K_1+1}, \ldots, v_{2K}) \in \mathbb{R}^{K_2}$, and $v_3, v_4 \in \mathbb{R}$ such that $\boldsymbol{v}^{\mathrm{T}} E[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]\boldsymbol{v} = 0$. It follows that, with probability 1, the score function along the submodel $\{\boldsymbol{\theta}_0 + \epsilon \boldsymbol{v}, \mathcal{A}_\epsilon(-\boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}^*)\}$ is zero. That is,

$$\begin{aligned}
\int_{\boldsymbol{b}} \Bigg(& \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(t) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(t) \right\} d\Lambda_{k0}(t) \\
& + v_3 \frac{\phi'_{\sigma_{10}^2}(b_1, \sigma_{10}^2)}{\phi(b_1; \sigma_{10}^2)} + v_4 \frac{\phi'_{\sigma_{20}^2}(b_2, \sigma_{20}^2)}{\phi(b_2; \sigma_{20}^2)} + \sum_{k=K_1+1}^{K} \Bigg[\Delta_k \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + v_{2k} b_1 - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(Y_k) \right\} \\
& + \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(t) + v_{2k} b_1 - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(t) \right\} d\Lambda_{k0}(t) \Bigg] \Bigg) \\
& \times J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \mathcal{A}_0) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0
\end{aligned}$$

with probability 1. For any $t_k \in \mathcal{U}_k$ for $k = K_1 + 1, \ldots, K$, we let $\Delta_k = 0$ and set $Y_k = t_k$ to obtain

$$\int_{\boldsymbol{b}} \Bigg[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(s, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s)$$

$$
\begin{aligned}
&+v_3\frac{\phi'_{\sigma^2_{10}}(b_1,\sigma^2_{10})}{\phi(b_1;\sigma^2_{10})}+v_4\frac{\phi'_{\sigma^2_{20}}(b_2,\sigma^2_{20})}{\phi(b_2;\sigma^2_{20})}\\
&+\left.\sum_{k=K_1+1}^{K}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)\right]\\
&\times\left(\prod_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right)\\
&\times\prod_{k=K_1+1}^{K}\exp\left\{-\int_0^{t} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.
\end{aligned}
$$

For any $k=1,\ldots,K_1$ and $m_k\in\{0,\ldots,M_k\}$, we sum over all possible $\delta_{k,m_k'}$ with $m_k'=m_k,\ldots,M_k$ to obtain

$$
\begin{aligned}
&\int_{\boldsymbol{b}}\left[\sum_{k=1}^{K_1}\int_0^{U_{k,m_k}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)+v_3\frac{\phi'_{\sigma^2_{10}}(b_1,\sigma^2_{10})}{\phi(b_1;\sigma^2_{10})}+v_4\frac{\phi'_{\sigma^2_{20}}(b_2,\sigma^2_{20})}{\phi(b_2;\sigma^2_{20})}\right.\\
&\qquad+\left.\sum_{k=K_1+1}^{K}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)\right]\\
&\times\exp\left\{-\sum_{k=1}^{K_1}\int_0^{U_{k,m_k}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)-\sum_{k=K_1+1}^{K}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.
\end{aligned}
$$

Because m_k is arbitrary, we can replace U_{k,m_k} in the above equation by any $t_k\in\mathcal{U}_k$. We apply the inverse Laplace transform to obtain

$$
\begin{aligned}
&\sum_{k=1}^{K_1}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)+v_3\frac{\phi'_{\sigma^2_{10}}(b_1,\sigma^2_{10})}{\phi(b_1;\sigma^2_{10})}+v_4\frac{\phi'_{\sigma^2_{20}}(b_2,\sigma^2_{20})}{\phi(b_2;\sigma^2_{20})}\\
&+\sum_{k=K_1+1}^{K}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)=0
\end{aligned}
$$

for any b_1 and b_2. Therefore, $\boldsymbol{v}_2=\boldsymbol{0}$, $v_3=v_4=0$, and

$$
\sum_{k=1}^{K}\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}\boldsymbol{v}_1^{\mathrm{T}}\left\{\boldsymbol{X}_k(s)-\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)=0.
$$

We differentiate both sides with respect to t_k to obtain $\boldsymbol{v}_1^{\mathrm{T}}\{\boldsymbol{X}_k(t_k)-\boldsymbol{h}_k^*(t_k)\}=0$ for $t_k\in\mathcal{U}_k$ and $k=1,\ldots,K$. By Condition 5, $\boldsymbol{v}_1=\boldsymbol{0}$. Hence, the matrix $\mathbb{P}\big[\big\{\boldsymbol{l}_\theta-l_{\mathcal{A}}(\boldsymbol{h}^*)\big\}^{\otimes 2}\big]$ is invertible.

Lemma 5. Under Conditions 1–5,

$$E\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\left\{\widehat{\Lambda}_k(U_{km})-\Lambda_{k0}(U_{km})\right\}^2+\sum_{k=K_1+1}^{K}\left\{\widehat{\Lambda}_k(Y_k)-\Lambda_{k0}(Y_k)\right\}^2\right]$$
$$= O_P(n^{-2/3})+O\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).$$

Proof. By Theorem 1, $\widehat{\mathcal{A}}$ is consistent for $\mathcal{A}_0$. Thus, there exists a finite constant M such that $\widehat{\Lambda}(\tau_k)\le M$. By the Donsker results in Lemma 2, $m(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})$ is in a Donsker class. Note that

$$\int_0^\delta\sqrt{1+\log N_{[]}(\epsilon,\mathcal{M},L_2(\mathbb{P}))}d\epsilon\le O(\delta^{1/2}).$$

In addition, by Lemma 1.3 of van der Geer (2000) and the mean-value theorem,

$$\mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-m\left(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}}\right)\right\}\le -cH^2\left\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\right\},$$

where c is a positive constant, and $H\{(\theta,\mathcal{A}),(\theta_0,\widetilde{\mathcal{A}})\}$ is the Hellinger distance, defined as

$$\left(\int\left[\exp\left\{\frac{\log L(\theta,\mathcal{A})}{2}\right\}-\exp\left\{\frac{\log L(\theta_0,\widetilde{\mathcal{A}})}{2}\right\}\right]^2d\mu\right)^{1/2}$$

with respect to the dominating measure μ. By Theorem 3.4.1 of van der Vaart and Wellner (1996), there exists r_n with $r_n^2\phi(1/r_n)\sim\sqrt{n}$ such that $H\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\theta_0,\widetilde{\mathcal{A}})\}=O_P(1/r_n)$. In particular, we choose r_n in the order of $n^{1/3}$ such that $H\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\}=O_P(n^{-1/3})$.

By the mean-value theorem,

$$E\left[\left\{\int_{\boldsymbol{b}}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}\right]\right)\right.\right.$$
$$\prod_{k=K_1+1}^{K}\left[\left\{e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\boldsymbol{\gamma}_kb_1+b_2}\widehat{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)(s)+\widehat{\gamma}_kb_1+b_2}d\widehat{\Lambda}_k(s)\right\}\right]\psi(\boldsymbol{b};\widehat{\boldsymbol{\Sigma}})d\boldsymbol{b}$$
$$-\int_{\boldsymbol{b}}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right)$$

$$
\prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2}\widetilde{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\widetilde{\Lambda}_k(s)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}\Bigg\}^2\Bigg]
$$

$$
=O_P(n^{-2/3}).
$$

Consequently, using the mean-value theorem again, we have

$$
\begin{aligned}
&O_P(n^{-2/3})+O(1)\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2+O(1)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+O(1)\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2\\
\geq &E\left(\left[\int_{\boldsymbol{b}}\left\{\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}\right]\right)\right.\right.\\
&\prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2}\widehat{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\widehat{\Lambda}_k(s)\right\}\right]\\
&-\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right)\\
&\left.\left.\prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2}\widetilde{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\widetilde{\Lambda}_k(s)\right\}\right]\right\}\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}\right]^2\right)\\
\geq &c_0E\left\{\left(\int_{\boldsymbol{b}}J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\mathcal{A}_0)\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}Q_1(t,U_{km},U_{k,m+1},b_1,\boldsymbol{X}_k;\boldsymbol{\beta}_0,\Lambda_{k0})d(\widehat{\Lambda}_k-\Lambda_{k0})(t)\right.\right.\right.\\
&\left.\left.\left.+\sum_{k=K_1+1}^{K}\left\{\Delta_k\left(\widehat{\Lambda}_k\{Y_k\}-\widetilde{\Lambda}_k\{Y_k\}\right)+\int_0^{\tau_k}Q_2(t,Y_k,\boldsymbol{b},\boldsymbol{X}_k;\boldsymbol{\beta}_0,\gamma_{k0})d(\widehat{\Lambda}_k-\widetilde{\Lambda}_k)(t)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}\right)^2\right\}
\end{aligned}
$$

for some positive constant c_0. We define a norm in $\mathcal{V}\equiv\prod_{k=1}^{K}BV(\mathcal{U}_k)$ such that for any $\boldsymbol{f}\equiv(f_1,\ldots,f_K)^{\mathrm{T}}\in\mathcal{V}$,

$$
\|\boldsymbol{f}\|_1=\left[E\left\{\sum_{k=1}^{K_1}\sum_{m=0}^{M}f_k(U_{km})^2+\sum_{k=K_1+1}^{K}f_k(Y_k)^2\right\}\right]^{1/2}.
$$

In addition, we define a seminorm

$$
\begin{aligned}
\|\boldsymbol{f}\|_2=&E\left\{\left(\int_{\boldsymbol{b}}J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\mathcal{A}_0)\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}Q_1(t,U_{km},U_{k,m+1},b_1,\boldsymbol{X}_k;\boldsymbol{\beta}_0,\Lambda_{k0})df_k(t)\right.\right.\right.\\
&\left.\left.\left.+\sum_{k=K_1+1}^{K}\left\{\Delta_kf_k(Y_k)+\int_0^{\tau_k}Q_2(t,Y_k,\boldsymbol{b},\boldsymbol{X}_k;\boldsymbol{\beta}_0,\gamma_{k0})df_k(t)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}\right)^2\right\}^{1/2}.
\end{aligned}
$$

Note that if $\|\boldsymbol{f}\|_2 = 0$ for some $\boldsymbol{f} \in \mathcal{V}$, then

$$\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) df_k(t) \right.$$
$$\left. + \sum_{k=K_1+1}^{K} \left\{ \Delta_k f_k(Y_k) + \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) df_k(t) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0 \qquad \text{(S.10)}$$

with probability 1.

Consider $k = K_1+1, \ldots, K$. For $\Delta_k = 0$, we set $Y_k = \tau_k$ in (S.10) to obtain an equation; for $\Delta_k = 1$, we integrate Y_k from 0 to τ_k in (S.10) to obtain another equation. We add all the equations for $k = K_1 + 1, \ldots, K$ to obtain

$$\int_{b_1} \sum_{k=1}^{K_1} \left\{ \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} - \exp\left\{ -\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} \right] \right) \right.$$
$$\left. \times \left\{ \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) df_k(t) \right\} \right\} \phi(b_1; \sigma_{10}^2) db_1 = 0.$$

For any $k \in \{1, \ldots, K_1\}$ and any $m_k \in \{0, \ldots, M_k\}$, we set $U_{k'm} = 0$ for $k' \neq k$ and sum over all possible $\delta_{km'}$ with $m' = m_k, \ldots, M_k$ to obtain

$$\int_{b_1} \left\{ \int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(t) + b_1} df_k(t) \right\} \exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} \phi(b_1; \sigma_{10}^2) db_1 = 0.$$

Therefore, $f_k(U_{km}) = 0$ for $k = 1, \ldots, K_1$. Because m is arbitrary, $f_k(t_k) = 0$ for any $t_k \in \mathcal{U}_k$ for $k = 1, \ldots, K_1$. In addition, we sum over (S.10) with all possible δ_{km} for $k = 1, \ldots, K_1$ and $m = 0, \ldots, M_k$ to obtain

$$\int_{\boldsymbol{b}} \sum_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k0} b_1 + b_2} \Lambda_{k0}\{Y_k\} \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \right.$$
$$\left. \times \left\{ \Delta_k f_k(Y_k) + \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) df_k(t) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0.$$

For $k = K_1 + 1, \ldots, k$, we let $\Delta_k = 0$ and set $Y_k = t_k \in \mathcal{U}_k$ to obtain

$$\int_{\boldsymbol{b}} \sum_{k=K_1+1}^{K} \left[\left\{ \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+} df_k(s) \right\} \exp\left\{ - \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0.$$

We set $t_{k'} = 0$ for $k' \neq k$ to obtain

$$\int_{\boldsymbol{b}} \left\{ \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+} df_k(s) \right\} \exp\left\{ - \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0.$$

Therefore, $f_k(t_k) = 0$ for any $t_k \in \mathcal{U}_k$ for $k = K_1 + 1, \ldots, K$. We obtain $\boldsymbol{f} = \boldsymbol{0}$, implying that $\| \cdot \|_2$ is a norm in $\mathcal{V}$.

By the Cauchy-Schwarz inequality, for any $\boldsymbol{f} \in \mathcal{V}$,

$$\begin{aligned}
\|\boldsymbol{f}\|_2 \leq & \Bigg(E \Bigg[\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \Bigg\{ \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) dt \\
& + \sum_{k=K_1+1}^{K} \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) dt \Bigg\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \Bigg]^2 \\
& \times E \left\{ \sum_{k=1}^{K_1} \sum_{m=0}^{M} f_k(U_{km})^2 + \sum_{k=K_1+1}^{K} f_k(Y_k)^2 \right\} \Bigg)^{1/2} \\
\leq & \; c_1 \|\boldsymbol{f}\|_1,
\end{aligned}$$

where c_1 is a finite constant. By the bounded inverse theorem in the Banach space, we have $\|\boldsymbol{f}\|_2 \geq c_1' \|\boldsymbol{f}\|_1$ for some constant c_1'. Therefore,

$$\begin{aligned}
& O_P(n^{-2/3}) + O\left(\left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + \left\| \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0 \right\|^2 \right) \\
\geq & \; c_0 {c_1'}^2 E \left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \left\{ \widehat{\Lambda}_k(U_{km}) - \Lambda_{k0}(U_{km}) \right\}^2 + \sum_{k=K_1+1}^{K} \left\{ \widehat{\Lambda}_k(Y_k) - \Lambda_{k0}(Y_k) \right\}^2 \right].
\end{aligned}$$

The lemma thus holds.

References

Elbers, C., and Ridder, G. (1982), "True and Spurious Duration Dependence: the Identifiability of the Proportional Hazard Model," *The Review of Economic Studies*, 49, 403–409.

Murphy, S. A. (1994), "Consistency in a Proportional Hazards Model Incorporating a Random Effect," *The Annals of Statistics*, 22, 712–731.

van der Geer, S. A. (2000), *Empirical Processes in M-estimation*, Cambridge: Cambridge University Press.

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

Zeidler, E. (1995), *Applied Functional Analysis – Applications to Mathematical Physics*, New York: Springer.

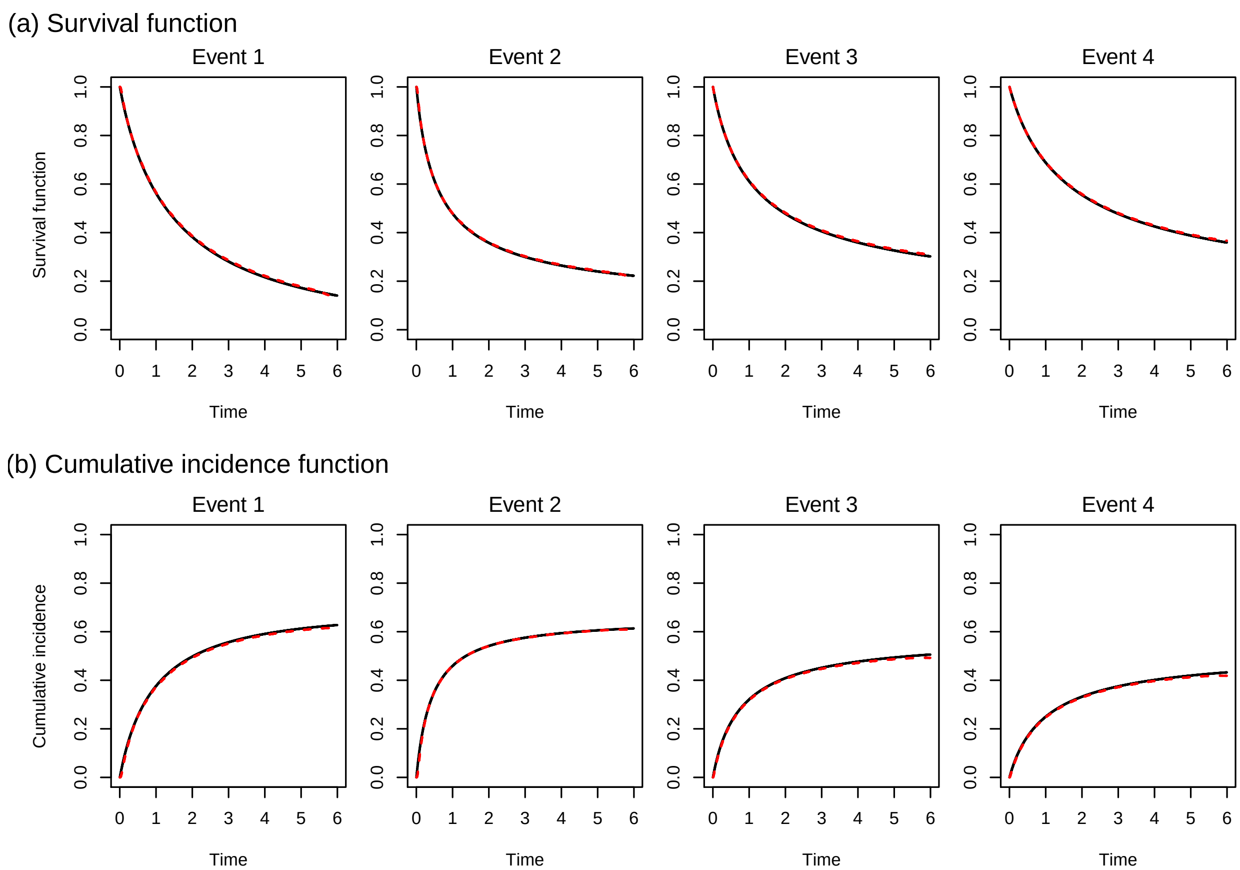

Figure S.1: Estimation of (a) the baseline survival function and (b) the baseline cumulative incidence function with $n = 200$. The solid black curve and dashed red curve pertain, respectively, to the true value and mean estimate from the proposed method.

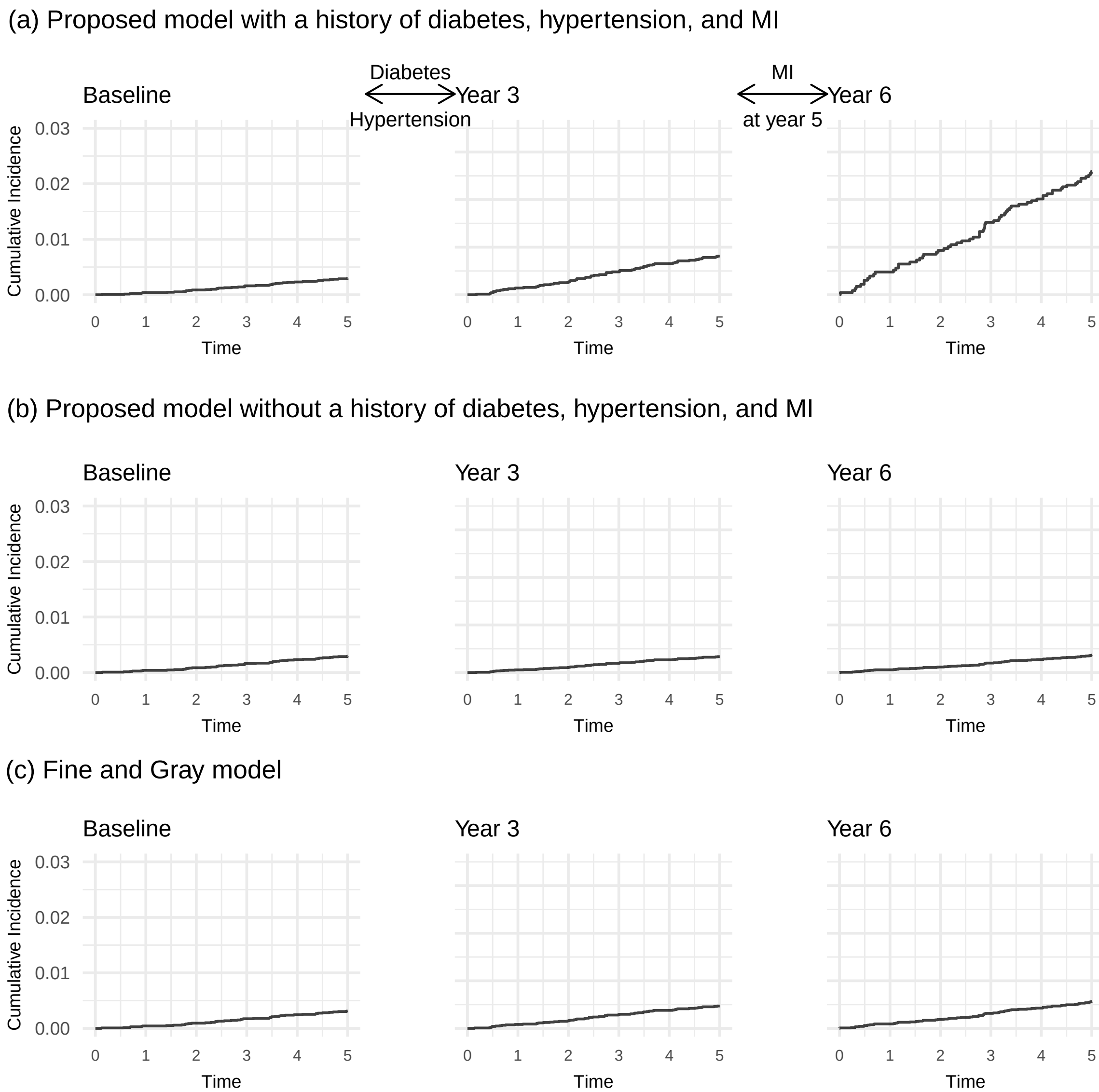

Figure S.2: Estimation of the cumulative incidence of stroke for a 50-year-old white female smoker residing in Forsyth County, NC, with BMI 40 kg/m^2, glucose 98 mg/dl, and systolic blood pressure 113 mmHg: (a) proposed model with MI developed at year 5 and diabetes and hypertension developed between baseline and year 3; (b) proposed model without MI, diabetes or hypertension by year 6; and (c) Fine and Gray model.

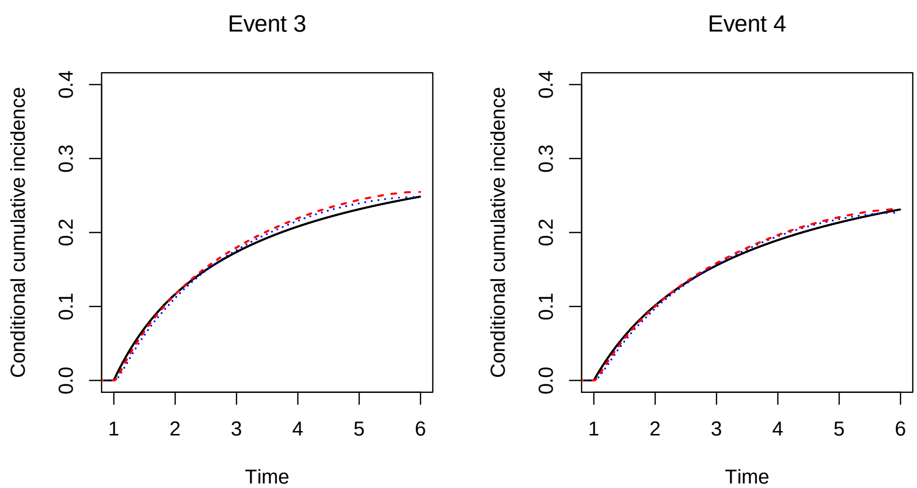

Figure S.3: Estimation of the baseline cumulative incidence function conditional on the event history under the mis-specified model. The solid black curve, dotted blue curve, and dashed red curve pertain, respectively, to the true value and the mean estimates from the proposed method with $n = 100$ and $n = 200$.

Table S.1: Summary Statistics for the Simulation Studies With a Terminal Event

	$n = 100$						$n = 200$					
			Profile		Bootstrap				Profile		Bootstrap	
	Bias	SE	SEE	CP	SEE	CP	Bias	SE	SEE	CP	SEE	CP
β_{11}	0.058	0.825	0.797	0.951	0.879	0.968	0.024	0.505	0.515	0.959	0.544	0.969
β_{12}	0.031	0.445	0.436	0.955	0.502	0.980	0.020	0.291	0.286	0.946	0.310	0.964
β_{21}	0.042	0.780	0.794	0.960	0.881	0.980	0.022	0.521	0.517	0.955	0.549	0.961
β_{22}	−0.019	0.452	0.437	0.948	0.498	0.978	−0.013	0.295	0.288	0.944	0.311	0.961
β_{31}	−0.019	0.675	0.724	0.968	0.694	0.962	−0.014	0.457	0.477	0.957	0.468	0.954
β_{32}	0.007	0.353	0.393	0.973	0.372	0.956	0.005	0.245	0.260	0.963	0.253	0.957
β_{41}	−0.027	0.716	0.775	0.971	0.749	0.959	−0.020	0.483	0.505	0.963	0.500	0.960
β_{42}	0.008	0.385	0.421	0.972	0.405	0.968	0.014	0.268	0.276	0.953	0.271	0.948
β_{51}	0.019	0.631	0.680	0.970	0.659	0.959	−0.004	0.440	0.451	0.959	0.446	0.955
β_{52}	−0.009	0.339	0.361	0.969	0.343	0.950	0.007	0.230	0.240	0.957	0.234	0.950
γ_1	0.004	0.313	0.331	0.972	0.490	0.985	−0.002	0.225	0.226	0.967	0.264	0.977
γ_2	0.009	0.374	0.410	0.982	0.549	0.980	0.008	0.261	0.277	0.977	0.311	0.980
γ_3	0.019	0.335	0.371	0.980	0.502	0.980	0.003	0.244	0.252	0.974	0.281	0.974
σ_1^2	0.134	0.598	0.933	0.969	0.753	0.950	0.077	0.424	0.573	0.970	0.453	0.945
σ_2^2	−0.129	0.390	0.519	0.993	0.532	0.974	−0.047	0.289	0.332	0.988	0.327	0.981

NOTE: SE and SEE denote, respectively, the empirical standard error and mean standard error estimator. CP stands for the empirical coverage probability of the 95% confidence interval based on the Wald method for the profile-likelihood approach and the 95% symmetric confidence interval for the bootstrap approach. For γ_1, γ_2, σ_1^2, and σ_2^2, bias and SEE are based on the median instead of the mean, and SE is based on the mean absolute deviation. For σ_1^2 and σ_2^2, the confidence intervals are based on the log transformation.

Table S.2: Estimation Results for the Random Effects in the ARIC Study

Parameter	Estimate	Std error	p-value
γ_{MI}	0.7145	0.1258	<0.0001
γ_{Stroke}	0.9045	0.1450	<0.0001
γ_{Death}	0.7184	0.1026	<0.0001
σ_1^2	0.5801	0.1215	<0.0001
σ_2^2	1.1465	0.1165	<0.0001