Web-based Supplementary Materials for Double robust semiparametric estimator for differences in restricted mean lifetimes in observational studies

Min Zhang* **and Douglas E. Schaubel**

Department of Biotatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

**email:* mzhangst@umich.edu

1 Notation and Regularity Conditions

We begin by listing notation used in the proofs, much of which has already been introduced in the main manuscript and is repeated here for continuity. For $i = 1, \ldots, n$ and $j = 0, 1$ we define

$$A_{ij} = I(A_i = j)$$
$$p_{ij}(\boldsymbol{\theta}) = \text{expit}\{(-1)^{j+1}\boldsymbol{\theta}^T \boldsymbol{X}_i\}$$
$$\boldsymbol{V}(\boldsymbol{\theta}) = E[\boldsymbol{X}^{\otimes 2} p_{i1}(\boldsymbol{\theta}) p_{i0}(\boldsymbol{\theta})]$$
$$\boldsymbol{R}_j^{(d)}(t;\boldsymbol{\beta}) = n^{-1}\textstyle\sum_{i=1}^n Y_{ij}(t)\boldsymbol{Z}_i^{\otimes d}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i), \qquad \boldsymbol{r}_j^{(d)}(t;\boldsymbol{\beta}) = E\{Y_{ij}(t)\boldsymbol{Z}_i^{\otimes d}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\}$$
$$\overline{\boldsymbol{Z}}_j(t;\boldsymbol{\beta}) = \frac{\boldsymbol{R}_j^{(1)}(t;\boldsymbol{\beta})}{R_j^{(0)}(t;\boldsymbol{\beta})}, \qquad \overline{\boldsymbol{z}}_j(t;\boldsymbol{\beta}) = \frac{\boldsymbol{r}_j^{(1)}(t;\boldsymbol{\beta})}{r_j^{(0)}(t;\boldsymbol{\beta})}$$
$$\boldsymbol{\Omega}_j(\boldsymbol{\beta}) = \textstyle\int_0^\tau \left\{\frac{\boldsymbol{r}_j^{(2)}(t;\boldsymbol{\beta})}{r_j^{(0)}(t;\boldsymbol{\beta})} - \overline{\boldsymbol{z}}_j(t;\boldsymbol{\beta})^{\otimes 2}\right\} E\{Y_{ij}(t)\lambda_{ij}(t)\}dt$$
$$d\Lambda_{0j}^*(t) = \frac{E\{dN_{ij}(t)\}}{r_j^{(0)}(t;\boldsymbol{\beta}_j^*)}$$
$$d\Lambda_{ij}^*(t) = \exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i)d\Lambda_{0j}^*(t)$$
$$dM_{ij}^*(t) = dN_{ij}(t) - Y_{ij}(t)d\Lambda_{ij}^*(t).$$

In addition, parallel to the notation defined above, we define a set of notation, with either superscript or subscript C, that will be used in proofs related to censoring C. Specifically, $\boldsymbol{R}_{Cj}^{(d)}(t;\boldsymbol{\alpha})$, $\boldsymbol{r}_{Cj}^{(d)}(t;\boldsymbol{\alpha})$, $\overline{\boldsymbol{Z}}_j^C(t;\boldsymbol{\alpha})$, $\overline{\boldsymbol{z}}_j^C(t;\boldsymbol{\alpha})$, $\boldsymbol{\Omega}_{Cj}(\boldsymbol{\alpha})$, $\Lambda_{0j}^{C*}(t)$, $\Lambda_{ij}^{C*}(t)$, and $dM_{ij}^{C*}(t)$ are defined as above, except that $\boldsymbol{Z}_i, \boldsymbol{\beta}, \lambda_{ij}(t), N_{ij}(t), \Lambda_{0j}$ are replaced by $\boldsymbol{Z}_i^C, \boldsymbol{\alpha}, \lambda_{ij}^C(t), N_{ij}^C(t), \Lambda_{0j}^C$ respectively. For example,

$$\boldsymbol{R}_{Cj}^{(d)}(t;\boldsymbol{\alpha}) = n^{-1}\textstyle\sum_{i=1}^n Y_{ij}(t)\boldsymbol{Z}_i^{C\otimes d}\exp(\boldsymbol{\alpha}^T\boldsymbol{Z}_i^C), \quad \boldsymbol{r}_{Cj}^{(d)}(t;\boldsymbol{\alpha}) = E\{Y_{ij}(t)\boldsymbol{Z}_i^{C\otimes d}\exp(\boldsymbol{\alpha}^T\boldsymbol{Z}_i^C)\}.$$

We assume the following regularity conditions for $i = 1, \ldots, n$, and $j = 0, 1$:

(a) $P(U_i \geq \tau) > 0$.

(b) $\boldsymbol{Z}_i$ is bounded almost surely.

(c) $\Lambda_{0j}(\tau) < \infty$.

(d) $\boldsymbol{\beta}_j^*$ is the unique solution to

$$\int_0^\tau E\{Y_{ij}(t)\boldsymbol{Z}_i\lambda_{ij}(t)\}dt - \int_0^\tau \overline{\boldsymbol{z}}_j(t;\boldsymbol{\beta})E\{Y_{ij}(t)\lambda_{ij}(t)\}dt = 0,$$

and $\boldsymbol{\Omega}_j(\boldsymbol{\beta}_j^*)$ is positive definite.

(e) $\boldsymbol{\theta}^*$ is the unique maximizer to $E\{A\boldsymbol{\theta}^T\boldsymbol{X} - \log(1 + e^{\boldsymbol{\theta}^T\boldsymbol{X}})\}$ and $\boldsymbol{V}(\boldsymbol{\theta}^*)$ is positive definite.

(f) $\Lambda_{0j}^C(\tau) < \infty$.

(g) $\boldsymbol{\alpha}_j^*$ is the unique solution to

$$\int_0^\tau E\{Y_{ij}(t)\boldsymbol{Z}_i^C\lambda_{ij}^C(t)\}dt - \int_0^\tau \overline{\boldsymbol{z}}_j^C(t;\boldsymbol{\alpha})E\{Y_{ij}(t)\lambda_{ij}^C(t)\}dt = 0,$$

and $\boldsymbol{\Omega}_{Cj}(\boldsymbol{\alpha}_j^*)$ is positive definite.

(h) $P(A_i = j|\boldsymbol{Z}_i)$ is bounded away from 0.

Conditions (a)-(d) ensure the convergence of $\widehat{\boldsymbol{\beta}}_j$ to $\boldsymbol{\beta}_j^*$ and the asymptotic normality of $\widehat{\boldsymbol{\beta}}_j$. Conditions (a),(b),(f) and (g) ensure the convergence of $\widehat{\boldsymbol{\alpha}}_j$ to $\boldsymbol{\alpha}_j^*$ and the asymptotic normality of $\widehat{\boldsymbol{\alpha}}_j$, while condition (g) ensures the convergence of $\widehat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}^*$ and its corresponding asymptotic normality. Condition (h) is the usual positivity assumption required for IPTW-based methods. The boundedness condition (b) ensures the convergence of the several stochastic integrals used in the proofs. Under condition (b) and in conjunction with the other assumptions, it can be shown that there exists a neighborhood $\mathcal{B}_j$ of $\boldsymbol{\beta}_j^*$ such that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}_j} ||\boldsymbol{R}_j^{(d)}(t;\boldsymbol{\beta}) - \boldsymbol{r}_j^{(d)}(t;\boldsymbol{\beta})|| \stackrel{p}{\longrightarrow} 0$$

as $n \to \infty$, where $\stackrel{p}{\longrightarrow}$ means "converges in probability to", with $\boldsymbol{r}_j^{(d)}(t;\boldsymbol{\beta})/r_j^{(0)}(t;\boldsymbol{\beta})$ bounded on $\mathcal{B}_j \times [0, \tau]$ (Lin and Wei, 1989). Similar results hold for $\boldsymbol{R}_{Cj}^{(d)}(t;\boldsymbol{\alpha})$ and $\boldsymbol{r}_{Cj}^{(d)}(t;\boldsymbol{\alpha})/r_{Cj}^{(0)}(t;\boldsymbol{\alpha})$. Condition (b) can be relaxed at the expense of additional technical difficulty.

2 Model for Event Time

The assumed model for the death time hazard is given by

$$\lambda_{ij}(t) \equiv \lambda(t|A_i = j, \boldsymbol{Z}_i) = \lambda_{0j}(t)\exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_i), \quad j = 0, 1.$$

We now provide a set of results pertinent to the asymptotic properties listed in Section 4 of the main manuscript.

2.1 $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)$

Lin and Wei (1989) proved that, under the assumed regularity conditions listed in Section 4 of the main manuscript, $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j^*$, and $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)$ is asymptotically normal with

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*) &= \boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}_i - \overline{\boldsymbol{z}}_j(t;\boldsymbol{\beta}_j^*)\} dM_{ij}^*(t) + o_p(1) \\ &= \boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*) + o_p(1), \end{aligned}$$

where $\boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*) = \int_0^\tau \{\boldsymbol{Z}_i - \overline{\boldsymbol{z}}_j(t;\boldsymbol{\beta}_j^*)\} dM_{ij}^*(t)$. When the assumed model for event time is correct, then $\boldsymbol{\beta}_j^*$ is equal to underlying true target value, $\boldsymbol{\beta}_j$.

2.2 $n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}(t) - \Lambda_{ij}^*(t)\}$

We make the following decomposition:

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}(t) - \Lambda_{ij}^*(t)\} &= n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}(t;\widehat{\boldsymbol{\beta}}_j) - \widehat{\Lambda}_{ij}(t;\boldsymbol{\beta}_j^*)\} && (1) \\ &\quad + n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}(t;\boldsymbol{\beta}_j^*) - \Lambda_{ij}^*(t)\}. && (2) \end{aligned}$$

Through a Taylor series expansion of $\Lambda_{ij}(t;\widehat{\boldsymbol{\beta}}_j)$ around $\boldsymbol{\beta}_j^*$, it is straightforward to show that

$$\begin{aligned} (1) &= \int_0^t \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}_j(u;\boldsymbol{\beta}_j^*)\}^T d\widehat{\Lambda}_{ij}(u;\boldsymbol{\beta}_j^*) n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*) + o_p(1) \\ &= \boldsymbol{K}_{ij}^T(t;\boldsymbol{\beta}_j^*)\boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*) + o_p(1), \end{aligned}$$

where

$$\boldsymbol{K}_{ij}(t;\boldsymbol{\beta}_j^*) = \int_0^t \{\boldsymbol{Z}_i - \overline{\boldsymbol{z}}_j(u;\boldsymbol{\beta}_j^*)\} d\Lambda_{ij}^*(u).$$

As for the second term,

$$\begin{aligned}(2) &= \exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i) n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{ij}^*(u)}{R_j^{(0)}(u;\boldsymbol{\beta}_j^*)} \\ &= \exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i) n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{ij}^*(u)}{r_j^{(0)}(u;\boldsymbol{\beta}_j^*)} + o_p(1).\end{aligned}$$

Combining these results, it then follows that

$$\begin{aligned}& n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}(t) - \Lambda_{ij}^*(t)\} \\ = \; & \boldsymbol{K}_{ij}^T(t;\boldsymbol{\beta}_j^*)\boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*) + \exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i) n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{ij}^*(u)}{r_j^{(0)}(u;\boldsymbol{\beta}_j^*)} + o_p(1).\end{aligned}$$

3 Model for Treatment Probability

The assumed model for treatment assignment is given by

$$\text{logit}\{P(A_i = 1|\boldsymbol{Z}_i)\} = \boldsymbol{\theta}^T \boldsymbol{X}_i,$$

where, as explained in the main manuscript, $\boldsymbol{X}_i$ consists of functions of the elements of $\boldsymbol{Z}_i$ plus an intercept. It is shown in Zeng and Chen (2009) that, under the assumed regularity conditions, $\widehat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}^*$ and

$$\begin{aligned}n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) &= \boldsymbol{V}^{-1}(\boldsymbol{\theta}^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{X}_i\{A_i - \text{expit}(\boldsymbol{\theta}^{*T}\boldsymbol{X}_i)\} + o_p(1) \\ &= \boldsymbol{V}^{-1}(\boldsymbol{\theta}^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{X}_i\{A_{ij} - p_{ij}(\boldsymbol{\theta}^*)\}(-1)^{j+1} + o_p(1),\end{aligned}$$

for $j = 0, 1$. Therefore,

$$\begin{aligned}& \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^T} w_{ij}(\boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right\} n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\ = \; & -w_{ij}(\boldsymbol{\theta}^*)\{1 - p_{ij}(\boldsymbol{\theta}^*)\} \boldsymbol{X}_i^T \boldsymbol{V}^{-1}(\boldsymbol{\theta}^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{X}_i\{A_{ij} - p_{ij}(\boldsymbol{\theta}^*)\} + o_p(1),\end{aligned}$$

where $w_{ij}(\boldsymbol{\theta}) = A_{ij}/p_{ij}(\boldsymbol{\theta})$. If the assumed model for $P(A_i = 1|\boldsymbol{Z}_i)$ is correct, then $\boldsymbol{\theta}^*$ is equal to the truth, $\boldsymbol{\theta}$.

4 Estimating Survival Probability for Censoring

The assumed model for the censoring time hazard is given by

$$\lambda_{ij}^C(t) \equiv \lambda^C(t|A_i = j, \boldsymbol{Z}_i) = \lambda_{0j}^C(t)\exp(\boldsymbol{\alpha}_j^T \boldsymbol{Z}_i^C), \quad j = 0, 1.$$

Parallel to results in Section 2, we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j^*) = \boldsymbol{\Omega}_{Cj}^{-1}(\boldsymbol{\alpha}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{U}_{ij}^C(\boldsymbol{\alpha}_j^*) + o_p(1),$$

where $\boldsymbol{U}_{ij}^C(\boldsymbol{\alpha}_j^*) = \int_0^\tau \{\boldsymbol{Z}_i^C - \overline{\boldsymbol{z}}_j^C(t; \boldsymbol{\alpha}_j^*)\} dM_{ij}^{C*}(t)$, and

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_{ij}^C(t) - \Lambda_{ij}^{C*}(t)\} \\ = \;& \boldsymbol{K}_{ij}^{C^T}(t; \boldsymbol{\alpha}_j^*)\boldsymbol{\Omega}_{Cj}^{-1}(\boldsymbol{\alpha}_j^*) n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{U}_{ij}^C(\boldsymbol{\alpha}_j^*) + \exp(\boldsymbol{\alpha}_j^{*T} \boldsymbol{Z}_i^C) n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{dM_{ij}^{C*}(u)}{r_{j\ \, Cj}^{(0)}(u; \boldsymbol{\alpha}_j^*)} + o_p(1), \end{aligned}$$

where $\boldsymbol{K}_{ij}^C(t; \boldsymbol{\alpha}_j^*) = \int_0^t \{\boldsymbol{Z}_i^C - \overline{\boldsymbol{z}}_j^C(u; \boldsymbol{\alpha}_j^*)\} d\Lambda_{ij}^{C*}(u)$.

5 Consistency

We proposed the following estimator for $\Lambda_j(t)$ in Section 3 of the main manuscript:

$$\widehat{\Lambda}_j(t) = \int_0^t \frac{n^{-1}\sum_{i=1}^n d\widehat{Q}_{ij}(u; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j)}{\widehat{D}_j(u; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j)}, \tag{3}$$

where

$$\begin{aligned} d\widehat{Q}_{ij}(u; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j) &= w_{ij}(\widehat{\boldsymbol{\theta}}) e^{\widehat{\Lambda}_j^C(u)} dN_{ij}(u) + e^{-\widehat{\Lambda}_{ij}(u)} d\widehat{\Lambda}_{ij}(u)\big\{1 - w_{ij}(\widehat{\boldsymbol{\theta}})\widehat{G}_{ij}(u)\big\}, \\ \widehat{D}_j(u; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j) &= n^{-1}\sum_{i=1}^n \big[w_{ij}(\widehat{\boldsymbol{\theta}}) e^{\widehat{\Lambda}_j^C(u)} Y_{ij}(u) + e^{-\widehat{\Lambda}_{ij}(u)}\big\{1 - w_{ij}(\widehat{\boldsymbol{\theta}})\widehat{G}_{ij}(u)\big\}\big], \\ \text{and} \qquad \widehat{G}_{ij}(u) &= 1 - \int_0^u e^{\widehat{\Lambda}_{ij}^C(s) + \widehat{\Lambda}_{ij}(s)} d\widehat{M}_{ij}^C(s). \end{aligned}$$

We now consider two scenarios. In the first scenario (Section 5.1), the coarsening mechanism, i.e., treatment assignment and censoring, is modeled correctly, although the death hazard model may be incorrectly specified. In the second scenario (Section 5.2), the model for the death time, T, is correct, but that for coarsening mechanism may be incorrect, i.e., one or both of the models for treatment assignment, A, and censoring, C, may be incorrect.

5.1 Models for A and C (coarsening) are Correct

If the model for $P(A_i = 1|\boldsymbol{Z}_i)$ is correct, then $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$ and therefore $p_{ij}(\widehat{\boldsymbol{\theta}}) \xrightarrow{p} p_{ij}(\boldsymbol{\theta}) = P(A_{ij} = 1|\boldsymbol{Z}_i)$. In addition, if the model for $C_i|A_i = j, \boldsymbol{Z}_i$ is correct, then $\widehat{\boldsymbol{\alpha}}_j \xrightarrow{p} \boldsymbol{\alpha}_j$ and $\widehat{\Lambda}_{ij}^C \xrightarrow{p} \Lambda_{ij}^C(t)$. As the model for T is possibly wrong, $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j^*$ and $\widehat{\Lambda}_{ij}(t) \xrightarrow{p} \Lambda_{ij}^*(t)$, where $\boldsymbol{\beta}_j^*$ and $\Lambda_{ij}^*(t)$ may or may not equal their respective true values.

Considering the denominator of (3),

$$\begin{aligned}
\widehat{D}_j(u; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j) \quad \xrightarrow{p} \quad & E\{w_{ij}(\boldsymbol{\theta}) e^{\Lambda_{ij}^C(u)} Y_{ij}(\mu)\} && (4)\\
& + E\big[e^{-\Lambda_{ij}^*(u)}\{1 - w_{ij}(\boldsymbol{\theta})\}\big] && (5)\\
& + E\Big\{e^{-\Lambda_{ij}^*(u)} w_{ij}(\boldsymbol{\theta}) \int_0^u e^{\Lambda_{ij}^C(s) + \Lambda_{ij}^*(s)} dM_{ij}^C(s)\Big\}. && (6)
\end{aligned}$$

Using iterated conditional expectation arguments by first conditioning on $\boldsymbol{Z}_i$, we show that

$$\begin{aligned}
(4) \quad &= \quad E\Big[E\Big\{\frac{A_{ij} I(T_i \geq u) I(C_i \geq u)}{p_{ij}(\boldsymbol{\theta}) e^{-\Lambda_{ij}^C(u)}} |\boldsymbol{Z}_i\Big\}\Big]\\
&= \quad E\Big\{\frac{p_{ij}(\boldsymbol{\theta})}{p_{ij}(\boldsymbol{\theta})} \frac{P(C_i \geq u|A_{ij} = 1, \boldsymbol{Z}_i)}{e^{-\Lambda_{ij}^C(u)}} P(T_i^j \geq u|\boldsymbol{Z}_i, A_{ij} = 1)\Big\}\\
&= \quad E\{P(T_i^j \geq u|\boldsymbol{Z}_i)\} = S_j(u),\\
(5) \quad &= \quad E\Big[e^{-\Lambda_{ij}^*(u)}\Big\{1 - \frac{p_{ij}(\boldsymbol{\theta})}{p_{ij}(\boldsymbol{\theta})}\Big\}\Big] = 0, \text{ and}\\
(6) \quad &= \quad E\Big[e^{-\Lambda_{ij}^*(u)} \frac{p_{ij}(\boldsymbol{\theta})}{p_{ij}(\boldsymbol{\theta})} E\Big\{\int_0^u e^{\Lambda_{ij}^C(s) + \Lambda_{ij}^*(s)} dM_{ij}^C(s)|\boldsymbol{Z}_i, A_{ij} = 1\Big\}\Big] = 0.
\end{aligned}$$

Combining the above results, the denominator of (3) $\xrightarrow{p} S_j(u)$.

Using similar techniques, one can obtain that the numerator of (3) converges in probability to $-dS_j(u)$ uniformly in $u \in [0, \tau]$. Combining results for the numerator and denominator, we obtain that $\widehat{\Lambda}_j(t) \xrightarrow{p} \Lambda_j(t)$ uniformly in $t \in [0, \tau]$. Therefore, by the continuous mapping theorem, $\widehat{S}_j(t) \xrightarrow{p} S_j(t)$ uniformly in $t \in [0, \tau]$ and in addition, $\widehat{\mu}_j$and $\widehat{\delta}$ are consistent for μ_j and δ respectively.

5.2 Model for T is correct

If the model for $T|(A, \boldsymbol{Z})$ is correct, then $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$ and $\widehat{\Lambda}_{ij}(t) \xrightarrow{p} \Lambda_{ij}(t)$ uniformly in $t \in [0, \tau]$. As the model for A or C is possibly wrong, the limiting values $\boldsymbol{\theta}^*$, $\boldsymbol{\alpha}_j^*$ and Λ_{ij}^{C*} may or may not equal to the corresponding truth.

Considering the denominator of (3), since $\kappa_i(u)Y_i^T(u) = Y_i(u)$, where $\kappa_i(u) = I(C_i \geq T_i \text{ or } C_i \geq u)$, $\widehat{D}_j(u;\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\beta}}_j,\widehat{\boldsymbol{\alpha}}_j)$ can be rewritten as

$$n^{-1}\sum_{i=1}^{n}\big[e^{-\widehat{\Lambda}_{ij}(u)} + \{w_{ij}(\widehat{\boldsymbol{\theta}})e^{\widehat{\Lambda}_{ij}^{C}(u)}\kappa_i(u)Y_i^T(u) - e^{-\widehat{\Lambda}_{ij}(u)}w_{ij}(\widehat{\boldsymbol{\theta}})\widehat{G}_{ij}(u)\}\big],$$

and it converges in probability to

$$E\big[e^{-\Lambda_{ij}(u)} + w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_j^{C*}(u)}\kappa_i(u)Y_i^T(u) - e^{-\Lambda_{ij}(u)}w_{ij}(\boldsymbol{\theta}^*)\{1-\int_0^u e^{\Lambda_{ij}^{C*}(s)+\Lambda_{ij}(s)}dM_{ij}^{C*}(s)\}\big]. \quad (7)$$

As explained in the main manuscript, in the context of estimating μ_j, the full data T_i^j, $i = 1,\ldots,n$, are monotonically coarensed due to treatment assignment and censoring. Applying Lemma 10.4 of Tsiatis (2006) on monotone coarsening and using the relationship between censoring and monotone coarsening (Chapter 9.3, Tsiatis, 2006), it follows that

$$\frac{A_{ij}\kappa_i(u)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(u)}} = \frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)} - \int_0^u \frac{A_{ij}dM_{ij}^{C*}(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}}. \quad (8)$$

Therefore, the second term inside the expectation of (7) can be rewritten as

$$\begin{aligned} w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_{ij}^{C*}(u)}\kappa_i(u)Y_i^T(u) &= \frac{A_{ij}\kappa_i(u)Y_i^T(u)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(u)}} \\ &= \Big\{\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)} - \int_0^u \frac{A_{ij}dM_{ij}^{C*}(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}}\Big\}I(T_i \geq u). \end{aligned}$$

Substituting the term back to (7), we have

$$\begin{aligned} (7) &= E\{e^{-\Lambda_{ij}(u)}\} + E\Big[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)}\big\{I(T_i\geq u) - e^{-\Lambda_{ij}(u)}\big\}\Big] & (9)\\ &\quad + E\Big[\int_0^u \frac{A_{ij}dN_{ij}^{C}(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}}\Big\{I(T_i\geq u) - \frac{e^{-\Lambda_{ij}(u)}}{e^{-\Lambda_{ij}(s)}}\Big\}\Big] & (10)\\ &\quad - E\Big[\int_0^u \frac{A_{ij}Y_{ij}(s)d\Lambda_{ij}^{C*}(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}}\Big\{I(T_i\geq u) - \frac{e^{-\Lambda_{ij}(u)}}{e^{-\Lambda_{ij}(s)}}\Big\}\Big]. & (11) \end{aligned}$$

It is easy to see that the first term in (9) equals $S_j(u)$ and the second term is equal to zero since

$$\begin{aligned} & E\Big[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)}\big\{I(T_i\geq u) - e^{-\Lambda_{ij}(u)}\big\}\Big] \\ &= E\Big\{E\Big[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)}\{I(T_i\geq u) - e^{-\Lambda_{ij}(u)}\}|A_{ij},\boldsymbol{Z}_i\Big]\Big\} \\ &= E\Big\{\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)}\big[E\{I(T_i\geq u)|\boldsymbol{Z}_i, A_{ij}=1\} - e^{-\Lambda_{ij}(u)}\big]\Big\} = 0. \end{aligned}$$

Applying iterated conditionalexpectation arguments again, we have

$$
\begin{aligned}
(10) \quad &= E\left\{\int_0^u \frac{A_{ij}dN_{ij}^C(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}} E\left[\left\{I(T_i \geq u) - \frac{e^{-\Lambda_{ij}(u)}}{e^{-\Lambda_{ij}(s)}}\right\}|\boldsymbol{Z}_i, A_{ij}=1, C_i=s, T_i \geq s\right]\right\} \\
&= E\Big\{\int_0^u \frac{A_{ij}dN_{ij}^C(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}} \times 0\Big\} = 0, \text{ and}
\end{aligned}
$$

$$
(11) \quad = E\left\{\int_0^u \frac{A_{ij}Y_{ij}(s)d\Lambda_{ij}^{C*}(s)}{p_{ij}(\boldsymbol{\theta}^*)e^{-\Lambda_{ij}^{C*}(s)}} E\left[\left\{I(T_i \geq u) - \frac{e^{-\Lambda_{ij}(u)}}{e^{-\Lambda_{ij}(s)}}\right\}|\boldsymbol{Z}_i, A_{ij}=1, C_i \geq s, T_i \geq s\right]\right\},
$$

which is equal to zero as well. Therefore, the denominator converges in probability to $S_j(u)$.

Similarly, we can show that the numerator, $n^{-1}\sum_{i=1}^n d\widehat{Q}_{ij}(\mu; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\alpha}}_j)$, converges in probability to $-dS_j(u)$ uniformly in $u \in [0, \tau]$. Therefore, the proposed estimators for $\Lambda_{ij}(t)$, μ_j and δ are consistent for the true values when the relationship between survival time and covariates is correctly modeled, even when the treatment assignment model or model for censoring is incorrect.

6 Asymptotic Normality

The proposed estimators are consistent for the true values when either the model for survival time or the models for treatment assignment and censoring are correct. In the proof of asymptotic normality, we do not specify explicitly which working model is correct and we generically denote that $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j^*$, $\widehat{\Lambda}_{ij}(t) \xrightarrow{p} \Lambda_{ij}^*(t)$, $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$, $\widehat{\boldsymbol{\alpha}}_j \xrightarrow{p} \boldsymbol{\alpha}_j^*$, $\widehat{\Lambda}_{ij}^C(t) \xrightarrow{p} \Lambda_{ij}^{C*}(t)$, keeping in mind that they converge to the truth when the corresponding working model is correct. Let us first consider $n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t) - \Lambda_j(t)\}$ which, as we will show, can be approximated by a scaled summation of independent and identically distributed variates. We make the following decomposition:

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t) - \Lambda_j(t)\} \quad &= n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_{ij}, \widehat{\Lambda}_j^C) - \Lambda_j(t)\} \\
= \quad & n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t; \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_{ij}, \widehat{\Lambda}_{ij}^C) - \widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \widehat{\Lambda}_{ij}, \widehat{\Lambda}_{ij}^C)\} & (12) \\
& + n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \widehat{\Lambda}_{ij}, \widehat{\Lambda}_{ij}^C) - \widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \Lambda_{ij}^*, \widehat{\Lambda}_{ij}^C)\} & (13) \\
& + n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \Lambda_{ij}^*, \widehat{\Lambda}_{ij}^C)\} - \widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \Lambda_{ij}^*, \Lambda_{ij}^{C*})\} & (14) \\
& + n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t; \boldsymbol{\theta}^*, \Lambda_{ij}^*, \Lambda_{ij}^{C*}) - \Lambda_j(t)\}. & (15)
\end{aligned}
$$

Considering (12), by a Taylor series expansion around $\boldsymbol{\theta}^*$ and substituting results in 3, we obtain that

$$(6.1) \quad = \quad \boldsymbol{B}_j^T(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\boldsymbol{V}^{-1}(\boldsymbol{\theta}^*)n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{X}_i\{A_{ij}-p_{ij}(\boldsymbol{\theta}^*)\}+o_p(1),$$

where

$$\begin{aligned}
&\boldsymbol{B}_j(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) = \\
&\int_0^t \frac{E\big[\{e^{\Lambda_{ij}^{C*}(u)}dN_{ij}(u)-e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)G_{ij}(u)\}\times(-1)w_{ij}(\boldsymbol{\theta}^*)\{1-p_{ij}(\boldsymbol{\theta}^*)\}\boldsymbol{X}_i\big]}{D_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&+\int_0^t E\big[\{e^{\Lambda_{ij}^{C*}(u)}Y_{ij}(u)-e^{-\Lambda_{ij}^{C*}(u)}G_{ij}(u)\}w_{ij}(\boldsymbol{\theta}^*)\{1-p_{ij}(\boldsymbol{\theta}^*)\}\boldsymbol{X}_i\big]\frac{dQ_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)}{D_j^2(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)},
\end{aligned}$$

with

$$\begin{aligned}
dQ_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) &= E\{d\widehat{Q}_{ij}(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\}, \\
D_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) &= E\{\widehat{D}_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\}, \\
G_{ij}(u) &= 1-\int_0^u e^{\Lambda_{ij}^{C*}(s)+\Lambda_{ij}^*(s)}dM_{ij}^{C*}(s).
\end{aligned}$$

Considering (13), by a Taylor series expansion of $\widehat{\Lambda}_{ij}(u)$ around $\Lambda_{ij}^*(u)$ and substituting results from 2.2, after a lot of algebra and interchanging orders of integration, we have

$$(13) = \boldsymbol{F}_j^T(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*)n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*)+n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t J_j(u,t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\frac{dM_{ij}^*(u)}{r_j^{(0)}(u;\boldsymbol{\beta}_j^*)}$$

plus a term that converges to zero, where

$$\begin{aligned}
&\boldsymbol{F}_j(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) = \\
&\int_0^t \frac{E\big\{e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)\big[\{1-G_{ij}(u)w_{ij}(\boldsymbol{\theta}^*)\}\{-\boldsymbol{K}_{ij}(u;\boldsymbol{\beta}_j^*)+\boldsymbol{Z}_i-\bar{z}_j(u;\boldsymbol{\beta}_j^*)\}+w_{ij}(\boldsymbol{\theta}^*)\boldsymbol{\mathcal{O}}_{ij}(u)\big]\big\}}{D_j(u;\,\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&+\int_0^t E\big\{e^{-\Lambda_{ij}^*(u)}\big[\{1-G_{ij}(u)w_{ij}(\boldsymbol{\theta}^*)\}\boldsymbol{K}_{ij}(u;\boldsymbol{\beta}_j^*)-w_{ij}(\boldsymbol{\theta}^*)\boldsymbol{\mathcal{O}}_{ij}(u)\big]\big\}\frac{dQ_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)}{D_j^2(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)},
\end{aligned}$$

$$
\begin{aligned}
&J_j(u,t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) = \\
&\int_u^t \frac{E\big[e^{-\Lambda_{ij}^*(s)} d\Lambda_{ij}^*(s)\exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i)\big\{-1+G_{ij}(s)w_{ij}(\boldsymbol{\theta}^*)+w_{ij}(\boldsymbol{\theta}^*)\mathcal{E}_{ij}(u,s)\big\}\big]}{D_j(s;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&+\frac{E\big[\{1-G_{ij}(u)w_{ij}(\boldsymbol{\theta}^*)\}e^{-\Lambda_{ij}^*(u)}\exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i)\big]}{D_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&+\int_u^t E\big[e^{-\Lambda_{ij}^*(s)}\exp(\boldsymbol{\beta}_j^{*T}\boldsymbol{Z}_i)\big\{1-G_{ij}(s)w_{ij}(\boldsymbol{\theta}^*)-w_{ij}(\boldsymbol{\theta}^*)\mathcal{E}_{ij}(u,s)\big\}\big]\frac{dQ_j(s;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)}{D_j^2(s;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)},
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{\mathcal{O}}_{ij}(u) &= \int_0^u e^{\Lambda_{ij}^{C*}(s)+\Lambda_{ij}^*(s)}\boldsymbol{K}_{ij}(s;\boldsymbol{\beta}_j^*)dM_{ij}^{C*}(s), \\
\mathcal{E}_{ij}(u,s) &= \int_u^s e^{\Lambda_{ij}^{C*}(v)+\Lambda_{ij}^*(v)}dM_{ij}^{C*}(v).
\end{aligned}
$$

Using similar techniques and substituting results in 4, we can represent (14) as

$$
(14) \quad = \quad \boldsymbol{P}_j^T(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\boldsymbol{\Omega}_{Cj}^{-1}(\boldsymbol{\alpha}_j^*)n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{U}_{ij}^C(\boldsymbol{\alpha}_j^*)+n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t H_j(u,t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)\frac{dM_{ij}^{C*}(u)}{r_{Cj}^{(0)}(u;\boldsymbol{\alpha}_j^*)},
$$

plus a term that converges to zero in probability, where

$$
\begin{aligned}
&\boldsymbol{P}_j(t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) = \\
&+\int_0^t \frac{E\big\{w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_{ij}^{C*}(u)}\boldsymbol{K}_{ij}^C(u;\boldsymbol{\alpha}_j^*)dN_{ij}(u)+w_{ij}(\boldsymbol{\theta}^*)\boldsymbol{\mathcal{R}}_{ij}(u)e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)\big\}}{D_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&-\int_0^t E\{w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_{ij}^{C*}(u)}\boldsymbol{K}_{ij}^C(u;\boldsymbol{\alpha}_j^*)Y_{ij}(u)+w_{ij}(\boldsymbol{\theta}^*)\boldsymbol{\mathcal{R}}_{ij}(u)e^{-\Lambda_{ij}^*(u)}\}\frac{dQ_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)}{D_j^2(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)},
\end{aligned}
$$

$$
\begin{aligned}
&H_j(u,t;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*) = \\
&\int_u^t \frac{E\big[w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_{ij}^{C*}(u)}e^{\boldsymbol{\alpha}_j^{*T}\boldsymbol{Z}_i^C}dN_{ij}(u)+w_{ij}(\boldsymbol{\theta}^*)e^{\boldsymbol{\alpha}_j^{*T}\boldsymbol{Z}_i^C}\mathcal{H}_{ij}(u,s)e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(s)\big]}{D_j(u;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)} \\
&-\int_u^t E\big\{w_{ij}(\boldsymbol{\theta}^*)e^{\Lambda_{ij}^{C*}(s)}e^{\boldsymbol{\alpha}_j^{*T}\boldsymbol{Z}_i^C}Y_{ij}(s)+w_{ij}(\boldsymbol{\theta}^*)e^{\boldsymbol{\alpha}_j^{*T}\boldsymbol{Z}_i^C}\mathcal{H}_{ij}(u,s)e^{-\Lambda_{ij}^*(s)}\big\}\frac{dQ_j(s;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)}{D_j^2(s;\boldsymbol{\theta}^*,\boldsymbol{\beta}_j^*,\boldsymbol{\alpha}_j^*)},
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{\mathcal{R}}_{ij}(u) &= \int_0^u e^{\Lambda_{ij}^{C*}(s)+\Lambda_{ij}^*(s)}[\boldsymbol{K}_{ij}^C(s;\boldsymbol{\alpha}_j^*)dM_{ij}^{C*}(s)-\{\boldsymbol{Z}_i^C-\bar{z}_j^C(s;\boldsymbol{\alpha}^*)\}Y_{ij}(s)d\Lambda_{ij}^{C*}(s)], \\
\mathcal{H}_{ij}(u,s) &= \int_u^s e^{\Lambda_{ij}^{C*}(v)+\Lambda_{ij}^*(v)}dM_{ij}^{C*}(v)-e^{\Lambda_{ij}^{C*}(u)+\Lambda_{ij}^*(u)}Y_{ij}(u).
\end{aligned}
$$

Finally, as for the last term, it is straightforward to show that

$$(15) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \frac{w_{ij}(\boldsymbol{\theta}^*) e^{\Lambda_{ij}^{C*}(u)} dM_{ij}^{\dagger}(u) + \{1 - w_{ij}(\boldsymbol{\theta}^*) G_{ij}(u)\} e^{-\Lambda_{ij}^*(u)} \{d\Lambda_{ij}^*(u) - d\Lambda_j(u)\}}{D_j(u; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*)}$$

plus a term that converges in probability to zero, where $dM_{ij}^{\dagger}(u) = dN_{ij}(u) - Y_{ij}(u) d\Lambda_j(u)$.

Combining the above results, we have shown that we can represent $n^{\frac{1}{2}}\{\widehat{\Lambda}_j(t) - \Lambda_j(t)\}$ as $n^{-\frac{1}{2}} \sum_{i=1}^{n} \varphi_{ij}(t)$ plus a term that converges in probability to zero, where

$$\begin{aligned}
\varphi_{ij}(t) \;=\; & \boldsymbol{B}_j^T(t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) \boldsymbol{V}^{-1}(\boldsymbol{\theta}^*) \boldsymbol{X}_i \{A_{ij} - p_{ij}(\boldsymbol{\theta}^*)\} \\
& + \boldsymbol{F}_j^T(t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) \boldsymbol{\Omega}_j^{-1}(\boldsymbol{\beta}_j^*) \boldsymbol{U}_{ij}(\boldsymbol{\beta}_j^*) + \int_0^t J_j(u, t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) \frac{dM_{ij}^*(u)}{r_j^{(0)}(u; \boldsymbol{\beta}_j^*)} \\
& + \boldsymbol{P}_j^T(t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) \boldsymbol{\Omega}_{Cj}^{-1}(\boldsymbol{\alpha}_j^*) \boldsymbol{U}_{ij}^C(\boldsymbol{\alpha}_j^*) + \int_0^t H_j(u, t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) \frac{dM_{ij}^{C*}(u)}{r_{Cj}^{(0)}(u; \boldsymbol{\alpha}_j^*)} \\
& + \int_0^t \frac{w_{ij}(\boldsymbol{\theta}^*) e^{\Lambda_{ij}^{C*}(u)} dM_{ij}^{\dagger}(u) + \{1 - w_{ij}(\boldsymbol{\theta}^*) G_{ij}(u)\} e^{-\Lambda_{ij}^*(u)} \{d\Lambda_{ij}^*(u) - d\Lambda_j(u)\}}{D_j(u; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*)}.
\end{aligned}$$

When one of the working models is correct, using techniques similar to those used to prove consistency of $\widehat{\Lambda}_{ij}(t)$ for $\Lambda_{ij}(t)$, it can be shown that $\varphi_{ij}(t)$ has mean zero and are identically and independently distributed across $i = 1, \ldots, n$. The expression for $\varphi_{ij}(t)$ seems complicated; however, this is derived without explicitly assuming which one of the working models is correct. In fact, when one or all of the models are correct, some of the terms will be identically equal to zero. For example, if the model for $T_i | \boldsymbol{Z}_i, A_i = j$ is correctly specified, then $\boldsymbol{B}_j(t; \boldsymbol{\beta}_j^*, \boldsymbol{\theta}^*, \boldsymbol{\alpha}_j^*)$ is equal to zero. To see this, we note that in this case, $\boldsymbol{\beta}_j^* = \boldsymbol{\beta}_j$ and $\Lambda_{ij}^*(t) = \Lambda_{ij}(t)$. Substituting the result (8), it follows that

$$\begin{aligned}
& E\left[\{e^{\Lambda_j^C(u)} Y_{ij}(u) - e^{-\Lambda_{ij}^*(u)} G_{ij}(u)\} w_{ij}(\boldsymbol{\theta}^*) \{1 - p_{ij}(\boldsymbol{\theta}^*)\} \boldsymbol{X}_i\right] \\
& = E\Bigg\{ \Big[\frac{A_{ij}}{p_{ij}(\boldsymbol{\theta}^*)} \{I(T_i \geq u) - e^{-\Lambda_{ij}^*(u)}\} - \int_0^u \frac{A_{ij} dM_{ij}^{C*}(s)}{p_{ij}(\boldsymbol{\theta}^*) e^{-\Lambda_{ij}^{C*}(s)}} \Big\{ I(T_i \geq u) - \frac{e^{-\Lambda_{ij}^*(u)}}{e^{-\Lambda_{ij}^*(s)}} \Big\} \Big] \\
& \times \{1 - p_{ij}(\boldsymbol{\theta}^*)\} \boldsymbol{X}_i \Bigg\},
\end{aligned}$$

which can be shown to be zero using the iterated conditional expectation arguments similar to that in section 5.2. Similarly, $E\big[\{e^{\Lambda_j^C(u)} dN_{ij}(u) - e^{-\Lambda_{ij}^*(u)} d\Lambda_{ij}^*(u) G_{ij}(u)\}(-1) w_{ij}(\boldsymbol{\theta}^*)\{1 - p_{ij}(\boldsymbol{\theta}^*)\} \boldsymbol{X}_i\big] = 0$. Therefore, $\boldsymbol{B}_j(t; \boldsymbol{\theta}^*, \boldsymbol{\beta}_j^*, \boldsymbol{\alpha}_j^*) = 0$ when the model for T is correct. In addition, if the models for $P(A_i = 1|\boldsymbol{Z}_i)$ and $C_i|\boldsymbol{Z}_i, A_i = j$ are correct, through arguments similar to those in Section 5.1, we can show that both $\boldsymbol{F}_j(t; \boldsymbol{\beta}^*, \boldsymbol{\theta}^*, \boldsymbol{\alpha}_j^*)$ and $J_j(u, t; \boldsymbol{\beta}_j^*, \boldsymbol{\theta}^*, \boldsymbol{\alpha}_j^*)$ are identically zero.

Considering estimation of μ_j, $n^{\frac{1}{2}}(\widehat{\mu}_j - \mu_j)$ can be written as

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\mu}_j - \mu_j) &= n^{\frac{1}{2}} \int_0^L \widehat{S}_j(u) - S_j(u) du \\ &= n^{\frac{1}{2}} \int_0^L e^{-\widehat{\Lambda}_j(u)} - e^{-\Lambda_j(u)} du \\ &= -n^{\frac{1}{2}} \int_0^L S_j(u)\{\widehat{\Lambda}_j(u) - \Lambda_j(u)\} du + o_p(1) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \phi_{ij} + o_p(1), \end{aligned}$$

where $\phi_{ij} = -\int_0^L S_j(u)\varphi_{ij}(u)du$, which is referred to as the ith influence function of $\widehat{\mu}_j$. When either the model for death time or the models for coarsening probability is correct, the ϕ_{ij} variates are independent and identically distributed with mean 0. Therefore, $n^{\frac{1}{2}}(\widehat{\mu}_j - \mu_j)$ converges to a normal distribution with mean 0 and variance $E(\phi_{ij}^2)$. It then follows that $n^{\frac{1}{2}}(\widehat{\delta} - \delta)$ is also asymptotically normal with mean 0 and variance $E(\phi_{i1} - \phi_{i0})^2$ with $n^{\frac{1}{2}}(\widehat{\delta} - \delta) = n^{-\frac{1}{2}} \sum_{i=1}^n (\phi_{i1} - \phi_{i0}) + o_p(1)$.

7 Additional Simulation

Additional simulation results are reported in Tables 1 and 2. Data are generated the same methods described in the main manuscript, except that a sample size of $n = 300$ is used.

ADDITIONAL REFERENCES

Fleming, T.R. and Harrington D.P. (1991). Counting Processes and Survival Analysis. New York: John Wiley and Sons.

Zeng, D. and Chen, Q. (2009). Adjustment for Missingness Using Auxiliary Information in Semiparametric Regression. *Biometrics* **66,** 115–122.

Table 1: Additional simulation with sample size n=300 and L=10. Entries as in Table 1 in the main manuscript.

Method	T	$\boldsymbol{Z}$	C	BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
				$\widehat{\mu}_0$ (μ_0=5.978)				$\widehat{\mu}_1$ (μ_1=6.949)			
Proposed	T	T	T	0.009	0.293	0.284	0.941	0.002	0.272	0.267	0.944
	T	F	F	0.025	0.295	0.283	0.929	-0.006	0.269	0.267	0.949
	F	T	T	0.009	0.301	0.315	0.948	-0.002	0.276	0.287	0.958
	F	F	F	0.405	0.313	0.300	0.687	-0.301	0.294	0.288	0.804
Hubbard	T	T	T	0.040	0.295	0.289	0.930	0.053	0.272	0.269	0.936
	T	F	F	0.060	0.295	0.284	0.923	0.052	0.270	0.268	0.937
	F	T	T	0.045	0.304	0.322	0.944	0.050	0.277	0.288	0.944
	F	F	F	0.462	0.315	0.302	0.620	-0.248	0.294	0.290	0.853
IPTW		T		-0.102	0.324	0.376	0.970	0.034	0.286	0.319	0.963
		F		0.258	0.328	0.352	0.889	-0.269	0.308	0.322	0.888
Chen & Tsiatis	T			0.012	0.285	0.277	0.931	0.009	0.263	0.259	0.949
	F			0.296	0.312	0.298	0.808	-0.326	0.285	0.282	0.772
				$\widehat{\delta} = \widehat{\mu}_1 - \widehat{\mu}_0$ (δ=0.871)							
Proposed	T	T	T	-0.007	0.314	0.310	0.946				
	T	F	F	-0.032	0.314	0.309	0.938				
	F	T	T	-0.011	0.326	0.375	0.966				
	F	F	F	-0.706	0.377	0.364	0.498				
Hubbard	T	T	T	0.013	0.316	0.315	0.943				
	T	F	F	-0.008	0.314	0.308	0.944				
	F	T	T	0.005	0.330	0.383	0.971				
	F	F	F	-0.710	0.379	0.366	0.500				
IPTW		T		0.136	0.356	0.494	0.988				
		F		-0.527	0.406	0.478	0.832				
Chen & Tsiatis	T			-0.003	0.301	0.293	0.933				
	F			-0.622	0.373	0.357	0.584				

Table 2: Additional simulation with sample size n=300 and L=20. Entries as in Table 1 in the main manuscript.

Method	T	$\boldsymbol{Z}$	C	BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
				$\widehat{\mu}_0$ (μ_0=9.806)				$\widehat{\mu}_1$ (μ_1=11.488)			
Proposed	T	T	T	0.007	0.564	0.576	0.949	0.020	0.582	0.576	0.947
	T	F	F	0.050	0.575	0.572	0.947	0.007	0.575	0.573	0.948
	F	T	T	0.001	0.579	0.628	0.971	0.014	0.594	0.623	0.955
	F	F	F	0.824	0.640	0.627	0.721	-0.677	0.616	0.610	0.789
Hubbard	T	T	T	0.024	0.573	0.598	0.955	0.127	0.583	0.580	0.937
	T	F	F	0.108	0.573	0.573	0.949	0.124	0.578	0.578	0.946
	F	T	T	0.025	0.590	0.651	0.970	0.121	0.596	0.629	0.955
	F	F	F	0.926	0.643	0.628	0.669	-0.573	0.618	0.615	0.832
IPTW		T		-0.416	0.609	0.723	0.934	0.131	0.639	0.715	0.969
		F		0.274	0.649	0.707	0.945	-0.569	0.654	0.697	0.889
Chen & Tsiatis	T			0.020	0.552	0.551	0.954	0.025	0.560	0.560	0.949
	F			0.441	0.611	0.598	0.877	-0.658	0.591	0.597	0.801
				$\widehat{\delta} = \widehat{\mu}_1 - \widehat{\mu}_0$ (δ=1.682)							
Proposed	T	T	T	0.013	0.637	0.642	0.949				
	T	F	F	-0.043	0.645	0.635	0.936				
	F	T	T	0.013	0.658	0.774	0.970				
	F	F	F	-1.501	0.792	0.762	0.492				
Hubbard	T	T	T	0.103	0.648	0.666	0.943				
	T	F	F	0.016	0.645	0.638	0.940				
	F	T	T	0.096	0.672	0.800	0.966				
	F	F	F	-1.498	0.795	0.767	0.497				
IPTW		T		0.547	0.731	1.020	0.974				
		F		-0.844	0.845	0.995	0.899				
Chen & Tsiatis	T			0.005	0.608	0.601	0.942				
	F			-1.100	0.754	0.730	0.677				