Propensity Score Weighting Analysis of Survival Outcomes Using Pseudo-observations

Shuxi Zeng[1], Fan Li [1], Liangyuan Hu [2], Fan Li [3,*]

[1]*Duke University,* [2]*Rutgers School of Public Health*

and [3]*Yale School of Public Health*

*fan.f.li@yale.edu

Supplementary Material

The online supplementary materials include the Web Appendix A-F with technical details and additional simulations, as well as Web Tables and Figures referenced in Section 4 and 5.

Web Appendix A Proof of theoretical properties

Proof of Theorem 1 (i) We first list the regularity assumptions needed for Theorem 1.

- (R1) We only consider time point $t < \bar{t}$ such that $G(\bar{t}) > \epsilon > 0$, where G is the survival function for the censoring time C_i. Namely, any time point of interest has a strictly positive probability of not being censored.

- (R2) The generalized propensity score model (GPS), $e_j(\mathbf{X}_i;\boldsymbol{\gamma})$, satisfies the regularity conditions specified in Theorem 5.1 of Lehmann and Casella (2006). The GPS $e_j(\mathbf{X}_i;\boldsymbol{\gamma})$ is correctly specified in the sense that there exists γ such that $e_j(\mathbf{X}_i) = e_j(\mathbf{X}_i;\boldsymbol{\gamma})$ and $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + o_p(1)$.

Next, we establish the consistency of estimator (5) in the main text. Let $D_{ij} = \mathbf{1}\{Z_i = j\}$. Notice that $h(\mathbf{X}_i) = h(\mathbf{X}_i;\widehat{\gamma})$ might be a function of the estimated GPS. With a first-order Taylor expansion, we can show that,

$$\begin{aligned}
w_j^h(\mathbf{X}_i;\widehat{\gamma}) = \frac{h(\mathbf{X}_i;\widehat{\gamma})}{e_j(\mathbf{X}_i;\widehat{\gamma})} &= \frac{h(\mathbf{X}_i;\boldsymbol{\gamma})}{e_j(\mathbf{X}_i;\boldsymbol{\gamma})} + \left[\frac{\partial}{\partial\boldsymbol{\gamma}}\left\{\frac{h(\mathbf{X}_i;\boldsymbol{\gamma})}{e_j(\mathbf{X}_i;\boldsymbol{\gamma})}\right\}\right]^T(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(1) \\
&= w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) + o_p(1)
\end{aligned}$$

where the last equation follows from the regularity assumption that $\frac{\partial}{\partial\boldsymbol{\gamma}}\left\{\frac{h(\mathbf{X}_i;\boldsymbol{\gamma})}{e_j(\mathbf{X}_i;\boldsymbol{\gamma})}\right\}$ as a function of $e_j(\mathbf{X}_i;\boldsymbol{\gamma})$ is uniformly bounded around a neighborhood as $\hat{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + o_p(1)$. Therefore, we can show that,

$$\begin{aligned}
\frac{\sum_{i=1}^N D_{ij}\widehat{\theta}_i^k(t)w_j^h(\mathbf{X}_i;\hat{\boldsymbol{\gamma}})}{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i;\hat{\boldsymbol{\gamma}})} &= \frac{\mathbb{E}\{D_{ij}\widehat{\theta}_i^k(t)w_j^h(\mathbf{X}_i;\boldsymbol{\gamma})\}}{\mathbb{E}(h(\mathbf{X}_i))} + o_p(1) \\
&= \frac{\mathbb{E}[\{D_{ij}\widehat{\theta}_i^k(t)w_j^h(\mathbf{X}_i;\boldsymbol{\gamma})|\mathbf{X}_i\}]}{\mathbb{E}(h(\mathbf{X}_i))} + o_p(1) \\
&= \frac{\mathbb{E}\{w_j^h(\mathbf{X}_i;\boldsymbol{\gamma})e_j(\mathbf{X}_i;\boldsymbol{\gamma})\,\mathbb{E}(v_k(T_i;t)|\mathbf{X}_i, D_{ij}=1)\}}{\mathbb{E}(h(\mathbf{X}_i))} + o_p(1) \\
&= \frac{\mathbb{E}\{w_j^h(\mathbf{X}_i;\boldsymbol{\gamma})e_j(\mathbf{X}_i;\boldsymbol{\gamma})\,\mathbb{E}(v_k(T_i(j);t)|\mathbf{X}_i)\}}{\mathbb{E}(h(\mathbf{X}_i))} + o_p(1) \\
&= \frac{\mathbb{E}\{h(\mathbf{X}_i)\,\mathbb{E}(v_k(T_i(j);t)|\mathbf{X}_i)\}}{\mathbb{E}(h(\mathbf{X}_i))} + o_p(1) \\
&= m_j^{k,h}(t) + o_p(1)
\end{aligned}$$

where the third equality follows from the fact that $\widehat{\theta}_i^k(t) = \theta^k(t) + \phi'_{k,i}(t) + o_p(1)$ and $\mathbb{E}(\theta^k(t) + \phi'_{k,i}(t)|\mathbf{X}_i, D_{ij} = 1) = \mathbb{E}(v_k(T_i;t)|\mathbf{X}_i, D_{ij} = 1)$ (Graw et al., 2009; Jacobsen and Martinussen, 2016) and the fourth equality follows from the unconfoundedness assumption (A1). Therefore, it follows that,

$$\frac{\sum_{i=1}^N D_{ij}\widehat{\theta}_i^k(t)w_j^h(\mathbf{X}_i)}{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)} - \frac{\sum_{i=1}^N D_{ij'}\widehat{\theta}_i^k(t)w_{j'}^h(\mathbf{X}_i)}{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)} \xrightarrow{p} m_j^{k,h}(t) - m_{j'}^{k,h}(t) = \tau_{j,j'}^{k,h}(t),$$

and thus the consistency of the weighting estimator (5) is proved.

(ii) Below we derive the asymptotic variance of estimator (5) using the von Mises expansion on the pseudo-observations (Jacobsen and Martinussen, 2016; Overgaard et al., 2017). Recall that estimator (5) is of the following form:

$$\hat{\tau}_{j,j'}^{k,h}(t) = \frac{\sum_{i=1}^N D_{ij}\widehat{\theta}_i^k(t)w_j^h(\mathbf{X}_i)}{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)} - \frac{\sum_{i=1}^N D_{ij'}\widehat{\theta}_i^k(t)w_{j'}^h(\mathbf{X}_i)}{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)} = \widehat{m}_j^{k,h}(t) - \widehat{m}_{j'}^{k,h}(t).$$

We can write the treatment-specific average potential outcome $\hat{m}_j^{k,h}(t)$ as the solution to the following estimating equation,

$$\sum_{i=1}^N D_{ij}\{\widehat{\theta}_i^k(t) - \widehat{m}_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) = 0.$$

A first-order Taylor expansion at the true value of $(m_j^{k,h}(t), \boldsymbol{\gamma})$ yields,

$$\sqrt{N}\{\widehat{m}_j^{k,h}(t) - m_j^{k,h}(t)\} = \bar{\omega}^{-1}\frac{1}{\sqrt{N}}\left\{\sum_{i=1}^N D_{ij}\{\widehat{\theta}_i^k(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) \right.$$
$$\left. + \mathbf{H}_j^T\sqrt{N}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right\} + o_p(1),$$

where $\bar{\omega} = \mathbb{E}(D_{ij} w_j^h(\mathbf{X}_i; \boldsymbol{\gamma})) = \mathbb{E}(h(\mathbf{X}_i))$ and

$$\begin{aligned}
\mathbf{H}_j &= \mathbb{E}\left\{ D_{ij}(\theta^k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t)) \frac{\partial}{\partial \boldsymbol{\gamma}} w_j^h(\mathbf{X}_i; \boldsymbol{\gamma}) \right\} \\
&= \mathbb{E}\left\{ D_{ij}(\theta^k(t) + \phi'_{k,i}(t) + \frac{1}{N-1} \sum_{l \neq i} \phi''_{k,(l,i)}(t) - m_j^{k,h}(t)) \frac{\partial}{\partial \boldsymbol{\gamma}} w_j^h(\mathbf{X}_i; \boldsymbol{\gamma}) \right\} \\
&= \mathbb{E}\left\{ D_{ij}(\widehat{\theta}_i^k(t) - m_j^{k,h}(t)) \frac{\partial}{\partial \boldsymbol{\gamma}} w_j^h(\mathbf{X}_i; \boldsymbol{\gamma}) \right\} + o_p(1).
\end{aligned}$$

The first line applies the centering property (equation 3.24 in Overgaard et al. (2017)) of the second order derivative $\mathbb{E}\{\phi''_{k,(l,i)}(t) | \mathcal{O}_i\} = 0$. The second line of the transformation for $\mathbf{H}_j$ follows from Von-Mises expansion of the pseudo-observations (equation (6) in the main text). Under the standard regularity conditions in Lehmann and Casella (2006), we have,

$$\sqrt{N}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} \mathbf{S}_{\boldsymbol{\gamma},i} + o_p(1).$$

Then we have

$$\sqrt{N}\{\widehat{m}_j^{k,h}(t) - m_j^{k,h}(t)\} = \bar{\omega}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ D_{ij}(\widehat{\theta}_i^k(t) - m_j^{k,h}(t)) w_j^h(\mathbf{X}_i; \boldsymbol{\gamma}) + \mathbf{H}_j^T \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \mathbf{S}_{\boldsymbol{\gamma},i} \right\} + o_p(1).$$

Applying the von Mises expansion of the pseudo-observations as in Jacobsen and Martinussen (2016) and Overgaard et al. (2017), we have,

$$\begin{aligned}
&\sqrt{N}\{\widehat{m}_j^{k,h}(t) - m_j^{k,h}(t)\} \\
=&\bar{\omega}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{ D_{ij} \left\{ \theta_k(t) + \phi'_{k,i}(t) + \frac{1}{N-1} \sum_{l \neq i} \phi''_{k,(l,i)}(t) - m_j^{k,h}(t) \right\} w_j^h(\mathbf{X}_i; \boldsymbol{\gamma}) \\
&+ \mathbf{H}_j^T \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \mathbf{S}_{\boldsymbol{\gamma},i} \} + o_p(1).
\end{aligned}$$

Similar expansions also apply to $\widehat{m}_{j'}^{k,h}(t)$, and thus we have,

$$\sqrt{N}\{\widehat{\tau}_{j,j'}^{k,h}(t) - \tau_{j,j'}^{k,h}(t)\} = \bar{\omega}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\psi_{ij} - \psi_{ij'}) + o_p(1),$$

$$\psi_{ij} = D_{ij}\left\{\theta_k(t) + \phi'_{k,i}(t) + \frac{1}{N-1}\sum_{l\neq i}\phi''_{k,(l,i)}(t) - m_j^{k,h}(t)\right\} w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma},i}.$$

Recall that the ith estimated pseudo-observation depends on the observed outcomes for the rest of sample. Due to the correlation between the estimated pseudo-observations, the usual Central Limit Theorem does not directly apply. Instead we reorganize the above expression into a sum of U-statistics of order 2 as follows,

$$\sum_{i=1}^{N}(\psi_{ij} - \psi_{ij'}) = \frac{N}{\binom{N}{2}}\sum_{i=1}^{N}\sum_{l<i}\frac{1}{2}g_{il},$$

where

$$\begin{aligned}
g_{il} =& D_{ij}\Big\{\theta_k(t) + \phi'_i(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
& - D_{ij'}\{\theta_k(t) + \phi'_i(t) - m_{j'}^{k,h}(t)\}w_{j'}^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_{j'}^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
& + D_{lj}\{\theta_k(t) + \phi'_l(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_l;\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},l} \\
& - D_{lj'}\{\theta_k(t) + \phi'_l(t) - m_{j'}^{k,h}(t)\}w_{j'}^h(\mathbf{X}_l;\boldsymbol{\gamma}) + \mathbf{H}_{j'}^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},l} \\
& + \phi''_{k,(l,i)}(t)\{D_{ij}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) - D_{ij'}w_{j'}^h(\mathbf{X}_i;\boldsymbol{\gamma}) + D_{lj}w_j^h(\mathbf{X}_l;\boldsymbol{\gamma}) - D_{lj'}w_{j'}^h(\mathbf{X}_l;\boldsymbol{\gamma})\Big\}.
\end{aligned}$$

Applying Theorem 12.3 in Van der Vaart (1998), we can show that the asymptotic variance of $\widehat{\tau}_{j,j'}^{k,h}(t)$ is,

$$\sqrt{N}\{\widehat{\tau}_{j,j'}^{k,h}(t) - \tau_{j,j'}^{k,h}(t)\} \xrightarrow{d} \mathcal{N}(0,\sigma^2), \quad \sigma^2 = \bar{\omega}^{-2}\,\mathbb{E}(g_{il}g_{im}),$$

where $\mathbb{E}(g_{il}g_{im}) = \mathbb{V}\{\Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t)\} = \mathbb{E}\{\Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t)\}^2$, and the scaled influence function for treatment j is

$$\begin{aligned}\Psi_j(\mathcal{O}_i;t) =& D_{ij}\{\theta_k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma})\\ &+ \frac{1}{N-1}\sum_{l\neq i}\phi''_{k,(l,i)}(t)D_{lj}w_j^h(\mathbf{X}_l,\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}.\end{aligned}$$

Hence, we have proved that the asymptotic variance of estimator (5) is $\mathbb{E}\{\Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t)\}^2/\{\mathbb{E}(h(\mathbf{X}_i)\}^2$. □

Explicit formulas for the functional derivatives We provide the explicit expression for the functional derivative $\phi'_{k,i}(t)$ and $\phi''_{k,(i,l)}(t)$ when the pseudo-observations are computed based on Kaplan-Meier estimator. We define three step functions in $\mathcal{E}:\mathcal{R}\rightarrow[0,1]$, that is, for each unit i: $Y_i(s) = \mathbf{1}\{\widetilde{T}_i \geq s\}, N_{i,0}(s) = \mathbf{1}\{\widetilde{T}_i \leq s, \Delta_i = 0\}, N_{i,1}(s) = \mathbf{1}\{\widetilde{T}_i \leq s, \Delta_i = 1\}$. Let $\widetilde{F}_N = N^{-1}\sum_{i=1}^N(Y_i, N_{i,0}, N_{i,1})$ be a vector of three step functions and its limit $\widetilde{F} = (H, H_0, H_1) \in \mathcal{E}^3$, where $H(s) = \Pr(\widetilde{T}_i \geq s), H_0(s) = \Pr(\widetilde{T}_i \leq s, \Delta_i = 0), H_1(s) = \Pr(\widetilde{T}_i \leq s, \Delta_i = 1)$ are the population analog of $(Y_i(s), N_{i,0}(s), N_{i,1}(s))$. Notice that for a given element in $\mathbf{D}$, the space of distribution, there is a unique image in $\mathcal{E}^3$. For example, $\delta_{\mathcal{O}_i}$ is mapped to $(Y_i, N_{i,0}, N_{i,1})$, F is mapped to $\widetilde{F}$, and F_N is mapped to $\widetilde{F}_N$.

We then introduce the Nelson-Aalen functional $\rho:\mathbf{D}\rightarrow\mathcal{R}$ at a fixed time point t as,

$$\rho(d;t) = \int_0^t \frac{\mathbf{1}\{h_* > 0\}}{h_*(s)} dh_1(s), \quad \widetilde{h} = (h_*, h_0, h_1) \in \mathcal{E}^3 \text{ is the unique image of } d \in \mathbf{D}$$

and the version using F and F_N as input,

$$\rho(F;t) = \int_0^t \frac{\mathbf{1}\{H(s) > 0\}}{H(s)} dH_1(s) = \Lambda_1(t), \rho(F_N;t) = \int_0^t \frac{\mathbf{1}\{Y(s) > 0\}}{Y(s)} dN_1(s) = \widehat{\Lambda}_1(t),$$

where $Y(s) = \sum_i Y_i(s), N_1(s) = \sum_i N_{1,i}(s)$. Also $\rho(F_N)$ actually corresponds to the Nelson-Aalen estimator of the cumulative hazard $\Lambda_1(t)$. Its first and second order derivative evaluated at F along the direction of sample i, l is given by James et al. (1997),

$$\rho'_i(t) = \int_0^t \frac{1}{H(s)} dM_{i,1}(s),$$
$$\rho''_{i,l}(t) = \int_0^t \frac{H(s) - Y_l(s)}{H(s)^2} dM_{i,1}(s) + \int_0^t \frac{H(s) - Y_i(s)}{H(s)^2} dM_{l,1}(s),$$

where $M_{i,1}(s) = N_{i,1}(s) - \int_0^s Y_i(u) d\Lambda_1(u)$ is a locally square integrable martingale for the counting process $N_{i,1}(s)$. The Kaplan-Meier estimator can then be represented as $\widehat{S}(t) = \phi_1(F_N;t)$, where $\phi_1(d;t)$ is defined as,

$$\phi_1(d;t) = \prod_0^t (1 - \rho(d;ds)), \quad d \in \mathbf{D}$$

where $\prod_0^{(\cdot)}$ is the product integral operator. Next, we fix the evaluation time point for the Kaplan-Meier functional and calculate its derivative along the direction of sample i at F,

$$\phi'_{1,i}(t) = -S(t)\rho'_i(t)$$

Similarly, we can take the second order derivative along the direction of sample (i, l) at F,

$$\phi''_{1,(i,l)}(t) = -S(t)\left\{\rho''_{(i,l)}(t) - \rho'_i(t)\rho'_l(t) + \mathbf{1}\{i = l\}\int_0^t \frac{1}{H^2(s)}dN_{i,1}(s)\right\}.$$

Now we have the expression for $\phi'_{1,i}(t), \phi''_{1,(i,l)}(t)$. Notice that the functional for the restricted mean survival time is the integral of the Kaplan-Meier functional,

$$\phi_2(d;t) = \int_0^t \phi_1(d;u)du, \quad d \in \mathbf{D}.$$

Then the functional derivative are given by,

$$\phi'_{2,i}(t) = \int_0^s \phi'_{1,i}(s)ds, \qquad \phi''_{2,(i,l)}(t) = \int_0^s \phi''_{1,(i,l)}(s)ds.$$

Notice that the above equality holds only if $\phi_1(d;t)$ is differentiable at any order in the p-variation setting (Dudley and Norvaiša, 1999) and its composition with the integration operator is also differentiable at any order, which is indeed the case for the Kaplan-Meier functional (Overgaard et al., 2017).

Proof of Remark 1: Without censoring, each pseudo-observation becomes $\widehat{\theta}_i^k(t) = \theta^k(t) + \phi'_{k,i}(t) = v_k(T_i;t)$ and $Q_i = 0$. Plugging these into the formula of the asymptotic variance in Theorem 1, we obtain the asymptotic variance derived in Li and Li (2019), replacing Y_i with $v_k(T_i;t)$. □

Proof of Remark 2: In this part, we prove that ignoring the "correlation term" between the pseudo-observations of different units will over-estimate the variance of the weighting estimator.

Treating each pseudo-observation as an "observed response variable" and ignoring the uncertainty associated with jackknifing will induce the following asymptotic variance,

$$\begin{aligned}
\sigma^{*2} =&\bar{\omega}^{-2}\,\mathbb{E}[D_{ij}\{\widehat{\theta}_i^k(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
&- D_{ij'}\{\widehat{\theta}_i^k(t) - m_{j'}^{k,h}(t)\}w_{j'}^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_{j'}^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}]^2 \\
=&\bar{\omega}^{-2}\,\mathbb{E}[D_{ij}\{\theta_k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t)\}w_j^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_j^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
&- D_{ij'}\{\theta_k(t) + \phi'_{k,i}(t) - m_{j'}^{k,h}(t)\}w_{j'}^h(\mathbf{X}_i;\boldsymbol{\gamma}) + \mathbf{H}_{j'}^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}]^2 \\
=&\bar{\omega}^{-2}\,\mathbb{E}\{\Psi_j^*(\mathcal{O}_i;t) - \Psi_{j'}^*(\mathcal{O}_i;t)\}^2,
\end{aligned}$$

where the first equality follows from Theorem 2 in Graw et al. (2009). We wish to show that,

$$\mathbb{E}\{\Psi_j^*(\mathcal{O}_i;t) - \Psi_{j'}^*(\mathcal{O}_i;t)\}^2 \geq \mathbb{E}\{\Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t)\}^2,$$

and hence $\sigma^{*2} \geq \sigma^2$. Notice that,

$$\begin{gathered}
\eta_i \triangleq \Psi_j^*(\mathcal{O}_i;t) - \Psi_{j'}^*(\mathcal{O}_i;t), \psi_i \triangleq \Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t) \\
\psi_i = \eta_i + \frac{1}{N-1}\sum_{l\neq i}\phi''_{k,(l,i)}(t)[D_{lj}w_j^h(\mathbf{X}_l,\boldsymbol{\gamma}) - D_{lj'}w_{j'}^h(\mathbf{X}_l,\boldsymbol{\gamma})]
\end{gathered}$$

Next, we plug the exact formula of $\phi'_{k,i}(t)$ and $\phi''_{k,(i,l)}(t)$ into the above equation, and obtain

$$
\begin{aligned}
&\mathbb{E}\{\eta_i(\psi_i-\eta_i)\} \\
=&-S^2(t)\,\mathbb{E}\left[\{D_{ij}w_j^h(\mathbf{X}_i,\boldsymbol{\gamma})-D_{ij'}w_{j'}^h(\mathbf{X}_i,\boldsymbol{\gamma})\}\int_0^t\frac{1}{H(s)}dM(s)\right.\\
&\left.\left\{\int_0^t\int_0^s\frac{1}{H(u)}dM(u)d\mu(s)-\int_0^t\left(1-\frac{Y(s)}{H(s)}\right)d\mu(s)\right\}\right],
\end{aligned}
$$

where $M(s)=N_1(s)-\int_0^s Y(t)d\Lambda_1(t)$, $d\mu(s)=\mathbb{E}\left\{\frac{D_{ij}w_j^h(\mathbf{X}_i,\boldsymbol{\gamma})-D_{ij'}w_{j'}^h(\mathbf{X}_i,\boldsymbol{\gamma})}{H(s)}dM_{i,1}(s)\right\}$. With the results established in the proof of Theorem 2 in Jacobsen and Martinussen (2016) (equation (22) in their Appendix, treating $D_{ij}w_j^h(\mathbf{X}_i,\boldsymbol{\gamma})-D_{ij'}w_{j'}^h(\mathbf{X}_i,\boldsymbol{\gamma})$ as the "$A(Z)$" in the equation), we can further simplify the above expression to,

$$
\mathbb{E}\{\eta_i(\psi_i-\eta_i)\}=-S^2(t)\int_0^t\int_0^t\int_0^{s\wedge u}\frac{\lambda_c(v)}{H(v)}dvd\mu(u)d\mu(s),
$$

where $\lambda_c(t)$ is the hazard function for the censoring time. Also, similar to equation (16) in the Appendix of Jacobsen and Martinussen (2016), we can show that,

$$
\begin{aligned}
\mathbb{E}\{\psi_i-\eta_i\}^2&=S^2(t)\int_0^t\int_0^t\int_0^{s\wedge u}\frac{\lambda_c(v)}{H(v)}dvd\mu(u)d\mu(s)\\
&=-\,\mathbb{E}\{\eta_i(\psi_i-\eta_i)\}
\end{aligned}
$$

Combining the above results, we obtain

$$
\begin{aligned}
\mathbb{E}\{\Psi_j(\mathcal{O}_i;t) - \Psi_{j'}(\mathcal{O}_i;t)\}^2 &= \mathbb{E}\{\psi_i\}^2 = \mathbb{E}\{\eta_i + \psi_i - \eta_i\}^2 \\
&= \mathbb{E}\{\eta_i\}^2 + \mathbb{E}\{\psi_i - \eta_i\}^2 + 2\,\mathbb{E}\{\eta_i(\psi_i - \eta_i)\} \\
&= \mathbb{E}\{\Psi_j^*(\mathcal{O}_i;t) - \Psi_{j'}^*(\mathcal{O}_i;t)\}^2 - \mathbb{E}\{\psi_i - \eta_i\}^2 \\
&\leq \mathbb{E}\{\Psi_j^*(\mathcal{O}_i;t) - \Psi_{j'}^*(\mathcal{O}_i;t)\}^2
\end{aligned}
$$

which completes the proof of this remark. □

Proof of Remark 3: Treating the generalized propensity score as known will remove the term $\mathbf{H}_{j'}^T \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \mathbf{S}_{\boldsymbol{\gamma},i}$ in $\Psi_j(\mathcal{O}_i;t)$. When $h(\mathbf{X}) = 1$ or equivalently under the IPW scheme, the asymptotic variance based on the known or fixed GPS in estimator (5), $\widetilde{\sigma}^2$, becomes:

$$
\begin{aligned}
\widetilde{\sigma}^2 = \mathbb{E}\Big[& D_{ij}\{\theta_k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t)\} e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1} - D_{ij'}\{\theta_k(t) + \phi'_{k,i}(t) - m_{j'}^{k,h}(t)\} e_{j'}(\mathbf{X}_i;\boldsymbol{\gamma})^{-1} \\
& + \frac{1}{N-1} \sum_{l \neq i} \phi''_{k,(l,i)}(t) \left\{ D_{lj} e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1} - D_{lj'} e_{j'}(\mathbf{X}_i;\boldsymbol{\gamma})^{-1} \right\} \Big]^2 .
\end{aligned}
$$

On the other hand, the asymptotic variance taking account of uncertainty in estimating the generalized propensity scores can be expressed as,

$$
\begin{aligned}
\sigma^2 =&\widetilde{\sigma}^2 + 2(\mathbf{H}_j - \mathbf{H}_{j'})^T \mathbf{I}_{\boldsymbol{\gamma\gamma}}^{-1} \mathbb{E}\Bigg[\Bigg\{D_{ij}(\theta_k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t))e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1} \\
&+ \frac{1}{N-1}\sum_{l\neq i}\phi''_{k,(l,i)}(t)D_{lj}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\Bigg\}\mathbf{S}_{\boldsymbol{\gamma},i} - \Bigg\{D_{ij'}(\theta_k(t) + \phi'_{k,i}(t) - m_{j'}^{k,h}(t))e_{j'}(\mathbf{X}_i;\boldsymbol{\gamma})^{-1} \\
&+\frac{1}{N-1}\sum_{l\neq i}\phi''_{k,(l,i)}(t)D_{lj'}e_{j'}(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\Bigg\}\mathbf{S}_{\boldsymbol{\gamma},i}\Bigg] + (\mathbf{H}_j - \mathbf{H}_{j'})^T\mathbf{I}_{\boldsymbol{\gamma\gamma}}(\mathbf{H}_j - \mathbf{H}_{j'}) \\
=&\widetilde{\sigma}^2 + 2(\mathbf{H}_j - \mathbf{H}_{j'})^T \mathbf{I}_{\boldsymbol{\gamma\gamma}}^{-1} \mathbb{E}\Bigg[D_{ij}(\theta_k(t) + \phi'_{k,i}(t) - m_j^{k,h}(t))e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
&+ \frac{1}{N-1}\sum_{l\neq i}\{\phi''_{k,(l,i)}(t)D_{lj}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\}\mathbf{S}_{\boldsymbol{\gamma},i} - D_{ij'}(\theta_k(t) + \phi'_{k,i}(t) - m_{j'}^{k,h}(t))e_{j'}(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i} \\
&-\frac{1}{N-1}\sum_{l\neq i}\{\phi''_{k,(l,i)}(t)D_{lj'}e_{j'}(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\}\mathbf{S}_{\boldsymbol{\gamma},i}\Bigg] + (\mathbf{H}_j - \mathbf{H}_{j'})^T\mathbf{I}_{\boldsymbol{\gamma\gamma}}(\mathbf{H}_j - \mathbf{H}_{j'}) \\
=&\widetilde{\sigma}^2 + 2I + II,
\end{aligned}
$$

where we applied the facts that $E(\mathbf{S}_{\boldsymbol{\gamma},i}\mathbf{S}_{\boldsymbol{\gamma},i}^T) = \mathbf{I}_{\boldsymbol{\gamma\gamma}}$. The score function $\mathbf{S}_{\boldsymbol{\gamma},i}$ can be expressed as,

$$
D_{ij}\mathbf{S}_{\boldsymbol{\gamma},i} = D_{ij}\sum_{k=1}^{J}\left\{\frac{\partial}{\partial\boldsymbol{\gamma}}e_k(\mathbf{X}_i;\boldsymbol{\gamma})\right\}D_{ik}/e_k(\mathbf{X}_i;\boldsymbol{\gamma}) = \left\{\frac{\partial}{\partial\boldsymbol{\gamma}}e_j(\mathbf{X}_i;\boldsymbol{\gamma})\right\}D_{ij}/e_j(\mathbf{X}_i;\boldsymbol{\gamma}).
$$

On the other hand, when $h(\mathbf{X}) = 1$, we have,

$$
\frac{\partial}{\partial\boldsymbol{\gamma}}w_j^h(\mathbf{X}_i,\boldsymbol{\gamma}) = \frac{\partial}{\partial\boldsymbol{\gamma}}\left\{e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\right\} = -e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-2}\frac{\partial}{\partial\boldsymbol{\gamma}}e_j(\mathbf{X}_i;\boldsymbol{\gamma}) = -D_{ij}e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}.
$$

Notice that,

$$\begin{aligned}
\mathbb{E}\{\phi''_{k,(l,i)}(t)D_{lj}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}\} &= \mathbb{E}\{\mathbf{S}_{\boldsymbol{\gamma},i}\,\mathbb{E}\{\phi''_{k,(l,i)}(t)D_{lj}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}|\mathcal{O}_i,\mathbf{X}_l\}\} \\
&= \mathbb{E}\{\mathbf{S}_{\boldsymbol{\gamma},i}\,\mathbb{E}\{D_{lj}|\mathbf{X}_l\}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\,\mathbb{E}\{\phi''_{k,(l,i)}(t)|\mathcal{O}_i,\mathbf{X}_l\}\} \\
&= \mathbb{E}\{\mathbf{S}_{\boldsymbol{\gamma},i}\,\mathbb{E}\{\phi''_{k,(l,i)}(t)|\mathcal{O}_i,\mathbf{X}_l\}\} \\
&= \mathbb{E}\{\phi''_{k,(l,i)}(t)\mathbf{S}_{\boldsymbol{\gamma},i}\} = \mathbb{E}\{\mathbf{S}_{\boldsymbol{\gamma},i}\,\mathbb{E}\{\phi''_{k,(l,i)}(t)|\mathcal{O}_i\}\} = 0,
\end{aligned}$$

where the second line follows from the weak unconfoundness assumption (A1), namely, $\phi''_{k,(l,i)}(t)$ is a function of $\widetilde{T}_l(j), \Delta_l(j)$ which independent of D_{lj} given $\mathbf{X}_l$. With the above equality, we can show,

$$\begin{aligned}
&\mathbb{E}\left\{D_{ij}(\theta^k(t)+\phi'_{k,i}(t)-m_j^{k,h}(t))e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}+\frac{1}{N-1}\sum_{l\neq i}\{\phi''_{k,(l,i)}(t)D_{lj}e_j(\mathbf{X}_l;\boldsymbol{\gamma})^{-1}\}\mathbf{S}_{\boldsymbol{\gamma},i}\right\} \\
=&\,\mathbb{E}\{D_{ij}(\theta^k(t)+\phi'_{k,i}(t)-m_j^{k,h}(t))e_j(\mathbf{X}_i;\boldsymbol{\gamma})^{-1}\mathbf{S}_{\boldsymbol{\gamma},i}\} \\
=&-\mathbb{E}\left\{(\theta^k(t)+\phi'_{k,i}(t)-m_j^{k,h}(t))\frac{\partial}{\partial\boldsymbol{\gamma}}w_j^h(\mathbf{X}_i,\boldsymbol{\gamma})\right\} = -\mathbf{H}_j
\end{aligned}$$

Hence, we have $2I = -2II$, and $\sigma^2 - \tilde{\sigma}^2 = -(\mathbf{H}_j - \mathbf{H}_{j'})^T\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\mathbf{H}_j - \mathbf{H}_{j'}) \leq 0$ since $\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ is semi-positive definite. As such, we have proved Remark 3. $\square$

Proof of Remark 4: First we will prove the consistency of estimator (5) in the main text under covariate dependent (conditional independent) censoring specified in Assumption (A4). We define the functional G by

$$G(f;s|X,Z) = \prod_0^{s^-}(1-\Lambda(f;du|X,Z)), \quad f\in\mathbf{D}$$

where Λ is the Nelson-Aalen functional for the cumulative hazard of censoring time C_i. And we define functional v as the $v_k(f;t) = v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}$, for $f \in \mathbf{D}$. Hence, we can view (5) using (8) in the main text as a functional from $\mathbf{D}$ to $\mathcal{R}$, which is given by,

$$\Theta_k(f) = \int \frac{v_k(f;t)}{G(f;\widetilde{T} \wedge t|\mathbf{X}, Z)} df.$$

According to Overgaard et al. (2019), functional Θ_k is measurable mapping and 2-times continuously differentiable with a Lipschitz continuous second-order derivative in a neighborhood of F. Assuming the censoring survival function G is consistently estimated, say, by a Cox proportional hazards model, we can establish a similar property as in the completely random censoring case that (Theorem 2 in Overgaard et al. (2019)),

$$\begin{aligned}
\mathbb{E}\{\widehat{\theta}_i^k(t)|\mathbf{X}_i, Z_i\} &= \mathbb{E}\left\{\frac{v_k(\widetilde{T}_i;t)\mathbf{1}\{C_i \geq \widetilde{T}_i \wedge t\}}{G(\widetilde{T}_i \wedge t|\mathbf{X}_i, Z_i)}|\mathbf{X}_i, Z_i\right\} + o_p(1) \\
&= \frac{\mathbb{E}(v_k(\widetilde{T}_i;t)|\mathbf{X}_i, Z_i)G(\widetilde{T}_i \wedge t|\mathbf{X}_i, Z_i)}{G(\widetilde{T}_i \wedge t|\mathbf{X}_i, Z_i)} + o_p(1) \\
&= \mathbb{E}(v_k(\widetilde{T}_i;t)|\mathbf{X}_i, Z_i) + o_p(1).
\end{aligned}$$

Therefore, we can show the consistency of estimator (5) based on (8) in the main text follows the exact same procedure as in the proof for Theorem 1 (i).

Moreover, the asymptotic normality of estimator (5) using (8) follows the same proof in Theorem 2 (ii) where we replace the derivative $\phi'_{k,i}(t)$ and $\phi''_{k,(i,l)}(t)$ with the one corresponding to the functional Θ_k. We omit the detailed steps for brevity, but present the specific forms of the functional derivatives. The first-order derivative of Θ_k at F along the direction of sample i is given by,

$$\Theta'_{k,i} = \int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F;\widetilde{T} \wedge t|\mathbf{X}, Z)} d\delta_{\mathcal{O}_i} - \int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F|\widetilde{T} \wedge t|\mathbf{X}, Z)^2} G'_F(\delta_{\mathcal{O}_i}; \widetilde{T} \wedge t|\mathbf{X}, Z) dF.$$

Note that $G'_F(g; s|\mathbf{X}, Z)$ is the derivative of functional G at F along direction g, which is,

$$G'_F(g; s|\mathbf{X}, Z) = -G(F; s|\mathbf{X}, Z) \int_0^{s^-} \frac{1}{1 - d\Lambda(F; u|\mathbf{X}, Z)} \Lambda'_F(g; du|\mathbf{X}, Z),$$

where $\Lambda'_F(g; du|\mathbf{X}, Z)$ is the functional derivative of the cumulative hazard evaluated at F along direction g. For example, if the censoring survival function is estimated by Cox proportional hazards model, the above functional derivative can be obtained by viewing it as a solution to a set of estimating equations for the Cox model and employing the implicit function theorem. Detailed derivation is provided in the proof of Proposition 2 in Overgaard et al. (2019). The second order derivative of Θ_k at F along the direction of sample i, l is given by,

$$
\begin{aligned}
\Theta''_{k,(i,l)} =& -\int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F;\widetilde{T} \wedge t|\mathbf{X}, Z)^2} G'_F(\delta_{\mathcal{O}_l};\widetilde{T} \wedge t|\mathbf{X}, Z) d\delta_{\mathcal{O}_i} \\
& -\int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F;\widetilde{T} \wedge t|\mathbf{X}, Z)^2} G'_F(\delta_{\mathcal{O}_i};\widetilde{T} \wedge t|\mathbf{X}, Z) d\delta_{\mathcal{O}_l} \\
& + 2\int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F;\widetilde{T} \wedge t|\mathbf{X}, Z)^3} G'_F(\delta_{\mathcal{O}_i};\widetilde{T} \wedge t|\mathbf{X}, Z) G'_F(\delta_{\mathcal{O}_l};\widetilde{T} \wedge t|\mathbf{X}, Z) dF \\
& -\int \frac{v_k(\widetilde{T};t)\mathbf{1}\{C \geq \widetilde{T} \wedge t\}}{G(F;\widetilde{T} \wedge t|\mathbf{X}, Z)^2} G''_F;(\delta_{\mathcal{O}_i}, \delta_{\mathcal{O}_l};\widetilde{T} \wedge t|\mathbf{X}, Z) dF.
\end{aligned}
$$

The second-order derivative of G at F along the direction of (g, h) is,

$$
\begin{aligned}
G''_F(g,h;s|\mathbf{X},Z) =& G(F;s|\mathbf{X},Z) \int_0^{s^-} \frac{1}{1-d\Lambda(F;u|\mathbf{X},Z)} \Lambda'_F(g;du|\mathbf{X},Z) \\
& \times \int_0^{s^-} \frac{1}{1-d\Lambda(F;u|\mathbf{X},Z)} \Lambda'_F(h;du|\mathbf{X},Z) \\
& - G(F;s|\mathbf{X},Z) \int_0^{s} \frac{d\Lambda'(g;|\mathbf{X},Z) d\Lambda'(h;|\mathbf{X},Z)}{(1-d\Lambda(F;u|\mathbf{X},Z))^2} \\
& - G(F;s|\mathbf{X},Z) \int_0^{s^-} \frac{\Lambda''_F(g,h;du|\mathbf{X},Z)}{1-d\Lambda(F;u|\mathbf{X},Z)}.
\end{aligned}
$$

The second-order derivative of the cumulative hazard for using the proportional hazard model, $\Lambda''_F(g,h;du|\mathbf{X},Z)$ is given in the Section 3 of the Appendix of Overgaard et al. (2019). □

Proof of Theorem 2 We proceed under the regularity conditions specified in the proof of Theorem 1. Let $\mathbf{c} = (c_1, c_2, \cdots, c_J) \in \{-1, 0, 1\}^J$ and define,

$$
\hat{\tau}(\mathbf{c};t)^{k,h} = \sum_{j=1}^{J} c_j \left\{ \frac{\sum_{i=1}^{N} D_{ij} \widehat{\theta}_i^k(t) w_j^h(\mathbf{X}_i)}{\sum_{i=1}^{N} D_{ij} w_j^h(\mathbf{X}_i)} \right\}.
$$

It is easy to show that when $c_j = 1, c_{j'} = -1, c_{j''} = 0, j'' \neq j, j'$, we have $\widehat{\tau}(\mathbf{c};t)^{k,h} = \widehat{\tau}^{k,h}_{j,j'}(t)$. Conditional on the collection of design points $\underline{\mathbf{Z}} = \{Z_1, \dots, Z_N\}$ and $\underline{\mathbf{X}} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$, the asymptotic variance of $\widehat{\tau}(\mathbf{c};t)^{k,h}$ is,

$$
\begin{aligned}
N\,\mathbb{V}(\widehat{\tau}(\mathbf{c};t)^{k,h}|\underline{\mathbf{X}},\underline{\mathbf{Z}}) =& N\sum_{j=1}^{J} c_j^2 \left[\frac{\sum_{i=1}^N D_{ij}\,\mathbb{V}\{\widehat{\theta}_i^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2}{\{\sum_{i=1}^N D_{ij} w_j^h(\mathbf{X}_i)\}^2}\right.\\
&\left.+\frac{\sum_{i\neq l} D_{ij}D_{lj}\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_j^h(\mathbf{X}_l)}{\{\sum_{i=1}^N D_{ij} w_j^h(\mathbf{X}_i)\}^2}\right]\\
&+N\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)}{\{\sum_{i=1}^N D_{ij} w_j^h(\mathbf{X}_i)\}\{\sum_{i=1}^N D_{ij'} w_{j'}^h(\mathbf{X}_i)\}}\\
=& A + B + C
\end{aligned}
$$

First, we consider the asymptotic behaviour of term C. Notice that with von Mises expansion (equation (6) in the main text),

$$
\begin{aligned}
\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\} =& \mathrm{Cov}\left\{\theta^k(t)+\phi'_{k,i}(t)+\frac{1}{N-1}\sum_{m\neq i}\phi''_{k,(m,i)},\right.\\
&\left.\theta^k(t)+\phi'_{k,l}(t)+\frac{1}{N-1}\sum_{n\neq l}\phi''_{k,(n,l)}|\underline{\mathbf{X}},\underline{\mathbf{Z}}\right\} + o_p(N^{-1/2})\\
=& \mathrm{Cov}\{\phi'_{k,i}(t),\phi'_{k,l}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\} + \frac{1}{N-1}\sum_{n\neq l}\mathrm{Cov}\{\phi'_{k,i}(t),\phi''_{k,(n,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\\
&+\frac{1}{N-1}\sum_{m\neq i}\mathrm{Cov}\{\phi'_{k,l}(t),\phi''_{k,(m,i)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\\
&+\frac{1}{(N-1)^2}\mathrm{Cov}\left\{\sum_{m\neq i}\phi''_{k,(m,i)}(t),\sum_{n\neq l}\phi''_{k,(n,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\right\} + o_p(N^{-1/2}).
\end{aligned}
$$

We view $\phi'_{k,i}(t)$ as a function of $(\widetilde{T}_i(j),\Delta_i(j))$ and $\phi''_{k,(i,l)}(t)$ as function of $(\widetilde{T}_i(j),\Delta_i(j),$

$\widetilde{T}_l(j'), \Delta_l(j'))$ (since we have $D_{ij}D_{lj'}$ as the multiplier). Due to the independence between $(\widetilde{T}_i(j), \Delta_i(j))$ and $(\widetilde{T}_l(j'), \Delta_l(j'))$ given $\underline{\mathbf{X}}, \underline{\mathbf{Z}}$, we can reduce the following covariance into zero,

$$\begin{aligned}
\mathrm{Cov}\{\phi'_{k,i}(t), \phi'_{k,l}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} &= 0, \text{when } i \neq l, \\
\mathrm{Cov}\{\phi'_{k,i}(t), \phi''_{k,(n,l)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} &= 0, \text{when } i \neq n, \\
\mathrm{Cov}\{\phi'_{k,l}(t), \phi''_{k,(m,i)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} &= 0, \text{when } l \neq m, \\
\mathrm{Cov}\{\phi''_{k,(m,i)}(t), \phi''_{k,(n,l)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} &= 0, \text{when } m \neq n.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathrm{Cov}\{\widehat{\theta}_i^k(t), \widehat{\theta}_l^k(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} =& \frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,i}(t), \phi''_{k,(i,l)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} + \frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,l}(t), \phi''_{k,(l,i)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} \\
&+ \frac{1}{(N-1)^2}\sum_{m\neq i, m\neq l}\mathrm{Cov}\{\phi''_{k,(m,i)}(t), \phi''_{k,(m,l)}(t)|\underline{\mathbf{X}}, \underline{\mathbf{Z}}\} + o_p(N^{-1/2}).
\end{aligned}$$

Note that we have,

$$\frac{1}{N}\sum_{i=1}^{N} D_{ij}w_j^h(\mathbf{X}_i) \xrightarrow{p} \int_{\mathcal{X}} \mathbb{E}(D_{ij}|\mathbf{X})/e_j(\mathbf{X})h(\mathbf{X})f(\mathbf{X})\mu(d\mathbf{X}) \triangleq C_h,$$

Then term C is asymptotically equals to,

$$
\begin{aligned}
&N\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)\}\{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)\}}\\
&=\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}\{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)/N\}}\\
&=\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,i}(t),\phi''_{k,(i,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}\{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)/N\}}\\
&+\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,l}(t),\phi''_{k,(l,i)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}\{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)/N\}}\\
&+\sum_{j\neq j'} c_j c_{j'} \frac{\sum_{i\neq l} D_{ij}D_{lj'}\frac{1}{(N-1)^2}\sum_{m\neq i,m\neq l}\mathrm{Cov}\{\phi''_{k,(m,i)}(t),\phi''_{k,(m,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_{j'}^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}\{\sum_{i=1}^N D_{ij'}w_{j'}^h(\mathbf{X}_i)/N\}}+o_p(1)\\
&=o_p(1)
\end{aligned}
$$

Next, we consider term B. Similarly, we have

$$
\begin{aligned}
&\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}=\mathrm{Cov}\{\phi'_{k,i}(t),\phi'_{k,l}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+\frac{1}{N-1}\sum_{n\neq l}\mathrm{Cov}\{\phi'_{k,i}(t),\phi''_{k,(n,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+\\
&\frac{1}{N-1}\sum_{m\neq i}\mathrm{Cov}\{\phi''_{k,(m,i)}(t),\phi'_{k,l}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+\frac{1}{(N-1)^2}\sum_{m\neq i,n\neq l}\mathrm{Cov}\{\phi'_{k,(m,i)}(t),\phi''_{k,(n,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+o_p(N^{-1/2})\\
&=\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,i}(t),\phi''_{k,(i,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,l}(t),\phi''_{k,(i,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\\
&+\frac{1}{(N-1)^2}\sum_{m\neq i,m\neq l}\mathrm{Cov}\{\phi'_{k,(m,i)}(t),\phi''_{k,(m,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+o_p(N^{-1/2}).
\end{aligned}
$$

Then the term B asymptotically equals,

$$N\frac{\sum_{i\neq l} D_{ij}D_{lj}\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_l^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_j^h(\mathbf{X}_l)}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)\}^2}$$
$$=\frac{\sum_{i\neq l} D_{ij}D_{lj}\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,i}(t),\phi''_{k,(i,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_j^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}$$
$$+\frac{\sum_{i\neq l} D_{ij}D_{lj}\frac{1}{N-1}\mathrm{Cov}\{\phi'_{k,l}(t),\phi''_{k,(i,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_j^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}$$
$$+\frac{\sum_{i\neq l} D_{ij}D_{lj}\frac{1}{(N-1)^2}\sum_{m\neq i,m\neq l}\mathrm{Cov}\{\phi'_{k,(m,i)}(t),\phi''_{k,(m,l)}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}w_j^h(\mathbf{X}_i)w_j^h(\mathbf{X}_l)/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}+o_p(1)=o_p(1)$$

Lastly, for term A, Note that we have,

$$\mathbb{V}\{\widehat{\theta}_i^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}=\mathrm{Cov}\{\widehat{\theta}_i^k(t),\widehat{\theta}_i^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}=\mathrm{Cov}\left\{\theta^k(t)+\phi'_{k,i}(t)+\frac{1}{N-1}\sum_{m\neq i}\phi''_{k,(m,i)},\right.$$
$$\left.\theta^k(t)+\phi'_{k,i}(t)+\frac{1}{N-1}\sum_{m\neq i}\phi''_{k,(m,i)}|\underline{\mathbf{X}},\underline{\mathbf{Z}}\right\}+o_p(N^{-1/2})$$
$$=\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+\frac{1}{(N-1)^2}\sum_{m\neq i}\mathrm{Cov}\{\phi''_{k,(m,i)}(t)^2|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}+o_p(N^{-1/2}).$$

Further observe that

$$N\frac{\sum_{i=1}^N D_{ij}\,\mathbb{V}\{\widehat{\theta}_i^k(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)\}^2}=\frac{\sum_{i=1}^N D_{ij}\,\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}$$
$$+\frac{\sum_{i=1}^N D_{ij}\sum_{m\neq i}\mathrm{Cov}\{\phi''_{k,(m,i)}(t)^2|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2/(N(N-1)^2)}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}+o_p(1)$$
$$=\frac{\sum_{i=1}^N D_{ij}\,\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2/N}{\{\sum_{i=1}^N D_{ij}w_j^h(\mathbf{X}_i)/N\}^2}+o_p(1).$$

Also, we have

$$\sum_{i=1}^{N} D_{ij}\,\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}\{w_j^h(\mathbf{X}_i)\}^2/N \underset{p}{\to} \int_{\mathcal{X}} \{\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\}/e_j(\mathbf{X})\}h(\mathbf{X})^2 f(\mathbf{X})\mu(d\mathbf{X}).$$

Therefore, assuming the generalized homoscedasticity condition such that $\mathbb{V}\{\phi'_{k,i}(t)|\underline{\mathbf{X}},\underline{\mathbf{Z}}\} = \mathbb{V}\{\phi'_{k,i}(t)|\mathbf{X}_i, Z_i\} = v$, the conditional asymptotic variance of $\widehat{\tau}(\mathbf{c};t)^{k,h}$ is,

$$\begin{aligned}
\lim_{N\to\infty} N\,\mathbb{V}\{\widehat{\tau}(\mathbf{c};t)^{k,h}|\underline{\mathbf{X}},\underline{\mathbf{Z}}\} &= \int_{\mathcal{X}} \sum_{j=1}^{J} c_j^2\{v/e_j(\mathbf{X})\}h(\mathbf{X})^2 f(\mathbf{X})\mu(d\mathbf{X})/C_h^2 \\
&= \frac{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X})\}}{C_h^2} v \\
&= \frac{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X})\}}{\mathbb{E}_{\mathcal{X}}[h(\mathbf{X})]^2} v \\
&\geq \frac{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X})\}}{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X})\}\,\mathbb{E}_{\mathcal{X}}\{(\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X}))^{-1}\}}.
\end{aligned}$$

The inequality follows from the Cauchy-Schwarz inequality and the equality is attained when $h(\mathbf{X}) \propto \{\sum_{j=1}^{J} c_j^2/e_j(\mathbf{X})\}^{-1}$. Consequently, the sum of the asymptotic variance of all pairwise comparisons is,

$$\sum_{j<j'} \lim_{N\to\infty} N\,\mathbb{V}(\hat{\tau}_{j,j'}(t)^{k,h}|\underline{\mathbf{X}},\underline{\mathbf{Z}}) = (J-1)\sum_{j=1}^{J} \frac{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})/e_j(\mathbf{X})\}}{\mathbb{E}_{\mathcal{X}}[h(\mathbf{X})]^2} v$$

We consider the variance of $\widehat{\tau}(\bar{\mathbf{c}};t)^{k,h}$ where $\bar{\mathbf{c}} = (1,1,1,\cdots,1)$. We can show that,

$$\lim_{N\to\infty} N\widehat{\tau}(\bar{\mathbf{c}};t)^{k,h} = \sum_{j=1}^{J} \frac{\mathbb{E}_{\mathcal{X}}\{h^2(\mathbf{X})/e_j(\mathbf{X})\}}{\mathbb{E}_{\mathcal{X}}[h(\mathbf{X})]^2} v$$

Therefore, $\sum_{j<j'} \lim_{N\to\infty} N\,\mathbb{V}(\widehat{\tau}_{j,j'}(t)^{k,h}|\underline{\mathbf{X}},\underline{\mathbf{Z}})$ attains its minimum when $\lim_{N\to\infty} N\widehat{\tau}(\bar{\mathbf{c}};t)^{k,h}$ are minimized. Notice that $c_j^2 = 1$ in $\bar{\mathbf{c}}$. Hence, when $h(\mathbf{X}) \propto \{\sum_{j=1}^{J} 1/e_j(\mathbf{X})\}^{-1}$,

the sum of the conditional asymptotic variance of all pairwise comparison is minimized, which completes the proof of Theorem 2. $\square$

Details on augmented weighting estimator In this part, we provide the outline on how to derive the variance estimator of the augmented weighting estimator using the pseudo-observations. Suppose the estimated parameter of the outcome model $\widehat{\boldsymbol{\alpha}}_j$ are the MLEs that solve the score functions $\sum_{i=1}^N \mathbf{1}\{Z_i = j\}S_j(\mathbf{X}_i, \widehat{\theta}_i^k; \boldsymbol{\alpha}_j) = 0$, then we can express the augmented weighting estimator based on the solution $(\widehat{\nu}_0, \widehat{\nu}_j, \widehat{\nu}_{j'}, \widehat{\boldsymbol{\alpha}}_1^T, \widehat{\boldsymbol{\alpha}}_2^T, \cdots, \widehat{\boldsymbol{\alpha}}_J^T, \widehat{\boldsymbol{\gamma}})^T$ to the following estimation equations $\sum_{i=1}^N U_i = 0$,

$$\sum_{i=1}^N U_i(\widehat{\nu}_0, \widehat{\nu}_j, \widehat{\nu}_{j'}, \widehat{\boldsymbol{\alpha}}_1^T, \widehat{\boldsymbol{\alpha}}_2^T, \cdots, \widehat{\boldsymbol{\alpha}}_J^T, \widehat{\boldsymbol{\gamma}}) = \sum_{i=1}^N \begin{bmatrix} h(\mathbf{X}_i; \boldsymbol{\gamma})\{m_j^k(\mathbf{X}_i; \boldsymbol{\alpha}_j) - m_{j'}^k(\mathbf{X}_i; \boldsymbol{\alpha}_j) - \nu_0\} \\ \mathbf{1}\{Z_i = j\}\{\widehat{\theta}_i^k - m_j^k(\mathbf{X}_i; \boldsymbol{\alpha}_j) - \nu_j\}w_j^h(\mathbf{X}_i) \\ \mathbf{1}\{Z_i = j'\}\{\widehat{\theta}_i^k - m_{j'}^k(\mathbf{X}_i; \boldsymbol{\alpha}_{j'}) - \nu_{j'}\}w_{j'}^h(\mathbf{X}_i) \\ \mathbf{1}\{Z_i = 1\}S_1(\mathbf{X}_i, \widehat{\theta}_i^k; \boldsymbol{\alpha}_1) \\ \cdots \\ \mathbf{1}\{Z_i = J\}S_J(\mathbf{X}_i, \widehat{\theta}_i^k; \boldsymbol{\alpha}_J) \\ \mathbf{S}_{\boldsymbol{\gamma}}(\mathbf{X}_i, Z_i; \boldsymbol{\gamma}) \end{bmatrix} = 0.$$

The augmented weighting estimator is $\widehat{\nu}_0 + \widehat{\nu}_j - \widehat{\nu}_{j'}$. The corresponding variance estimator can be obtained by applying Theorem 3.4 in Overgaard et al. (2017), which offers the asymptotic variance of the estimated parameters based on the estimating equations involving the pseudo-observations.

Web Appendix B Details on simulation design

Web Figure 1 illustrates the distribution of the true generalized propensity score (GPS) in the simulations that represent (i) a three-arm randomized controlled trial (RCT), (ii) observational study with relatively good covariate overlap between groups, and (iii) observational study with poor covariate overlap between groups. In the simulated RCT, the propensity for being assigned to three arms are the same (1/3) for each unit. In the simulated observational study, the GPS for three arms differ; the distributions of the GPS to each arm exhibit a larger difference when overlap is poor.

Below, we describe the details of the alternative estimators considered in the simulation studies.

1. Cox model with g-formula (Cox): We fit the Cox proportional hazard model with the hazard rate $\lambda(t|\mathbf{X}_i, Z_i)$,

$$\lambda(t|\mathbf{X}_i, Z_i) = \lambda_0(t) \exp\left(\mathbf{X}_i\boldsymbol{\alpha}^T + \sum_{j\in\mathcal{J}} \boldsymbol{\gamma}_j \mathbf{1}\{Z_i = j\}\right).$$

   Based on the estimated hazard rate, we can calculate the conditional survival probability function $\widehat{S}(t|\mathbf{X}_i, Z_i)$ and estimate $\widehat{\tau}^{k,h}_{j,j'}(t)$ when $h(\mathbf{X}) = 1$ with the usual g-formula,

$$\hat{\tau}^{1}_{j,j'}(t) = N^{-1}\sum_{i=1}^{N}\left\{\widehat{S}(t|\mathbf{X}_i, Z_i = j) - \widehat{S}(t|\mathbf{X}_i, Z_i = j')\right\},$$
$$\hat{\tau}^{2}_{j,j'}(t) = N^{-1}\sum_{i=1}^{N}\int_0^t\left\{\widehat{S}(u|\mathbf{X}_i, Z_i = j) - \sum_{i=1}^{N}\widehat{S}(u|\mathbf{X}_i, Z_i = j')\right\}du.$$

2. Cox with IPW (IPW-Cox): We first fit a multinomial logistic regression model for the GPS and construct the IPW, i.e. we assign stablized weights $w_{ij} = \Pr(Z_i = j)/\Pr(Z_i = j|\mathbf{X}_i)$ for each unit. Next, we fit a Cox proportional hazard model on the weighted sample with a hazard rate,

$$\lambda(t|\mathbf{X}_i, Z_i) = \lambda_0(t)\exp\left(\sum_{j\in\mathcal{J}}\boldsymbol{\gamma}_j\mathbf{1}\{Z_i = j\}\right).$$

We then calculate the survival probability $\widehat{S}(t|Z_i)$ specific to each arm and estimate $\widehat{\tau}^{k,h}_{j,j'}(t)$ when $h(\mathbf{X}) = 1$ using,

$$\widehat{\tau}^{1}_{j,j'}(t) = \widehat{S}(t|Z_i = j) - \widehat{S}(t|Z_i = j'),$$
$$\widehat{\tau}^{2}_{j,j'}(t) = \int_0^t\left\{\widehat{S}(u|Z_i = j) - \widehat{S}(u|Z_i = j')\right\}du.$$

3. Trimmed IPW-PO (T-IPW): this is the propensity score weighting estimator (5) with $h(\mathbf{X}) = 1$, but applied after trimming units with $\max_j\{e_j(\mathbf{X}_i)\} > 0.97$ and $\min_j\{e_j(\mathbf{X}_i)\} < 0.03$. We select this threshold so that the proportion of the sample being trimmed does not exceed 20%.

4. Unadjusted estimator based on the pseudo-observations (PO-UNADJ): we take the mean difference of the pseudo-observations between two arms.

This estimator is only unbiased for estimating the average causal effect under RCT. In observational studies, its bias quantifies the degree of unconfounding.

$$\tau^k_{j,j'}(t) = \frac{\sum_{i=1}^N \widehat{\theta}_i^k(t)\mathbf{1}\{Z_i = j\}}{\sum_{i=1}^N \mathbf{1}\{Z_i = j\}} - \frac{\sum_{i=1}^N \widehat{\theta}_i^k(t)\mathbf{1}\{Z_i = j'\}}{\sum_{i=1}^N \mathbf{1}\{Z_i = j'\}}.$$

5. Regression model using the pseudo-observations with the g-formula (PO-G): we fit the following regression model for the pseudo-observations on $\mathbf{X}_i$ and Z_i,

$$\mathbb{E}(\widehat{\theta}_i^k(t)|\mathbf{X}_i, Z_i) = g^{-1}\left(\mathbf{X}_i\alpha^T + \sum_{j\in\mathcal{J}} \boldsymbol{\gamma}_j \mathbf{1}\{Z_i = j\}\right),$$

where $g(\cdot)$ is the link function (we use log-link for RACE/ASCE and complementary log-log link for SPCE, and construct the estimator for $\widehat{\tau}^{k,h}_{j,j'}(t)$ with $h(\mathbf{X}) = 1$ using the g-formula,

$$\hat{\tau}^k_{j,j'}(t) = N^{-1}\sum_{i=1}^N \{\mathbb{E}(\widehat{\theta}_i^k(t)|\mathbf{X}_i, Z_i = j) - \mathbb{E}(\widehat{\theta}_i^k(t)|\mathbf{X}_i, Z_i = j')\}.$$

6. Augmented weighting estimator (AIPW, OW): we use equation (9) in the main text with IPW or OW.

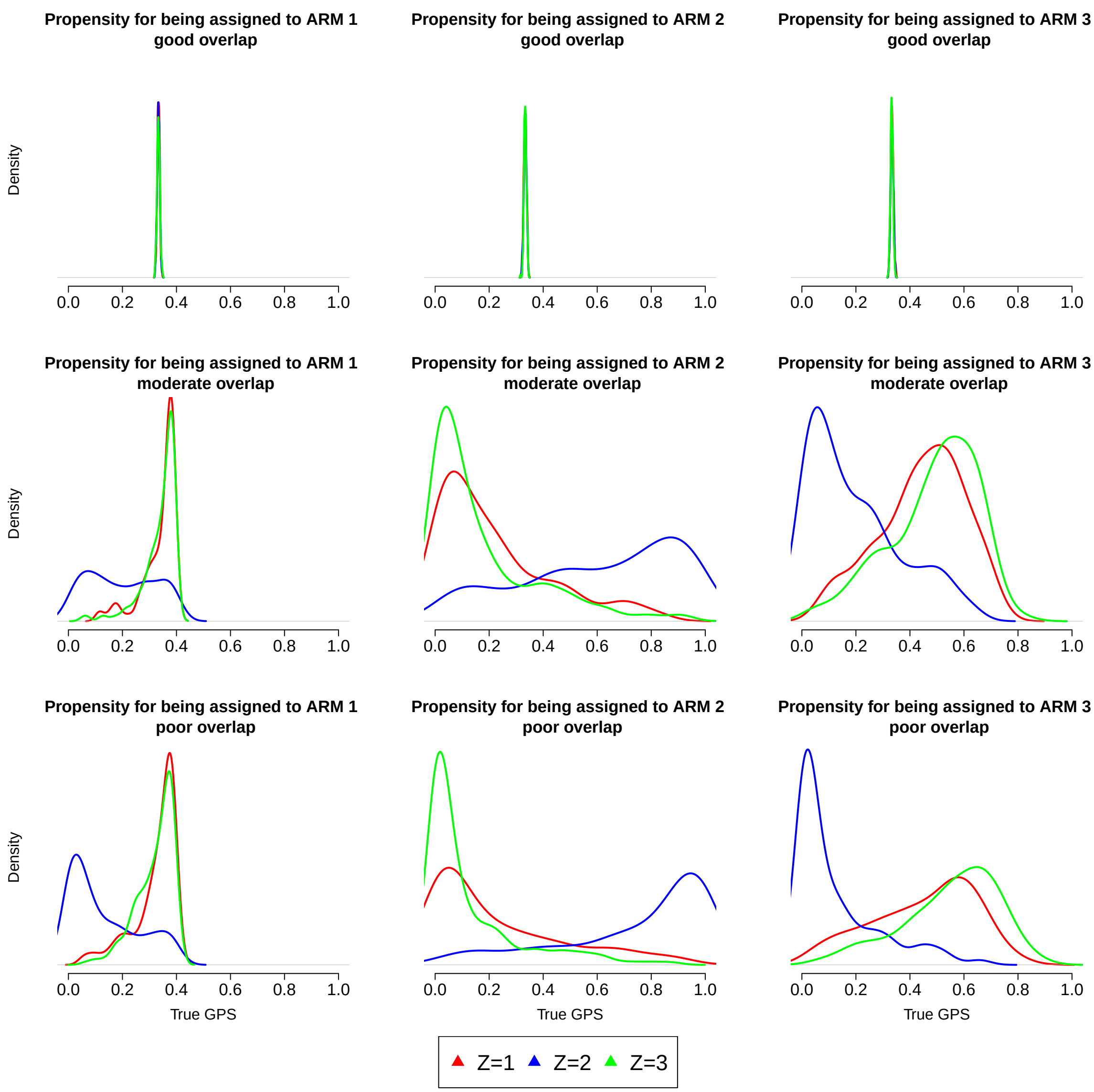
7. Propensity score weighted Cox model estimator in Mao et al. (2018) (IPW-MAO, OW-MAO): we employ the estimator proposed in Mao et al. (2018) combining IPW or OW in fitting the Cox model with the weighted likeli-

hood,

$$l_i = \sum_{j=1}^{J} D_{ij} w_j^h(\mathbf{X}_i) \left(\Delta_i \mathbf{U}(\widetilde{T}_i)^T \mathbf{a}_j - \int_0^{\widetilde{T}_i} \exp(\mathbf{U}(t)^T \mathbf{a}_j) dt \right),$$

where $\mathbf{U}(t) = (\zeta(t)^T, \lambda(t))^T)^T$ are the known truncated power basis function of degree L with K knots. $\zeta(t) = (1, 2, \cdots, t^L)^T, \lambda(t) = (\lambda_1^L(t), \lambda_2^L(t), \cdots, \lambda_K^L(t))^T$. We place the "$w_i''$ in the equation (9) of Mao et al. (2018) with $w_j^h(\mathbf{X}_i)$ to accommodate the multi-arm scenario. We estimate the survival function in each arm $\widehat{S}_j(t), j \in \mathcal{J}$ separately using the same procedure in Mao et al. (2018),

$$\widehat{S}_j(t) = \exp\left\{ - \int_0^t \mathbf{U}(s)^T \widehat{\mathbf{a}}_j ds \right\}.$$



Web Figure 1: Generalized propensity score distribution under different overlap conditions across three arms in the simulation studies. First row: randomized controlled trials (RCT); second row: observational study with relatively good covariate overlap; third row: observational study with poor covariate overlap.

Web Appendix C   Additional simulation results

Additional comparisons under poor covariate overlap Web Figure 2 shows the comparison of different estimators in the simulated data with good covariate overlap between treatment arms. The OW estimator achieves lower bias and RMSE compared with other estimators in most cases (except that the adjusted Cox outcome model has smaller RMSE when the outcome model is correctly specified). Moreover, coverage of the 95% confidence interval of the OW estimator is close to the nominal level while the other estimators can exhibit poor coverage especially in estimating the ASCE.

Comparison with trimmed IPW In Web Figure 3, we compare the performance of the trimmed IPW estimator (T-IPW) with other estimators under poor overlap. Firstly, we notice that T-IPW substantially reduces RMSE and absolute bias compared to the un-trimmed IPW estimator in Figure 1 of the main text. Moreover, coverage rate of T-IPW estimator become closer to the nominal level. Nonetheless,T-IPW is still consistently more biased and less efficient than OW.

Comparison with regression on pseudo-observations Web Figure 4 shows the comparison with the estimators using regression on the pseudo-observations. When there is good overlap, the regression adjusted estimator (PO-G) achieves a similar RMSE and bias to the IPW estimator and is slightly better when estimating ASCE. However, PO-G has larger bias and RMSE compared OW across all scenarios. The coverage of PO-G is poor compared with the weighting estimators, which might be due to misspecification of the regression models. Performance of PO-G deteriorates when the covariate overlap is poor, leading to larger bias and RMSE, and lower coverage rate.

Comparison with augmented weighting estimator In Web Figure 5, we compare the proposed estimators with two augmented weighting estimators, augmented IPW (AIPW) and augmented OW (AOW), under good and poor overlap, respectively. The AOW achieves a lower bias and RMSE than the AIPW. Compared with the IPW estimator, the AIPW estimator has substantially smaller bias and higher efficiency. The improvement due to augmenting IPW estimator with an outcome model is more pronounced under poor overlap. On the other hand, AOW and OW are nearly indistinguishable, regardless of the degree of overlap.

Comparison with the estimators in Mao et al. (2018) Web Figure 6 compares our estimators with the estimators proposed in Mao et al. (2018) (their original estimators were extended to accommodate multiple treatments) in the simulations. The OW estimator based on pseudo-observations usually has smaller bias than OW-MAO, while OW-MAO can have slightly smaller RMSE due to the almost correct outcome model specification, except for the estimation on ASCE. The IPW-MAO estimator has a smaller bias and RMSE than the IPW estimator but is not comparable to the OW estimator in all cases. However, the coverage of both estimators, especially the IPW-MAO, is lower than the nominal level. The under-coverage can be substantial under poor overlap or when the target estimand is the ASCE. We also replicate the simulations when the outcome is generated from model B. In this case, the proportional hazards assumption in the OW-MAO and IPW-MAO estimators fails to hold, and the performance of these estimators quickly deteriorate, while the performance of the OW estimator based on pseudo-observations remains almost unaffected. These additional results are omitted for brevity.

Simulation results with non-zero treatment effect Web Figure 7 draws the comparison among estimators when the true treatment effect is not zero ($j = 1$, $j' = 2$). For a fair comparison, we scale the bias and RMSE by the absolute value of the true estimand $\tau_{1,2}(t)^{k,h}$ for different choices of $h(\mathbf{X})$. The pattern under good or poor overlap is similar to the counterpart with zero treatment effect. OW has the smallest bias across all scenarios. OW also has the smallest RMSE except when comparing with a correctly specified Cox g-formula estimator for estimating SPCE. Additionally, we find that the coverage rate of the Cox and IPW-Cox estimator using the bootstrap method is surprisingly low for ASCE, which is similar to our findings under zero treatment effect. In Web Table 1, we report the performance of different estimators under covariate dependent censoring or when the proportional hazards assumption is violated. The pattern is similar to Table 1 in the main text with OW performing the best under covariate dependent censoring or with the violation of proportional hazards assumption. The coverage rate for OW estimator is occasionally below the nominal level (but is still closest to nominal among its competitors) for estimating ASCE under covariate dependent censoring since the presentation in Web Table 1 assumes a limited sample size $N = 300$. The coverage rate of OW further improves when N further increases (results not shown).

Results with for simulated RCT In Web Figure 8, we present the results in the simulations when $e_j(\mathbf{X}) = 1/3$ for all $\mathbf{X}$, namely in a three-arm RCT. The bias and RMSE of different "covariate-adjusted" estimators perform similar and are usually more efficient than the unadjusted estimator based on psuedo-observations (PO-UNADJ); in particular, the correctly specified Cox g-formula estimator achieves the smallest RMSE. Furthermore, we observe that the weighting estimators using IPW and OW show a similar bias yet a lower RMSE compared to the PO-UNADJ. Compared to IPW, OW has a smaller RMSE especially when estimating RACE and ASCE. This demonstrates the

Web Table 1: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 1$ versus $j = 2$ under different degrees of overlap. In the "proportional hazards" scenario, the survival outcomes are generated from a Cox model (model A), and in the "non-proportional hazards" scenario, the survival outcomes are generated from an accelerated failure time model (model B). The sample size is fixed at $N = 300$.

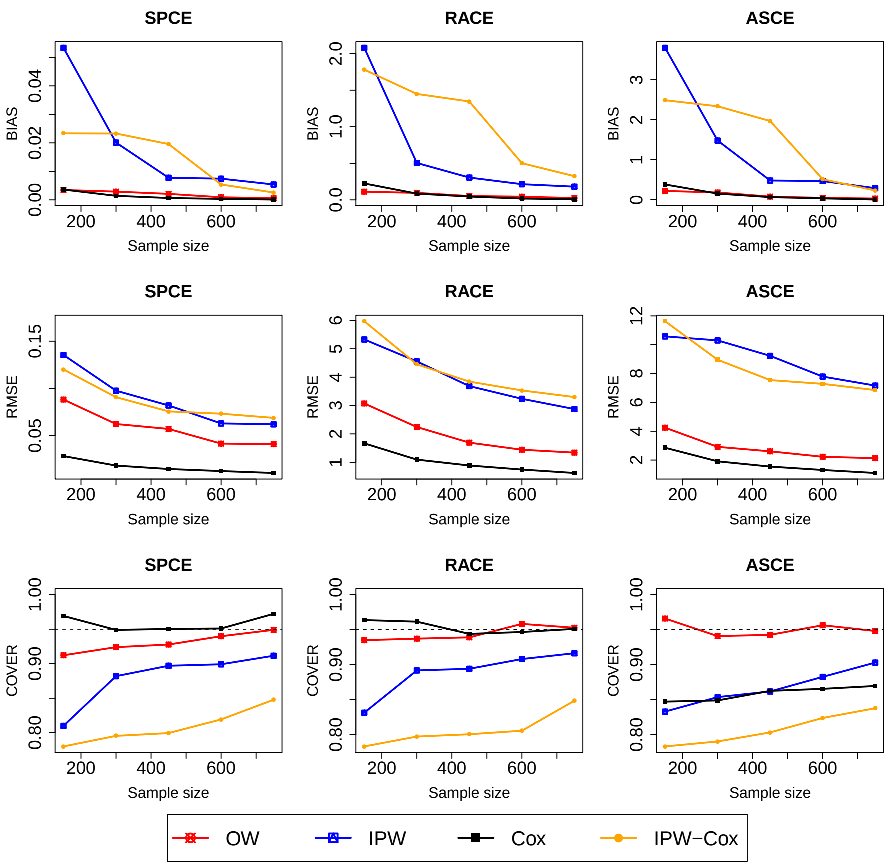
	Degree of	Absolute bias				RMSE				95% Coverage			
	overlap	OW	IPW	Cox	IPW-Cox	OW	IPW	Cox	IPW-Cox	OW	IPW	Cox	IPW-Cox
						Model A, completely random censoring							
SPCE	Good	0.003	0.008	0.002	0.032	0.068	0.096	0.020	0.098	0.930	0.897	0.941	0.769
	Poor	0.004	0.008	0.003	0.078	0.083	0.100	0.048	0.145	0.917	0.892	0.942	0.569
RACE	Good	0.062	0.528	0.119	1.598	1.456	3.169	1.144	4.689	0.937	0.910	0.924	0.792
	Poor	0.071	1.378	0.140	4.049	1.953	3.722	2.762	7.213	0.937	0.897	0.918	0.597
ASCE	Good	1.398	3.629	10.979	3.362	6.792	6.065	5.927	9.107	0.960	0.895	0.052	0.722
	Poor	2.432	5.782	21.485	3.726	9.413	6.663	12.852	12.540	0.888	0.890	0.008	0.666
						Model B, completely random censoring							
SPCE	Good	0.004	0.007	0.007	0.049	0.072	0.105	0.079	0.181	0.943	0.936	0.753	0.784
	Poor	0.004	0.027	0.021	0.130	0.086	0.125	0.198	0.250	0.942	0.933	0.730	0.656
RACE	Good	0.089	0.184	0.702	1.740	2.667	4.256	4.366	8.605	0.956	0.928	0.743	0.789
	Poor	0.112	1.035	1.883	5.573	2.948	5.224	10.762	11.610	0.939	0.928	0.734	0.686
ASCE	Good	2.730	6.266	9.288	7.313	7.832	9.237	12.051	17.280	0.930	0.893	0.531	0.799
	Poor	3.549	8.221	19.251	8.771	8.595	9.271	26.296	19.850	0.862	0.860	0.477	0.628
						Model A, covariate dependent censoring							
SPCE	Good	0.002	0.001	0.004	0.081	0.056	0.085	0.055	0.146	0.953	0.927	0.908	0.708
	Poor	0.005	0.009	0.006	0.144	0.069	0.086	0.060	0.194	0.957	0.895	0.907	0.539
RACE	Good	0.115	0.112	0.296	3.387	2.185	3.778	3.104	6.118	0.954	0.931	0.907	0.746
	Poor	0.204	0.217	0.300	6.250	2.505	4.481	3.407	8.536	0.956	0.930	0.883	0.567
ASCE	Good	0.723	2.298	21.155	12.682	10.300	8.797	12.718	47.799	0.948	0.942	0.653	0.765
	Poor	0.989	4.029	21.787	26.463	16.998	10.508	12.859	48.089	0.955	0.926	0.635	0.599
						Model B, covariate dependent censoring							
SPCE	Good	0.020	0.037	0.001	0.028	0.066	0.085	0.091	0.110	0.727	0.730	0.709	0.866
	Poor	0.028	0.062	0.018	0.036	0.084	0.105	0.241	0.161	0.685	0.690	0.711	0.840
RACE	Good	0.703	1.122	0.451	0.829	5.329	7.251	4.807	7.489	0.934	0.933	0.726	0.848
	Poor	1.129	2.086	0.585	3.511	6.286	9.705	11.919	10.292	0.929	0.929	0.711	0.798
ASCE	Good	4.688	9.188	11.283	10.353	11.343	12.044	13.293	14.549	0.759	0.719	0.528	0.666
	Poor	6.111	15.091	19.178	12.482	12.905	13.497	27.745	15.250	0.755	0.703	0.526	0.531

efficiency gain from covariates adjustment through weighting in RCT, similar to the findings in (Zeng et al., 2020) but under the censored outcome scenario. Moreover, all estimators include the simple PO-UNADJ achieve the coverage rates close to the nominal level.

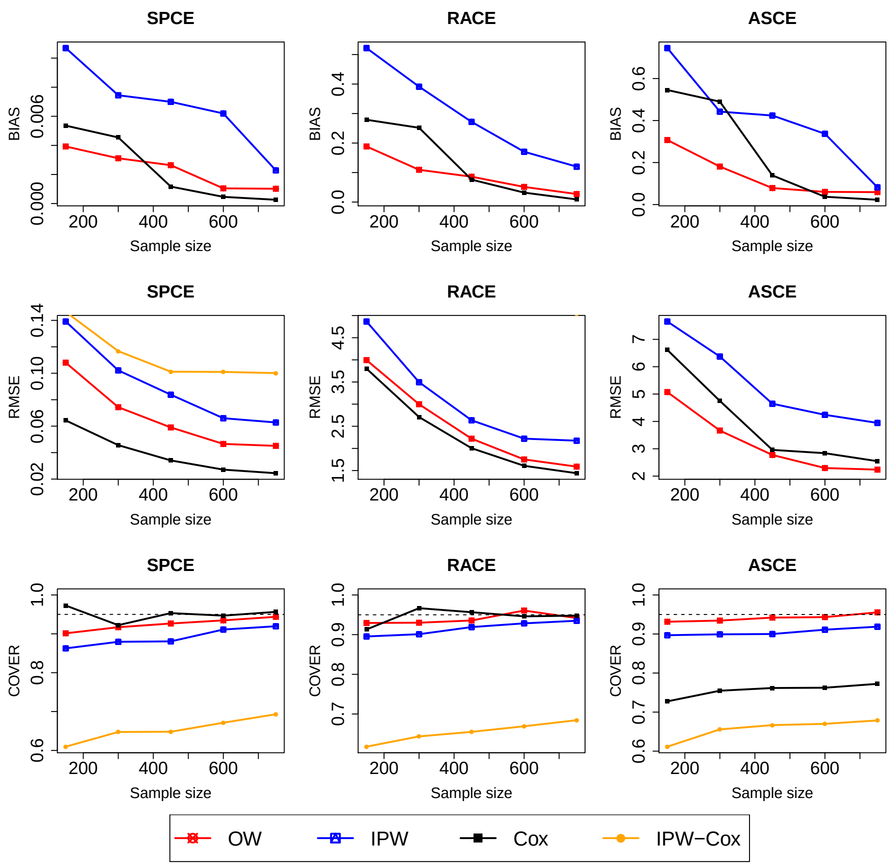
On the other hand, when the survival outcomes are generated from model B under the RCT configuration, Web Table 2 shows that the OW estimator can lead to substantially smaller RMSE than the Cox g-formula and COX-IPW estimators and maintain nominal coverage throughout. This is because the proportional hazards assumption does not hold under data generating model B, and violations of this assumption can lead to efficiency loss even when the treatment is completely randomized. The two Cox-model based estimators are also prone to notable under-coverage in this scenario. Throughout, the OW estimator frequently improves the efficiency over the unadjusted estimator, and performs consistently better than IPW in terms of both bias and RMSE. This observation echoes the findings in Zeng et al. (2020), and supports estimator (5) in the main text as covariate-adjusted estimators for analyzing RCTs.

Web Table 2: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 2$ versus $j = 3$ in simulated RCT. In the "proportional hazards" scenario, the survival outcomes are generated from a Cox model (model A), and in the "non-proportional hazards" scenario, the survival outcomes are generated from an accelerated failure time model (model B). The sample size is fixed at $N = 300$.

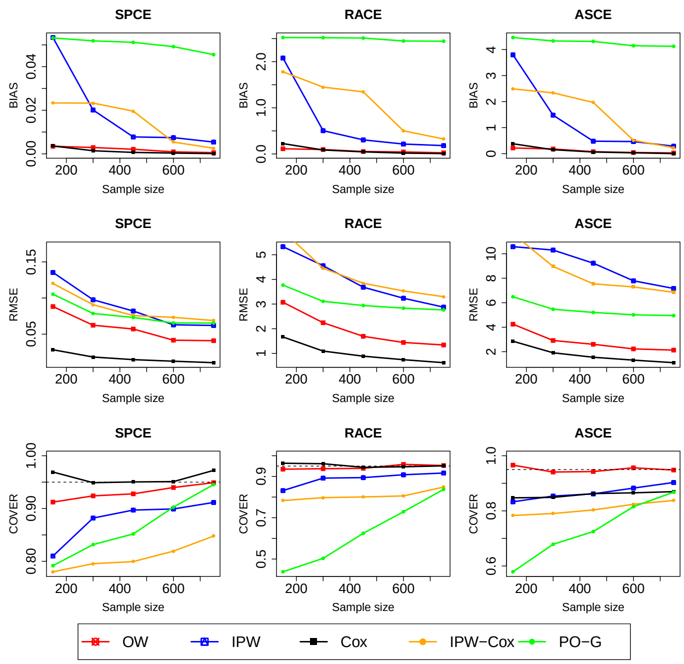
	Absolute bias					RMSE					95% Coverage				
	OW	IPW	Cox	IPW-Cox	UNADJ	OW	IPW	Cox	IPW-Cox	UNADJ	OW	IPW	Cox	IPW-Cox	UNADJ
Model A, completely random censoring															
SPCE	0.002	0.003	0.001	0.003	0.002	0.052	0.052	0.011	0.029	0.065	0.944	0.945	0.960	0.960	0.945
RACE	0.056	0.057	0.047	0.137	0.138	1.570	1.550	0.651	1.413	2.651	0.945	0.953	0.941	0.970	0.954
ASCE	0.096	0.176	0.090	0.269	0.193	2.174	2.917	1.139	2.766	4.592	0.957	0.958	0.937	0.968	0.938
Model B, completely random censoring															
SPCE	0.002	0.005	0.002	0.006	0.004	0.069	0.069	0.042	0.081	0.074	0.946	0.947	0.759	0.840	0.952
RACE	0.072	0.101	0.137	0.314	0.064	2.432	2.418	2.400	4.096	3.062	0.955	0.957	0.758	0.836	0.937
ASCE	0.107	0.223	0.244	0.605	0.095	3.455	4.229	4.173	7.600	5.074	0.943	0.958	0.941	0.835	0.934
Model A, covariate dependent censoring															
SPCE	0.001	0.005	0.001	0.000	0.002	0.044	0.044	0.039	0.039	0.059	0.953	0.955	0.912	0.965	0.943
RACE	0.003	0.027	0.065	0.022	0.100	2.257	2.248	2.315	1.717	2.995	0.948	0.949	0.922	0.968	0.958
ASCE	0.007	0.187	0.163	0.188	0.366	2.716	6.887	4.899	10.564	11.279	0.952	0.951	0.955	0.979	0.949
Model B, covariate dependent censoring															
SPCE	0.000	0.000	0.001	0.000	0.002	0.037	0.037	0.055	0.059	0.043	0.949	0.948	0.706	0.904	0.896
RACE	0.005	0.068	0.064	0.136	0.106	4.700	4.671	2.944	5.310	4.371	0.951	0.953	0.722	0.856	0.959
ASCE	0.002	0.080	0.166	0.268	0.105	3.523	4.391	4.761	6.548	4.375	0.951	0.949	0.936	0.856	0.960



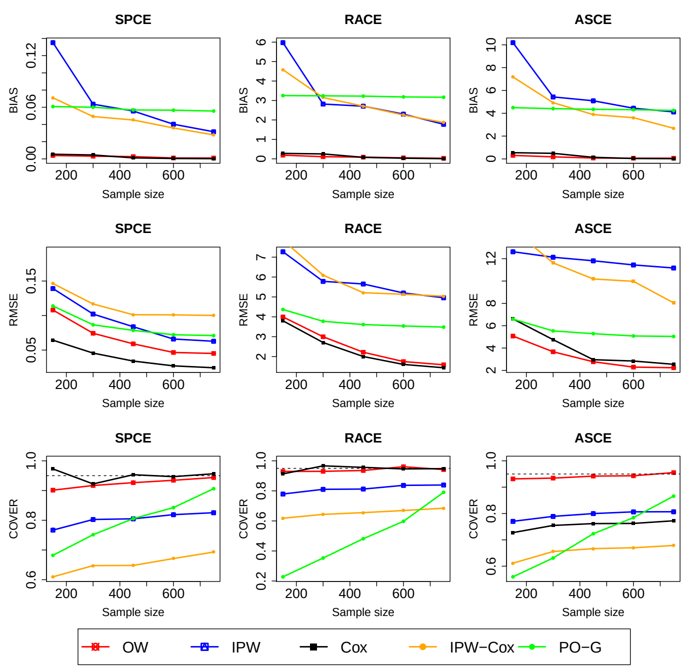
Web Figure 2: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 2$ versus $j = 3$ under good overlap, when the survival outcomes are generated from Model A and censoring is completely independent.



Web Figure 3: Absolute bias, root mean squared error (RMSE) and coverage for comparing treatment $j = 2$ versus $j = 3$ under poor overlap, when survival outcomes are generated from model A and censoring is completely independent. Additional comparison with T-IPW.

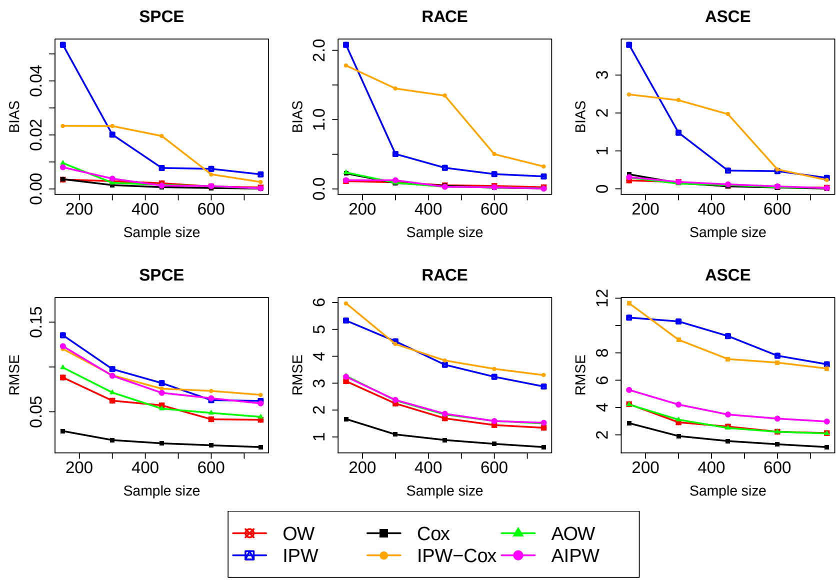


(a) Comparison under good overlap

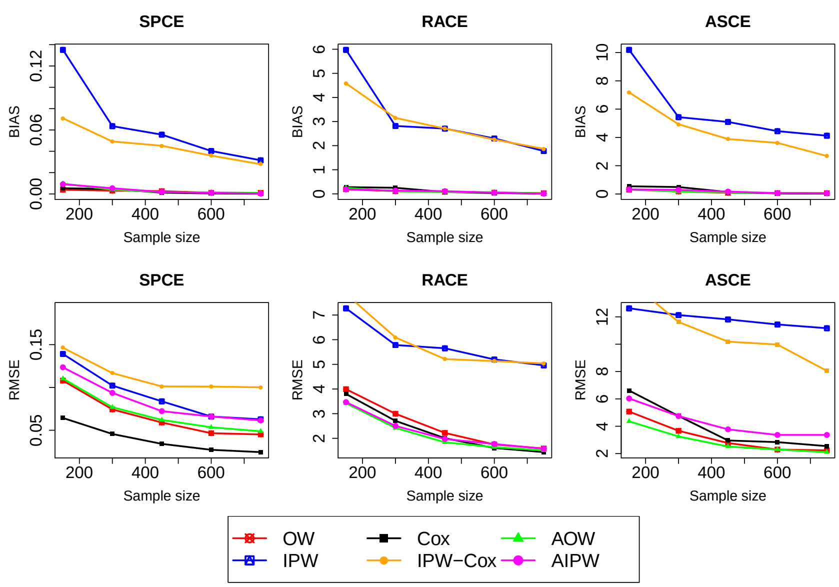


(b) Comparison under poor overlap

Web Figure 4: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for $j = 2$ versus $j = 3$ comparison, when survival outcomes are generated from model A and censoring is completely independent. Additional comparison with PO-G.

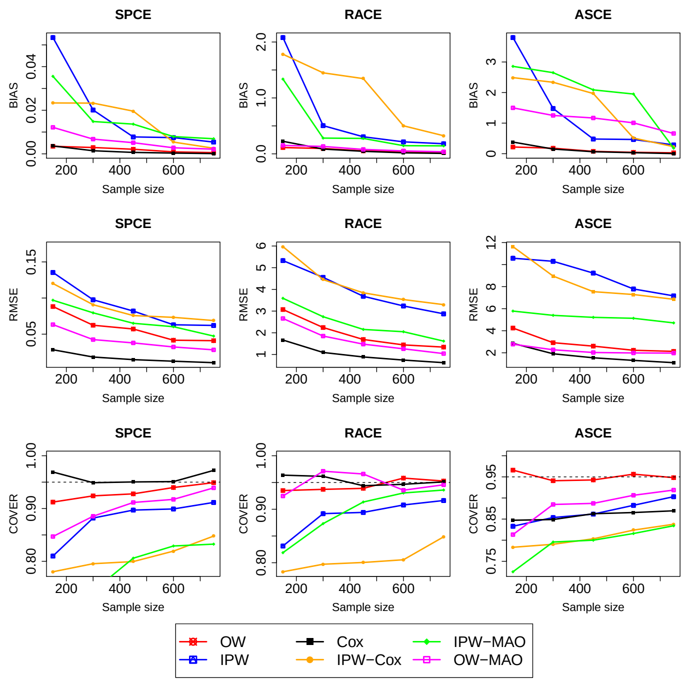


(a) Comparison under good overlap

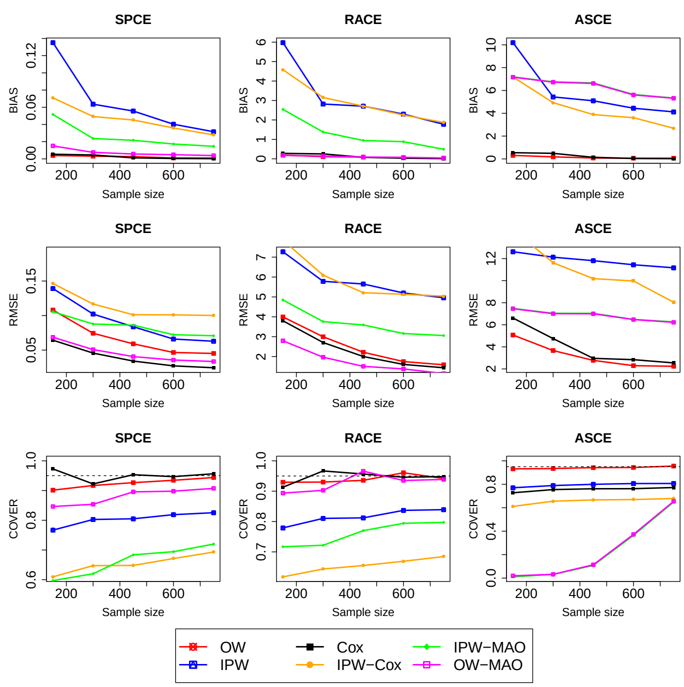


(b) Comparison under poor overlap

Web Figure 5: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 2$ versus $j = 3$, when survival outcomes are generated from model A and censoring is completely independent. Additional comparison with augmented weighting estimators.

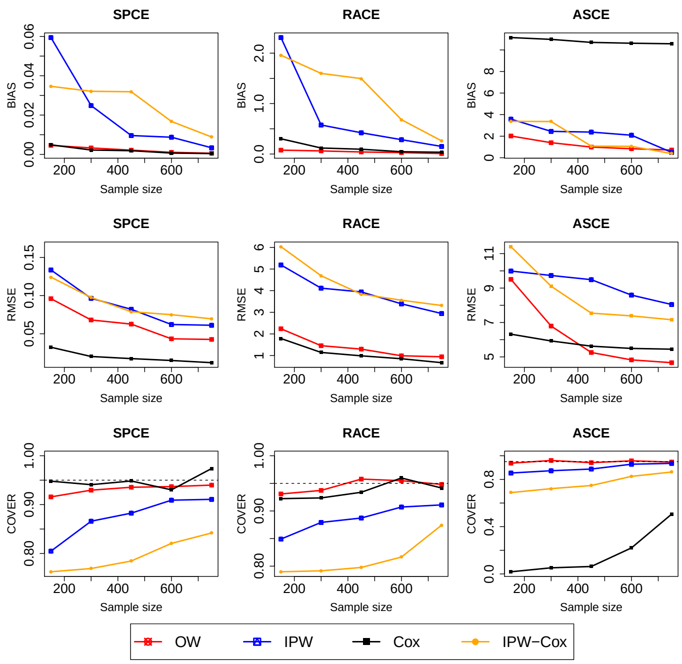


(a) Comparison under good overlap

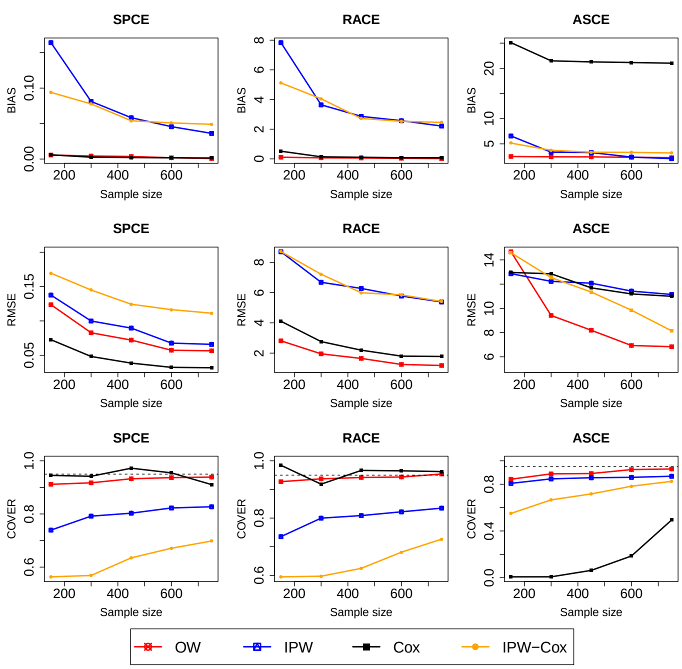


(b) Comparison under poor overlap

Web Figure 6: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing $j = 2$ versus $j = 3$, when survival outcomes are generated from model A and censoring is completely independent, including IPW-MAO, OW-MAO.

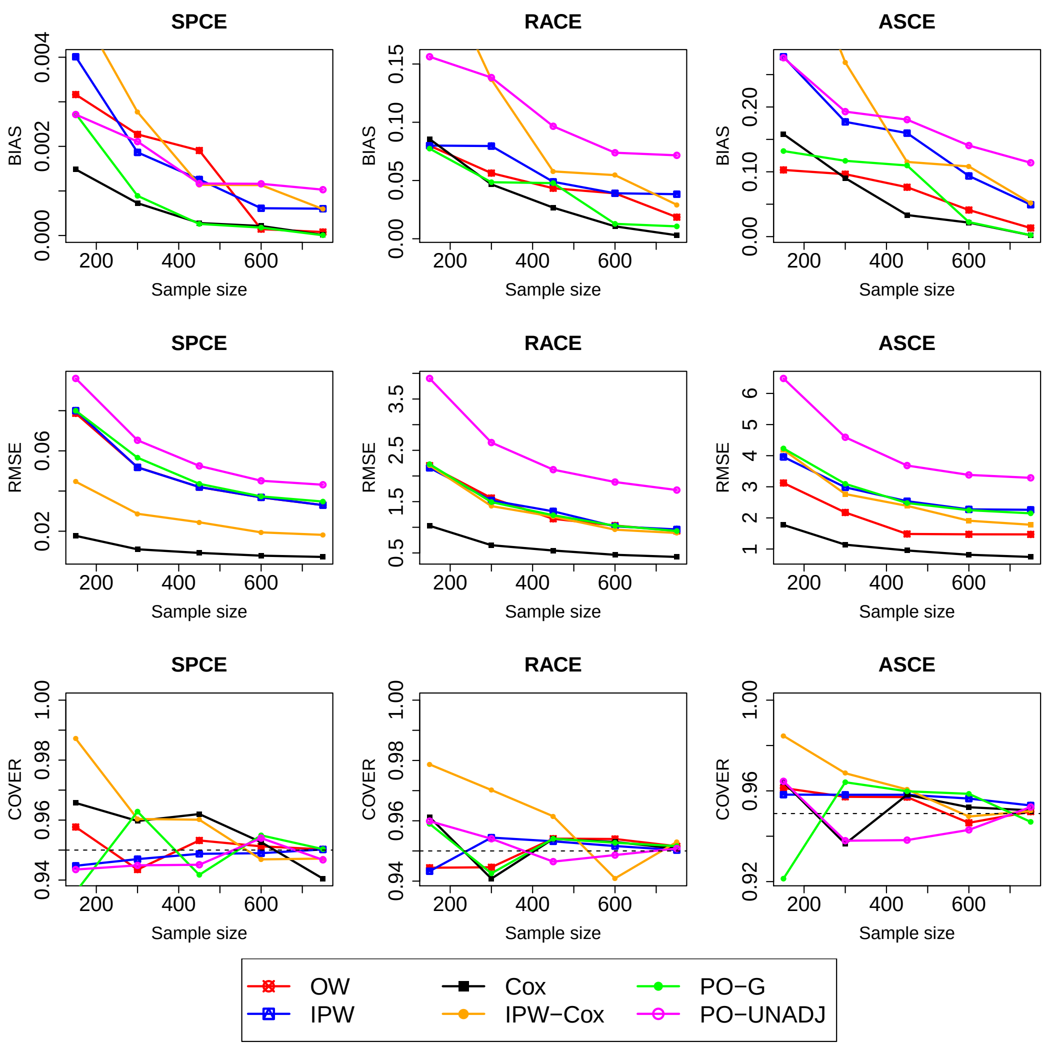


(a) Comparison under good overlap



(b) Comparison under poor overlap

Web Figure 7: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 1$ versus $j = 2$, when survival outcomes are generated from model A and censoring is completely independent.



Web Figure 8: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 2$ versus $j = 3$ in the simulate RCT, when the survival outcomes are generated from Model A and censoring is completely independent. It also shows additional comparison with PO-G and PO-UNADJ.

Web Appendix D Simulations with a misspecified censoring model when the censoring mechanism depends on covariates

To investigate the performance of OW and IPW with pseudo-observations with a misspecified censoring model under Assumption (A4) – covariate-dependent censoring, we generate the censoring time $C_i(j) = C_i \ \forall$ from a log-normal distribution given covariates

$$\log(C_i) \sim \mathcal{N}(\mu_i^c, \sigma^2 = 0.64), \quad \mu_i^c = 3.5 + \mathbf{X}_i^T \alpha_c,$$

where $\alpha_c = (0.2, 0.1, -0.1, 0.1)^T$ captures the dependence of censoring time on covariates. For analyzing the simulated data sets, however, we assume a Cox proportional hazards model for the censoring time when calculating the pseudo-observations, rendering the covariate-dependent censoring model *misspecified*. We summarize the additional simulation results in Web Table 3.

When the proportional hazards assumption holds (model A), the Cox g-formula estimator shows a smaller bias and RMSE compared to the OW and IPW estimator. This is expected because the confounder-adjusted Cox model is the true outcome model and the usual "independent censoring" assumption holds for the Cox model under our data generating process. However, IPW and OW with pseudo-observations have close to nominal coverage (based on the proposed variance estimators) across all three estimands, while the Cox g-formula estimator and IPW-Cox exhibit substantial under-coverage when the estimand of interest is ASCE. Regardless of overlap, OW has smaller RMSE and closer to nominal coverage than IPW.

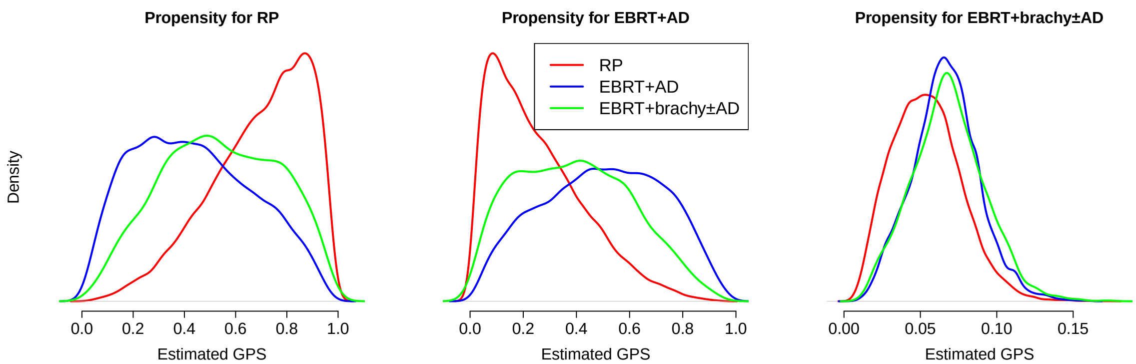
Under the non-proportional hazards scenario (model B), both the Cox g-formula and Cox-IPW estimators are largely compromised, exhibiting a much larger absolute bias and RMSE than OW and IPW under a misspecified censoring model. Again, OW outperforms IPW in terms of bias, RMSE and coverage regardless of overlap, which is consistent with observations in our main simulations. When the proportional hazards assumption is violated, both OW and IPW lead to coverage rates close to the nominal level except when the target estimand is SPCE or ASCE, but the coverage rates for RACE is surprisingly robust for both weighting estimators with a misspecified censoring model. Overall, this additional simulation study reveals that (i) OW outperforms IPW even when the censoring model is misspecified across all scenarios; (ii) OW can exhibit an increasingly greater advantage over the Cox g-formula estimator when the proportional hazards assumption is violated and/or the target estimand is the ASCE.

Web Table 3: Absolute bias, root mean squared error (RMSE) and coverage of the 95% confidence interval for comparing treatment $j = 1$ versus $j = 2$ under different degrees of overlap when the censoring model is misspecified under the covariate-dependent censoring scenario. In the "proportional hazards" scenario, the survival outcomes are generated from a Cox model (model A), and in the "non-proportional hazards" scenario, the survival outcomes are generated from an accelerated failure time model (model B). The sample size is fixed at $N = 300$.

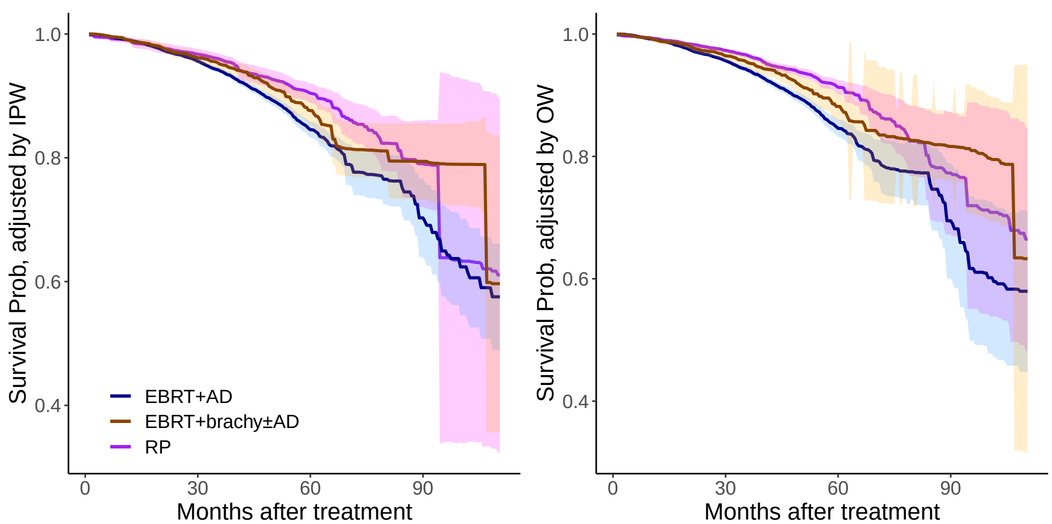
	Degree of	Absolute bias				RMSE				95% Coverage			
	overlap	OW	IPW	Cox	IPW-Cox	OW	IPW	Cox	IPW-Cox	OW	IPW	Cox	IPW-Cox
						Model A, covariate-dependent censoring							
SPCE	Good	0.006	0.005	0.002	0.022	0.196	0.272	0.145	0.108	0.964	0.952	0.958	0.939
	Poor	0.003	0.001	0.004	0.064	0.211	0.297	0.160	0.167	0.970	0.919	0.959	0.864
RACE	Good	0.474	0.330	0.022	2.349	13.219	15.112	8.662	7.640	0.927	0.903	0.941	0.846
	Poor	0.493	0.557	0.126	6.107	13.015	16.650	9.608	12.047	0.932	0.910	0.951	0.689
ASCE	Good	0.364	0.201	0.067	2.630	10.282	16.280	11.650	9.734	0.947	0.917	0.635	0.848
	Poor	0.615	0.783	0.080	7.125	8.701	20.947	12.791	15.012	0.934	0.923	0.644	0.695
						Model B, covariate-dependent censoring							
SPCE	Good	0.002	0.006	0.044	0.091	0.188	0.172	0.490	0.288	0.822	0.880	0.747	0.798
	Poor	0.010	0.012	0.027	0.174	0.213	0.215	0.526	0.369	0.833	0.714	0.755	0.708
RACE	Good	1.074	1.488	2.922	4.831	12.862	13.300	29.210	13.548	0.944	0.937	0.943	0.808
	Poor	1.601	1.701	1.760	8.642	12.514	15.227	31.294	17.708	0.947	0.940	0.953	0.714
ASCE	Good	5.701	9.695	14.281	12.553	23.034	29.036	81.629	73.041	0.745	0.725	0.653	0.804
	Poor	6.804	10.885	24.735	27.326	21.860	30.429	87.814	86.107	0.756	0.754	0.652	0.703

Web Appendix E Additional information on the application

Web Table 4 reports summary statistics of the covariates in the application on prostate cancer (Section 5) before and after weighting adjustment. The $\text{MPASD}^{\text{IPW}}$ and MPASD^{OW} is smaller than the unadjusted difference $\text{MPASD}^{\text{UNADJ}}$. Of note, while OW with logistic propensity scores leads to exact covariate balance for $J = 2$ groups (Li et al., 2018), OW with multinomial logistic propensity scores do not guarantee exact covariate balance among $J \geq 3$ groups (Li and Li, 2019). Nonetheless, Web Table 4 shows that OW still leads to consistently smaller MPASD compared to IPW, with values below the usual 0.1 threshold across all covariates. Web Figure 9 illustrates the distributions of the estimated generalized propensity scores in the three treatments; it indicates a good covariate overlap between the groups. Web Figure 10 contains the estimated counterfactual survival curves of three treatments of prostate cancer based on weighting estimator using IPW or OW. Web Table 5 further shows the point and interval estimations for SPCE and RACE (at 60 months) regarding the pairwise comparison for three treatments of prostate cancer, using different methods.



Web Figure 9: Marginal distributions of the estimated generalized propensity scores for three arms from a multinomial logistic regression in the prostate cancer application.



Web Figure 10: Estimated counterfactual survival curves of the three treatments of prostate cancer (Section 5) estimated from the pseudo-observations-based weighting estimator, using IPW (left) and OW (right).

Web Table 4: Descriptive statistics of the baseline covariates in the comparative effectiveness study on prostate cancer described in Section 5 and maximized pairwise absolute standardized difference (MPASD) of each covariate across three arms before and after weighting.

	Overall	RP	EBRT+AD	EBRT+brachy±AD	$MPASD^{UNADJ}$	$MPASD^{IPW}$	$MPASD^{OW}$
No (%)	44551(100)	26474 (59.42)	15435 (34.65)	2642(5.93)			
Continuous covariates, mean and standard deviation (in parenthesis).							
Age	65.32 (8.19)	62.61 (7.02)	69.66 (8.19)	67.15 (7.72)	0.919	0.105	0.096
PSA	201.89 (223.42)	189.20 (214.84)	225.77 (238.08)	189.577 (207.46)	0.166	0.055	0.029
Categorial covariates, number of units in each class.							
Race							
Black	7127	3632	3000	495	0.151	0.032	0.036
Other	1524	903	522	99	0.020	0.012	0.004
Spanish or Hispanic	1963	1135	703	125	0.021	0.020	0.013
Insure type							
Not insured	986	555	402	29	0.110	0.004	0.009
Private insurance	19522	14608	3925	989	0.629	0.014	0.015
Medicaid	1284	598	612	74	0.100	0.030	0.033
Medicare	1026	436	553	37	0.149	0.013	0.006
Government	482	235	211	36	0.044	0.020	0.006
Income level ($)							
<30000	5533	2954	2234	345	0.099	0.034	0.018
30000-34999	7628	4330	2858	440	0.057	0.024	0.013
35000-45999	12436	7317	4458	661	0.087	0.003	0.009
Education level							
>29	6776	3719	2651	406	0.086	0.024	0.021
20-28.9	9707	5461	3690	556	0.079	0.005	0.004
14-19.9	10706	6299	3806	601	0.045	0.014	0.005
Charlson Comorbidity Index							
1	7008	4575	2101	332	0.134	0.002	0.011
≥ 2	1211	631	517	63	0.060	0.003	0.003
Gleason score							
≤ 6	3493	2769	553	171	0.274	0.030	0.007
7	9347	5964	2837	546	0.103	0.023	0.016
9	11781	6130	4968	683	0.204	0.012	0.007
10	932	348	532	52	0.144	0.008	0.004
Clinical T stage							
$\leq$ cT3	5723	2785	2529	409	0.169	0.008	0.025
Year of diagnosis							
2004-2007	330	127	167	36	0.090	0.012	0.013
2008-2010	11582	6665	4082	835	0.144	0.009	0.005

Web Table 5: Pairwise treatment effect estimates of the three treatments of prostate cancer (Section 5 using four methods, on the scale of restricted average causal effect (RACE) and survival probability causal effect (SPCE) at 60 months/5 years post-treatment.

Method	Estimate	Standard error	95% Confidence interval	p-value
		EBRT-AD vs. RP comparison		
		Restricted average causal effect		
OW	-1.277	0.150	(-1.524, -1.031)	0.000
IPW	-0.917	0.264	(-1.351, -0.484)	0.001
COX	-1.342	0.126	(-1.549, -1.136)	0.000
MSM	-0.931	0.220	(-1.294, -0.568)	0.000
		Survival probability causal effect		
OW	-0.062	0.009	(-0.076, -0.048)	0.000
IPW	-0.067	0.009	(-0.083, -0.052)	0.000
COX	-0.059	0.006	(-0.068, -0.050)	0.000
MSM	-0.039	0.010	(-0.056, -0.023)	0.000
		EBRT+brachy±AD vs. RP comparison		
		Restricted average causal effect		
OW	-0.562	0.236	(-0.950, -0.174)	0.017
IPW	-0.309	0.331	(-0.855, 0.236)	0.350
COX	-0.802	0.214	(-1.155, -0.450)	0.000
MSM	-0.363	0.317	(-0.885, 0.158)	0.252
		Survival probability causal effect		
OW	-0.032	0.013	(-0.054, -0.010)	0.016
IPW	-0.031	0.013	(-0.053, -0.009)	0.021
COX	-0.036	0.009	(-0.051, -0.020)	0.000
MSM	-0.015	0.014	(-0.038, 0.007)	0.256
		EBRT+brachy±AD vs. EBRT+AD comparison		
		Restricted average causal effect		
OW	0.715	0.240	(0.321, 1.109)	0.003
IPW	0.710	0.242	(0.195, 1.021)	0.015
COX	0.540	0.216	(0.184, 0.896)	0.012
MSM	0.568	0.246	(0.163, 0.973)	0.021
		Survival probability causal effect		
OW	0.030	0.014	(0.006, 0.053)	0.036
IPW	0.036	0.014	(0.013, 0.059)	0.011
COX	0.024	0.009	(0.008, 0.039)	0.013
MSM	0.024	0.010	(0.007, 0.041)	0.021

Web Appendix F Details for reproducing the results

In this appendix, we include the details to reproduce all the results reported in the paper. Please download the codebase from `https://github.com/zengshx777/OW_Survival_CodeBase`.

R Scripts in folder `estimators`

`fast_pseudo_calculation.R`	Function for calculating pseudo-observations, which is faster than Rpackage **pseudo**.
`PSW_pseudo.R`	Function for IPW-PO and OW-PO.
`cox_model.R`	Function for estimator Cox and Cox-IPW.
`pseudo_G.R`	Function for PO-UNADJ and PO-G.
`Mao_Method_func.R`	Function for estimators in Mao et al. (2018).
`AIPW_pseudo.R`	Function for AIPW and AOW.

R Scripts in folder `simulation`

`simu_main.R`	Main script for running simulations.
`simu_utils.R`	Utility function for simulations.
`simu_data_gen.R`	Utility function for generating simulated data.
`simu_exe.sh`	Bash script to run simulations in all settings.

R Scripts in folder `data_application`

`data_preprocessing.R`	Data pre-processing for application.
`data_application.R`	Analysis function for data application.

To run the simulations in the paper, you can run the following command and set `simulation` as the working directory.:

```
git clone https://github.com/zengshx777/OW_Survival_CodeBase
```

```
R CMD BATCH --vanilla '--args dependent.censoring=F multi.arm=T
prop.hazard=F good_overlap=1 sample_size=150' simu_main.R R1.out
```

where `dependent.censoring` controls whether the censoring is independent of the covariates; `multi.arm` controls the number of arms in the data (`T` for $J = 3$, `F` for $J = 2$); `prop.hazard` controls whether the proportional hazard assumption is correct; `good_overlap` control the overlap conditions (`1` for RCT, `2` for observational study with good overlap, `3` for observational study with poor overlap); `sample_size` controls the sample size. One simple way to run many simulations in different settings in parallel is to run the `simu_exe.sh` directly (you can customize the scenario in this file). The current `simu_main.R` will run all estimators mentioned in the paper by default, which might be time-consuming. You can comment out certain estimators to speed up.

The results will be saved in the folder `simulation_results`. To output similar Figures and Tables as in the paper, please refer to the scripts in folder `output_utils`.

The NCDB data used in the case study is publicly available upon request to and approval of the NCDB Participant User File application.

Bibliography

Dudley, R. M. and R. Norvaiša (1999). *Differentiability of six operators on nonsmooth functions and p-variation, Lecture Notes in Math.* 1703. Springer, Berlin.

Graw, F., T. A. Gerds, and M. Schumacher (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis 15*(2), 241–255.

Jacobsen, M. and T. Martinussen (2016). A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. *Scandinavian Journal of Statistics 43*(3), 845–862.

James, L. F. et al. (1997). A study of a class of weighted bootstraps for censored data. *Annals of Statistics 25*(4), 1595–1621.

Lehmann, E. L. and G. Casella (2006). *Theory of point estimation.* Springer Science & Business Media.

Li, F. and F. Li (2019). Propensity score weighting for causal inference with multiple treatments. *The Annals of Applied Statistics 13*(4), 2389–2415.

Li, F., K. L. Morgan, and A. M. Zaslavsky (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association 113*(521), 390–400.

Mao, H., L. Li, W. Yang, and Y. Shen (2018). On the propensity score weighting analysis with survival outcome: Estimands, estimation, and inference. *Statistics in Medicine 37*(26), 3745–3763.

Overgaard, M., E. T. Parner, and J. Pedersen (2019). Pseudo-observations under covariate-dependent censoring. *Journal of Statistical Planning and Inference 202*, 112–122.

Overgaard, M., E. T. Parner, J. Pedersen, et al. (2017). Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations. *The Annals of Statistics 45*(5), 1988–2015.

Van der Vaart, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probablistic Mathematics*, Volume 3. Cambridge university press.

Zeng, S., F. Li, R. Wang, and F. Li (2020). Propensity score weighting for covariate adjustment in randomized clinical trials. *Statistics in Medicine 40*(4), 842–858.