Web Appendices for "Using Overlap Weights to Address Extreme Propensity Scores in Estimating Restricted Mean Counterfactual Survival Times"

Zhiqiang Cao[1], Lama Ghazi[2], Claudia Mastrogiacomo[3,4], Laura Forastiere[3,4], F.Perry Wilson[5], Fan Li[3,4*]

[1]Department of Mathematics, College of Big Data and Internet, Shenzhen Technology University, Guangdong, China.
[2]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA.
[3]Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.
[4]Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven, Connecticut, USA.
[5]Clinical and Translational Research Accelerator, Department of Medicine, Yale School of Medicine, New Haven, Connecticut, USA.

1 Web Appendix 1: Definition of target estimands $\triangle_w(L)$

Before defining the causal estimand of interest $\triangle_w(L)$ mathematically, we introduce some notation. $A(=a)$ is defined as treatment or control group (i.e., $a = 0$ or 1), and $\mathbf{X}$ as baseline covariates. Let T denote survival time, which is subject to right censoring C, $U = \min(T, C)$ and $\delta = I(T \leq C)$. Furthermore, let $T^{(a)}$ $(a = 0, 1)$ denote the potential (or counterfactual) lifetime and $C^{(a)}$ denote the potential censoring time of each unit if, possibly contrary to fact, s/he received treatment $A = a$.

We assume the marginal density of $\mathbf{X}$ (from the combined treatment and control groups) exists and denote it by $f(\mathbf{X})$. Under the balancing weight framework, we could generally represent the target population density by $f(\mathbf{X})h(\mathbf{X})/\text{const}$, where $h(\mathbf{X})$ is a pre-specified tilting function, and const is a normalization constant such that the density is well-defined and integrates to unity. Each choice of the tilting function h corresponds to a target estimand: the average treatment effect based on the restricted mean counterfactual survival times among the target population

$$\triangle_w(L) \quad = \quad E_w\{\min(T^{(1)}, L)\} - E\{\min(T^{(0)}, L)\} = \frac{E[h(\mathbf{X})(\mu^{(1)}(\mathbf{X}; L) - \mu^{(0)}(\mathbf{X}; L))]}{E[h(\mathbf{X})]},$$

where both expectations are taken over the original density $f(\mathbf{X})$, and $\mu^{(1)}(\mathbf{X}; L) = E\{\min(T^{(1)}, L)|\mathbf{X}\}$, $\mu^{(0)}(\mathbf{X}; L) = E\{\min(T^{(0)}, L)|\mathbf{X}\}$ are conditional means of L-restricted counterfactual survival time given pre-treatment covari-

*Correspondence to Fan Li, PhD, Department of Biostatistics, Center for Methods in Implementation and Prevention Science, Yale School of Public Health, Suite 200, Room 229, 135 College Street, New Haven, Connecticut 06510 (email: fan.f.li@yale.edu)

ates. The quantity $\triangle_w(L)$ is called the weighted average treatment effect (WATE) on the restricted survival times. Equivalently, we can also express the restricted mean counterfactual survival times estimands as integrals of the counterfactual survival functions, and therefore,

$$\triangle_w(L) \quad = \quad \int_0^L \{S_w^{(1)}(t) - S_w^{(0)}(t)\}dt = \frac{\int_0^L E\big[h(\mathbf{X})\big(S^{(1)}(t|\mathbf{X}) - S^{(0)}(t|\mathbf{X})\big)\big]}{E[h(\mathbf{X})]},$$

where $S^{(1)}(t|\mathbf{X}) = P(T^{(1)} \geq t|\mathbf{X})$ and $S^{(0)}(t|\mathbf{X}) = P(T^{(0)} \geq t|\mathbf{X})$ are two conditional counterfactual survival functions. For this entire class of WATE estimands, when the weighting function $h(\mathbf{X}) = 1$ for all values of $\mathbf{X}$, $\triangle_w(L) = \triangle_{IPTW}(L) = E\{\mu^{(1)}(\mathbf{X}; L) - \mu^{(0)}(\mathbf{X}; L)\} = E\{\min(T^{(1)}, L)|\mathbf{X}\} - E\{\min(T^{(0)}, L)|\mathbf{X}\}$, which is the standard L-restricted average treatment effect (ATE) for the combined treatment and control population. With such a choice of the weighting function, treatment comparison from each unit contributes equally. When $h(\mathbf{X}) = I_{\alpha<e(\mathbf{X})<1-\alpha}$, $\triangle_w(L)$ is the L-restricted ATE for the target population after symmetric trimming. Notice that h can be analogously defined to characterize the L-restricted ATE for the target population after asymmetric trimming. Finally, when $h(\mathbf{X}) = e(\mathbf{X})(1 - e(\mathbf{X}))$, $\triangle_w(L) = \triangle_{OW}(L)$ is the L-restricted ATE for the overlap population, i.e., the population emphasizing clinical equipoise whose treatment decisions remain most uncertain. With this choice of the weighting function, $h(\mathbf{X})$ is maximized when the value of PS is 0.5, and decreases to zero as PS becomes extreme. Therefore, OW up-weights patients who have a substantial probability to receive either treatment and smoothly down-weights the patients in the tails of the PS distribution. Even though the truncation corresponds to a choice of $h(\mathbf{X})$ as a function of the propensity score, in our simulations, we still consider the true estimand for truncation to be the average treatment effect based on the restricted mean counterfactual survival times among the combined population, that is, the exact same estimand for IPTW. This is because, when used in practice, truncation only serves as a way to stabilize the estimated propensity scores among extreme units, but the intended target population is often the usual ATE. It simply aims to reduce the variance of the treatment effect estimates at the cost of potentially inflated bias (Cole and Hernán, 2008).

This Appendix is organized as follows: we show consistency of $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$ under the weighting scheme w in Web Appendix 2. Then, we prove the optimality of the overlap weighting in asymptotic efficiency in Web Appendix 3 and provide their corresponding variance estimation in Web Appendix 4. R tutorial for calculating RMCST based on each weighting scheme w is presented in Web Appendix 5. To further illustrate the use of our methods, we analyze the observational data of Right-Heart Catheterization Study (RHC Study) by the proposed RMCST estimators in Web Appendix 6. Web Appendix 7 applies the proposed estimators to the Yale New Haven Hospital Hypertension data. Web Tables and Figures are shown in Web Appendix 8 of this Appendix.

2 Web Appendix 2: Consistency of $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$

As is shown in the main manuscript, propensity score model $e(\mathbf{X}) = e(\mathbf{X}, \boldsymbol{\beta})$ is assumed to be a logistic model, i.e.,

$$e(\mathbf{X}, \boldsymbol{\beta}) = \{1 + \exp(-\boldsymbol{\beta}^T\mathbf{X})\}^{-1}.$$

For the censoring time C, which follows a Cox model, that is,

$$\begin{aligned} K_C^{(a)}(t, \mathbf{X}) &= P(C \geq t|\mathbf{X}, A = a) = \exp(-\Lambda_C^{(a)}(t|\mathbf{X})), \\ \Lambda_C^{(a)}(t|\mathbf{X}) &= \int_0^t \lambda_C^{(a)}(u|\mathbf{X})du = \int_0^t \lambda_0^{(a)}(u)\exp(\boldsymbol{\theta}_a^T\mathbf{X})du, \qquad a = 0, 1, \end{aligned}$$

where $\lambda_0^{(a)}(t)$ is an unspecified baseline hazard function.

As is shown in the main manuscript, we adopt the following four standard assumptions in causal survival analysis:

- (A1) consistency and no-interference, which ensures $T = T^{(a)}$ and $C = C^{(a)}$ for $A = a$;
- (A2) conditional exchangeability, i.e., $\{T^{(a)}, C^{(a)}\} \perp A|\mathbf{X}$, which assumes away unmeasured confounders;
- (A3) covariate-dependent censoring, i.e., $T^{(a)} \perp C^{(a)}|\{\mathbf{X}, A = a\}$ for $a = 0, 1$, which means failure time is independent of censoring time given observed covariates in each group;
- (A4) positivity, such that the conditional probability of treatment assignment is bounded away from 0 and 1, and the conditional survival probability of censoring is larger than 0.

Under each balancing weights w, $\mu_w^{(a)}(L)$ and $\triangle_w(L)$ can be rewritten as

$$\mu_w^{(a)}(L) = \int_0^L \exp(-\Lambda_w^{(a)}(t))dt \quad \text{and} \quad \triangle_w(L) = \int_0^L \left\{ \exp(-\Lambda_w^{(1)}(t)) - \exp(-\Lambda_w^{(0)}(t)) \right\} dt = \mu_w^{(1)}(L) - \mu_w^{(0)}(L), \tag{1}$$

respectively, where $\Lambda_w^{(a)}(t)$ is the corresponding cumulative hazard function of $T^{(a)}$. If we can prove the estimator of $\Lambda_w^{(a)}(t)$ (i.e., $\hat{\Lambda}_w^{(a)}(t)$) is consistent, then, consistency of $\hat{\mu}_w^{(a)}(L)$ and $\hat{\triangle}_w(L)$ are straightforward by formula (1).

Recall our balancing weights estimator for $\Lambda_w^{(a)}(t)$ is given by (see (3) of manuscript)

$$\begin{aligned} \hat{\Lambda}_w^{(a)}(t) &= \int_0^t \frac{\sum_{i=1}^n [I(A_i = a)\{\hat{w}_i / \hat{K}_C^{(a)}(u, \mathbf{X}_i)\} dN_{ia}(u)]}{\sum_{i=1}^n [I(A_i = a)\{\hat{w}_i / \hat{K}_C^{(a)}(u, \mathbf{X}_i)\} Y_{ia}(u)]}, \\ &= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n [I(A_i = a)\hat{w}_i e^{\hat{\Lambda}_C^{(a)}(u|\mathbf{X}_i)} dN_{ia}(u)]}{\frac{1}{n}\sum_{i=1}^n [I(A_i = a)\hat{w}_i e^{\hat{\Lambda}_C^{(a)}(u|\mathbf{X}_i)} Y_{ia}(u)]}, \end{aligned} \tag{2}$$

where $\hat{K}_C^{(a)}(t, \mathbf{X}_i) = \exp(-\hat{\Lambda}_C^{(a)}(t|\mathbf{X}_i))$, $\hat{\Lambda}_C^{(a)}(t|\mathbf{X}_i) = \hat{\Lambda}_0^{(a)}(t)\exp(\hat{\boldsymbol{\theta}}_a^T \mathbf{X}_i)$, $N_{ia}(u) = I(A_i = a)N_i(u)$, $N_i(u) = I(U_i \leq u, \delta_i = 1)$, $Y_{ia}(u) = I(A_i = a)Y_i(u)$, $Y_i(u) = I(U_i \geq u)$, $i = 1, \ldots, n$ represents each individual. Under regularity conditions of Zhang and Schaubel (2012), we have $\hat{\boldsymbol{\beta}} \xrightarrow{\mathcal{P}} \boldsymbol{\beta}$ and $\hat{\boldsymbol{\theta}}_a \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_a$. When $a = 1$, for the denomerator of (2), we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \left[I(A_i = 1)\hat{w}_i e^{\hat{\Lambda}_C^{(1)}(u|\mathbf{X}_i)} Y_{i1}(u) \right] \\ \xrightarrow{\mathcal{P}} \; & E\left[\frac{I(A = 1)h(\mathbf{X})I(C^{(1)} \geq u)I(T^{(1)} \geq u)}{e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})} \right] \quad \text{(by the weak law of large numbers)} \\ = \; & E\left[E\left[\frac{I(A = 1)h(\mathbf{X})I(C^{(1)} \geq u)I(T^{(1)} \geq u)}{e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})} \right] |\mathbf{X} \right] \\ = \; & E\left[\frac{h(X)E[I(A = 1)|\mathbf{X}]E[I(C^{(1)} \geq u)I(T^{(1)} \geq u)|\mathbf{X}]}{e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})} \right] \quad \text{(because } \{T^{(a)}, C^{(a)}\} \perp A|\mathbf{X}) \\ = \; & E\left[\frac{h(X)Pr(A = 1|\mathbf{X})E[I(C^{(1)} \geq u)|\mathbf{X}]E[I(T^{(1)} \geq u)|\mathbf{X}]}{e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})} \right] \quad \text{(because } \{T^{(a)} \perp C^{(a)}|\mathbf{X}, A = a\}) \\ = \; & E\left[\frac{h(X)e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})E[I(T^{(1)} \geq u)|\mathbf{X}]}{e(\mathbf{X}, \boldsymbol{\beta})K_C^{(1)}(u, \mathbf{X})} \right] \\ = \; & E\{h(\mathbf{X})Pr(T^{(1)} \geq u|\mathbf{X})\} \quad \text{(because } 0 < \mathrm{e}(\mathbf{X}, \boldsymbol{\beta}) < 1 \;\; \& \;\; K_C^{(a)}(u, \mathbf{X}) > 0) \\ = \; & E[h(\mathbf{X})S^{(1)}(u|\mathbf{X})], \end{aligned}$$

where $S^{(1)}(u|\mathbf{X}) = P(T^{(1)} \geq u|\mathbf{X})$. Similarly, we can obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left[I(A_i=1)\hat{w}_i e^{\hat{\Lambda}_C^{(1)}(u|\mathbf{X}_i)}dN_{i1}(u)\right] \xrightarrow{\mathcal{P}} -dE[h(\mathbf{X})S^{(1)}(u|\mathbf{X})].$$

Therefore, we have $\hat{\Lambda}_w^{(1)}(t) \xrightarrow{\mathcal{P}} \Lambda_w^{(1)}(t)$, i.e., $\hat{S}_w^{(1)}(t) \xrightarrow{\mathcal{P}} S_w^{(1)}(t)$. Using the same technique, we can show that $\hat{\Lambda}_w^{(0)}(t) \xrightarrow{\mathcal{P}} \Lambda_w^{(0)}(t)$ and $\hat{S}_w^{(0)}(t) \xrightarrow{\mathcal{P}} S_w^{(0)}(t)$, respectively. Thus, $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$ are consistent for $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$, respectively.

3 Web Appendix 3: Optimality of the overlap weighting in asymptotic efficiency

In this Section, we prove that OW achieves the smallest asymptotic pointwise variance for estimating $\triangle_w(L)$ among the class of balancing weights under certain homoscedasticity conditions. In the following proof, we will assume that the propensity score $e(\mathbf{X})$ and the censoring score $K_c^{(a)}(t,\mathbf{X})$ are known and therefore do not further consider the uncertainty for estimating these nuisance functions to establish the optimality of the OW estimator..

Result 1. (Optimal variance) Define $U^{(a)} = \min(T^{(a)}, C^{(a)})$ as the right-censored outcome that would have been observed under treatment and control assignment if $a = 1$ and 0, respectively. If the variance of the "pseudo-outcome" $\frac{I(U_i^{(a)}\geq t)}{K_C^{(a)}(t,\mathbf{X}_i)}$ $(i = 1,\ldots,n)$ is homoskedastic across both treatment and control groups, i.e.,

$$Var\left(\frac{I(U_i^{(1)}\geq t)}{K_C^{(1)}(t,\mathbf{X}_i)}|\mathbf{X}_i\right) = Var\left(\frac{I(U_i^{(0)}\geq t)}{K_C^{(0)}(t,\mathbf{X}_i)}|\mathbf{X}_i\right) = c,$$

for some constant $c > 0$, then the OW with $\triangle_{OW}(L)$ gives the smallest asymptotic variance for weighted estimator $\triangle_w(L)$ among all $h(\mathbf{X}_i)$.

Proof. Frist notice that,

$$\triangle_w(L) = \int_0^L \big\{\exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big)\big\}dt. \tag{3}$$

If OW gives the smallest asymptotic variance for $\exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big)$ among all $h(\mathbf{X}_i)$, then conclusion in result 1 is straightforward because RMST estimand is a monotone transformation, so by minimizing the variance of the counterfactual survival function estimator, we can end up with the optimal RMST estimator. Recall that

$$\begin{aligned}
\Lambda_w^{(a)}(t) &= \int_0^t \frac{\sum_{i=1}^n [I(A_i=a)\{w_i/K_C^{(a)}(u,\mathbf{X}_i)\}dN_{ia}(u)]}{\sum_{i=1}^n [I(A_i=a)\{w_i/K_C^{(a)}(u,\mathbf{X}_i)\}Y_{ia}(u)]} \\
&= -\log\left(\sum_{i=1}^n [I(A_i=a)\{w_i/K_C^{(a)}(u,\mathbf{X}_i)\}Y_{ia}(u)\right)|_0^t \\
&= -\log\left(\frac{\sum_{i=1}^n [I(A_i=a)\{w_i/K_C^{(a)}(t,\mathbf{X}_i)\}Y_{ia}(t)}{\sum_{i=1}^n I(A_i=a)w_i}\right). \quad (K_C^{(a)}(0,\mathbf{X})=1, Y_{ia}(0)=1)
\end{aligned}$$

Note that $S_w^{(a)}(t) = \exp\big(-\Lambda_w^{(a)}(t)\big)$, thus, we have

$$S_w^{(a)}(t) = \frac{\sum_{i=1}^n [I(A_i=a)\{w_i/K_C^{(a)}(t,\mathbf{X}_i)\}Y_{ia}(t)}{\sum_{i=1}^n I(A_i=a)w_i} = \frac{\sum_{i=1}^n \frac{w_i I(A_i=a)I(U_i\geq t)}{K_C^{(a)}(t,\mathbf{X}_i)}}{\sum_{i=1}^n I(A_i=a)w_i} = \frac{\sum_{i=1}^n \frac{w_i I(A_i=a)I(U_i^{(a)}\geq t)}{K_C^{(a)}(t,\mathbf{X}_i)}}{\sum_{i=1}^n I(A_i=a)w_i}, \tag{4}$$

where the last equality holds since $I(U_i \geq t) = I(U_i^{(a)} \geq t)$ under $A_i = a$. And for $A_i = 1$, $w_i = \frac{h(\mathbf{X}_i)}{e(\mathbf{X}_i)}$; $A_i = 0$, $w_i = \frac{h(\mathbf{X}_i)}{1-e(\mathbf{X}_i)}$. Conditional on the sample $\tilde{\mathbf{X}} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ and $\tilde{\mathbf{A}} = \{A_1, \ldots, A_n\}$, and noticing the assumption $\{T_i^{(a)}, C_i^{(a)}\} \perp A_i | \mathbf{X}_i$ and only $\frac{I(U_i^{(a)} \geq t)}{K_C^{(a)}(t, \mathbf{X}_i)}$ are random in (4), we obtain

$$
\begin{aligned}
& Var\Big\{ \exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big) | \tilde{\mathbf{X}}, \tilde{\mathbf{A}} \Big\} \\
= \; & Var\big\{ S_w^{(1)}(t) - S_w^{(0)}(t) | \tilde{\mathbf{X}}, \tilde{\mathbf{A}} \big\} \\
= \; & \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i) A_i}{e^2(\mathbf{X}_i)} Var\left(\frac{I(U_i^{(1)} \geq t)}{K_C^{(1)}(t, \mathbf{X}_i)} | \mathbf{X}_i, A_i \right)}{\{\sum_{i=1}^n A_i h(\mathbf{x}_i)/e(\mathbf{x}_i)\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i)(1-A_i)}{(1-e(\mathbf{X}_i))^2} Var\left(\frac{I(U_i^{(0)} \geq t)}{K_C^{(0)}(t, \mathbf{X}_i)} | \mathbf{X}_i, A_i \right)}{\{\sum_{i=1}^n (1-A_i) h(\mathbf{x}_i)/(1-e(\mathbf{x}_i))\}^2} \\
= \; & \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i) A_i}{e^2(\mathbf{X}_i)} Var\left(\frac{I(U_i^{(1)} \geq t)}{K_C^{(1)}(t, \mathbf{X}_i)} | \mathbf{X}_i \right)}{\{\sum_{i=1}^n A_i h(\mathbf{x}_i)/e(\mathbf{x}_i)\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i)(1-A_i)}{(1-e(\mathbf{X}_i))^2} Var\left(\frac{I(U_i^{(0)} \geq t)}{K_C^{(0)}(t, \mathbf{X}_i)} | \mathbf{X}_i \right)}{\{\sum_{i=1}^n (1-A_i) h(\mathbf{x}_i)/(1-e(\mathbf{x}_i))\}^2} \quad (\{T_i^{(a)}, C_i^{(a)}\} \perp A_i | \mathbf{X}_i) \\
= \; & \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i) A_i}{e^2(\mathbf{X}_i)} c_1(\mathbf{X}_i)}{\{\sum_{i=1}^n A_i h(\mathbf{x}_i)/e(\mathbf{x}_i)\}^2} + \frac{\sum_{i=1}^n \frac{h^2(\mathbf{X}_i)(1-A_i)}{(1-e(\mathbf{X}_i))^2} c_0(\mathbf{X}_i)}{\{\sum_{i=1}^n (1-A_i) h(\mathbf{x}_i)/(1-e(\mathbf{x}_i))\}^2},
\end{aligned}
$$

where $c_a(\mathbf{X}_i) = Var\left(\frac{I(U_i^{(1)} \geq t)}{K_C^{(a)}(t, \mathbf{X}_i)} | \mathbf{X}_i \right)$ for $a = 1, 0$. Averaging the above first over the distribution of A (using $E[A_i/e(\mathbf{X}_i)] = E[(1-A_i)/(1-e(\mathbf{X}_i))] = 1$), then over the distribution of $\mathbf{X}$, and applying Slutsky's theorem, we have

$$
n \times Var\Big\{ \exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big) | \tilde{\mathbf{X}}, \tilde{\mathbf{A}} \Big\} \to \frac{\int \big(\frac{c_1(\mathbf{X})}{e(\mathbf{X})} + \frac{c_0(\mathbf{X})}{1-e(\mathbf{X})} \big) h(\mathbf{X})^2 f(\mathbf{X}) \mu(d\mathbf{X})}{(\int h(\mathbf{X}) f(\mathbf{X}) \mu(d\mathbf{X}))^2},
$$

where $f(\mathbf{X})$ is the population density function of $\mathbf{X}$. If the pseudo-outcome is homoskedastic, i.e., $c_1(\mathbf{X}) = c_0(\mathbf{X}) = c$, then the above formula simplifies to

$$
\begin{aligned}
n \times Var\Big\{ \exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big) | \tilde{\mathbf{X}}, \tilde{\mathbf{A}} \Big\} \; & \to \; \frac{c \int \big(\frac{1}{e(\mathbf{X})} + \frac{1}{1-e(\mathbf{X})} \big) h(\mathbf{X})^2 f(\mathbf{X}) \mu(d\mathbf{X})}{(\int h(\mathbf{X}) f(\mathbf{X}) \mu(d\mathbf{X}))^2} \\
& = \; c/C_w \int \frac{h(\mathbf{X})^2 f(\mathbf{X})}{e(\mathbf{X})(1-e(\mathbf{X}))} \mu(d\mathbf{X}), \qquad (5)
\end{aligned}
$$

where $C_w = (\int h(\mathbf{X}) f(\mathbf{X}) \mu(d\mathbf{X}))^2$. Then, by applying the Cauchy-Schwarz inequality, we have that

$$
\begin{aligned}
C_w \; & = \; \left(\int \frac{h(\mathbf{X})}{\sqrt{e(\mathbf{X})(1-e(\mathbf{X}))}} \sqrt{e(\mathbf{X})(1-e(\mathbf{X}))} f(\mathbf{X}) \mu(d\mathbf{X}) \right)^2 \\
& \leq \; \int \frac{h^2(\mathbf{X})}{e(\mathbf{X})(1-e(\mathbf{X}))} f(\mathbf{X}) \mu(d\mathbf{X}) \times \int e(\mathbf{X})(1-e(\mathbf{X})) f(\mathbf{X}) \mu(d(\mathbf{X})).
\end{aligned}
$$

The above equality is achieved when $\frac{h(\mathbf{X})}{\sqrt{e(\mathbf{X})(1-e(\mathbf{X}))}} \propto \sqrt{e(\mathbf{X})(1-e(\mathbf{X}))}$, or equivalently $h(\mathbf{X}) \propto e(\mathbf{X})(1-e(\mathbf{X}))$. Therefore, we obtain OW gives the smallest asymptotic variance for $\exp\big(-\Lambda_w^{(1)}(t)\big) - \exp\big(-\Lambda_w^{(0)}(t)\big)$ by applying the above to the right-hand side of (5). Thus, result 1 is obtained by the definition of $\triangle_w(L)$ in (3).

4 Web Appendix 4: Variance estimation for $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$

In this section, we will derive the variance estimator for $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$ based on the empirical sandwich method, when the PS and censoring process are estimated by a logistic regression and Cox regression, respectively.

The derivation is made up of three parts. In parts 1 and 2, we will derive the estimating equation for the PS and censoring process model, respectively. Then, we will propose the variance estimator for $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$ in part 3.

4.1 Part 1: Propensity score

In main manuscript, we apply a logistic model to describe the propensity score, that is, $e(\mathbf{X}_i;\boldsymbol{\beta}) = Pr(A=1|\mathbf{X}_i) = \{1+\exp(-\mathbf{X}_i^T\boldsymbol{\beta})\}^{-1}$ for $i=1,\ldots,n$. Then, the score function of $\hat{\boldsymbol{\beta}}$ is

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\hat{\boldsymbol{\beta}})] = 0.$$

By using Taylor expansion, we have

$$\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\hat{\boldsymbol{\beta}})] \\
&= n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})] + n^{-1/2}\frac{\partial\{\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})]\}}{\partial\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})] - \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T e(\mathbf{X}_i;\boldsymbol{\beta})\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1).
\end{aligned}$$

Next, we expand the above score equation around the true parameter $\boldsymbol{\beta}$, which leads to

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T e(\mathbf{X}_i;\boldsymbol{\beta})\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})] + o_p(1) \\
&= E_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})] + o_p(1),
\end{aligned}$$

where $E_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T e(\mathbf{X}_i;\boldsymbol{\beta})\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\} \xrightarrow{\mathcal{P}} E[\mathbf{X}^{\otimes 2}e(\mathbf{X};\boldsymbol{\beta})\{1-e(\mathbf{X};\boldsymbol{\beta})\}]$ and $\otimes 2$ denotes Kronecker product.

4.2 Part 2: Censoring process

For censoring time C, we consider Cox regression, that is,

$$K_C^{(a)}(t,\mathbf{X}) = P(C \geq t|\mathbf{X}, A=a) = \exp(-\Lambda_C^{(a)}(t|\mathbf{X})), \qquad a=0,1.$$

where

$$\Lambda_C^{(a)}(t|\mathbf{X}) = \int_0^t \lambda_0^{(a)}(u)\exp(\mathbf{X}^T\boldsymbol{\theta}_a)du. \tag{6}$$

Since the technique used to estimate $\Lambda_C^{(a)}(t|\mathbf{X})$ for $a=0,1$ is the same. For simplicity, we omit C and a in $\Lambda_C^{(a)}(t|\mathbf{X})$ and related notations in this subsection.

By definition of Cox model, estimation of model (6) can be rewritten as

$$\widehat{\Lambda}(t_j|\mathbf{X}_j) = \int_0^{t_j}\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_j(u)\}d\widehat{\Lambda}_0(u), \qquad j=1,\ldots,n.$$

According to Fleming and Harrington (1991) (pp 152), we have

$$d\widehat{\Lambda}_0(u) = \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_i(u)\}}.$$

Thus,

$$\widehat{\Lambda}(t_j) = \int_0^{t_j} \exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_j(u)\}\frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_i(u)\}}.$$

By Fleming and Harrington (1991), we have

$$N_i(u) = M_i(u) + \int_0^u Y_i(s)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(s)\}d\Lambda_0(s).$$

Therefore,

$$\widehat{\Lambda}(t_j) = \int_0^{t_j} \frac{\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_i(u)\}}\sum_{i=1}^n\Big[dM_i(u) + Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}d\Lambda_0(u)\Big].$$

By using Taylor expansion, we obtain

$$\begin{aligned}\frac{\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\hat{\theta}^T\mathbf{X}_i(u)\}} &= \frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}} + \frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}\mathbf{X}_j^T(u)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}}\\ &- \frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{[\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}]^2}\Big[\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}\mathbf{X}_i^T(u)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\Big] + o_p(1).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\widehat{\Lambda}(t_j) &= \int_0^{t_j}\Big[\frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\theta^T\mathbf{X}_i(u)\}} + \frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}}\Big\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\Big\}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\Big]\times\\ &\quad\sum_{i=1}^n\Big[dM_i(u) + Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}d\Lambda_0(u)\Big] + o_p(1)\\ &= \sum_{i=1}^n\Big[\int_0^{t_j}\frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}}dM_i(u)\Big] + \int_0^{t_j}\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}d\Lambda_0(u)\\ &\quad + \int_0^{t_j}\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}\Big\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\Big\}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})d\Lambda_0(u) + o_p(1),\end{aligned}$$

where $\sum_{i=1}^n\int_0^{t_j}\frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}}\Big\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\Big\}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})dM_i(u) = 0$ and
$\bar{\mathbf{X}}(u;\boldsymbol{\theta}) = \frac{E[1/n\sum_{i=1}^n Y_i(u)\mathbf{X}_i^T(u)\exp(\boldsymbol{\theta}^T\mathbf{X}_i(u))]}{E[1/n\sum_{i=1}^n Y_i(u)\exp(\boldsymbol{\theta}^T\mathbf{X}_i(u))]}$.

Thus,

$$\begin{aligned}\widehat{\Lambda}(t_j) - \Lambda(t_j) &= \sum_{i=1}^n\Big[\int_0^{t_j}\frac{\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}}dM_i(u)\Big]\\ &\quad + \int_0^{t_j}\exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}\Big\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\Big\}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})d\Lambda_0(u) + o_p(1). \qquad (7)\end{aligned}$$

By the formula (2.6) of Andersen and Gill (1982) (pp 1103), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = \{n^{-1}\Phi(\boldsymbol{\theta}^*)\}^{-1}\{n^{-1/2}U(\boldsymbol{\theta})\},$$

where $\boldsymbol{\theta}^*$ is between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$. Note that

$$\begin{aligned}\Phi(\boldsymbol{\theta}^*) \approx \Phi(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\int_0^\tau \left[\frac{S^{(2)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)} - \left\{\frac{S^{(1)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)}\right\}^{\otimes 2}\right] dN_i(t) \\ &= \int_0^\tau \left[\frac{S^{(2)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)} - \left\{\frac{S^{(1)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)}\right\}^{\otimes 2}\right] nS^{(0)}(\boldsymbol{\theta},t)d\Lambda_0(t),\end{aligned}$$

where $d\widehat{\Lambda}_0(u) = \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp\{\hat{\boldsymbol{\theta}}^T\mathbf{X}_i(u)\}}$ and $S^{(d)}(\boldsymbol{\theta},t) = \frac{1}{n}\sum_{i=1}^n Y_i(t)\mathbf{X}_i^{\otimes d}(u)\exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}$ for $d = 0, 1, 2$, and τ is the Supremum of time interval.

We note that $n^{-1}\Phi(\boldsymbol{\theta}^*)$ can be approximated by the following expression, i.e.,

$$\int_0^\tau \big[a_2(\boldsymbol{\theta},u) - \bar{\mathbf{X}}(u;\theta)^{\otimes 2}a_0(\boldsymbol{\theta},u)\big]d\Lambda_0(u),$$

where $a_d(\boldsymbol{\theta},u) = E[1/n\sum_{i=1}^n Y_i(u)\mathbf{X}_i^{\otimes d}(u)\exp(\boldsymbol{\theta}^T\mathbf{X}_i(u))]$. According to the definitions of $\bar{\mathbf{X}}(u;\boldsymbol{\theta})$ and $a_d(\boldsymbol{\theta},u)$, we can obtain that $\bar{\mathbf{X}}(u;\boldsymbol{\theta}) = \frac{a_1(\boldsymbol{\theta},u)}{a_0(\boldsymbol{\theta},u)}$.

By the second formula in Andersen and Gill (1982) (pp 1103), we have

$$U(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \left[\mathbf{X}_i^T(u) - \frac{S^{(1)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)}\right] dM_i(u).$$

Since $\bar{\mathbf{X}}(u;\boldsymbol{\theta}) = \frac{a_1(\boldsymbol{\theta},u)}{a_0(\boldsymbol{\theta},u)}$, so we have $\frac{S^{(1)}(\boldsymbol{\theta},t)}{S^{(0)}(\boldsymbol{\theta},t)} = \bar{\mathbf{X}}(u;\boldsymbol{\theta}) + o_p(1)$. Then, we can get the second term of (7), which can be approximated by

$$\begin{aligned}\int_0^{t_j} \exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}\left\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\right\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})d\Lambda_0(u) &\approx \int_0^{t_j} \exp\{\boldsymbol{\theta}^T\mathbf{X}_j(u)\}\left\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\right\}d\Lambda_0(u)\{n^{-1}\Phi(\boldsymbol{\theta}^*,\tau)\}^{-1} \\ &\quad \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\mathbf{X}_j^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\right\} dM_i(u)\end{aligned}$$

Denote $G_i(t;\boldsymbol{\theta}) = \int_0^t \exp\{\boldsymbol{\theta}^T\mathbf{X}_i(u)\}\left\{\mathbf{X}_i^T(u) - \bar{\mathbf{X}}(u;\theta)\right\}d\Lambda_0(u) = \int_0^t \left\{\mathbf{X}_i^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\right\}d\Lambda_i(u)$ and $U_i(\boldsymbol{\theta}) = \int_0^\tau \left\{\mathbf{X}_i^T(u) - \bar{\mathbf{X}}(u;\boldsymbol{\theta})\right\} dM_i(u)$. Therefore, we have

$$\begin{aligned}\sqrt{n}\{\widehat{\Lambda}(t) - \Lambda(t)\} &= G_i^T(t;\boldsymbol{\theta})\Phi(\boldsymbol{\theta})^{-1}n^{-1/2}\sum_{i=1}^n U_i(\boldsymbol{\theta}) + \exp(\boldsymbol{\theta}^T\mathbf{X}_i)n^{-1/2}\sum_{i=1}^n\int_0^t \frac{dM_i(u)}{a_0(\boldsymbol{\theta},u)} + o_p(1), \\ \sqrt{n}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} &= \Phi(\boldsymbol{\theta})^{-1}n^{-1/2}\sum_{i=1}^n U_i(\boldsymbol{\theta}) + o_p(1).\end{aligned}$$

4.3 Part 3: Variance estimation

Next, we consider $\sqrt{n}\{\hat{\Lambda}_w^{(a)}(t) - \Lambda_w^{(a)}(t)\}$ and make the following decomposition:

$$\begin{aligned}\sqrt{n}\{\hat{\Lambda}_w^{(a)}(t) - \Lambda_w^{(a)}(t)\} &= n^{\frac{1}{2}}\left\{\hat{\Lambda}_w^{(a)}(t;\hat{\boldsymbol{\beta}},\widehat{\Lambda}_C^{(a)}) - \hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\widehat{\Lambda}_C^{(a)})\right\} &(8)\\ &\quad + n^{\frac{1}{2}}\left\{\hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\widehat{\Lambda}_C^{(a)}) - \hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\Lambda_C^{(a)})\right\} &(9)\\ &\quad + n^{\frac{1}{2}}\left\{\hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\Lambda_C^{(a)}) - \Lambda_w^{(a)}(t)\right\}. &(10)\end{aligned}$$

By using Taylor expansion, we can show

$$\begin{aligned}\hat{\Lambda}_w^{(a)}(t;\hat{\boldsymbol{\beta}},\widehat{\Lambda}_C^{(a)}) &= \hat{\Lambda}_w^{(a)}(t;\beta,\widehat{\Lambda}_C^{(a)}) + \Bigg[\int_0^t \frac{E[I(A_i=a)e^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dN_{ia}(u)w_i(-1)\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\mathbf{X}_i]}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)} + \\ &\int_0^t E[I(A_i=a)e^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}Y_{ia}(u)w_i\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\mathbf{X}_i]\frac{dQ_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)^2}\Bigg](\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a) &= E[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}Y_{ia}(u)],\\ dQ_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a) &= E[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dN_{ia}(u)].\end{aligned}$$

Denote

$$\begin{aligned}B_a(t;\boldsymbol{\beta},\boldsymbol{\theta}_a) &= \int_0^t \frac{E[I(A_i=a)e^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dN_{ia}(u)w_i(-1)\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\mathbf{X}_i]}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}\\ &+\int_0^t E[I(A_i=a)e^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}Y_{ia}(u)w_i\{1-e(\mathbf{X}_i;\boldsymbol{\beta})\}\mathbf{X}_i]\frac{dQ_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)^2},\end{aligned}$$

Then, we can obtain

$$(8) = B_a(t;\boldsymbol{\beta},\boldsymbol{\theta}_a)^T E_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})].$$

For (9), by using Taylor expansion, we have

$$\begin{aligned}\hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\widehat{\Lambda}_C^{(a)}) &= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dN_{ia}(u)]}{\frac{1}{n}\sum_{i=1}^n[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}Y_{ia}(u)]}\\ &+\Bigg[\int_0^t\frac{dQ_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)} - \int_0^t\frac{dQ_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}{D_a^2(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}\Bigg]\{\widehat{\Lambda}_C^{(a)}(u|\mathbf{X}_i) - \Lambda_C^{(a)}(u|\mathbf{X}_i)\} + o_p(1)\\ &= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dN_{ia}(u)]}{\frac{1}{n}\sum_{i=1}^n[I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}Y_{ia}(u)]} + o_p(1).\end{aligned}$$

Thus, the variance of (9) is zero, which means estimating censoring process does not lead to change in the asymptotic variance estimator of $\hat{\Lambda}_w^{(a)}(t)$. This result is interesting and yet useful in a sense that it greatly decreases the complexity of variance estimation for $\hat{\Lambda}_w^{(a)}(t)$.

For (10), it is straightforward to show that

$$\sqrt{n}\left\{\hat{\Lambda}_w^{(a)}(t;\boldsymbol{\beta},\Lambda_C^{(a)}) - \Lambda_w^{(a)}(t)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t\frac{I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dM_{ia}(u)}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)} + o_p(1),$$

where $dM_{ia}(u) = dN_{ia}(u) - Y_{ia}(u)d\Lambda_w^{(a)}(u)$.

Combining the aforementioned results, we have shown that $\sqrt{n}\{\hat{\Lambda}_w^{(a)}(t) - \Lambda_w^{(a)}(t)\}$ can be represented as $n^{-\frac{1}{2}}\sum_{i=1}^n\psi_i^{(a)}(t)$ plus a term that converges in probability to zero, where

$$\psi_i^{(a)}(t) = B_a(t;\boldsymbol{\beta},\boldsymbol{\theta}_a)^T E_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})] + \int_0^t\frac{I(A_i=a)w_ie^{\Lambda_C^{(a)}(u|\mathbf{X}_i)}dM_{ia}(u)}{D_a(u;\boldsymbol{\beta},\boldsymbol{\theta}_a)}.$$

When both the PS model and the censoring model are correctly specified, it can be seen that $\psi_i^{(a)}(t)$ has mean zero and are identically and independently distributed across $i = 1, \ldots, n$.

Considering the estimation of $\mu_w^{(a)}(L)$, $n^{\frac{1}{2}}\big(\hat{\mu}_w^{(a)}(L) - \mu_w^{(a)}(L)\big)$ can be written as

$$\begin{aligned} n^{\frac{1}{2}}\big(\hat{\mu}_w^{(a)}(L) - \mu_w^{(a)}(L)\big) = n^{\frac{1}{2}} \int_0^L \left\{\hat{S}_w^{(a)}(u) - S_w^{(a)}(u)\right\} du &= n^{\frac{1}{2}} \int_0^L \left\{e^{-\hat{\Lambda}_w^{(a)}(u)} - e^{-\Lambda_w^{(a)}(u)}\right\} du \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \phi_{ia}(L) + o_p(1), \end{aligned}$$

where $\phi_{ia}(L) = -\int_0^L S_w^{(a)}(u)\psi_i^{(a)}(u)du$. When both the propensity score model and the censoring model are correctly specified, the $\phi_{ia}(L)$ variates are independent and identically distributed with mean 0. Therefore, $n^{\frac{1}{2}}\big(\hat{\mu}_w^{(a)}(L) - \mu^{(a)}(L)\big)$ converges to a normal distribution with mean 0 and variance $E(\phi_{ia}^2(L))$. It then follows that $n^{\frac{1}{2}}(\hat{\triangle}_w(L) - \triangle_w(L))$ is also asymptotically normal with mean 0 and variance $E(\phi_{i1}(L) - \phi_{i0}(L))^2$ and $n^{\frac{1}{2}}(\hat{\triangle}_w(L) - \triangle_w(L)) = n^{-\frac{1}{2}} \sum_{i=1}^n I_{\triangle,i}^2 + o_p(1)$, where

$$\begin{aligned} I_{\triangle,i} &= \phi_{i1}(L) - \phi_{i0}(L) = I_{(\triangle,\boldsymbol{\beta}),i} + I_{(\triangle,\boldsymbol{\theta}),i}, \\ I_{(\triangle,\boldsymbol{\beta}),i} &= -\int_0^L \big\{S_w^{(1)}(u)B_1(t;\boldsymbol{\beta},\boldsymbol{\theta}_1)^T - S_w^{(0)}(u)B_0(t;\boldsymbol{\beta},\boldsymbol{\theta}_0)^T\big\}E_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathbf{X}_i[A_i - e(\mathbf{X}_i;\boldsymbol{\beta})]du, \\ I_{(\triangle,\boldsymbol{\theta}),i} &= -\int_0^L \left[S_w^{(1)}(u)\int_0^u \frac{I(A_i=1)w_i e^{\Lambda_C^{(1)}(t|\mathbf{X}_i)}dM_{i1}(t)}{D_1(t;\boldsymbol{\beta},\boldsymbol{\theta}_1)} - S_w^{(0)}(u)\int_0^u \frac{I(A_i=0)w_i e^{\Lambda_C^{(0)}(t|\mathbf{X}_i)}dM_{i0}(t)}{D_0(t;\boldsymbol{\beta},\boldsymbol{\theta}_0)}\right]dt. \end{aligned}$$

In practice, once $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}_a$ are estimated, $\widehat{E}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}$ and $e^{\hat{\Lambda}_C^{(a)}(u|\mathbf{X}_i)}$ are easily obtained from standard software. By counting process and martingale theory (Fleming and Harrington, 1991), it is not difficult to compute $\hat{B}_a(t;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}_a)$ and $\int_0^t \frac{I(A_i=a)\hat{w}_i e^{\hat{\Lambda}_C^{(a)}(u|\mathbf{X}_i)}d\hat{M}_{ia}(u)}{D_a(u;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}_a)}$. We have produced R code to implement the proposed methods of variance estimators (see Web Appendix 5).

5 Web Appendix 5: R tutorial

5.1 Aim

In this Appendix, we provide a step-by-step tutorial for implementation of the proposed propensity score weighting approaches to estimate RMCST. We shall demonstrate our proposed methodologies by using the simulated data set surv.csv, which is available at Github page https://github.com/Zhiqiangcao/Rcode_OW_RMST.

5.2 Dataset

We will use the simulated dataset surv.csv to demonstrate the proposed methods step-by-step. This data set is a one-row-per-patient dataset with time-fixed baseline covariates. The outcome time is patient's survival time (unit: month), which is defined as the difference between date of death and the study admission date and then divided it by 30. Since the outcome is subject to right censoring, we only observed date of first occurrence of last follow-up and death. The censoring indicator is delta, delta = 1 means the patient is alive; otherwise, delta = 0. z is a binary variable representing treated group with value 1 and controlled group with value 0 in this example. In addition,

there are six pre-treatment covariates denoted as x1-x6 in this example. Now we load the dataset and identify those variables.

```
#1. Load Data set
data = read.csv("https://raw.githubusercontent.com/Zhiqiangcao/Rcode_OW_RMST/main/surv.csv")[,-1]
#2. Identity treatment, survival outcome, and censoring indicator
SurvTime = "time"
Status = "delta"
Treatment = "z"
#3. Identity column names of the pre-treatment covariates
Covariates = c("x1","x2","x3","x4","x5","x6")
#4. For convenience, we reorder the dataset by observed survival time
data.sort = data[order(data[,SurvTime]),]
#5. summary of these variables
summary(data.sort[,c(Treatment,SurvTime,Status,Covariates)])
```

```
##        z               time             delta             x1
##  Min.   :0.0000   Min.   :  0.010   Min.   :0.000   Min.   :-3.35000
##  1st Qu.:0.0000   1st Qu.:  1.240   1st Qu.:0.000   1st Qu.:-0.70000
##  Median :0.0000   Median :  3.130   Median :1.000   Median :-0.01500
##  Mean   :0.4955   Mean   :  4.987   Mean   :0.732   Mean   :-0.04249
##  3rd Qu.:1.0000   3rd Qu.:  6.532   3rd Qu.:1.000   3rd Qu.: 0.63000
##  Max.   :1.0000   Max.   :114.080   Max.   :1.000   Max.   : 3.11000
##        x2                  x3                 x4               x5
##  Min.   :-4.470000   Min.   :-3.88000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:-0.670000   1st Qu.:-0.68000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :-0.020000   Median :-0.04000   Median :0.0000   Median :0.000
##  Mean   : 0.000075   Mean   :-0.03687   Mean   :0.4885   Mean   :0.494
##  3rd Qu.: 0.680000   3rd Qu.: 0.60000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   : 3.470000   Max.   : 3.89000   Max.   :1.0000   Max.   :1.000
##        x6
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :1.0000
##  Mean   :0.5035
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

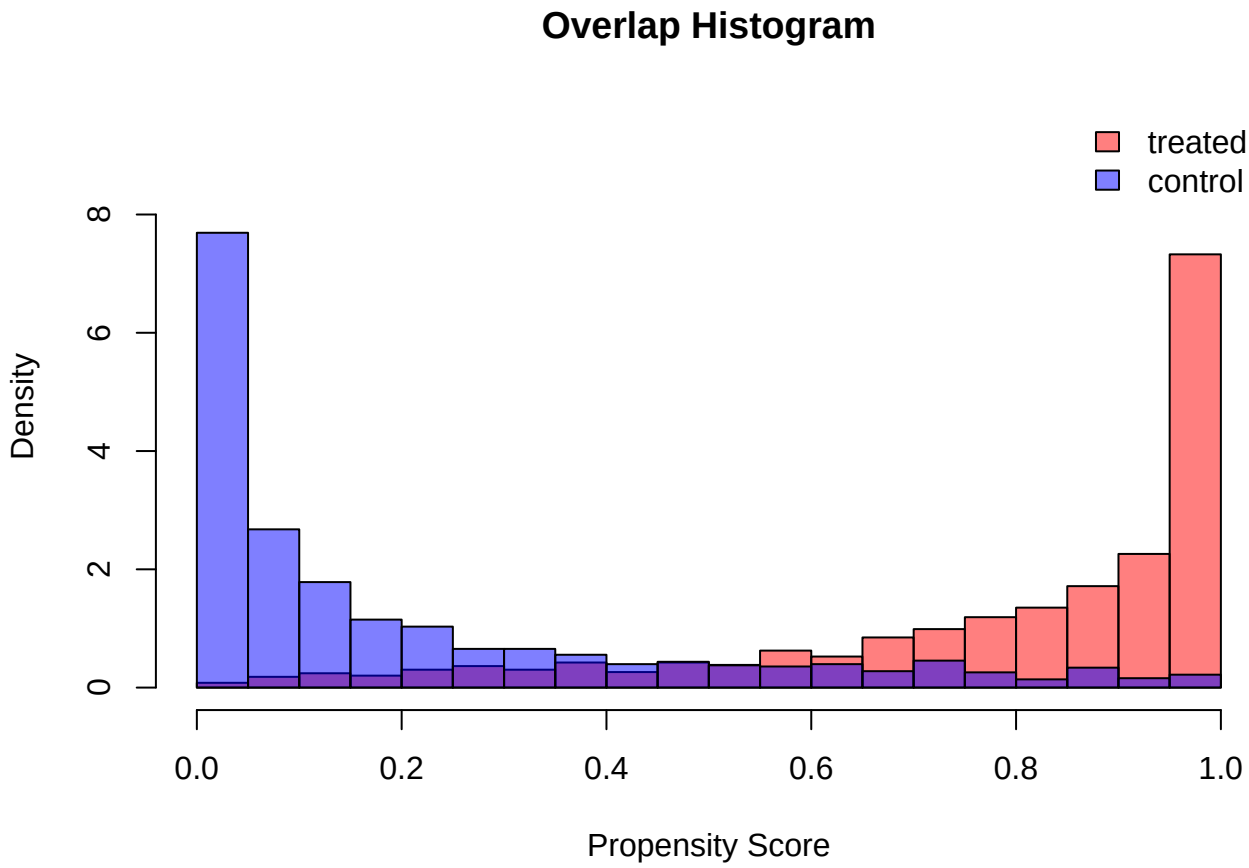
5.3 Propensity score modeling

We use the logistic regression to estimate the propensity score. In this example, we consider all of the six pre-treatment covariates (x1-x6) are included in the analysis:

```
# 1. Construct the logistic regression formula
PS.formula = as.formula(paste(Treatment,"~",paste(Covariates,collapse="+"),sep=""))
# the PS.formula is shown as below:
# z ~ x1 + x2 + x3 + x4 + x5 + x6
# 2. run the logistic regression
PS.model = glm(PS.formula, data=data.sort, family=binomial(link="logit"))
# 3. estimated propensity score
PS.est = 1/(1+exp(-c(PS.model$linear.predictors)))
```

The distributions of the estimated propensity scores in the treated and control groups are visualized as below:

```
hist(PS.est[data.sort[,Treatment]==1], breaks=20,col=rgb(1,0,0,0.5),xlim=c(0,1),
     ylim=c(0,9.5),main="Overlap Histogram",freq=F,xlab="Propensity Score")
hist(PS.est[data.sort[,Treatment]==0], breaks=20,col=rgb(0,0,1,0.5), add=T,freq=F)
legend("topright", c("treated", "control"), bty = "n",
       fill=c(rgb(1,0,0,0.5),rgb(0,0,1,0.5)))
```



5.4 Censoring score modeling

We apply the Cox proportional hazards model to model the censoring process; more details about this model can be seen in *Combining PS weighting and IPCW for estimating RMCST* section of the manuscript. Also, we include all of the 6 pre-treatment covariates in the Cox model analysis.

It is noted that if we choose to implement a single Cox regression model on all the observations, we may need to include full interaction terms between treatment A and covariates $\mathbf{X}$ to accommodate all related effects, which may be tedious especially when the dimensional of $\mathbf{X}$ is high. Since the same task can be achieved by fitting separate models for treatment and control groups, we model the censoring process for the treated group and control group separately. See the code below for details.

```
# 1. library survival package
library("survival")
# 2. Construct the Cox model formula
Censor.formula = as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
```

```
                                    paste(Covariates,collapse="+"),sep=""))
# Censor.formula is shown below
# Surv(time, I(1 - delta)) ~ x1 + x2 + x3 + x4 + x5 + x6
# 3. Cox Model for the treated group
data.trt = subset(data.sort,data.sort[,Treatment]==1)
Censor.trt.model = coxph(Censor.formula,data=data.trt)
# 4. Cox Model for the control group
data.con = subset(data.sort,data.sort[,Treatment]==0)
Censor.con.model = coxph(Censor.formula,data=data.con)
```

5.5 Overlap weighting

Next, we calculate the restricted mean survival time up to 8-month (i.e., $L = 8$) with overlap weighting, i.e., $\triangle_{OW}(8)$. RMCST up to 8 month using other balancing weights (IPTW, symmetric, asymmetric trimming and truncation) can be similarly obtained and the code will be introduced in the next subsection. To estimate RMCST, we first calculate estimated survival probability for each unit at observed survival time U_i ($\hat{S}_w^{(a)}(U_i)$). Since $\hat{S}_w^{(a)}(U_i) = \exp(-\hat{\Lambda}_w^{(a)}(U_i))$, we turn to compute $\hat{\Lambda}_w^{(a)}(U_i)$. Therefore, we need to compute the balancing weights with estimated propensity scores, and then combine with the estimated censoring probabilities to create the final weight for each unit. Once $\hat{S}_w^{(a)}(U_i)$ is obtained for each unit, we can use discrete-time summation to approximate the continuous-time integral for calculating RMCST.

The following code demonstrate how to obtain RMCST up to $L = 8$ with OW:

```
# 1. Define a function to calculate E{min(T^{(a)}, L)}, i.e., \mu^{(a)}(L)=\int_0^L(\hat S_w^{(a)}(t)dt)
### Input: 1. timevec: A vector of time (say, t)
###        2. L: restricted survival time (numerical value)
###        3. sf.est: estimated survival function (corresponds to sorted observed survival time)
###        4. n: number of observations
### Output: Restricted mean counterfactual survival time up to L.
cal.rmst = function(timevec,L,sf.est,n){
  us = timevec[order(timevec)]
  uss = c(0,us)
  survh = c(1,sf.est)
  nii = sum(uss <= L)
  auc = 0
  #using the following summation to approximate integral for calculating RMCST
  if(nii<(n+1)){
    for(j in 1:nii){
      difft = min(uss[j+1]-uss[j],L-uss[j])
      auc = auc + difft*survh[j]
```

```
    }
  }else{ #in this case, L=max(time)
    difft = diff(uss)
    auc = sum(difft*sf.est)
  }
  return (auc)
}
# 2. Define a function to calculate the predicted censoring score of z group
### Input: 1. Coxmodel: A Cox model object
###        2. deltaz: censoring indicator of z group
###        3. n: number of observation in group z
### Output: Probability of P(c_i>u_j|X_i) for i=1,...,n and j=1,...,n
###         note: observed survival time u has been sorted, i.e., u_1<u_2<...<u_n)
CensorScore.est = function(Coxmodel,deltaz,n){
  bc_est = Coxmodel$coef
  xc = model.matrix(Coxmodel)
  s2 = cumsum(exp(xc[n:1,]%*%bc_est))
  dlc = (1-deltaz)/s2[n:1]
  dlc[(1:length(dlc))[is.nan(dlc)]] = 0
  ssc = matrix(0, n, n)
  ssc[,1] = dlc[1]*exp(as.matrix(xc)%*%bc_est)
  otc = outer(c(exp(xc%*%bc_est)), dlc, "*")
  otc[,1] = rep(0, n)
  otc_cum = t(apply(otc,1,cumsum))
  ssc_1 = outer(ssc[,1],rep(1,n))
  cumhcij = ssc_1 + otc_cum
  #for patient i with covariates x[i], compute censoring score at observed time u[j]
  ssc = exp(-cumhcij)
  return(ssc)
}
# 3. Define a function to calculate the estimated survival function for group z
### Input: 1. CensorScore: censoring score matrix (CensorScore[i,j] for unit i at u_j)
###        2. bw: balancing weight
###        3. deltaz: censoring indicator of z group
###        4. n: number of observations in group z
### Output: estimated survival function S^{(a)}(u_i) for i=1,2,...,n, where u_1<u_2<...<u_n
survf.est = function(CensorScore,bw,deltaz,n){
  ssc = CensorScore
  ld = lower.tri(matrix(1, n, n), diag=TRUE)
```

```
  denom1 = (bw/ssc)*ld
  #result for denominator of \hat \Lambda^{(a)}_w(t) in (3), i.e., denom[u_i], i=1,...,n
  denom = apply(denom1,2,sum)
  ud = matrix(1, n, n)-ld
  num2 = bw*deltaz/ssc
  denom2 = matrix(rep(denom,n),n,n)
  otr = num2/denom2
  #note that for patient i, \Lambda^{(a)}_w(u_j|X_i)=\Lambda^{(a)}_w(u_i|X_i) for all j>=i,
  cumhij0 = otr*ld+diag(otr)*ud
  udd = diag(n)+ud
  cumhij = cumhij0*udd
  cumhij[is.na(cumhij)] = 0
  cumhz = apply(cumhij,2,sum)
  sf.est = exp(-cumhz)
  return(sf.est)
}
# 4. Calculate the censoring scores
z = data.sort$z #z=1 for treated group; z=0 for control group
delta = data.sort$delta
n1 = dim(data.trt)[1] # censoring scores in the treated group
deltaz1 = delta[z==1]
Kc.trt.est = CensorScore.est(Censor.trt.model,deltaz1,n1)
n0 = dim(data.con)[1] # censoring scores in the control group
deltaz0 = delta[z==0]
Kc.con.est = CensorScore.est(Censor.con.model,deltaz0,n0)
# 5. Calculate the counterfactual survival probability
bw.trt = 1-PS.est[z==1] # balancing weight in the treated group
sf.est.trt = survf.est(Kc.trt.est,bw.trt,deltaz1,n1)
bw.con = PS.est[z==0]  # balancing weight in the control group
sf.est.con = survf.est(Kc.con.est,bw.con,deltaz0,n0)
# 6. CalculateE{min(T^{(a)}, L)}, i.e., \mu^{(a)}(L)=\int_0^L(\hat S_w^{(a)}(t)dt)
auc1 = cal.rmst(data.sort$time[z==1],8,sf.est.trt,n1)
auc0 = cal.rmst(data.sort$time[z==0],8,sf.est.con,n0)
# 7. Finally, obtain the RMCST for L=8
Delta.OW = auc1-auc0
cat(paste("The estimated RMCST up to L=8 is",round(Delta.OW,4)))

## The estimated RMCST up to L=8 is 1.2037
```

5.6 Other balancing weights and condidence interval construction

Similar to the previous subsection, we can estimate RMCST based on the IPTW, symmetric weight, asymmetric weight and truncation weight. The corresponding R code for implementation of all balancing weights are summarized in a unified function RMST.Effect.BW, available at https://github.com/Zhiqiangcao/Rcode_OW_RMST. Usage of this function is demonstrated as follows:
RMST.Effect.BW(Data,L,Treatment,SurvTime,Status,ps.formula,censor.formula, Method="IPTW",alpha,q)
Arguments are:

- Data: a data frame
- L: interested restricted mean lifetime time
- Treatment: treatment variable
- SurvTime: observed survival time
- Status: censoring indicator
- ps.formula: regression formula for the propensity score; see the *Propensity score modeling* subsection for more details
- censor.formula: regression formula for the Cox model for describing censoring process; see *the Censoring score modeling* subsection for more details
- Method: balancing weights; IPTW for IPTW, OW for overlap weighting, Symmetric for symmetric weighting, Asymmetric for asymmetric weighting, Truncation for truncation weighting
- alpha: the triming threshold for symmetric weighting, i.e., α
- q: the triming threshold for asymmetric weighting or truncated threshold for truncation weighting, i.e., q

The output of this function is a data.frame, which includes the point estimate, and the standard error and 95% normality-based confidence interval given by the robust sandwich variance approach of $\hat{\mu}_w^{(1)}(L)$, $\hat{\mu}_w^{(0)}(L)$ and $\hat{\triangle}_w(L)$.

We now calculate RMCST based on IPTW to illustrate usage of this function:

```
# 1. Load RMST.Effect.BW function
source("https://raw.githubusercontent.com/Zhiqiangcao/Rcode_OW_RMST/main/RMST_functions_v3.R")
# 2. define PS and Cox model formulas
ps.formula = as.formula(paste(Treatment,"~",paste(Covariates,collapse="+"),sep=""))
censor.formula = as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                  paste(Covariates,collapse="+"),sep=""))
Data = data
# 3. IPTW estimator
Delta.IPTW = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                            censor.formula,Method="IPTW")
cat("Estimation of RMCST based on IPTW is: \n")
```

```
## Estimation of RMCST based on IPTW is:

round(Delta.IPTW,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.3459 0.1660   4.0206   4.6711
## mu0     3.4284 0.3845   2.6747   4.1821
## Delta   0.9175 0.4272   0.0802   1.7548
```

Next, we calculate RMCST based on symmetric trimming with trimming threshold $\alpha = 0.1$:

```
Delta.SW = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                          censor.formula,Method="Symmetric",alpha=0.1)
cat("Estimation of RMCST based on symmetric trimming is: \n")

## Estimation of RMCST based on symmetric trimming is:

round(Delta.SW,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6657 0.1633   4.3457   4.9857
## mu0     3.5430 0.1654   3.2187   3.8672
## Delta   1.1227 0.2315   0.6690   1.5764
```

Then, we calculate RMCST based on asymmetric trimming with trimming threshold $q = 0.01$:

```
Delta.AW = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                          censor.formula,Method="Asymmetric",q=0.01)
cat("Estimation of RMCST based on asymmetric trimming is: \n")

## Estimation of RMCST based on asymmetric trimming is:

round(Delta.AW,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6251 0.1539   4.3234   4.9267
## mu0     3.2508 0.1737   2.9103   3.5913
## Delta   1.3743 0.2364   0.9110   1.8376
```

And we calculate RMCST based on truncation IPTW with truncated threshold $q = 0.1$:

```
Delta.TR = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                          censor.formula,Method="Truncation",q=0.1)
cat("Estimation of RMCST based on truncation IPTW is: \n")
```

```
## Estimation of RMCST based on truncation IPTW is:

round(Delta.TR,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.3455 0.1658   4.0205   4.6705
## mu0     3.6547 0.1911   3.2802   4.0291
## Delta   0.6908 0.2631   0.1751   1.2066
```

Finally, we calculate RMCST based on overlap weighting to reproduce results in subsection 5.5.

```
Delta.OW = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                          censor.formula,Method="OW")
cat("Estimation of RMCST based on OW is: \n")

## Estimation of RMCST based on OW is:

round(Delta.OW,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6245 0.1453   4.3398   4.9093
## mu0     3.4209 0.1348   3.1566   3.6851
## Delta   1.2037 0.1991   0.8134   1.5939
```

5.7 Some other scenarios when using the program

In the above illustration, it is noted that we use the same covariates in both propensity score model and censoring model. For some scenarios, covariates in both propensity score model and censoring model are different, in this case, all we need to do is to reset different covariates in these two models. For example, assume there are 6 covariates in propensity score model and only 1 covariate (i.e., x1) in censoring model, then the following code can help us compute corresponding RMCST estimators results.

```
#update censor formula
Covariates_censor = c("x1")
censor.formula = as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                  paste(Covariates_censor,collapse="+"),sep=""))
Data = data
# 1. IPTW estimator of RMCST
Delta.IPTW1 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,
                             ps.formula,censor.formula,Method="IPTW")
cat("Estimation of RMCST based on IPTW is: \n")

## Estimation of RMCST based on IPTW is:
```

```
round(Delta.IPTW1,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.3407 0.1682   4.0110   4.6703
## mu0     3.4638 0.3887   2.7021   4.2256
## Delta   0.8768 0.4318   0.0305   1.7232

# 2. OW estimator
Delta.OW1 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="OW")
cat("Estimation of RMCST based on OW is: \n")

## Estimation of RMCST based on OW is:

round(Delta.OW1,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6150 0.1458   4.3291   4.9008
## mu0     3.4244 0.1343   3.1612   3.6877
## Delta   1.1905 0.1989   0.8007   1.5804

# 3. Symmetric trimming estimator
Delta.SW1 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Symmetric",alpha=0.1)
cat("Estimation of RMCST based on symmetric trimming is: \n")

## Estimation of RMCST based on symmetric trimming is:

round(Delta.SW1,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6606 0.1647   4.3377   4.9834
## mu0     3.5441 0.1640   3.2225   3.8656
## Delta   1.1165 0.2315   0.6628   1.5703

# 4. Asymmetric trimming estimator
Delta.AW1 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Asymmetric",q=0.01)
cat("Estimation of RMCST based on asymmetric trimming is: \n")

## Estimation of RMCST based on asymmetric trimming is:

round(Delta.AW1,4)
```

```
##       Estimate     SE CI.lower CI.upper
## mu1     4.6174 0.1559   4.3119   4.9229
## mu0     3.2547 0.1728   2.9160   3.5934
## Delta   1.3627 0.2372   0.8979   1.8276

# 5. Truncation IPTW estimator
Delta.TR1 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Truncation",q=0.1)
cat("Estimation of RMCST based on truncation IPTW is: \n")

## Estimation of RMCST based on truncation IPTW is:

round(Delta.TR1,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.3403 0.1680   4.0110   4.6696
## mu0     3.6876 0.1849   3.3252   4.0500
## Delta   0.6527 0.2598   0.1436   1.1618
```

From the above results, it can be seen that if only one covariate is used for adjusting censoring model, all four RMCST estimators are different, but not much.

Furthemore, if no covariate shall be included in the censoring model (this is precisely the case when the censoring is completely independent of any baseline covariates, or when the censoring is administrative due to end of follow-up), our program RMST.Effect.BW can also handle this case. Our strategy is to first assign a "none" covariate (make sure original data set has no covariate called none. Users can also assign another name that is not used in original data set) with 0 values in dataset, then put the none covariate in censoring model. Since estimated coefficient of the covariate "none" from Cox model will be NA, we next set the estimated coefficient (i.e., NA) to be 0 in Cox model. After that, other estimating procedure is the same.

If the propensity score model still has 6 covariates, and no covariate in censoring model, then the corresponding code of computing RMCST is

```
#assign a new covariate with 0 values in original data set
data$none = 0
Covariates_censor = c("none")
#modify censor formula
censor.formula = as.formula(paste("Surv(",SurvTime,",I(1-",Status,"))","~",
                                  paste(Covariates_censor,collapse="+"),sep=""))
Data = data
# 1. IPTW estimator of RMCST
Delta.IPTW2 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,
                             ps.formula,censor.formula,Method="IPTW")
```

```r
cat("Estimation of RMCST based on IPTW is: \n")
```

```
## Estimation of RMCST based on IPTW is:
```

```r
round(Delta.IPTW2,4)
```

```
##       Estimate     SE CI.lower CI.upper
## mu1     4.3396 0.1679   4.0105   4.6687
## mu0     3.4694 0.3894   2.7062   4.2325
## Delta   0.8702 0.4323   0.0229   1.7176
```

```r
# 2. OW estimator
Delta.OW2 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="OW")
cat("Estimation of RMCST based on OW is: \n")
```

```
## Estimation of RMCST based on OW is:
```

```r
round(Delta.OW2,4)
```

```
##       Estimate     SE CI.lower CI.upper
## mu1     4.6142 0.1458   4.3285   4.8998
## mu0     3.4267 0.1343   3.1634   3.6899
## Delta   1.1875 0.1988   0.7978   1.5772
```

```r
# 3. Symmetric trimming estimator
Delta.SW2 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Symmetric",alpha=0.1)
cat("Estimation of RMCST based on symmetric trimming is: \n")
```

```
## Estimation of RMCST based on symmetric trimming is:
```

```r
round(Delta.SW2,4)
```

```
##       Estimate     SE CI.lower CI.upper
## mu1     4.6620 0.1649   4.3388   4.9851
## mu0     3.5327 0.1643   3.2106   3.8547
## Delta   1.1293 0.2318   0.6751   1.5835
```

```r
# 4. Asymmetric trimming estimator
Delta.AW2 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Asymmetric",q=0.01)
cat("Estimation of RMCST based on asymmetric trimming is: \n")
```

```
## Estimation of RMCST based on asymmetric trimming is:
```

```
round(Delta.AW2,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.6195 0.1562   4.3134   4.9256
## mu0     3.2424 0.1731   2.9031   3.5818
## Delta   1.3770 0.2376   0.9114   1.8427

# 5. Truncation IPTW estimator
Delta.TR2 = RMST.Effect.BW(Data,L=8,Treatment,SurvTime,Status,ps.formula,
                           censor.formula,Method="Truncation",q=0.1)
cat("Estimation of RMCST based on truncation IPTW is: \n")

## Estimation of RMCST based on truncation IPTW is:

round(Delta.TR2,4)

##       Estimate     SE CI.lower CI.upper
## mu1     4.3392 0.1677   4.0105   4.6680
## mu0     3.6926 0.1845   3.3310   4.0542
## Delta   0.6467 0.2592   0.1386   1.1547
```

According to the above results, we can see that although no covariate is included in censoring model, RMCST estimators are similar to those based on 6 covariates or 1 covariate.

In summary, we observe some differences in the RMCST estimates from four different balancing weights; in particular, the SE of OW is the smallest while SE of IPTW is the largest among all scenarios. The empirical results based on this simulated data set are consistent with our theoretical results.

6 Web Appendix 6: Application of Right-Heart Catheterization Study

6.1 Dataset and Pre-processing

In this section, we analyze the Right-Heart Catheterization Study (RHC Study) as an additional illustrative data example. Introduction of this datset is available at https://hbiostat.org/data/repo/rhc.html and Connors et al.(1996). The "rhc.csv" dataset available at https://hbiostat.org/data/repo/rhc.csv will be used. The treatment variable "swang1" is binary, which takes one of "RHC" and "No RHC", indicating, respectively, treated and untreated in this illustrative example. The outcome is patient's survival time, which is defined as the difference between date of death and the study admission date. However, the outcome is subject to right censoring such that we only observed date of first occurance of last follow-up and death. The censoring indicator is "death", which takes "Yes" or "No" to denote the patient is alive or death on the previously given date. Summary statistics of swang1, survival time (denoted as survtime) and death are

```
##      swang1          survtime          death
##  Min.   :0.0000   Min.   :   2.0   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:  16.0   1st Qu.:0.000
##  Median :0.0000   Median : 166.0   Median :1.000
##  Mean   :0.3808   Mean   : 186.4   Mean   :0.649
##  3rd Qu.:1.0000   3rd Qu.: 232.0   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1943.0   Max.   :1.000
```

Next, we clean the dataset before analyzing. Although the treatment and outcome variables do not have missing data, two covariates, "adld3p" and "urin1", have missing values. Here, we simply replace those missing values with the median of the remaining obervations. In addition, we observed that several categories of the covariate "cat1" (include 9 categories of admission diagnosis) have very few samples. To avoid separation when we estimate the variance estimator through bootstrap, we will combine (i) categories "Cirrhosis", "Colon Cancer", and "Lung Cancer" into a single category "Cirrhosis or Cancer" and (ii) categories "MOSF w/Malignancy" and "MOSF w/Sepsis" into a single category "MOSF".

6.2 RMCST estimators of RHC data by balancing weights

We consider the following 50 pre-treatment covariates (32 categorical variables and 18 numerical variables) that may determine choices of treatments, that is:

```
# 1. identify column names of the pre-treatment covariates
Covariates=c("age","sex","race","edu","income","ninsclas","cat1","resp","card",
             "neuro","gastr","renal","meta","hema","seps","trauma","adld3p","das2d3pc",
             "dnr1","ca","surv2md1","aps1","scoma1","wtkilo1","temp1","meanbp1","resp1",
             "hrt1","pafi1","paco21","ph1","wblc1","hema1","sod1","pot1","crea1","bili1",
             "alb1","urin1","cardiohx","chfhx","dementhx","psychhx","chrpulhx","renalhx",
             "liverhx","gibledhx","immunhx","transhx","amihx")
```

Detailed definition of these covariates can be found in Table 1 at https://hbiostat.org/data/repo/rhc.html and will not be described here. We estimate the propensity scores using logistic regression with linear terms on all covariates (and no interactions). Web Figure 1 presents the histogram of the estimated propensity scores by each treatment group. From this histogram, we observe some degree of lack of overlap; the shape of the propensity score histogram appears similar to the moderate overlap scenario in our simulation study.

Next, we calculate the restricted mean survival time up to 8 restriction times, i.e., L=60,120,180,240,300,360,420,1943 (in days), with the first 7 representing each two months and the last one representing the end of study. The 95% confidence intervals for all estimates were obtained by the proposed closed-form variance estimators. Corresponding estimated results based on four weighting methods are shown in Web Table 1.

According to Web Table 1, it can be seen that RMCST estimators for $\triangle_{OW}(L)$ from four methods are negative and significantly different from zero except for L=1943, which includes zero in its 95% confidence interval. This could be because there are few subjects that survive through the end of the study or not censored at the end

Web Table 1: Estimated Difference in the RMCST for eight restriction times of RHC Study.

	IPTW		OW		Symmetric		Asymmetric		Truncation	
L(in days)	$\hat{\triangle}_w(L)$	95%CI	$\hat{\triangle}_w(L)$	95%CI	$\hat{\triangle}_w(L)$	95%CI	$\hat{\triangle}_w(L)$	95%CI	$\hat{\triangle}_w(L)$	95%CI
60 (2 months)	-2.26	(-3.74,-0.78)	-2.59	(-3.93,-1.25)	-2.85	(-4.26,-1.43)	-2.65	(-4.09,-1.21)	-2.59	(-3.98,-1.20)
120(4 months)	-5.18	(-8.41,-1.96)	-6.05	(-9.01,-3.08)	-6.78	(-9.91,-3.66)	-6.22	(-9.41,-3.02)	-5.80	(-8.88,-2.72
180(6 months)	-7.87	(-12.90,-2.84)	-9.51	(-14.16,-4.87)	-10.60	(-15.50,-5.70)	-9.83	(-14.84,-4.82)	-8.77	(-13.59,-3.94
240(8 months)	-10.94	(-17.76,-4.12)	-13.32	(-19.64,-6.99)	-14.84	(-21.50,-8.18)	-13.96	(-20.76,-7.15)	-12.06	(-18.62,-5.51)
300(10 months)	-13.76	(-22.34,-5.18)	-16.72	(-24.69,-8.75)	-18.80	(-27.19,-10.41)	-17.66	(-26.23,-9.09)	-15.01	(-23.27,-6.74)
360(12 months)	-16.18	(-26.59,-5.77)	-19.60	(-29.30,-9.91)	-22.48	(-32.67,-12.29)	-20.96	(-31.36,-10.57)	-17.45	(-27.50,-7.39)
420(14 months)	-18.48	(-30.81,-6.16)	-22.21	(-33.70,-10.72)	-25.83	(-37.92,-13.74)	-24.10	(-36.41,-11.79)	-19.73	(-31.66,-7.81)
1943(MFU)	-23.16	(-65.85,19.53)	-21.23	(-61.50,19.04)	-28.47	(-71.56,14.61)	-28.91	(-72.68,14.87)	-24.67	(-66.36,17.02)

Note: MFU referst to Maximum follow-up.

of the study, so standarded error of estimated survival functions for both treament and control groups are large, resulting a wide 95% confidence interval for $\hat{\triangle}_{OW}(1943)$. Nevertheless, at the 90 percentile of the survival time (i.e., $L \leq 420$), the result shows that, after adjustment for pre-treatment covariates, RHC treatment decreases the survival probabilities for the study population at clinical equipoise, which is consistent with conclusions in Connors et al.(1996)(https://jamanetwork.com/journals/jama/fullarticle/407990).

7 Web Appendix 7: Application of the Yale New Haven Hospital Hypertension data

We analyzed the observational data obtained from the Yale New Haven Health System (YNHHS) network hospitals to study the average causal effect of intravenous antihypertensives on clinical outcomes for hospitalized patients who develop severe hypertension; the study details were described elsewhere (Ghazi et al., 2023). To ensure a less heterogeneous population, we considered the analysis of 2,427 patients who were admitted for cardiovascular and cerebrovascular disease based on admission ICD-10 codes. Among them, 250 received intravenous antihypertensives and the remaining 2,177 were not treated with antihypertensives.

We used the time (in hours) from treatment to development of acute kidney injury (AKI) as a survival outcome, and censored patients without events at 30 days (setting 717 hours as the end of follow-up). Approximately 75% patients are administratively censored due to maximum follow-up and censoring is considered to be independent of covariates. The PS was estimated using logistic regression adjusting for 25 baseline characteristics including demographics, hospital location (e.g., surgical vs medical unit), comorbidities, lab information on admission and relevant medications that affect blood pressure. Unweighted baseline patient characteristics and weighted baseline patient characteristics for combined (under IPTW) and overlap (under OW) target populations are shown in Web Tables 2 and 3.

The baseline comparison suggests that treated patients are slightly older, more likely to be admitted to the surgical ward, have higher mean arterial pressure on admission and at time of severe hypertension development, lower BMI, but are less likely to be on relevant medications that affect blood pressure before the onset of severe inpatient hypertension. The target populations for IPTW and OW are roughly similar, except that, for example,

the overlap population has slightly higher blood pressure, slightly lower elixhauser score and BMI on average. The histogram of the estimated PS (Web Figure 2) shows reasonable overlap between groups, but the distributions of PS are slightly shifted toward 0. We estimated the causal effect of intravenous antihypertensives using IPTW and OW. Since the censoring is only due to end of follow-up, the censoring weights were estimated from Cox regression within each group without any covariates; in this special case, the censoring survival function estimates are equivalent to those given by the nonparametric Nelson-Aalen estimator.

Web Table 4 presents the causal effects in RMCST for developing AKI at 5 restriction times (L=168,336,504,672,717 hours), with the first 4 representing each whole week and the last one representing the end of study. The 95% confidence intervals for all estimates were obtained by the proposed closed-form variance estimators. The results consistently suggest that intravenous antihypertensives lead to earlier development of AKI (i.e., intravenous antihypertensive treatment is harmful as it results in earlier occurrence of the outcome), although the 95% confidence intervals include zero. However, a noticeable difference between IPTW and OW is that the latter is associated with a narrower confidence interval at each time point. Web Figure 3 presents the estimated difference in RMCST under IPTW and OW, respectively. From this figure, we confirm the greater efficiency of OW by emphasizing the target population at clinical equipoise, although the point estimates under OW are slightly smaller in magnitude compared with IPTW. The corresponding RMCST curves over follow-up time are also presented in Web Figure 3.

References

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10(4), 1100–1120.

Cole, S. R. and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6), 656–664.

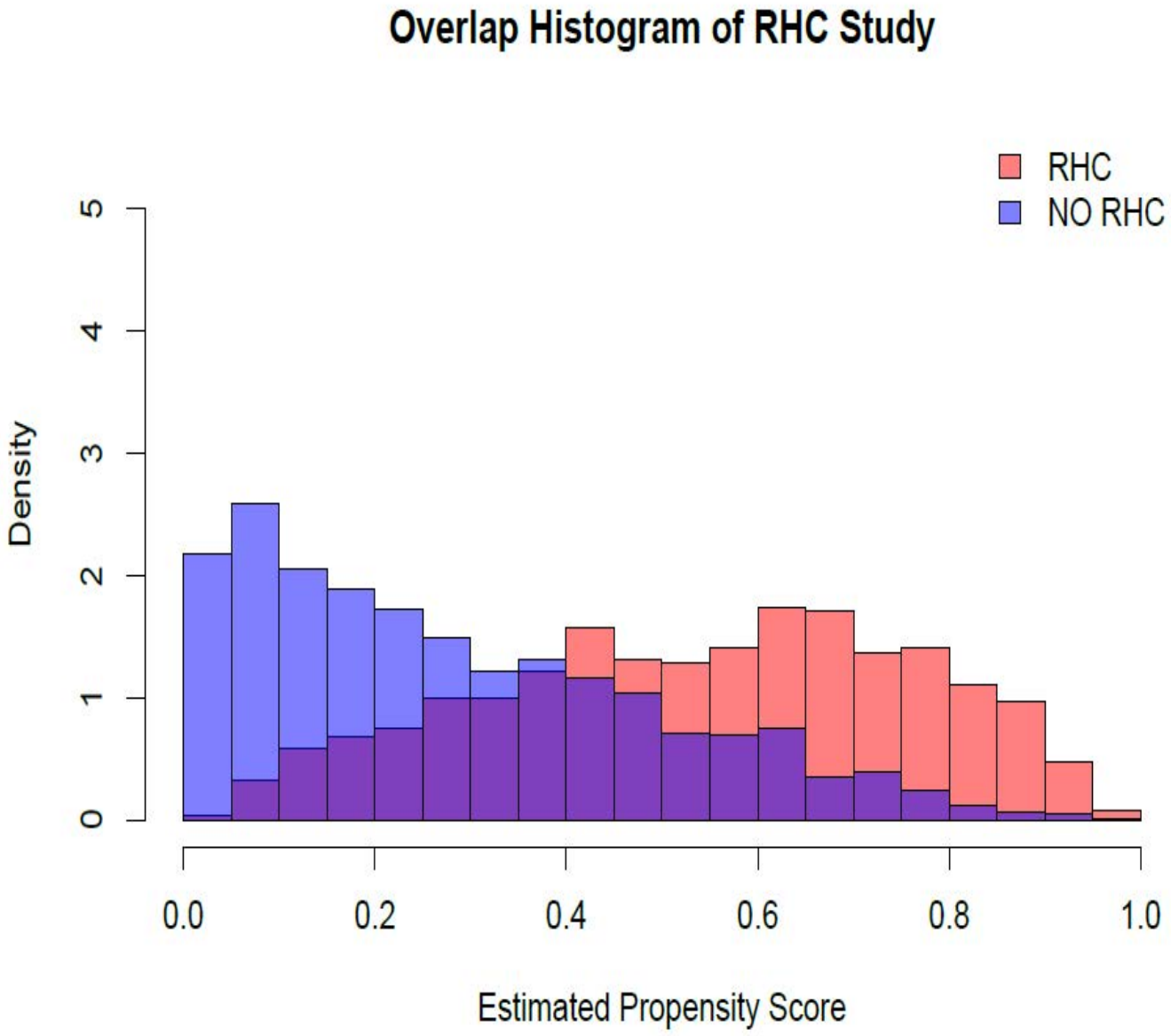
Connors, A.F., Speroff, T., Dawson, N.V., et al. (1996). The Effectiveness of Right Heart Catheterization in the Initial Care of Critically III Patients. *JAMA*, 276(11), 889–897.

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

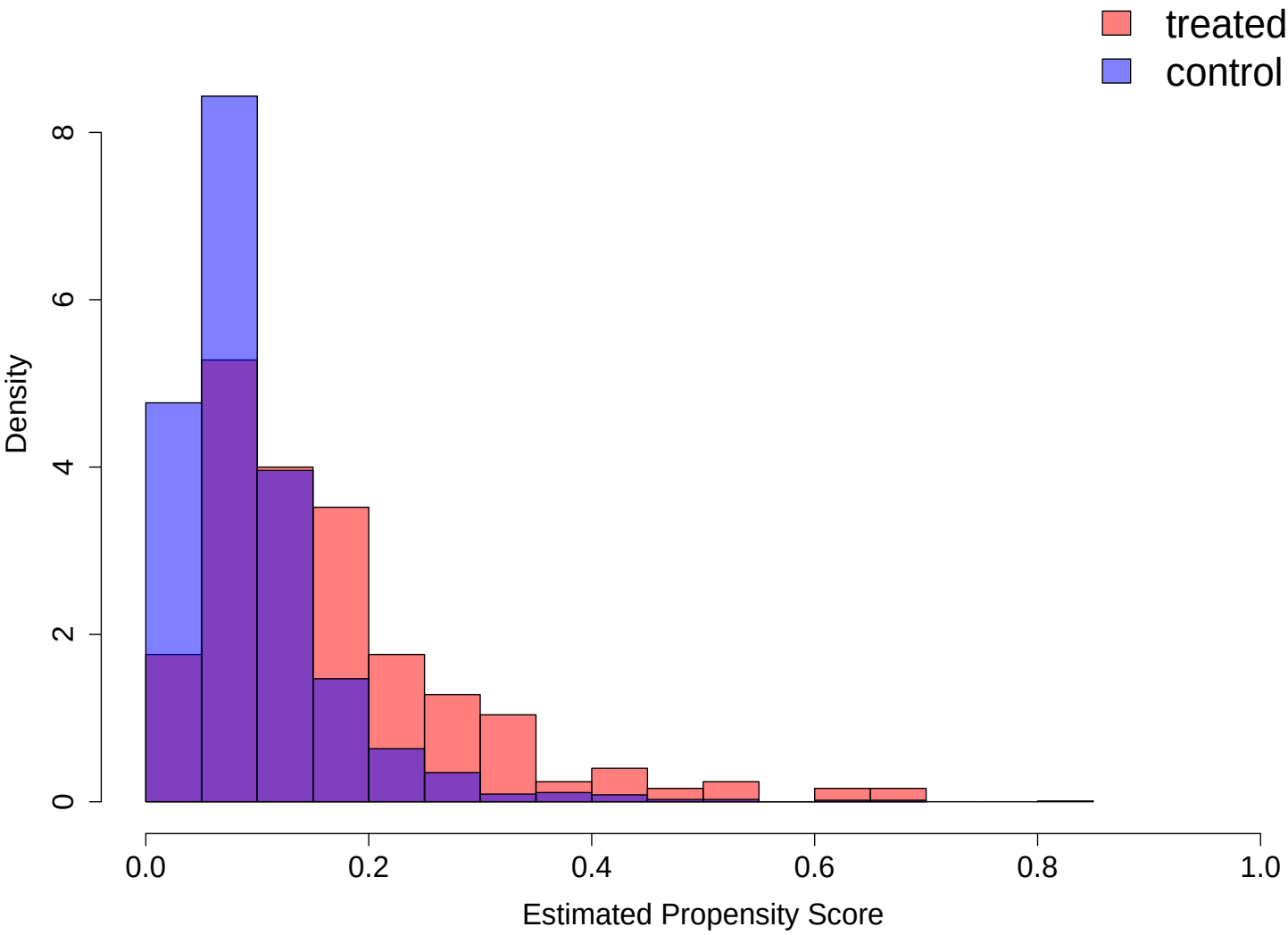
Ghazi, L., F. Li, M. Simonov, Y. Yamamoto, J.T. Nugent, J.H. Greenberg, C.Y. Bakhoum, A.J. Peixoto, and F. P. (2023). Wilson. Effect of intravenous antihypertensives on outcomes of severe hypertension in hospitalized patients without acute target organ damage. *Journal of Hypertension*, 41(2), 288–294.

Zhang, M. and Schaubel, D.E. (2012). Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics*, 68(4), 999–1009.

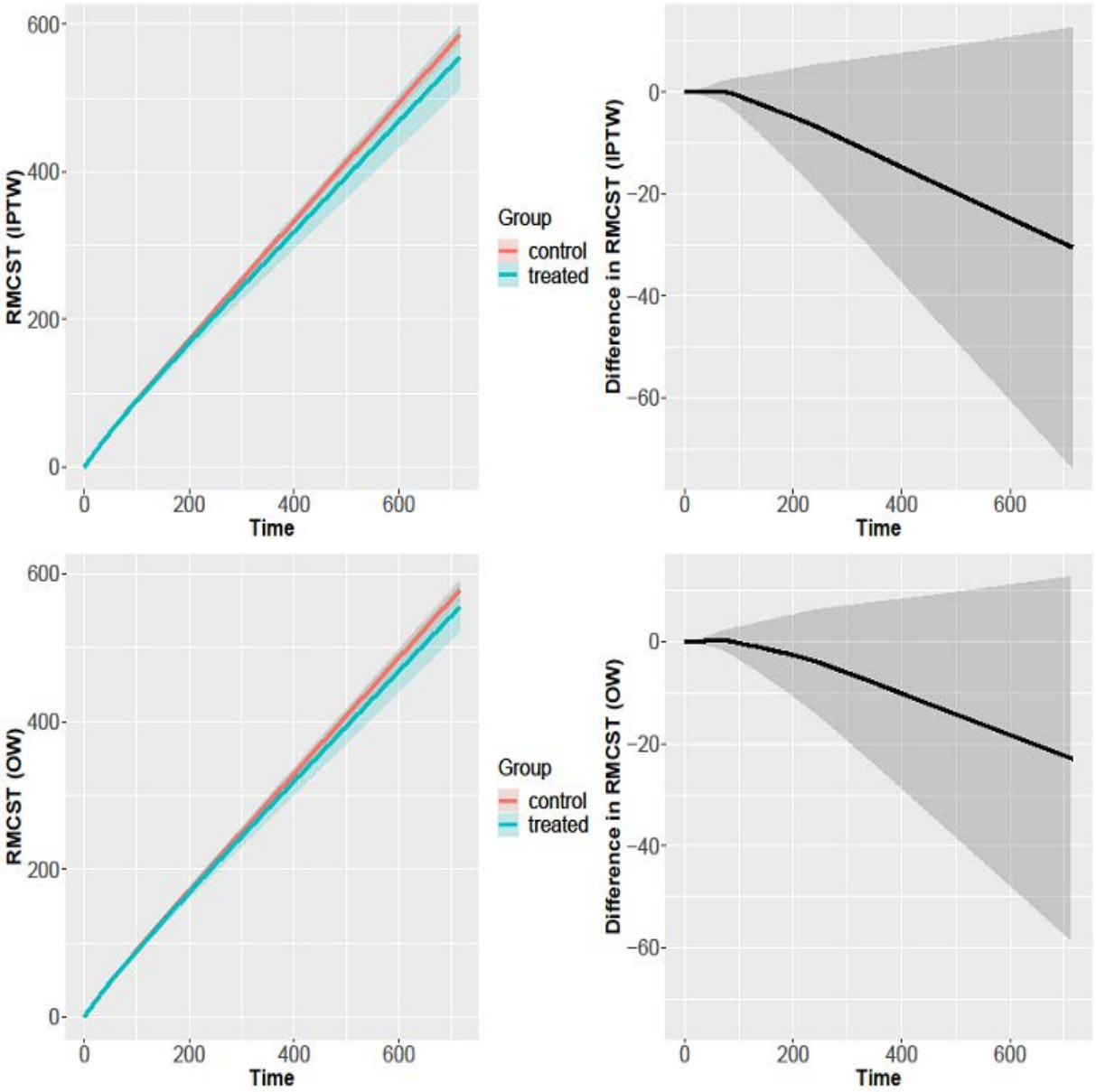
8 Web Appendix 8: Web figures and tables



Web Figure 1: Histogram of the estimated PS based on linear terms on all covariates (and no interactions) in the RHC study.



Web Figure 2: Histogram of the estimated PS based on 25 baseline characteristics in the Yale New Haven Hospital Hypertension data of empirical study.



Web Figure 3: Results obtained from the empirical study of the Yale New Haven Hospital Hypertension Data, 2016-2020. The left panel show estimated restricted mean counterfactual survival times to develop acute kidney injury between intravenous antihypertensives (blue lines) and control (red lines) patients in the first 30 days after treatment, derived using IPTW (upper plot) and OW (lower plot). The right panel show estimated differences in restricted mean counterfactual survival times, derived using IPTW (upper plot) and OW (lower plot). The 95% confidence intervals (shaded region) were obtained with the proposed closed-form variance estimators. We used grid points between 0 and 715 hours with an increment of 5 hours.

Web Table 2: Unweighted baseline characteristics by treatment groups and absolute standardized mean differences (ASDs) for each characteristic from the empirical study of the Yale New Haven Hospital Hypertension Data, 2016-2020. Mean (standard deviation) and proportion (standard deviation) are presented for continuous and binary characteristics, respectively. An asterisk indicates that the ASD exceeds the usual threshold 0.1.

	Characteristics	Treated($n = 250$)	Untreated($n = 2177$)	ASD
Demographics				
	Age	75.07 (15.12)	73.76 (14.98)	0.09
	Black	0.22 (0.42)	0.19 (0.39)	0.07
	Male sex	0.41 (0.49)	0.46 (0.5)	0.09
	Hispanic or Latino	0.07 (0.26)	0.09 (0.29)	0.07
Location				
	Surgical ward	1.28 (0.45)	1.2 (0.4)	0.21*
Comorbidities				
	Elixhauser score	5.22 (4.06)	6.03 (4.19)	0.2*
Labs on admission				
	Sodium	139.2 (3.87)	139.03 (3.8)	0.04
	Potassium	4.15 (0.54)	4.21 (0.56)	0.1*
	Chloride	101.86 (4.98)	102.11 (5.02)	0.05
	Bicarbonate	24.66 (3.66)	24.59 (3.81)	0.02
	Blood urea nitrogen	25.53 (16.69)	25.24 (15.94)	0.02
	eGFR	61.69 (28.63)	62.37 (28.45)	0.02
	White blood cell count	9.16 (3.64)	9.23 (3.88)	0.02
	Platelet count	233.24 (82.17)	225.43 (82.23)	0.09
	Hemoglobin	12.42 (2.02)	12.13 (2.06)	0.14*
	Hematocrit	38.13 (5.63)	37.36 (5.93)	0.13*
Medications given 6 hours before onset of severe inpatient hypertension				
	NSAID	0 (0.06)	0.01 (0.09)	0.05
	Crystalloid	0.08 (0.28)	0.12 (0.33)	0.13*
	Narcotics	0.07 (0.25)	0.09 (0.29)	0.1
	Sedatives	0.02 (0.14)	0.04 (0.19)	0.11*
	Steroids	0.01 (0.09)	0.02 (0.15)	0.13*
Admission characteristics				
	Mean arterial pressure (MAP)	117.91 (22.36)	107.55 (18.3)	0.51*
	Heart rate	82.25 (21.98)	81.04 (21.86)	0.05
	BMI	27.08 (7.01)	29.03 (7.46)	0.27*
	MAP at time of severe inpatient hypertension development	129.98 (14.76)	122.83 (11.82)	0.53*

Note: NSAID refers to non-steroidal anti-inflammatory medications.

Web Table 3: Baseline patient characteristics among the propensity score weighted populations and absolute standardized mean differences (ASDs) for each characteristic from the empirical study of the Yale New Haven Hospital Hypertension Data, 2016-2020. Mean (standard deviation) and proportion (standard deviation) are presented for continuous and binary characteristics, respectively.

		IPTW			OW		
	Characteristics	Treated($n = 250$)	Untreated($n = 2177$)	ASD	Treated($n = 250$)	Untreated($n = 2177$)	ASD
Demographics							
	Age	73.77 (15.75)	73.9 (15.04)	0.01	74.9 (15.32)	74.9 (15.16)	0
	Black	0.21 (0.41)	0.2 (0.4)	0.04	0.22 (0.41)	0.22 (0.41)	0
	Male sex	0.47 (0.5)	0.45 (0.5)	0.03	0.42 (0.49)	0.42 (0.49)	0
	Hispanic or Latino	0.08 (0.27)	0.09 (0.29)	0.04	0.08 (0.26)	0.08 (0.26)	0
Location							
	Surgical ward	1.23 (0.42)	1.2 (0.4)	0.06	1.27 (0.45)	1.27 (0.45)	0
Comorbidities							
	Elixhauser score	5.84 (4.2)	5.93 (4.2)	0.02	5.3 (4.07)	5.3 (4.13)	0
Labs on admission							
	Sodium	139.06 (3.85)	139.05 (3.8)	0.00	139.25 (3.85)	139.25 (3.85)	0
	Potassium	4.22 (0.56)	4.2 (0.56)	0.04	4.16 (0.54)	4.16 (0.57)	0
	Chloride	101.96 (5.11)	102.08 (5.02)	0.02	101.94 (4.99)	101.94 (4.99)	0
	Bicarbonate	24.45 (3.8)	24.6 (3.82)	0.04	24.62 (3.67)	24.62 (3.79)	0
	Blood urea nitrogen	26.46 (16.71)	25.29 (16.01)	0.07	25.66 (16.82)	25.66 (16.56)	0
	eGFR	61.69 (29.73)	62.29 (28.43)	0.02	61.75 (28.79)	61.75 (28.21)	0
	White blood cell count	9.5 (4.04)	9.23 (3.87)	0.07	9.22 (3.7)	9.22 (3.8)	0
	Platelet count	222.01 (80.1)	226.24 (82.24)	0.05	232.15 (81.88)	232.15 (82.71)	0
	Hemoglobin	12.16 (2.07)	12.15 (2.06)	0.00	12.35 (2.03)	12.35 (2.04)	0
	Hematocrit	37.53 (5.81)	37.44 (5.93)	0.02	37.97 (5.68)	37.97 (5.88)	0
Medications given 6 hours before onset of severe inpatient hypertension							
	NSAID	0 (0.07)	0.01 (0.09)	0.04	0 (0.07)	0 (0.07)	0
	Crystalloid	0.11 (0.32)	0.12 (0.33)	0.02	0.09 (0.29)	0.09 (0.29)	0
	Narcotics	0.09 (0.28)	0.09 (0.29)	0.01	0.07 (0.26)	0.07 (0.26)	0
	Sedatives	0.03 (0.17)	0.04 (0.19)	0.03	0.02 (0.15)	0.02 (0.15)	0
	Steroids	0.02 (0.16)	0.02 (0.15)	0.02	0.01 (0.1)	0.01 (0.1)	0
Admission characteristics							
	Mean arterial pressure (MAP)	106.79 (20.33)	108.77 (19.34)	0.10	115.47 (21.34)	115.47 (20.88)	0
	Heart rate	80.9 (24.03)	81.19 (21.86)	0.01	81.9 (22.15)	81.9 (21.5)	0
	BMI	29.08 (7.47)	28.82 (7.4)	0.03	27.37 (7.06)	27.37 (6.61)	0
	MAP at time of severe inpatient hypertension development	123.78 (13.04)	123.74 (12.99)	0.00	128.39 (13.8)	128.39 (15.32)	0

Note: NSAID refers to non-steroidal anti-inflammatory medications.

Web Table 4: Estimated Difference in the RMCST to develop acute kidney injury Between Patients Receiving Intravenous Antihypertensives and Patients not Receiving Treatment in Five Different Restriction Times, Yale New Haven Hospital Hypertension Data, 2016-2020.

	IPTW		OW	
Restriction Time (in hours)	Estimated Difference in RMCST	95% Confidence Interval	Estimated Difference in RMCST	95% Confidence Interval
$L = 168$ hours (Week 1)	-3.61	(-11.02,3.81)	-1.81	(-8.12,4.50)
$L = 336$ hours (Week 2)	-11.46	(-29.57,6.65)	-7.48	(-22.51,7.56)
$L = 504$ hours (Week 3)	-19.96	(-49.09,9.16)	-14.39	(-38.52,9.74)
$L = 672$ hours (Week 4)	-28.34	(-68.56,11.88)	-21.19	(-54.55,12.17)
$L = 717$ hours (Maximum follow-up)	-30.54	(-73.67,12.60)	-22.97	(-58.76,12.81)

Web Table 5: True Values of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ for Three RMCST at Strong Moderate and Weak Levels of Covariates Overlap (Simulation Study).

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	1.504	2.694	3.564	1.777	3.805	6.129	-0.273	-1.112	-2.565
	3	1.520	2.709	3.517	1.778	3.809	6.132	-0.259	-1.100	-2.616
	5	1.527	2.717	3.493	1.779	3.810	6.134	-0.251	-1.093	-2.641
IPTW	1	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
	3	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
	5	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
Symmetric Trimming										
	1	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
$\alpha = 0.05$	3	1.514	2.702	3.536	1.778	3.808	6.132	-0.264	-1.105	-2.595
	5	1.526	2.715	3.497	1.779	3.810	6.133	-0.253	-1.095	-2.636
	1	1.498	2.688	3.580	1.777	3.804	6.128	-0.278	-1.115	-2.548
$\alpha = 0.1$	3	1.522	2.710	3.512	1.778	3.809	6.133	-0.257	-1.099	-2.621
	5	1.530	2.720	3.484	1.779	3.811	6.134	-0.249	-1.091	-2.650
	1	1.500	2.690	3.576	1.777	3.804	6.128	-0.277	-1.115	-2.552
$\alpha = 0.15$	3	1.526	2.716	3.496	1.779	3.810	6.133	-0.252	-1.095	-2.637
	5	1.532	2.722	3.476	1.779	3.811	6.135	-0.247	-1.089	-2.658
Asymmetric Trimming										
	1	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
$q = 0$	3	1.499	2.689	3.580	1.777	3.804	6.128	-0.278	-1.115	-2.548
	5	1.501	2.690	3.574	1.777	3.805	6.129	-0.276	-1.114	-2.554
	1	1.506	2.694	3.555	1.777	3.807	6.131	-0.271	-1.112	-2.576
$q = 0.01$	3	1.519	2.707	3.513	1.778	3.810	6.134	-0.259	-1.103	-2.621
	5	1.528	2.716	3.487	1.779	3.811	6.135	-0.251	-1.095	-2.648
	1	1.518	2.704	3.516	1.778	3.809	6.134	-0.261	-1.105	-2.618
$q = 0.05$	3	1.530	2.717	3.478	1.779	3.812	6.136	-0.250	-1.094	-2.658
	5	1.534	2.723	3.466	1.780	3.812	6.136	-0.246	-1.090	-2.670
Truncation										
	1	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
	3	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547
	5	1.498	2.688	3.581	1.776	3.804	6.128	-0.278	-1.115	-2.547

Note: Values of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ are evaluated using a sufficiently large sample with 1000,000 observations. In this additional simulation, PS term is omitted in generating the potential survival outcomes in the simulation design of the main manuscript.

Web Table 6: Bias of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study).

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	0.11	0.28	0.76	0.01	0.05	0.02	-0.51	-0.52	-1.02
	3	0.04	0.21	0.62	0.03	0.18	0.26	-0.05	0.11	-0.22
	5	0.01	0.20	0.69	0.04	0.14	0.30	0.21	0.00	-0.21
IPTW	1	0.11	0.31	0.84	0.01	0.03	-0.05	-0.51	-0.64	-1.30
	3	0.06	0.24	0.69	0.05	0.21	0.10	0.02	0.15	-0.72
	5	-0.73	-1.11	-1.21	-0.08	-0.08	-0.30	3.42	2.40	0.99
Symmetric Trimming										
	1	0.11	0.31	0.84	0.01	0.03	-0.05	-0.52	-0.64	-1.30
$\alpha = 0.05$	3	0.09	0.34	0.83	0.02	0.19	0.27	-0.40	-0.17	-0.48
	5	0.05	0.32	0.98	0.05	0.14	0.27	0.09	-0.32	-0.69
	1	0.12	0.31	0.80	0.01	0.04	-0.02	-0.56	-0.59	-1.18
$\alpha = 0.1$	3	0.03	0.23	0.79	0.01	0.12	0.27	-0.13	-0.15	-0.43
	5	-0.05	0.11	0.70	0.07	0.22	0.50	0.77	0.49	0.24
	1	0.15	0.33	0.77	0.02	0.05	0.00	-0.72	-0.64	-1.09
$\alpha = 0.15$	3	0.04	0.24	0.85	0.02	0.16	0.30	-0.12	-0.02	-0.45
	5	0.07	0.27	0.90	0.04	0.20	0.37	-0.20	0.03	-0.31
Asymmetric Trimming										
	1	0.29	0.42	0.80	0.05	0.10	0.04	-1.23	-0.69	-1.03
$q = 0$	3	1.42	2.11	2.63	0.26	0.64	0.81	-5.96	-2.91	-1.75
	5	2.25	3.73	4.57	0.46	1.04	1.50	-9.32	-5.45	-2.80
	1	0.15	0.35	0.82	0.02	0.07	0.05	-0.70	-0.59	-1.01
$q = 0.01$	3	0.19	0.49	1.07	0.01	0.14	0.26	-1.02	-0.71	-0.82
	5	0.02	0.25	0.92	0.08	0.14	0.33	0.39	-0.15	-0.45
	1	0.09	0.28	0.80	-0.02	0.02	0.05	-0.64	-0.59	-0.97
$q = 0.05$	3	-0.01	0.13	0.65	0.06	0.21	0.40	0.45	0.41	0.07
	5	0.02	0.28	1.09	0.01	0.21	0.43	-0.05	0.04	-0.42
Truncation										
	1	-0.06	-0.02	0.30	-0.03	-0.04	-0.16	0.14	-0.10	-0.79
$q = 0.025$	3	-0.23	-0.40	-0.42	-0.02	0.06	-0.10	1.12	1.19	0.35
	5	-0.79	-1.28	-1.50	-0.07	-0.07	-0.29	3.78	2.83	1.41
	1	-0.24	-0.37	-0.26	-0.06	-0.12	-0.28	0.89	0.50	-0.30
$q = 0.05$	3	-0.54	-1.07	-1.53	-0.08	-0.07	-0.32	2.40	2.33	1.39
	5	-0.97	-1.75	-2.33	-0.11	-0.16	-0.42	4.49	3.67	2.26
	1	-0.64	-1.13	-1.40	-0.14	-0.28	-0.53	2.51	1.78	0.69
$q = 0.1$	3	-1.20	-2.44	-3.72	-0.23	-0.39	-0.78	5.00	4.56	3.36
	5	-1.58	-3.18	-4.77	-0.27	-0.53	-0.95	6.78	5.84	4.43

Web Table 7: Relative Efficiency of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study).

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	1.00	1.00	1.03	1.03	1.08	1.10	1.01	1.05	1.08
	3	1.61	2.09	2.72	1.74	1.93	1.82	1.70	2.10	2.25
	5	3.55	4.87	6.34	3.81	3.64	2.92	3.73	4.40	4.41
IPTW	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Symmetric Trimming										
	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha = 0.05$	3	1.49	1.81	2.10	1.37	1.44	1.35	1.48	1.70	1.71
	5	2.78	3.83	4.80	2.94	2.75	2.13	2.87	3.36	3.25
	1	1.00	1.00	1.01	1.00	1.01	1.01	1.00	1.01	1.01
$\alpha = 0.1$	3	1.38	1.82	2.32	1.44	1.57	1.51	1.44	1.80	1.92
	5	2.84	3.95	5.08	2.86	2.77	2.24	2.85	3.33	3.36
	1	1.00	1.01	1.03	1.00	1.03	1.03	1.00	1.02	1.03
$\alpha = 0.15$	3	1.21	1.60	2.22	1.48	1.66	1.62	1.30	1.65	1.88
	5	2.67	3.70	4.75	2.66	2.55	2.08	2.74	3.15	3.09
Asymmetric Trimming										
	1	0.99	1.00	0.99	1.01	1.01	1.01	0.99	1.01	1.01
$q = 0$	3	1.00	1.04	1.08	0.98	0.99	0.99	0.96	0.99	1.01
	5	0.95	0.95	0.96	0.94	0.94	0.92	0.91	0.90	0.90
	1	0.92	0.93	0.97	1.00	1.06	1.06	0.95	0.99	1.02
$q = 0.01$	3	1.35	1.61	2.03	1.36	1.45	1.43	1.38	1.57	1.68
	5	2.51	3.40	4.37	2.94	2.64	2.09	2.68	3.05	3.03
	1	0.77	0.74	0.79	0.84	0.96	1.01	0.79	0.85	0.93
$q = 0.05$	3	1.02	1.36	1.90	1.25	1.46	1.45	1.09	1.37	1.61
	5	1.90	2.79	3.80	2.00	1.99	1.77	2.02	2.42	2.57
Truncation										
	1	1.00	1.01	1.02	1.02	1.04	1.03	1.01	1.02	1.03
$q = 0.025$	3	1.19	1.31	1.43	1.12	1.13	1.10	1.17	1.23	1.25
	5	1.13	1.16	1.14	1.14	1.05	1.03	1.13	1.11	1.09
	1	1.00	1.01	1.03	1.03	1.07	1.07	1.01	1.04	1.06
$q = 0.05$	3	1.35	1.54	1.73	1.31	1.31	1.24	1.35	1.44	1.44
	5	1.39	1.52	1.53	1.31	1.18	1.12	1.36	1.38	1.35
	1	0.99	1.02	1.06	1.05	1.11	1.12	1.02	1.07	1.10
$q = 0.1$	3	1.57	1.89	2.23	1.57	1.62	1.49	1.61	1.80	1.80
	5	2.17	2.57	2.80	1.96	1.73	1.52	2.11	2.21	2.15

Web Table 8: Coverage Rates (%) of the 95% Confidence Intervals for $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study).

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	96.3	95.9	95.8	95.1	93.3	92.0	95.4	94.9	95.4
	3	95.2	95.0	94.4	95.0	94.6	93.4	94.8	95.4	93.8
	5	95.1	93.8	93.7	93.6	94.5	93.6	93.6	95.0	93.4
IPTW	1	96.2	96.3	96.2	95.0	92.9	91.1	95.6	95.3	94.6
	3	92.7	92.6	89.3	90.5	91.0	87.4	90.8	90.4	86.8
	5	88.0	82.4	76.4	84.9	85.7	84.7	83.6	78.9	78.7
Symmetric Trimming										
	1	96.2	96.2	96.2	95.0	93.0	91.1	95.6	95.3	94.6
$\alpha = 0.05$	3	95.6	96.5	95.0	92.1	92.3	91.2	94.3	93.9	93.0
	5	93.8	93.4	93.0	91.9	93.4	91.0	93.8	93.5	92.1
	1	95.9	96.0	96.0	95.1	93.1	91.2	95.8	95.3	94.9
$\alpha = 0.1$	3	94.4	95.5	94.9	93.5	93.5	93.0	94.8	95.1	94.3
	5	94.5	94.0	93.2	92.4	92.9	91.7	93.9	93.2	92.8
	1	96.2	96.2	96.2	94.5	92.9	91.7	95.5	94.8	94.9
$\alpha = 0.15$	3	94.4	94.0	93.9	94.7	95.0	93.6	93.9	94.6	94.4
	5	95.2	94.9	93.3	93.2	94.7	92.4	94.4	94.0	92.8
Asymmetric Trimming										
	1	95.8	96.0	96.2	95.0	92.9	91.1	95.0	95.0	94.6
$q = 0$	3	88.4	91.5	90.9	89.3	90.0	86.9	90.2	89.6	86.5
	5	80.4	83.5	82.0	80.9	82.5	83.0	81.1	78.0	77.4
	1	95.9	95.3	95.6	95.2	93.4	91.9	95.3	95.5	94.4
$q = 0.01$	3	93.1	93.7	93.6	92.5	93.6	92.0	93.4	93.8	92.5
	5	92.2	92.4	92.1	92.6	93.1	92.6	92.9	92.3	91.3
	1	95.4	94.7	93.1	94.7	93.8	93.1	94.8	94.0	93.7
$q = 0.05$	3	92.7	93.3	93.5	94.6	94.2	94.0	93.6	93.8	93.7
	5	93.9	94.0	93.0	92.4	92.9	93.3	93.6	93.5	92.8
Truncation										
	1	96.0	96.4	95.8	95.4	93.2	91.7	95.5	95.6	95.0
$q = 0.025$	3	94.1	93.7	90.4	91.7	92.0	88.1	92.7	91.7	88.1
	5	88.5	82.9	76.8	85.0	86.0	84.9	84.3	79.3	79.3
	1	96.2	96.6	96.4	95.6	93.4	91.8	95.2	95.6	95.3
$q = 0.05$	3	94.7	93.8	91.2	92.2	92.6	89.5	93.6	92.3	89.6
	5	90.5	84.9	78.7	86.3	87.4	85.7	86.5	81.9	81.3
	1	95.7	95.5	96.0	95.6	93.6	92.7	95.1	95.5	95.7
$q = 0.1$	3	94.3	91.6	88.5	93.8	92.9	91.0	93.7	92.0	91.4
	5	92.2	86.4	79.7	89.7	91.5	88.3	89.1	86.6	86.4

Web Table 9: Bias of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study) when Censoring Process is Generated from Log-logistic AFT Model but Estimated by Cox Model.

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	-0.04	0.05	0.58	0.06	0.13	0.11	0.63	0.32	-0.54
	3	-0.02	0.04	0.51	0.07	0.17	0.17	0.60	0.49	-0.29
	5	-0.01	0.17	0.74	0.07	0.16	0.24	0.57	0.15	-0.43
IPTW	1	-0.07	0.01	0.60	0.07	0.14	0.10	0.82	0.45	-0.61
	3	-0.13	-0.38	-0.38	0.10	0.20	0.14	1.33	1.62	0.87
	5	-0.97	-2.27	-3.82	-0.13	-0.06	-0.26	4.40	5.27	4.74
Symmetric Trimming										
	1	-0.07	0.01	0.60	0.07	0.14	0.10	0.81	0.44	-0.62
$\alpha = 0.05$	3	-0.03	0.00	0.44	0.07	0.23	0.24	0.66	0.77	-0.02
	5	-0.05	0.21	0.89	0.06	0.13	0.27	0.73	-0.05	-0.57
	1	-0.05	0.01	0.55	0.07	0.14	0.12	0.74	0.45	-0.48
$\alpha = 0.1$	3	-0.07	-0.01	0.43	0.04	0.13	0.18	0.68	0.48	-0.15
	5	-0.06	0.13	0.79	0.08	0.15	0.29	1.00	0.21	-0.36
	1	0.00	0.07	0.53	0.07	0.14	0.11	0.45	0.30	-0.49
$\alpha = 0.15$	3	-0.03	0.12	0.53	0.05	0.13	0.19	0.54	0.17	-0.25
	5	0.08	0.28	0.95	0.08	0.19	0.36	0.13	-0.04	-0.43
Asymmetric Trimming										
	1	0.12	0.14	0.52	0.11	0.21	0.21	0.05	0.39	-0.23
$q = 0$	3	1.24	1.51	1.55	0.29	0.61	0.81	-4.80	-1.55	-0.23
	5	2.06	2.62	2.00	0.38	1.03	1.53	-8.76	-2.83	0.87
	1	0.03	0.11	0.42	0.07	0.15	0.18	0.29	0.24	-0.14
$q = 0.01$	3	0.10	0.25	0.71	0.05	0.15	0.18	-0.22	-0.10	-0.55
	5	-0.05	0.16	0.86	0.09	0.15	0.31	0.96	0.12	-0.42
	1	0.00	0.17	0.57	0.03	0.10	0.13	0.22	-0.06	-0.47
$q = 0.05$	3	-0.04	0.14	0.66	0.09	0.20	0.28	0.87	0.33	-0.22
	5	0.04	0.33	1.08	0.07	0.25	0.44	0.27	0.04	-0.41
Truncation										
	1	-0.23	-0.31	0.05	0.03	0.05	-0.05	1.42	0.93	-0.18
$q = 0.025$	3	-0.38	-0.88	-1.13	0.02	0.03	-0.16	2.15	2.23	1.20
	5	-0.99	-2.31	-3.88	-0.12	-0.07	-0.31	4.57	5.32	4.71
	1	-0.41	-0.67	-0.53	-0.01	-0.03	-0.19	2.16	1.51	0.29
$q = 0.05$	3	-0.68	-1.53	-2.14	-0.04	-0.10	-0.41	3.40	3.33	2.03
	5	-1.13	-2.54	-4.18	-0.13	-0.14	-0.43	5.27	5.67	4.86
	1	-0.80	-1.43	-1.70	-0.08	-0.20	-0.47	3.78	2.76	1.25
$q = 0.1$	3	-1.33	-2.84	-4.16	-0.18	-0.40	-0.90	6.02	5.49	3.69
	5	-1.66	-3.57	-5.57	-0.23	-0.46	-0.98	7.50	7.04	5.47

Note: In this additional simulation, true values of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ are the same as those in Web Table 5. PS term is omitted in generating the potential survival outcomes in the simulation design of the main manuscript. The censoring model is generated from the log-logistic AFT model: $\log(C^a) = -1.6 + \boldsymbol{\theta}_a^T \mathbf{X} + \epsilon^a$, where $\boldsymbol{\theta}_a = \boldsymbol{\theta} = (-0.3, 0.5, 0.5, 0.2, -0.4, -0.5)^T$ and ϵ^a follows standard logistic distribution.

Web Table 10: Relative Efficiency of $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study) when Censoring Process is Generated from Log-logistic AFT Model but Estimated by Cox Model.

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	1.00	1.05	1.19	1.01	1.03	1.05	1.00	1.05	1.12
	3	1.68	2.52	3.39	1.63	1.72	1.89	1.67	2.26	2.62
	5	3.06	3.73	4.15	4.43	3.93	3.82	3.68	4.11	4.11
IPTW	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Symmetric Trimming										
	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha = 0.05$	3	1.53	1.92	2.23	1.32	1.44	1.58	1.45	1.77	1.91
	5	2.40	2.76	2.96	3.55	2.99	2.82	2.85	3.02	2.90
	1	1.00	1.01	1.03	1.00	1.01	1.02	1.00	1.01	1.01
$\alpha = 0.1$	3	1.45	2.24	2.98	1.34	1.42	1.57	1.41	1.95	2.28
	5	2.44	2.89	3.41	3.43	3.14	3.00	2.85	3.16	3.25
	1	1.00	1.02	1.06	0.99	1.00	1.01	0.99	1.01	1.03
$\alpha = 0.15$	3	1.28	1.99	2.99	1.32	1.35	1.47	1.29	1.77	2.09
	5	2.31	2.92	3.63	3.06	2.85	2.73	2.68	2.99	3.10
Asymmetric Trimming										
	1	0.99	1.01	1.05	1.00	1.01	1.03	0.99	1.02	1.03
$q = 0$	3	0.99	1.03	1.06	0.98	0.99	1.02	0.95	0.97	1.00
	5	0.94	0.95	0.95	0.94	0.94	0.95	0.92	0.91	0.92
	1	0.92	0.96	1.09	0.97	0.99	1.02	0.95	0.98	1.03
$q = 0.01$	3	1.42	1.93	2.49	1.31	1.38	1.53	1.37	1.75	1.99
	5	2.20	2.54	2.74	3.49	2.95	2.84	2.68	2.78	2.71
	1	0.78	0.87	1.13	0.82	0.86	0.88	0.81	0.89	1.00
$q = 0.05$	3	1.11	1.69	2.69	1.19	1.27	1.39	1.13	1.53	1.92
	5	1.62	2.12	2.71	2.25	2.05	2.00	1.98	2.23	2.38
Truncation										
	1	1.00	1.03	1.09	1.01	1.00	1.01	1.00	1.02	1.04
$q = 0.025$	3	1.20	1.27	1.30	1.09	1.17	1.24	1.16	1.23	1.27
	5	1.06	1.04	1.04	1.14	1.08	1.08	1.09	1.06	1.06
	1	1.01	1.06	1.14	1.01	1.01	1.01	1.01	1.03	1.07
$q = 0.05$	3	1.38	1.57	1.63	1.27	1.33	1.41	1.33	1.46	1.52
	5	1.21	1.18	1.15	1.46	1.32	1.29	1.33	1.25	1.22
	1	1.01	1.10	1.24	1.02	1.01	1.01	1.02	1.05	1.11
$q = 0.1$	3	1.65	2.07	2.27	1.51	1.59	1.68	1.59	1.87	1.99
	5	1.80	1.73	1.58	2.35	2.09	2.03	2.04	1.91	1.77

Web Table 11: Coverage Rates (%) of the 95% Confidence Intervals for $\mu_w^{(1)}(L)$, $\mu_w^{(0)}(L)$ and $\triangle_w(L)$ in the Presence of Increasing Strong Tails in the Propensity Score Distribution (Simulation Study) when Censoring Process is Generated From Log-logistic AFT Model but Estimated by Cox Model.

		$\mu_w^{(1)}(L)$			$\mu_w^{(0)}(L)$			$\triangle_w(L)$		
Estimator	γ	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$	$L=2$	$L=5$	$L=10$
Overlap Weighting	1	95.6	94.9	93.5	94.7	94.2	94.7	95.5	95.6	95.9
	3	94.2	95.9	94.5	94.7	94.9	94.7	94.0	94.8	95.6
	5	95.1	94.9	94.2	93.0	95.0	95.8	94.6	95.2	95.2
IPTW	1	95.3	94.9	92.3	94.6	94.4	95.3	95.1	95.1	94.5
	3	92.6	88.6	81.6	91.9	92.2	92.5	90.9	87.5	86.1
	5	88.4	82.0	73.2	84.5	85.5	86.7	83.8	79.0	79.3
Symmetric Trimming										
	1	95.3	94.9	92.3	94.7	94.4	95.4	95.1	95.2	94.5
$\alpha = 0.05$	3	94.2	94.8	90.2	93.0	94.2	94.5	93.9	93.3	92.1
	5	93.5	95.0	92.8	92.1	92.7	95.0	93.6	93.5	93.4
	1	95.3	95.0	92.6	94.7	94.3	95.5	95.1	95.4	94.2
$\alpha = 0.1$	3	94.5	95.3	94.1	93.1	94.3	94.7	93.9	93.5	94.7
	5	94.2	94.5	93.4	93.1	94.5	95.2	93.9	93.6	94.0
	1	95.6	95.7	92.7	94.5	94.2	94.9	95.3	95.5	94.5
$\alpha = 0.15$	3	94.4	94.4	93.3	94.7	94.0	94.5	93.6	94.2	94.1
	5	94.6	95.0	93.5	93.2	95.0	95.5	94.5	95.1	94.2
Asymmetric Trimming										
	1	95.3	94.8	92.6	94.6	94.2	95.1	95.4	95.9	94.3
$q = 0$	3	87.8	87.9	84.0	90.7	91.9	91.6	87.8	86.9	85.7
	5	80.0	83.2	80.8	80.1	83.2	86.6	80.6	79.7	79.8
	1	95.5	94.9	93.5	94.9	94.6	95.5	95.5	94.7	94.4
$q = 0.01$	3	93.6	95.2	91.7	93.6	94.1	94.7	93.0	93.6	93.1
	5	92.6	92.7	91.3	92.6	93.3	95.3	93.1	92.2	91.6
	1	95.2	94.6	94.1	94.3	94.5	95.1	94.8	95.1	95.4
$q = 0.05$	3	93.2	94.7	93.7	94.9	95.2	95.5	93.8	94.1	94.6
	5	93.8	95.1	93.3	92.2	93.2	93.6	93.8	93.7	93.0
Truncation										
	1	95.3	95.0	93.8	94.6	94.5	95.5	95.0	95.2	95.0
$q = 0.025$	3	93.8	90.5	83.3	93.0	93.1	93.5	92.1	89.9	88.9
	5	88.4	82.1	73.1	84.8	86.0	87.2	84.3	79.3	79.6
	1	94.8	94.4	93.4	94.6	94.6	95.6	94.9	95.4	94.7
$q = 0.05$	3	94.5	91.8	84.5	93.6	93.8	94.0	93.5	90.9	90.6
	5	89.4	83.1	73.4	86.5	87.2	88.2	86.6	80.3	81.0
	1	94.1	93.8	92.2	94.7	94.7	95.7	93.8	95.2	95.0
$q = 0.1$	3	94.0	90.2	83.8	94.7	95.2	94.3	92.9	91.5	91.9
	5	91.3	84.2	73.9	90.2	91.5	91.7	89.6	84.8	84.5