Web Appendix for

"Addressing Extreme Propensity Scores via the Overlap Weights"

Fan Li, Laine E. Thomas and Fan Li

TABLE OF CONTENTS

WEB APPENDIX 1: DEFINITION OF TARGET ESTIMANDS

We assume the marginal density of $\boldsymbol{X}$ (from the combined treatment and control groups) exists and denote it by $f(\boldsymbol{X})$. We could generally represent the target population density by $f(\boldsymbol{X})\ h(\boldsymbol{X})/C$, where $h(\boldsymbol{X})$ is a pre-specified weighting or tilting function, and C is a normalization constant such that the density integrates to unity. Each choice of h corresponds to a target estimand—the average treatment effect among the target population

$$\Delta_w = \frac{E[h(\boldsymbol{X})(\mu_1(\boldsymbol{X}) - \mu_0(\boldsymbol{X}))]}{E[h(\boldsymbol{X})]}$$

where both expectations are taken over the original density $f(\boldsymbol{X})$, and $\mu_1(\boldsymbol{X}) = E[Y(1)|\boldsymbol{X}]$, $\mu_0(\boldsymbol{X}) = E[Y(0)|\boldsymbol{X}]$ are conditional means of potential outcomes given pre-treatment covariates. The quantity Δ_w is often called the weighted average treatment effect (WATE). When the weighting function $h(\boldsymbol{X}) = 1$ for all values of $\boldsymbol{X}$, $\Delta_w = \Delta_{IPW} = E[\mu_1(\boldsymbol{X}) - \mu_0(\boldsymbol{X})] = E[Y(1) - Y(0)]$, which is the standard 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 $h(\boldsymbol{X}) = I_{\{\alpha<e(\boldsymbol{X})<1-\alpha\}}$, Δ_w is the ATE for the target population after symmetric trimming. Notice that h can be analogously defined to characterize the ATE for the target population after asymmetric trimming. Finally, when $h(\boldsymbol{X}) = e(\boldsymbol{X})\big(1 - e(\boldsymbol{X})\big)$, $\Delta_w = \Delta_{OW}$ is the 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, $h(\boldsymbol{X})$ 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.

WEB APPENDIX 2: EXACT BALANCE PROPERTY OF OVERLAP WEIGHTING

The absolute standardized difference (ASD) is typically used as a measure of covariate balance. To assess how effective each weighting method is in removing potential confounding, we define ASD for the kth covariate X_k as the absolute weighted mean difference between groups scaled by the square root of the pooled within-group variance:

$$ASD_k = \left| \frac{\sum_{i=1}^{n} Z_i X_{ik} w_i}{\sum_{i=1}^{n} Z_i w_i} - \frac{\sum_{i=1}^{n} (1 - Z_i) X_{ik} w_i}{\sum_{i=1}^{n} (1 - Z_i) w_i} \right| \Big/ \sqrt{\frac{s_0^2 + s_1^2}{2}}$$

where w_i is the weight for unit i, and s_1^2, s_0^2 are unweighted sample variance of X_k in the treated and control groups. Similar definitions have also been suggested in Austin and Stuart [1], with the unweighted variance replaced by the weighted counterpart. Here, we use the unweighted sample variance to set a common denominator and therefore facilitate comparisons across different weighting methods. The ASD in the original sample is the unweighted mean difference with $w_i = 1$, and measures the degree of potential confounding bias (imbalance). The ASD for IPW and trimming could also be calculated by replacing $w_i$with the corresponding weights.

OW possesses the exact balance property for any covariate X_k entering the logistic propensity score model. This is because the logistic MLE happens to exploit the balancing moment constraints in the weighted population. More specifically, the MLE $\widehat{\boldsymbol{\beta}}$ satisfies the logistic score equations: $\sum_{i=1}^{n} Z_i (1 - \hat{e}_i) = \sum_{i=1}^{n} \hat{e}_i (1 - Z_i)$, and $\sum_{i=1}^{n} Z_i X_{ik} (1 - \hat{e}_i) = \sum_{i=1}^{n} X_{ik} \hat{e}_i (1 - Z_i)$ for all k. If we replace $w_i = 1 - \hat{e}_i$ for a treated unit and $w_i = \hat{e}_i$ for a control unit in the numerator of the ASD definition, it is immediate that $ASD_k = 0$ for all k as required by the score equations. Finally, as the exact balance property is granted by the nature of logistic-binomial likelihood, it holds for all covariates included in the logistic PS model regardless of whether the PS model

coincides with the true assignment mechanism. A more complete and technical proof can be found in Theorem 3 of Li et al. [2] and discussions therein.

WEB APPENDIX 3: VARIANCE ESTIMATION FOR $\widehat{\boldsymbol{\Delta}}_{\boldsymbol{OW}}$

For a logistic propensity score of the form $e(\boldsymbol{X}_i) = 1/(1+\exp(-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}))$ estimated by maximum likelihood, we can expand the score equations around the true parameter value:

$$\sqrt{n}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big) = \boldsymbol{E}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-\mathbf{1}}\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\big(Z_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})\big)e_{\boldsymbol{\beta}}/(e(\boldsymbol{X}_i;\boldsymbol{\beta})(1-e(\boldsymbol{X}_i;\boldsymbol{\beta}))) + o_p(1)$$

$$= \boldsymbol{E}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-\mathbf{1}}\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\boldsymbol{X}_i(Z_i - e(\boldsymbol{X}_i;\boldsymbol{\beta})) + o_p(1)$$

where we use $o_p(1)$ to represent a term that vanishes to zero in probability as the sample size n gets larger, and $e_{\boldsymbol{\beta}} = \partial e/\partial\boldsymbol{\beta} = \boldsymbol{X}_i(e(\boldsymbol{X}_i;\boldsymbol{\beta})(1-e(\boldsymbol{X}_i;\boldsymbol{\beta})))$ is the gradient of the logistic propensity score. Recall that the OW estimator is

$$\widehat{\Delta}_{OW} = \hat{\tau}_1 - \hat{\tau}_0 = \frac{\sum_{i=1}^{n} Z_i Y_i(1-\hat{e}_i)}{\sum_{i=1}^{n} Z_i(1-\hat{e}_i)} - \frac{\sum_{i=1}^{n}(1-Z_i)Y_i\hat{e}_i}{\sum_{i=1}^{n}(1-Z_i)\hat{e}_i}.$$

From this, we can view $\hat{\tau}_1$ as the solution to the unbiased estimating equation $\sum_{i=1}^{n} Z_i(1-\hat{e}_i)(Y_i-\hat{\tau}_1) = 0$. Denote $\theta = E\big[e(\boldsymbol{X})\big(1-e(\boldsymbol{X})\big)\big]$, we expand the estimating equation around $\tau_1 = \theta^{-1}E\left(e(\boldsymbol{X})\big(1-e(\boldsymbol{X})\big)\mu_1(\boldsymbol{X})\right)$ and the true PS e_i to obtain

$$\sqrt{n}(\hat{\tau}_1-\tau_1) = \theta^{-1}\left[\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n} Z_i(Y_i-\tau_1)(1-e_i) - E\big(Z_i(Y_i-\tau_1)e_{\boldsymbol{\beta}}\big)^{T}\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right] + o_p(1).$$

Similarly, if we define $\tau_0 = \theta^{-1}E\left(e(\boldsymbol{X})\big(1-e(\boldsymbol{X})\big)\mu_0(\boldsymbol{X})\right)$, we have

$$\sqrt{n}(\hat{\tau}_0 - \tau_0) = \theta^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-Z_i)(Y_i-\tau_0)e_i + E\big((1-Z_i)(Y_i-\tau_0)e_{\boldsymbol{\beta}}\big)^T\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right] + o_p(1)$$

Differencing the above two expansions to get

$$\sqrt{n}\big(\widehat{\Delta}_{OW} - \Delta_{OW}\big) = \theta^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Z_i(Y_i-\tau_1)(1-e_i) - (1-Z_i)(Y_i-\tau_0)e_i] - \boldsymbol{H}_{\boldsymbol{\beta}}^{\boldsymbol{T}}\sqrt{n}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)\right] + o_p(1) = \theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i + o_p(1)$$

where

$$I_i = Z_i(Y_i-\tau_1)(1-e_i) - (1-Z_i)(Y_i-\tau_0)e_i - (Z_i-e_i)\boldsymbol{H}_{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{E}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-\mathbf{1}}\boldsymbol{X}_i$$

and $\boldsymbol{H}_{\boldsymbol{\beta}} = E[Z_i(Y_i-\tau_1)e_{\boldsymbol{\beta}} + (1-Z_i)(Y_i-\tau_0)e_{\boldsymbol{\beta}}]$. The above expansion indicates that the influence function of $\widehat{\Delta}_{OW}$ is I_i/θ, which is consistently estimated by $\hat{I}_i/\hat{\theta}$. The empirical sandwich method uses the sample variance of the influence function $n^{-2}\sum_{i=1}^{n}(\hat{I}_i/\hat{\theta})^2$ to estimate the variance of $\widehat{\Delta}_{OW}$ [3]. The explicit variance estimator is provided in the main text and is not repeated here for brevity. Further, the asymptotic variance expression of $\widehat{\Delta}_{OW}$ can also be calculated in closed-form by the second moment of the influence function $Var(I_i/\theta) = E(I_i^2/\theta^2)$. The resulting expression is tedious and omitted for brevity, but is available upon request from the first author. Finally, we comment that although the derivation here focuses on the most common case with a logistic propensity score model, it can be easily generalized to an arbitrary PS model 1) where the PS is differentiable in the PS model parameters and 2) that admits a regular and asymptotically linear estimator for the PS model parameters. In that case, the influence function of the OW estimator, I_i/θ, can be obtained by replacing $\sqrt{n}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)$ with

the suitable stochastic expansion, from which the empirical sandwich method can be applied in the same way to estimate the variance of $\widehat{\Delta}_{OW}$.

WEB APPENDIX 4: SIMULATION RESULTS FOR MODIFIED ASYMMETRIC TRIMMING

In our simulations with no unmeasured confounding, asymmetric trimming did not perform well and we provide a further explanation here. When there are strong tails, asymmetric trimming with $q = 0$ ensures overlap but does not fully address the extreme tails since there may still be treated patients with $\hat{e}_i \approx 1$ and control patients with $\hat{e}_i \approx 0$. In this case, we observe large absolute standardized differences (ASD) of potential confounders for both asymmetric trimming and IPW in Figure 2 and Web Figure 2, which underlie the apparent bias of asymmetric trimming with $q = 0$. If the trimming threshold $q > 0$, one then intentionally creates PS non-overlap in the trimmed sample. In particular, as the treated patients with small $\hat{e}_i$ are excluded, there are no corresponding treated patients for control patients with small $\hat{e}_i$, and vice versa. The post-trimming non-overlap compromises the ability of PS to balance covariates, as shown in Figure 2 and Web Figure 2. Further, as the trimming threshold q increases, the degree of PS non-overlap increases, the bias of treatment estimates increases and coverage deteriorates.

We have carried out an additional simulation exercise to confirm our interpretation. Specifically, we modify the asymmetric trimming rule by further ensuring overlap after the asymmetric trimming step. In other words, we exclude both treated and control patients whose $\hat{e}_i$'s are below the q quantile of the PS distribution among the treated and above the $(1 - q)$ quantile of the PS distribution among the control. As shown in Web Tables 5 to 7, the bias, relative efficiency and

coverage of the modified asymmetric trimming are much improved (the improved ASD results are also confirmed and omitted here for brevity), suggesting the necessity of ensuring overlap.

REFERENCES

1. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679.
2. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc*. 2018;113(521):390-400.
3. Stefanski LA, Boos DD. The calculus of M-estimation. *The American Statistician*. 2002;56(1):29-38.

Web Table 1. Bias of the Estimators in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 500$).

Estimator	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
Crude	−2.02	−3.18	−3.76	−4.06
Overlap Weighting	0.00	0.00	0.00	0.00
IPW				
No Trimming	−0.01	−0.03	−0.24	−0.50
Symmetric Trimming				
$\alpha = 0.05$	−0.01	−0.02	−0.01	−0.02
$\alpha = 0.1$	−0.01	−0.01	0.00	0.00
$\alpha = 0.15$	−0.01	−0.01	−0.01	0.00
Asymmetric Trimming				
$q = 0$	0.05	0.25	0.47	0.63
$q = 0.01$	−0.08	−0.26	−0.40	−0.49
$q = 0.05$	−0.40	−0.98	−1.33	−1.46
Treatment Prevalence = 0.1				
Crude	−2.07	−3.46	−4.20	−4.58
Overlap Weighting	0.00	0.00	−0.01	−0.01
IPW				
No Trimming	−0.04	−0.35	−1.02	−1.70
Symmetric Trimming				
$\alpha = 0.05$	0.00	−0.01	−0.03	−0.05
$\alpha = 0.1$	−0.01	−0.03	−0.03	−0.02
$\alpha = 0.15$	−0.02	−0.02	−0.02	−0.03
Asymmetric Trimming				
$q = 0$	0.26	0.55	0.61	0.56
$q = 0.01$	0.06	0.06	−0.05	−0.14
$q = 0.05$	−0.11	−0.42	−0.69	−0.83

Web Table 2. Relative Efficiency of the Estimators in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 500$).

Estimator	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
Crude	1.00	1.00	1.00	1.00
Overlap Weighting	3.58	2.25	1.55	1.13
IPW				
No Trimming	2.22	0.25	0.09	0.04
Symmetric Trimming				
$\alpha = 0.05$	2.40	1.09	0.86	0.65
$\alpha = 0.1$	2.76	1.65	1.08	0.80
$\alpha = 0.15$	3.06	1.76	1.03	0.79
Asymmetric Trimming				
$q = 0$	2.70	0.51	0.15	0.07
$q = 0.01$	2.47	0.82	0.49	0.30
$q = 0.05$	1.77	0.63	0.36	0.22
Treatment Prevalence = 0.1				
Crude	1.00	1.00	1.00	1.00
Overlap Weighting	3.78	2.57	1.62	1.21
IPW				
No Trimming	0.64	0.11	0.06	0.05
Symmetric Trimming				
$\alpha = 0.05$	1.90	1.28	0.87	0.71
$\alpha = 0.1$	1.81	1.33	0.94	0.76
$\alpha = 0.15$	1.18	1.14	0.87	0.70
Asymmetric Trimming				
$q = 0$	0.98	0.23	0.13	0.11
$q = 0.01$	1.00	0.32	0.20	0.18
$q = 0.05$	0.67	0.25	0.16	0.14

Web Table 3. 95% Confidence Interval Coverage of the Estimators in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 500$).

Estimator	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
Crude	0.0	0.0	0.0	0.0
Overlap Weighting	95.3	94.4	95.3	94.3
IPW				
No Trimming	94.3	84.1	69.5	56.6
Symmetric Trimming				
$\alpha = 0.05$	94.9	92.2	93.5	92.8
$\alpha = 0.1$	94.8	94.0	93.9	93.9
$\alpha = 0.15$	95.6	94.4	94.4	93.7
Asymmetric Trimming				
$q = 0$	95.7	95.2	96.4	95.6
$q = 0.01$	91.5	69.5	59.9	51.8
$q = 0.05$	42.7	3.7	1.5	1.6
Treatment Prevalence = 0.1				
Crude	0.4	0.0	0.0	0.0
Overlap Weighting	94.2	95.2	93.3	93.1
IPW				
No Trimming	86.8	62.3	37.7	24.2
Symmetric Trimming				
$\alpha = 0.05$	92.7	91.8	90.2	89.0
$\alpha = 0.1$	93.1	91.3	91.5	90.3
$\alpha = 0.15$	92.4	91.5	90.8	91.5
Asymmetric Trimming				
$q = 0$	94.9	95.3	93.9	91.8
$q = 0.01$	91.9	85.7	78.4	75.9
$q = 0.05$	82.0	65.6	53.5	46.8

Web Table 4. 95% Bootstrap Confidence Interval Coverage of Overlap Weighting in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 500$). For each simulated dataset, 1,000 bootstrap sample is used to estimate the variance of OW estimator and to construct the confidence interval.

CI Estimator	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
Normality-Based CI	95.5	94.9	96.0	94.9
Quantile-Based CI	95.7	94.3	96.0	94.7
Treatment Prevalence = 0.1				
Normality-Based CI	94.7	95.6	93.9	93.8
Quantile-Based CI	94.8	95.2	94.0	93.7

Web Table 5. Bias of Modified Asymmetric Trimming in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 2{,}000$).

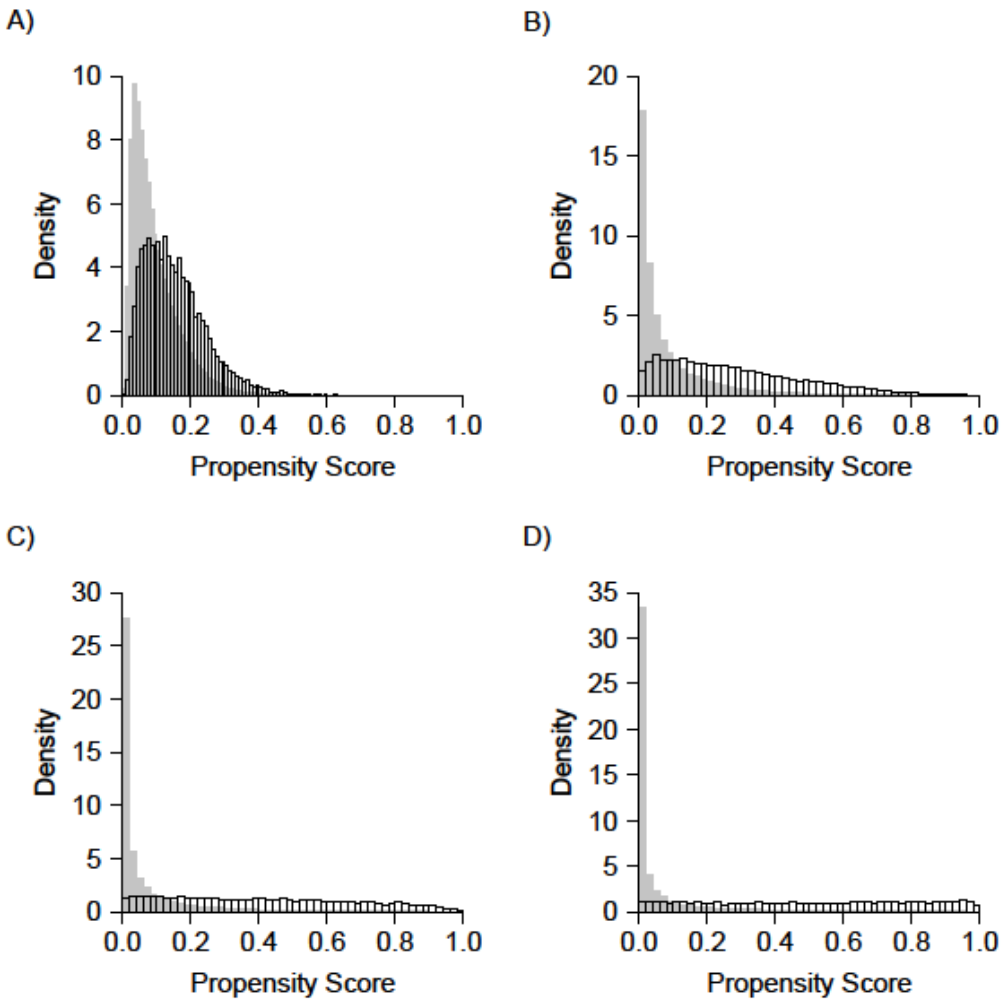
Threshold	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
$q = 0.01$	0.01	0.01	0.01	0.01
$q = 0.05$	0.01	0.00	0.00	0.01
Treatment Prevalence = 0.1				
$q = 0.01$	0.03	0.08	0.08	0.04
$q = 0.05$	0.01	0.02	0.03	0.02

Web Table 6. Relative Efficiency of Modified Asymmetric Trimming in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 2{,}000$).

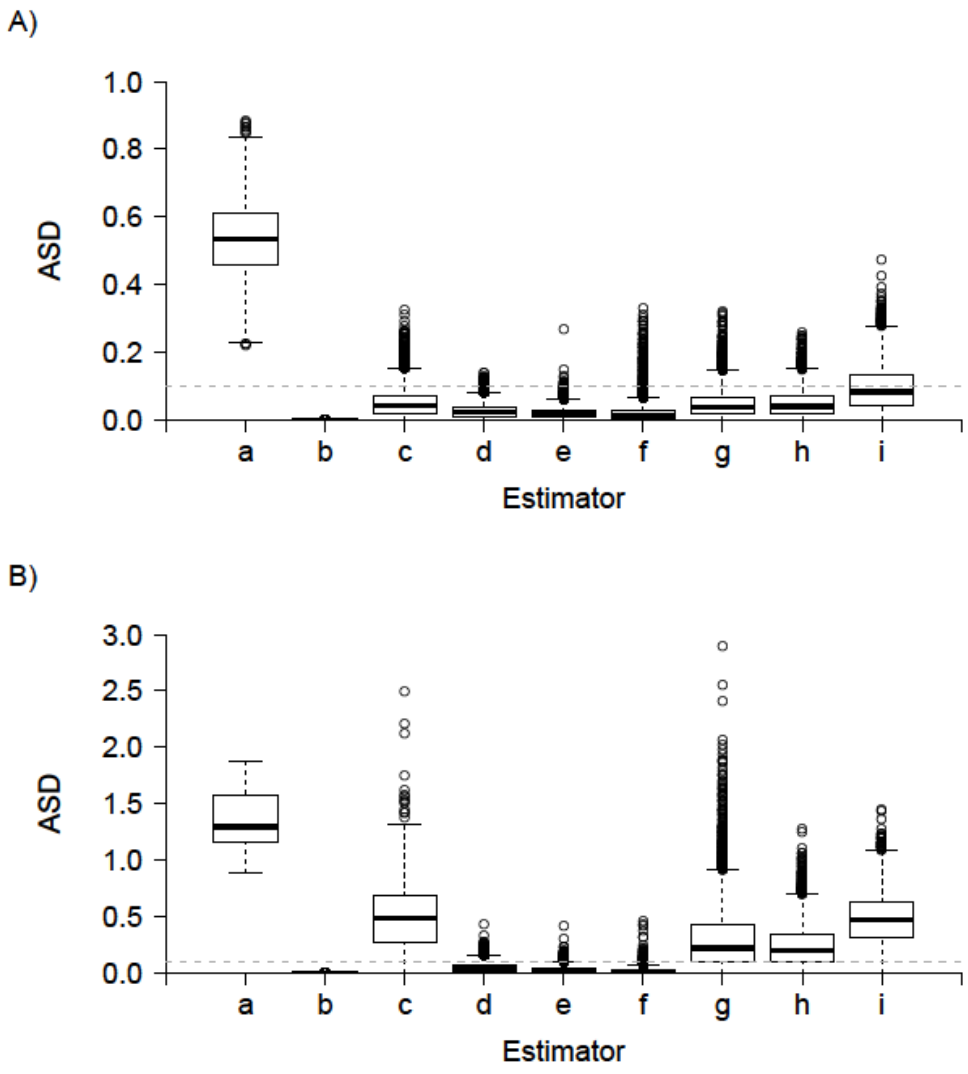
Threshold	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
$q = 0.01$	3.01	1.61	0.97	0.69
$q = 0.05$	2.80	1.71	0.97	0.61
Treatment Prevalence = 0.1				
$q = 0.01$	1.29	0.44	0.28	0.25
$q = 0.05$	1.75	0.96	0.62	0.35

Web Table 7. 95% Confidence Interval Coverage of Modified Asymmetric Trimming in the Presence of Increasingly Strong Tails in the Propensity Score Distributions ($n = 2{,}000$).

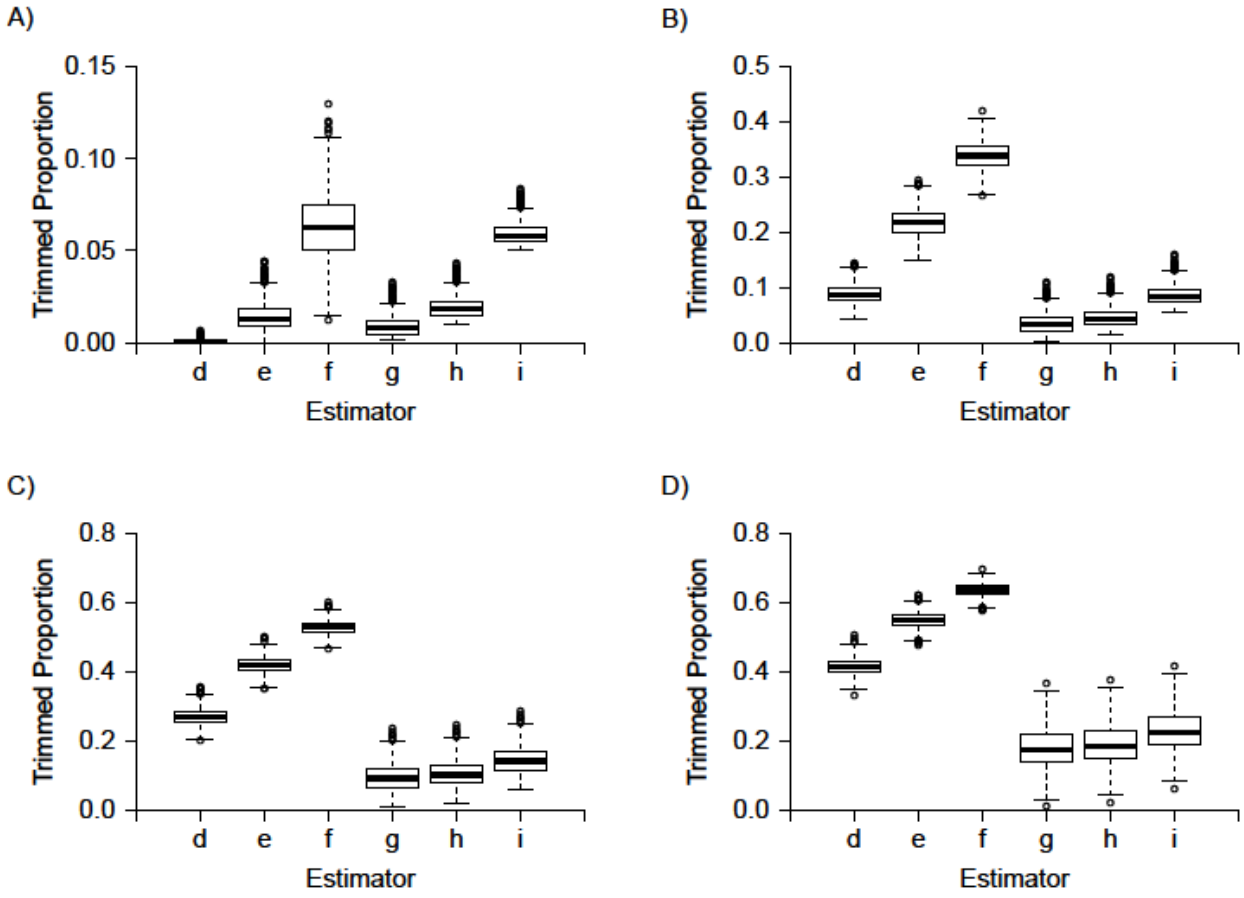
Threshold	**Data Generating Model**			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Treatment Prevalence = 0.4				
$q = 0.01$	94.6	95.2	94.9	95.1
$q = 0.05$	93.5	95.2	93.2	95.1
Treatment Prevalence = 0.1				
$q = 0.01$	94.7	93.8	94.1	91.3
$q = 0.05$	93.6	94.8	93.8	91.4



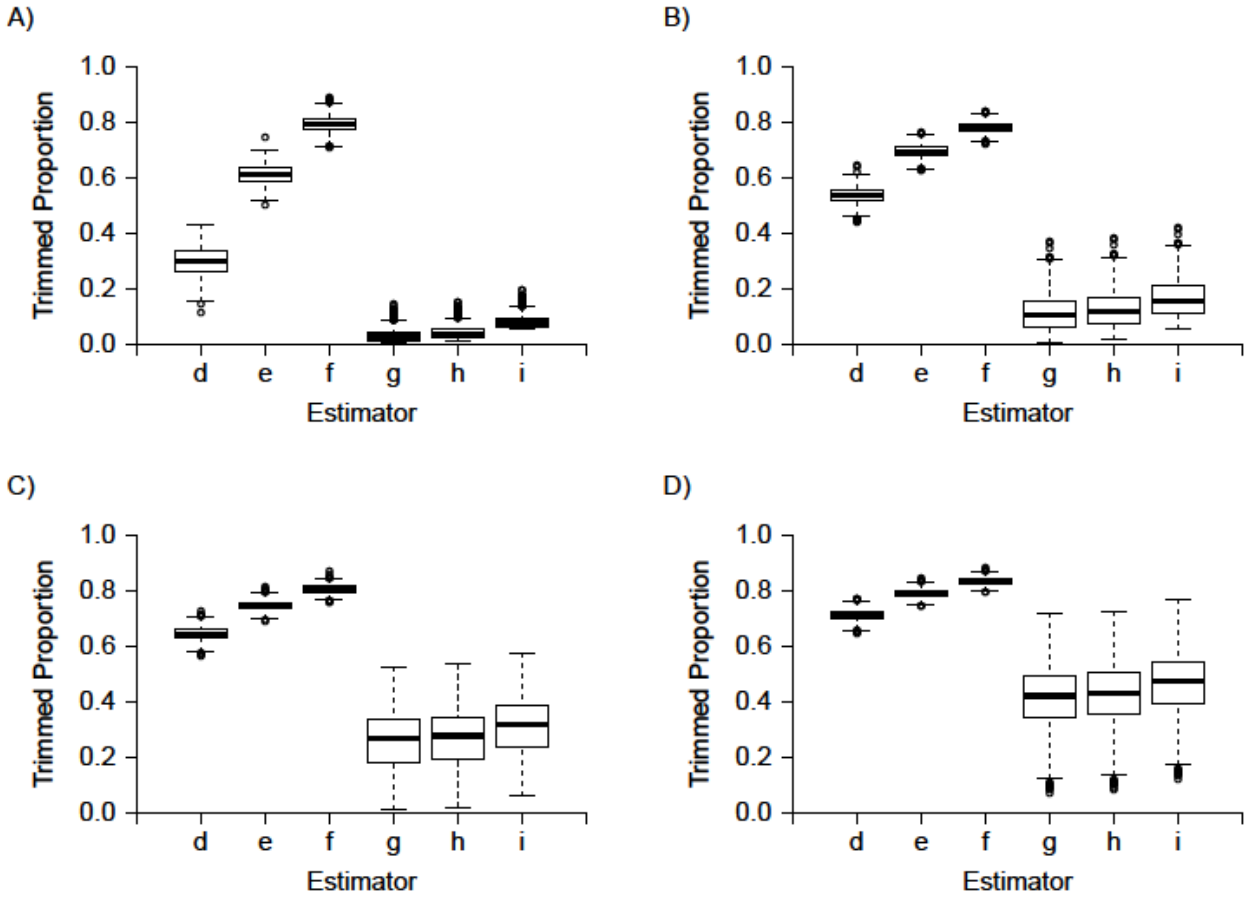
Web Figure 1. Distributions of Propensity Scores Associated with 4 Different Data-generating Process with Overall Treatment Prevalence = 0.1. The Unshaded Bars Indicate the Treated Group; the Gray/Shaded Bars Indicate the Control Group. (A) When $\gamma = 1$; (B) When $\gamma = 2$; (C) When $\gamma = 3$; (D) When $\gamma = 4$.



Web Figure 2. Boxplots of Absolute Standardized Difference (ASD) When Treatment Prevalence = 0.1 and Sample Size = 2,000. The Horizontal Dashed Line Indicates Adequate Balance (ASD = 0.1). (A) When $\gamma = 1$; (B) When $\gamma = 4$. Estimator: Crude (a); OW (b); IPW without Trimming (c); Symmetric Trimming with Threshold 0.05 (d), 0.1 (e), 0.15 (f); Asymmetric Trimming with Threshold 0 (g), 0.01 (h), 0.05 (i).



Web Figure 3. Boxplots of Trimmed Proportions When Treatment Prevalence = 0.4 and Sample Size = 2,000. (A) When $\gamma = 1$; (B) When $\gamma = 2$; (C) When $\gamma = 3$; (D) When $\gamma = 4$. Estimator: Symmetric Trimming with Threshold 0.05 (d), 0.1 (e), 0.15 (f); Asymmetric Trimming with Threshold 0 (g), 0.01 (h), 0.05 (i).



Web Figure 4. Boxplots of Trimmed Proportions When Treatment Prevalence = 0.1 and Sample Size = 2,000. (A) When $\gamma = 1$; (B) When $\gamma = 2$; (C) When $\gamma = 3$; (D) When $\gamma = 4$. Estimator: Symmetric Trimming with Threshold 0.05 (d), 0.1 (e), 0.15 (f); Asymmetric Trimming with Threshold 0 (g), 0.01 (h), 0.05 (i).