Average treatment effect on the treated, under lack of positivity

Yi Liu

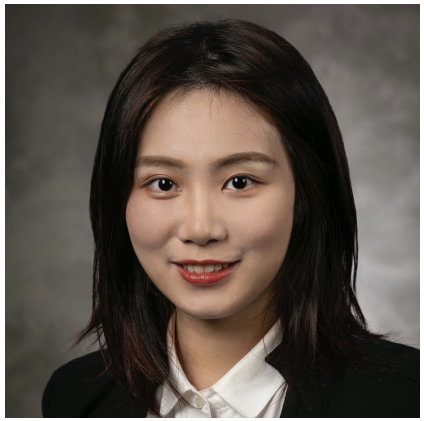
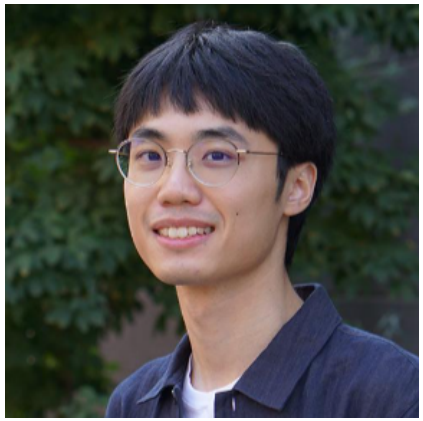
Department of Statistics

NC STATE UNIVERSITY

ENAR Spring Meeting

March 11, 2024

Collaborators

- Roland Matsouaka, PhD (Duke Biostat & Bioinfo and DCRI, Durham, NC)
- Huiyue Li (Baim Institute for Clinical Research, Boston, MA)
- Yunji Zhou (Univ. Washington Biostat, Seattle, WA)

Publication

- **Liu Y**, Li H, Zhou Y, and Matsouaka RA (2024). Average treatment effect on the treated, under lack of positivity. *Statistical Methods in Medical Research*. Under Revision.

Overview

- The **positivity** assumption plays a key role in identification of causal effects in observational studies.
- Each participant should have non-zero subject-specific probability to receive either treatment or control.
  - In randomized trials, this probability is known by design.
  - In observational studies, we have to model and estimate this probability subject-specifically.
- In this talk, we focus on the positivity violation when the interest is identifying the average treatment effect on the treated (ATT), i.e.,

$$\mathbb{E}\{\text{treatment effect} \mid \text{participants who received the active treatment}\}.$$

Overview

Literature often defines two types of violation of positivity.[1]

- Random violation: due to chance, small sample size, model misspecifications, etc.
- Structural violation: due to the inherent characteristics of the target population.
  - ATT is technically **not** identifiable.
- We proposed a method that addressed both violations.

[1] Petersen, M. L. *et al.* Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research* **21,** 31–54 (2012).

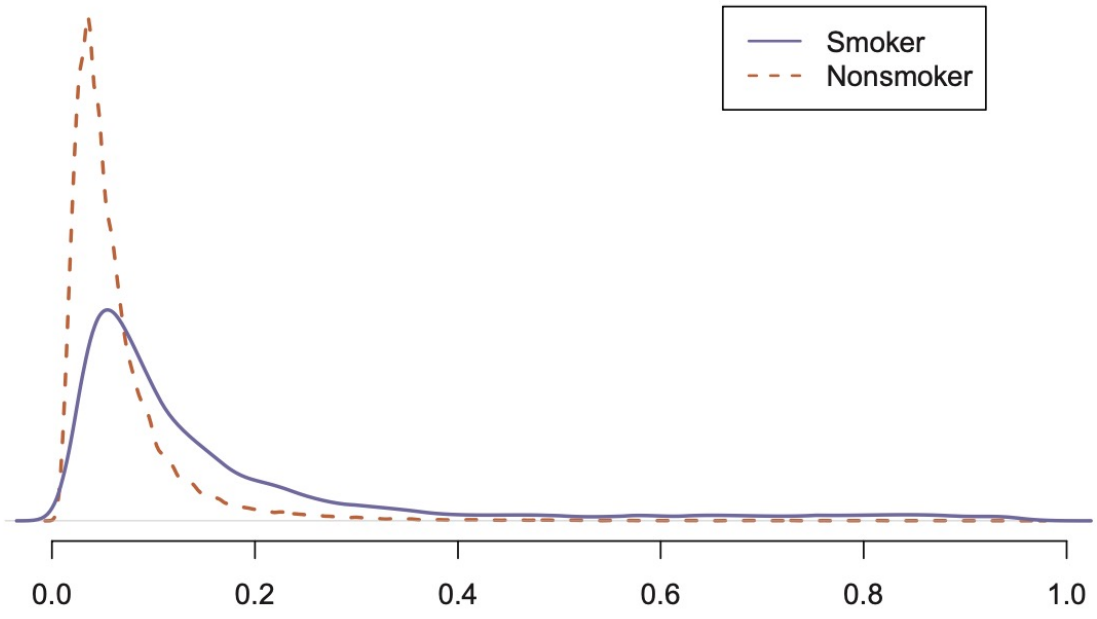
Notations and Definitions

- We assume the full-data of an observational study are from a super-population model $f(x, z, y(0), y(1))$.
- X: covariate; $Z \in \{0, 1\}$: binary treatment; $Y(0)$ and $Y(1)$: two potential outcomes.
- However, for each participant, we can only observe the outcome associated with their received treatment Z, i.e., $Y = Y(Z)$.
- Assume also $Y(z) \perp\!\!\!\perp Z \mid X$ for both $z = 0, 1$.
- **Positivity**: the propensity score $e(X) = P(Z = 1 \mid X)$ (a subject-specific score) must satisfy $0 < e(X) < 1$ w.p.1.
- Or **strict positivity**[2]: $0 < c_1 \leq e(X) \leq c_2 < 1$ w.p.1. for some constants c_1 and c_2.

[2] Hirano, K. *et al.* Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71,** 1161–1189 (2003), D'Amour, A. *et al.* Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221,** 644–654 (2021).

Positivity violation example

Extreme propensity scores: NC birth weights data[3]



	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Smoker	0.0044	0.06	0.10	0.17	0.18	0.98
Nonsmoker	0.0014	0.03	0.05	0.07	0.08	0.97

[3] Zhou, Y. *et al.* Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research* **29,** 3721–3756 (2020).

Trimming or truncating extreme weights

Two common practices for excluding/capping extreme weights:

- Trimming: exclude participants with estimated $e(X)$ outside a range $[c_1, c_2]$, where $0 < c_1 < c_2 < 1$.
- Truncation: a weight capping, i.e., assign c_1 as the new propensity score to those $e(X) < c_1$ and c_2 to those $e(X) > c_2$.

*Moving the **goalposts**...*

- In fact, they moved the target of inference (goalposts).[4] For example, the trimming targets $\mathcal{O}(X) = \{X : c_1 \leq e(X) \leq c_2\} \subset$ full support.

[4] Crump, R. *et al.* *Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand.* 2006.

Average treatment effect on the treated (ATT)

- ATT is defined by
  $\tau_{att} = \mathbb{E}\{Y(1) - Y(0) \mid Z = 1\} = \mathbb{E}\{Y \mid Z = 1\} - \mathbb{E}\{Y(0) \mid Z = 1\}$.
- $Y(0)$ is missing (unobserved) for $Z = 1$.
- We can re-write ATT using the propensity score $e(X)$:

$$\tau_{att} = \frac{\mathbb{E}(ZY)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{w_0(X)(1-Z)Y\}}{\mathbb{E}\{w_0(X)(1-Z)\}},$$

  where $w_0(X) = \dfrac{e(X)}{1 - e(X)}$.

- Extreme weights occur when $e(X) \approx 1$ on control participants.

Moving the goalposts: weighted ATT (WATT)

The WATT is defined by:

$$\tau^h_{watt} = \frac{\mathbb{E}(ZY)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{\omega_{0h}(X)(1-Z)Y\}}{\mathbb{E}\{\omega_{0h}(X)(1-Z)\}}, \quad \text{with } \omega_{0h}(x) = w_0(x)h(x) = \frac{e(x)h(x)}{1-e(x)}.$$

- $h(x)$ is a tilting function. It generalizes the weights on control and thus generalizes the estimand.
- A weighting estimator for ATT:

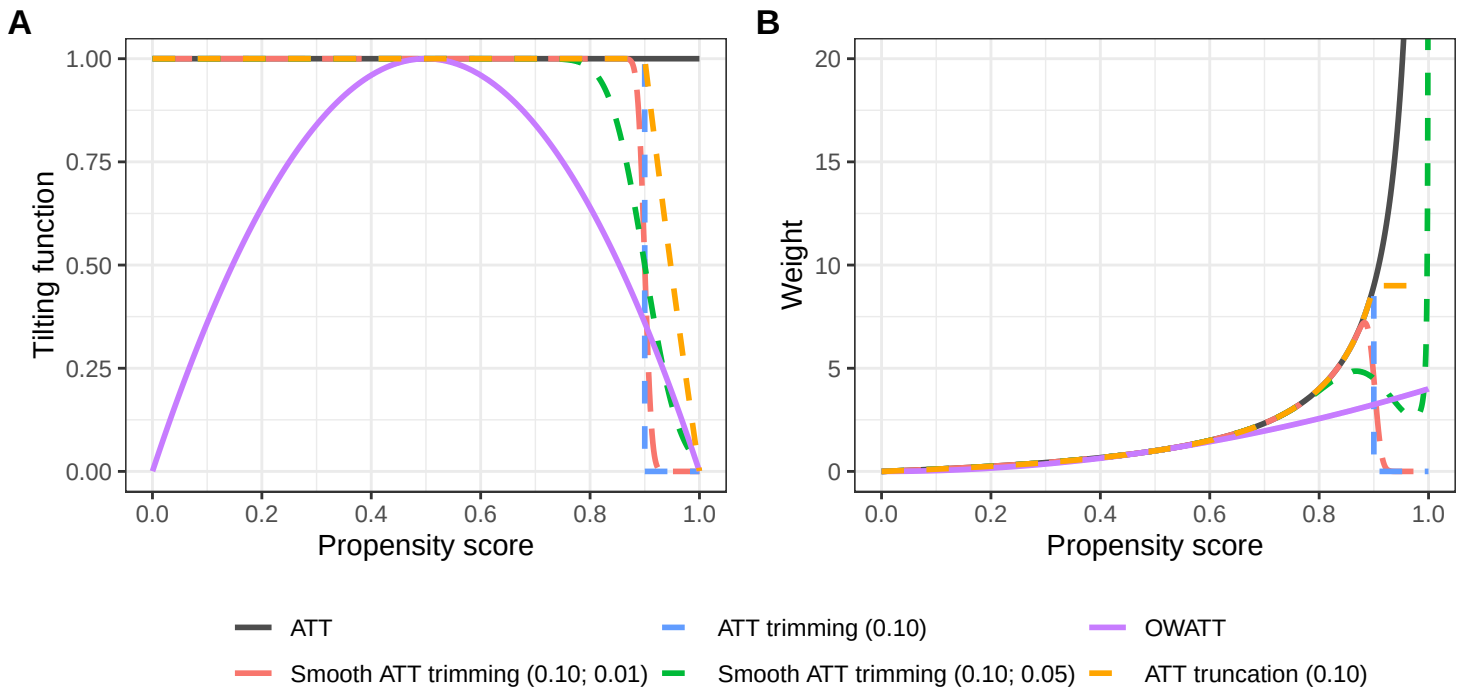
$$\widehat{\tau}^h_{watt} = \frac{\sum_{i=1}^N Z_i Y_i}{\sum_{i=1}^N Z_i} - \frac{\sum_{i=1}^N (1-Z_i)\widehat{\omega}_{0h}(X_i) Y_i}{\sum_{i=1}^N (1-Z_i)\widehat{\omega}_{0h}(X_i)}.$$

- The idea of defining this WATT is motivated by the idea of weighted average treatment effect (WATE).[5]

[5] Hirano, K. *et al.* Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71,** 1161–1189 (2003).

Overlap weighted ATT (OWATT)

A: $h(x)$ vs. $e(x)$, and **B**: weights $\omega_{0h}(x)$ on the controls.



The purple curves correspond to $h(x) = e(x)\{1 - e(x)\}$ (**overlap** function).[6] We call the WATT when choosing the overlap function as the tilting functions by **"overlap weighted ATT (OWATT)"**.

[6] Li, F. *et al.* Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* **113,** 390–400 (2018).

Inference

Assuming we use a GLM for the propensity score $e(x) = e(x'\beta)$, the estimator $\widehat{\tau}^h_{watt}$ is regular and asymptotic linear (RAL), with

$$\sqrt{N}(\widehat{\tau}^h_{watt} - \tau^h_{watt}) \to_d \mathcal{N}(0, \sigma^2 + b_1' \mathcal{I}(\beta^*)^{-1} b_1 - b_2' \mathcal{I}(\beta^*)^{-1} b_2),$$

where $\sigma^2 = \sum_{z=0}^{1} \mathbb{E}\left\{\eta_z(X)\{\mu\{z, e(X)\}^2 + \sigma^2\{z, e(X)\} + \sigma^2(z, X)\}\right\}$ with

$$\eta_1(X) = \frac{e(X)}{\mathbb{E}\{e(X)\}^2}, \quad \eta_0(X) = \frac{\omega_{0h}(X)^2\{1 - e(X)\}}{\mathbb{E}\{e(X)h(X)\}^2},$$

$$\mu\{z, e(X)\} = \mathbb{E}\{Y \mid e(X), Z = z\},$$

$$\sigma^2\{z, e(X)\} = \text{var}\{Y \mid e(X), Z = z\},$$

$$\sigma^2(z, X) = \text{var}\{Y \mid X, Z = z\}, \quad \text{for } z = 0, 1,$$

where $\mathcal{I}(\beta^*)$ is the Fisher's information matrix of β, with β^* the truth of β, and

$$b_1' = \mathbb{E}\left\{\frac{\partial}{\partial \beta'}\left[\frac{e(X'\beta^*)}{\mathbb{E}\{e(X'\beta^*)\}}\right]\mu(1, X) - \frac{\partial}{\partial \beta'}\left[\frac{e(X'\beta^*)h(X'\beta^*)}{\mathbb{E}\{e(X'\beta^*)h(X'\beta^*)\}}\right]\mu(0, X)\right\},$$

$$b_2' = \mathbb{E}\left\{\left[\frac{\mathbb{E}\{X\mu(1, X) \mid e(X)\}}{\mathbb{E}\{e(X)\}} + \frac{\omega_{0h}(X)\mathbb{E}\{X\mu(0, X) \mid e(X)\}}{\mathbb{E}\{e(X)h(X)\}}\right] f(X)\right\}.$$

Inference

Remarks:

- The asymptotic linearity allows the use of bootstrap for variance estimation.
- In the asymptotic variance term, $\eta_0(X) = \frac{\omega_{0h}(X)^2\{1-e(X)\}}{\mathbb{E}\{e(X)h(X)\}^2}$. Thus,
  - when $h(x) \propto 1$ (ATT), $\eta_0(X) \propto e(x)^2/\{1-e(x)\}$, which can still be extreme.
  - when $h(x) \propto e(x)\{1-e(x)\}$ (OWATT), $\eta_0(x) \propto e(x)^4\{1-e(x)\}$, which is always bounded.

Inference

We demonstrated that, when the propensity score is possibly misspecified and converges to a limit $\widetilde{e}(x)$, the asymptotic biases of estimating ATT and OWATT are, respectively,

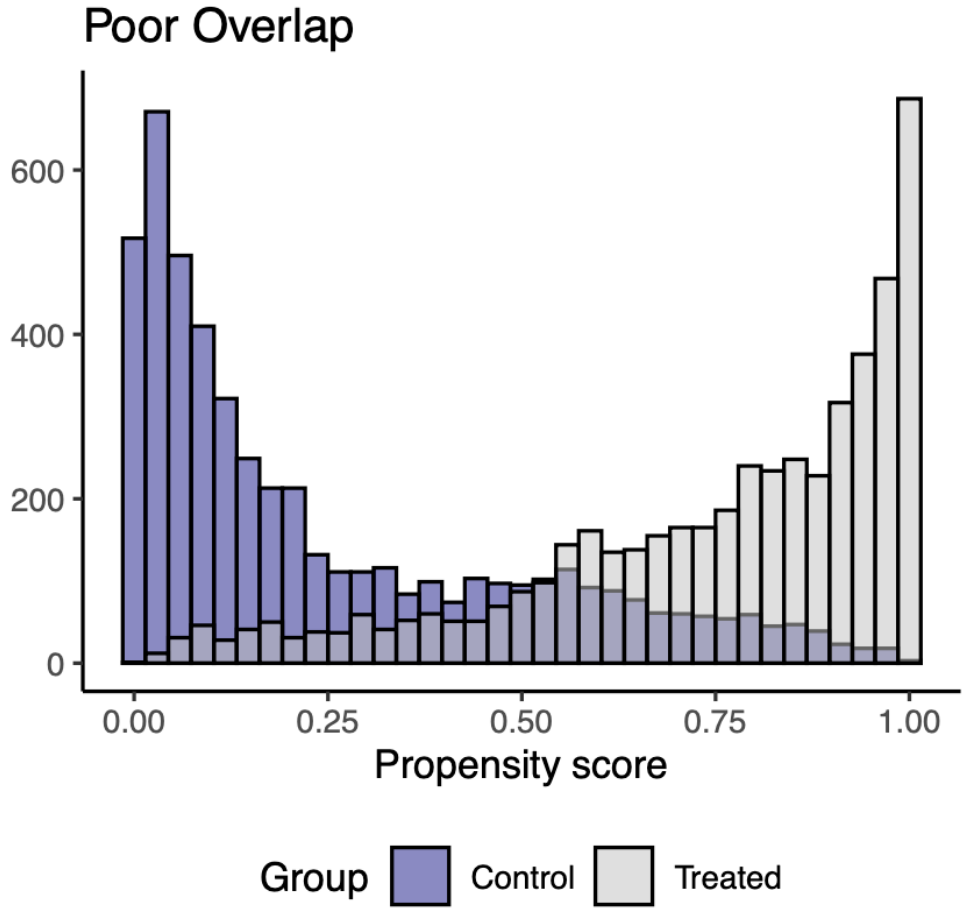
$$\text{ABias}(\widehat{\tau}_{att}) = \frac{\mathbb{E}\{e(X)m_0(X)\}}{\mathbb{E}\{e(X)\}} - \frac{\mathbb{E}\left\{\dfrac{\widetilde{e}(X)}{1-\widetilde{e}(X)}\{1-e(X)\}m_0(X)\right\}}{\mathbb{E}\left\{\dfrac{\widetilde{e}(X)}{1-\widetilde{e}(X)}\{1-e(X)\}\right\}},$$

$$\text{ABias}(\widehat{\tau}_{owatt}) = \frac{\mathbb{E}\{e(X)^2\{1-e(X)\}m_0(X)\}}{\mathbb{E}\{e(X)^2\{1-e(X)\}\}} - \frac{\mathbb{E}\left\{\widetilde{e}(X)^2\{1-e(X)\}m_0(X)\right\}}{\mathbb{E}\left\{\widetilde{e}(X)^2\{1-e(X)\}\right\}}.$$

The teal parts can incur extreme values when $\widetilde{e}(x) \to 1$.

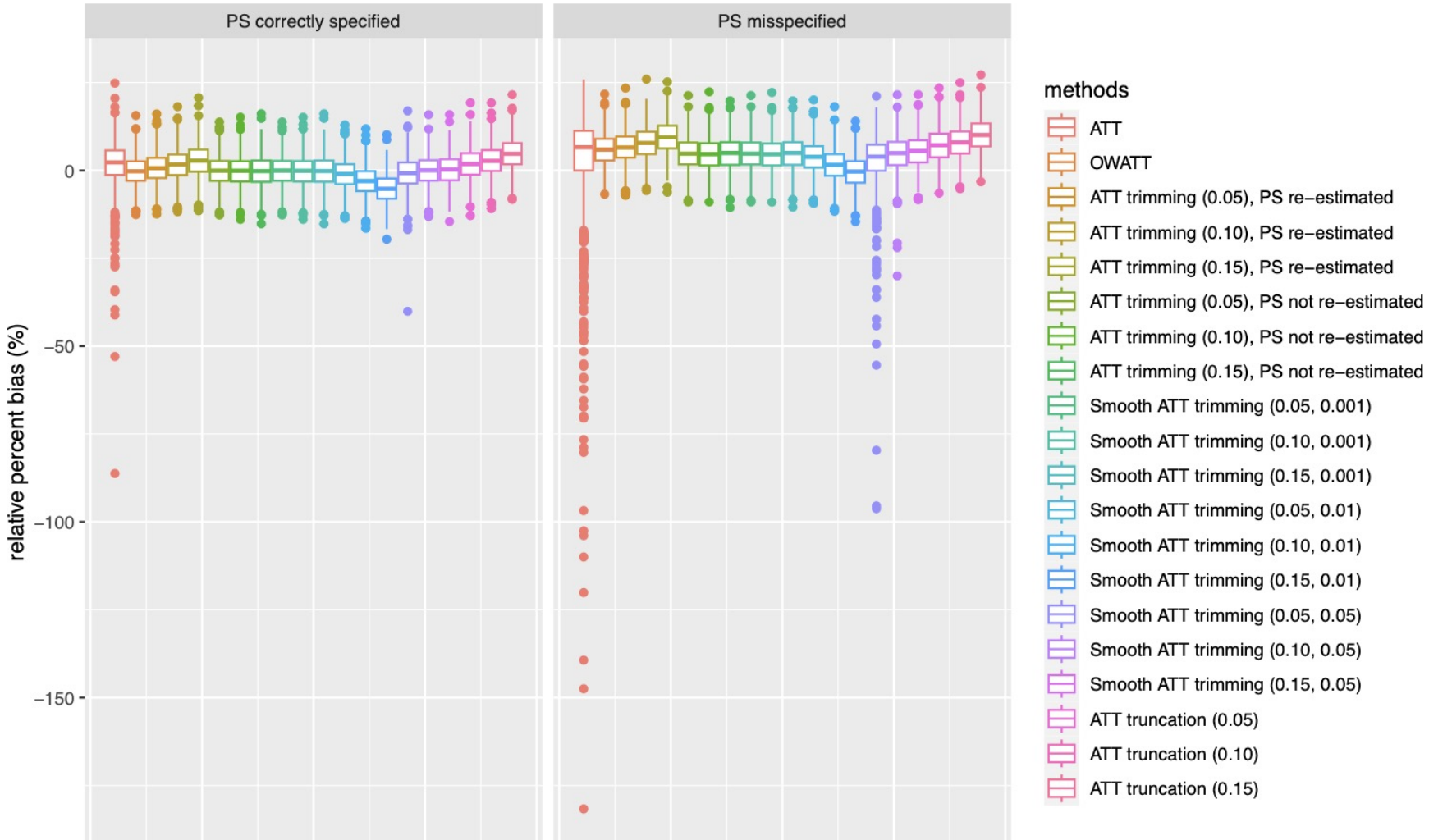
Simulation study

We conducted a simulation study with propensity score model such that the overlap is as follows. There are some extreme weights by this model.



Simulation study

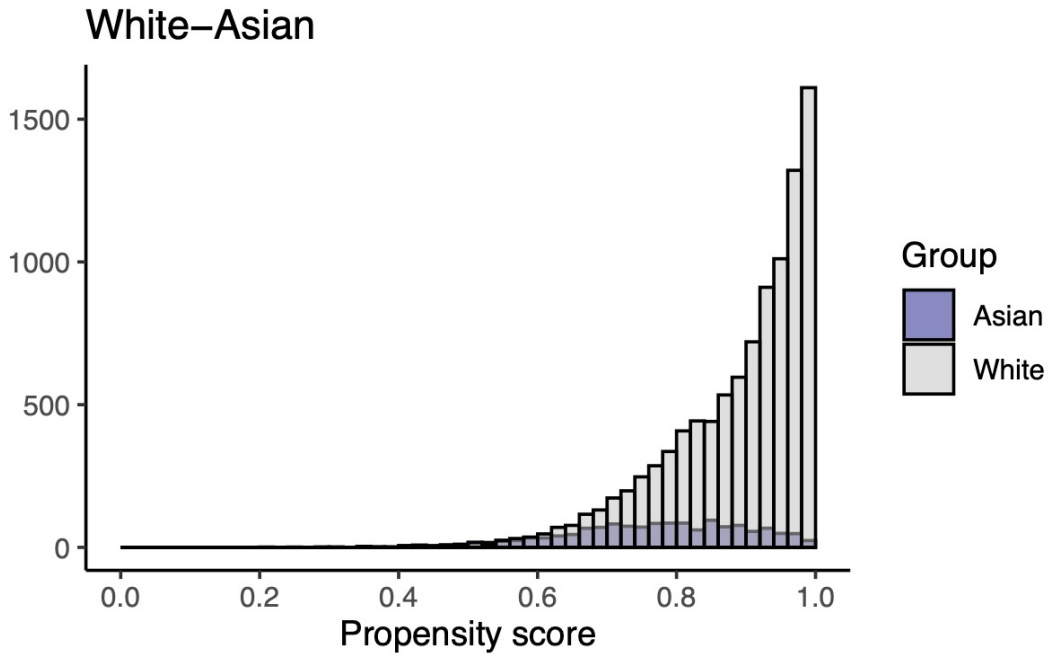
Boxplots of relative biases:



For smooth ATT trimming methods, the parameter in the bracket is (α, ε), i.e., trimming threshold and standard error of the normal cdf in the tilting function, respectively.

Racial disparities in health care expenditure

- Data from the Medical Expenditure Panel Survey (MEPS): `https://www.meps.ahrq.gov/mepsweb/`
- We include 11276 individuals, with 9830 (87.18%) non-Hispanic White as treated and 1446 (12.82%) Asian as control. We included 31 covariates, and considered the health care expenditure as the outcome of interest.



Racial disparities in health care expenditure

Method	Point estimate	Standard error	p-value
ATT	2399.32	787.37	0.002
OWATT	**2511.91**	**255.20**	< 0.001
ATT trimming ($\alpha = 0.05$), PS re-estimated	2363.09	403.42	< 0.001
ATT trimming ($\alpha = 0.10$), PS re-estimated	2666.13	356.62	< 0.001
ATT trimming ($\alpha = 0.15$), PS re-estimated	3054.09	352.98	< 0.001
ATT trimming ($\alpha = 0.05$), PS not re-estimated	2487.25	352.16	< 0.001
ATT trimming ($\alpha = 0.10$), PS not re-estimated	2928.39	286.52	< 0.001
ATT trimming ($\alpha = 0.15$), PS not re-estimated	3286.90	270.04	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.001$)	2488.98	348.88	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.001$)	2926.52	285.92	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.001$)	3291.05	268.68	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.01$)	2419.59	327.68	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.01$)	2881.88	277.57	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.01$)	3229.41	259.47	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.05$)	2337.55	373.65	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.05$)	2638.19	250.78	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.05$)	3014.23	232.06	< 0.001
ATT truncation ($\alpha = 0.05$)	1945.35	385.00	< 0.001
ATT truncation ($\alpha = 0.10$)	2211.56	307.63	< 0.001
ATT truncation ($\alpha = 0.15$)	2419.23	271.39	< 0.001

Discussion

Summary

- We proposed overlap weighted ATT (OWATT) under lack of positivity, as an alternative to ATT.
- OWATT has some practical advantages:
  - No selection on any threshold parameters.
  - Makes use of information from all samples.
  - Statistically sound and efficient under lack of positivity.

Limitation

- We may require some relatively strong assumptions on the propensity score estimation for our method, but we also demonstrated that under lack of positivity, when the propensity score is misspecified, OWATT is more robust.

Future research

- Augmented estimator, sandwich variance estimation, extensions to multi-valued treatment data, survival data, etc.

References I

1. Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y. & van der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research* **21,** 31–54 (2012).
2. Hirano, K., Imbens, G. W. & Ridder, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71,** 1161–1189 (2003).
3. D'Amour, A., Ding, P., Feller, A., Lei, L. & Sekhon, J. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221,** 644–654 (2021).
4. Zhou, Y., Matsouaka, R. A. & Thomas, L. Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research* **29,** 3721–3756 (2020).
5. Crump, R., Hotz, V. J., Imbens, G. & Mitnik, O. *Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand.* 2006.
6. Li, F., Morgan, K. L. & Zaslavsky, A. M. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* **113,** 390–400 (2018).

Thank you!

Email: yliu297@ncsu.edu

https://yiliu1998.github.io/