

# Master's Project Preliminary Defense

*Overlap, Inverse Probability, and Matching Weights: What Are  
We Weighting For?*

**Yi Liu**

Advisor: Dr. Roland Matsouaka  
Biostatistics & Bioinformatics, Duke University

January 19, 2022

# Outline

- ▶ Background
- ▶ Question
- ▶ Simulation
- ▶ Data Example

## Background

Propensity score (PS) weights are commonly used in mitigating covariate imbalance of different treatment groups in *non-randomized studies*. Controversies have emerged about the goals of these methods (weights), their estimands, and their underlying targeted populations.

# Background

## Notation

- ▶ Treatment:  $Z = z$  (0 for control, 1 for treatment)
- ▶ Covariates:  $X = (X_1, \dots, X_p)'$
- ▶ Propensity score:  $e(\mathbf{x}) = P(Z = 1 | X = \mathbf{x})$ 
  - ▶ Positivity assumption:  $0 < e(\mathbf{x}) < 1$
- ▶ Outcome:  $Y$ 
  - ▶ Potential outcome  $Y(z), z = 0, 1$
  - ▶ Observed outcome  $Y = ZY(1) + (1 - Z)Y(0)$
  - ▶ Unconfoundedness assumption:  
 $\{Y(0), Y(1)\} \perp\!\!\!\perp Z | X \Rightarrow E(Y(z)|X, Z = z) = E(Y(z)|X), z = 0, 1$

## Background

Goal: estimate  $\tau = E[\tau(X)]$ , the average treatment effect (ATE), where  $\tau(\mathbf{x}) = E[Y(1) - Y(0)|X = \mathbf{x}]$

A generalized class of *weighted* ATE (WATE)<sup>1</sup>:

$$\tau_h = \frac{E[h(\mathbf{X})\tau(\mathbf{X})]}{E[h(\mathbf{X})]} = \frac{\int h(\mathbf{x})f(\mathbf{x})\tau(\mathbf{x})d\mu(\mathbf{x})}{\int h(\mathbf{x})f(\mathbf{x})d\mu(\mathbf{x})}$$

$f(\mathbf{x})$ : density of the covariates;  $h(\mathbf{x})$ : *tilting function* which re-distributes the covariates.  $\tau$  is the special case when  $h \equiv 1$ .

---

<sup>1</sup>F. Li, K. L. Morgan, and A. M. Zaslavsky, “Balancing covariates via propensity score weighting,” *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 390–400, 2018.

## Background

- ▶ Denote  $f_z(\mathbf{x}) = P(X = \mathbf{x} | Z = z)$ ,  $z = 0, 1$
- ▶ The balancing weights ( $w_0, w_1$ ):

$$\begin{cases} w_0(\mathbf{x}) \propto \frac{h(\mathbf{x})}{1 - e(\mathbf{x})} \\ w_1(\mathbf{x}) \propto \frac{h(\mathbf{x})}{e(\mathbf{x})} \end{cases}$$

balance the distributions of the covariates in comparison groups  
 $f_0(\mathbf{x})w_0(\mathbf{x}) = f_1(\mathbf{x})w_1(\mathbf{x}) = f(\mathbf{x})h(\mathbf{x})$

# Background

**Table:** Choices of  $h$  and Corresponding Target Population and Causal Estimands

Target	$h(\mathbf{x})$	Estimand	Method
overall	1	ATE	IPW
treated	$e(\mathbf{x})$	ATT	IPWT
control	$1 - e(\mathbf{x})$	ATC	IPWC
restricted	$\mathbf{1}\{\alpha \leq e(\mathbf{x}) \leq 1 - \alpha\}$	ATE	IPW Trimming
overlap	$e(\mathbf{x})(1 - e(\mathbf{x}))$	ATO	OW
overlap	$\min\{e(\mathbf{x}), 1 - e(\mathbf{x})\}$	ATM	MW
overlap	$-[e(\mathbf{x}) \ln(e(\mathbf{x})) + (1 - e(\mathbf{x})) \ln(1 - e(\mathbf{x}))]$	ATEN	EW

IPW: inverse probability weight; OW: overlap weight; MW: matching weight; EW: entropy weight

We choose  $\alpha = 0.05, 0.1$  and  $0.15$

## Background

Observed data  $\{(X_i, Y_i, Z_i), i = 1, \dots, N\}$ .  $\tau_h$  can be estimated by the *weighted estimator*

$$\widehat{\tau}_h = \frac{\sum_{i=1}^N Z_i \widehat{w}_1(\mathbf{x}_i) Y_i}{\sum_{i=1}^N Z_i \widehat{w}_1(\mathbf{x}_i)} - \frac{\sum_{i=1}^N (1 - Z_i) \widehat{w}_0(\mathbf{x}_i) Y_i}{\sum_{i=1}^N (1 - Z_i) \widehat{w}_0(\mathbf{x}_i)}$$

$\widehat{w}_z(\mathbf{x})$ ,  $z = 0, 1$  is calculated by plugging in the estimated propensity score  $\widehat{e}(\mathbf{x})$  (usually by logistic regression).

## Background

Observed data  $\{(X_i, Y_i, Z_i), i = 1, \dots, N\}$ . The *augmented estimator* of  $\tau_h$  is given by

$$\hat{\tau}_h^{aug} = \frac{\sum_{i=1}^N h(\mathbf{x}_i) \{ \widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i) \}}{\sum_{i=1}^N h(\mathbf{x}_i)} +$$
$$\frac{\sum_{i=1}^N Z_i \widehat{w}_1(\mathbf{x}_i) \{ Y_i - \widehat{m}_1(\mathbf{x}_i) \}}{\sum_{i=1}^N Z_i \widehat{w}_1(\mathbf{x}_i)} - \frac{\sum_{i=1}^N (1 - Z_i) \widehat{w}_0(\mathbf{x}_i) \{ Y_i - \widehat{m}_0(\mathbf{x}_i) \}}{\sum_{i=1}^N (1 - Z_i) \widehat{w}_0(\mathbf{x}_i)}$$

where  $m_z(X) = E(Y(z)|X)$ ,  $z = 0, 1$ , an outcome regression (OR) model.

## Question

- ▶ What to expect when using overlap weights (OW), matching weights (MW), or entropy weights (EW) and compare to IPW weights
- ▶ What is the role of the proportion of participants in the treatment groups ( $p = P(Z = 1)$ ) in estimating these quantities
  - ▶ We know:  $\text{ATE} = p\text{ATT} + (1 - p)\text{ATC}$  (weight ATT/ATC more when  $p$  is high/small)
  - ▶ Our hunch: Overlap estimators (ATO, ATM, and ATEN) weight on the *opposite* direction
  - ▶ Our hunch: When  $p \approx 0.5$  and positivity satisfied, they all have similar estimates
- ▶ We target to show these results under finite sample sizes through simulations

# Question

Why do we care?

- ▶ Either ATT or ATC can be a primary interest in different questions, especially for policy-related purposes
  - ▶ ATT: more informative to assess the treatment on the treated participants
  - ▶ ATC: help to evaluate the impact of rolling out a treatment to those who do not receive it

# Question

Why do we care?

- ▶ What to expect for a WATE when we have different objectives on ATC/ATT and population of interest
  - ▶ E.g. When exposure is rare (small  $p$ ) and the treated population is of interest, ATT is encouraged and ATE reaches extreme values and larger biases<sup>2</sup>
  - ▶ Then who (in the WATE class) are closer to ATT which might also be useful in this case?
- ▶ Want to investigate the sensitivity of these estimands to model misspecifications (the truth of PS and OR model are unknown)

---

<sup>2</sup>D. Hajage, F. Tubach, P. G. Steg, *et al.*, “On the use of propensity scores in case of rare exposure,” *BMC medical research methodology*, vol. 16, no. 1, p. 38, 2016.

# Simulation

Data generating process (DGP). We follow Li and Li<sup>3</sup> for generating covariates:

- ▶  $X_4 \sim \text{Bern}(0.5)$ ,  $X_3 \sim \text{Bern}(0.4 + 0.2X_4)$
- ▶  $(X_1, X_2)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  where

$$\boldsymbol{\mu} = (X_4 - X_3 + 0.5X_3X_4, X_3 - X_4 + X_3X_4)',$$

$$\boldsymbol{\Sigma} = X_3 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} + X_4 \begin{pmatrix} 2 & 0.25 \\ 0.25 & 2 \end{pmatrix}$$

- ▶  $X_5 = X_1^2$ ,  $X_6 = X_1X_2$ , and  $X_7 = X_2^2$

---

<sup>3</sup>Y. Li and L. Li, “Propensity score analysis methods with balancing constraints: A monte carlo study,” *Statistical Methods in Medical Research*, vol. 30, no. 4, pp. 1119–1142, 2021.

# Simulation

Data generating process (DGP):

- ▶ Treatment:  $Z \sim \text{Bern}(\text{expit}(X\beta))$ , where  
 $\beta = (\beta_0, 0.3, 0.4, 0.4, 0.4, -0.1, -0.1, 0.1)'$
- ▶  $\beta_0$  is varying for having different proportions of treatment in different simulated data
- ▶ Outcome regression:  
$$Y(0) = 0.5 + X_1 + 0.6X_2 + 2.2X_3 - 1.2X_4 + (X_1 + X_2)^2 + \varepsilon \text{ and}$$
$$Y(1) = Y(0) + \delta(X), \text{ for } \varepsilon \sim N(0, 4)$$
  - ▶ Constant treatment effect:  $\delta(X) = 4$
  - ▶ Heterogeneous treatment effect:  $\delta(X) = 4 + 3(X_1 + X_2)^2 + X_1 X_3$
- ▶  $M = 2000$  iterations, size  $N = 1000$  for each simulated data

# Simulation

Models:

- ▶ Propensity score (PS) models
  - ▶ Misspecified model: exclude term  $X_5$  to  $X_7$
- ▶ Outcome regression (OR) models
  - ▶ Misspecified model: exclude term  $(X_1 + X_2)^2$

Misspecified models are considered only if we use augmented estimators, and here are 4 cases:

- ▶ Both PS and OR models are correctly specified
- ▶ Only PS model is correctly specified
- ▶ Only OR model is correctly specified
- ▶ Both PS and OR models are misspecified

Overlap weights (OW, MW, EW) are more robust to model misspecification than IPW when using weighted estimator<sup>4</sup>.

---

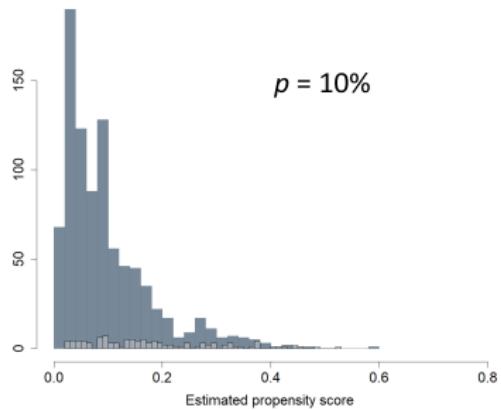
<sup>4</sup>Y. Zhou, R. A. Matsouaka, and L. Thomas, "Propensity score weighting under limited overlap and model misspecification," *Statistical Methods in Medical Research*, vol. 29, no. 12, pp. 3721–3756, 2020.

## Simulation

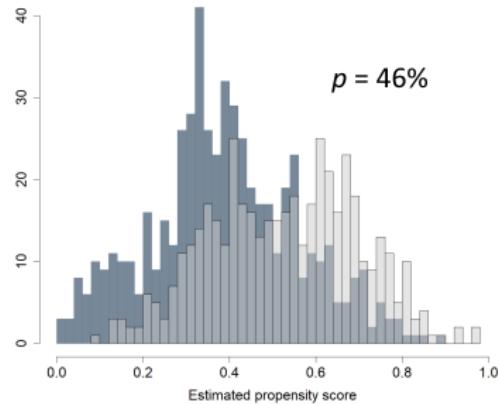
We present results from 3 PS models that simulate different proportions  $p$  of subjects in treatment group:

- ▶ Model 1:  $\beta_0 = -3.07$ ,  $p = 10\%$
- ▶ Model 2:  $\beta_0 = -0.78$ ,  $p = 46\%$
- ▶ Model 3:  $\beta_0 = 1.86$ ,  $p = 89\%$

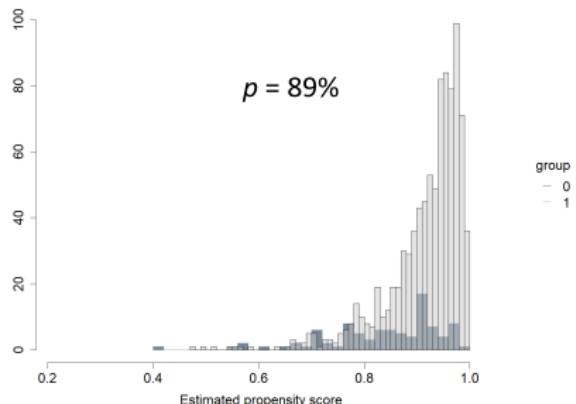
# Simulation - Propensity Score



$p = 10\%$



$p = 46\%$



$p = 89\%$

group  
— 0  
— 1

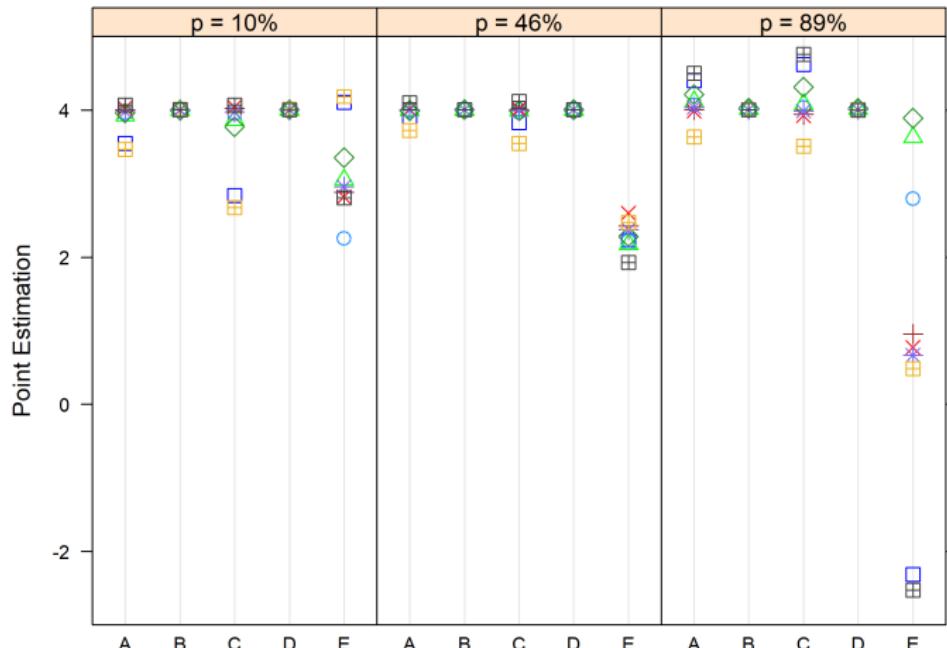
# Simulation - Point Estimation

Constant treatment effect:

- Truth:  $\tau_h = 4$

ATE      ATE (0.05)      ATE (0.1)  
□      ○      ▲  
ATO      ATM  
△      ✕

ATEN      ATC      ATT  
\*      ■      ▨  
A: Weighted Estimator  
B: Augmented - Both correct  
C: Augmented - PS correct  
D: Augmented - OR correct  
E: Augmented - Both misspecified



# Simulation - Point Estimation

Heterogeneous treatment effect:

► Truth:

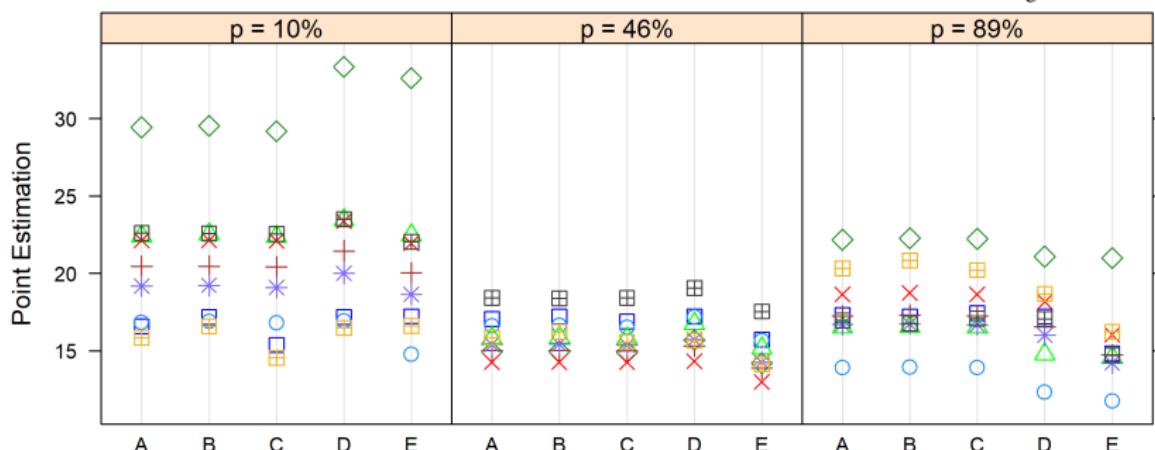
$p$	ATE	ATE (0.05)	ATE (0.1)	ATE (0.15)	ATO	ATM	ATEN	ATC	ATT
10%	17.22	16.61	22.79	30.46	20.53	22.28	19.26	16.62	22.59
46%	17.22	16.74	15.98	15.06	15.07	14.25	15.48	16.26	18.34
89%	17.22	13.67	16.59	22.78	17.36	18.84	16.78	20.79	16.78

ATE  
ATE (0.05)  
ATE (0.1)

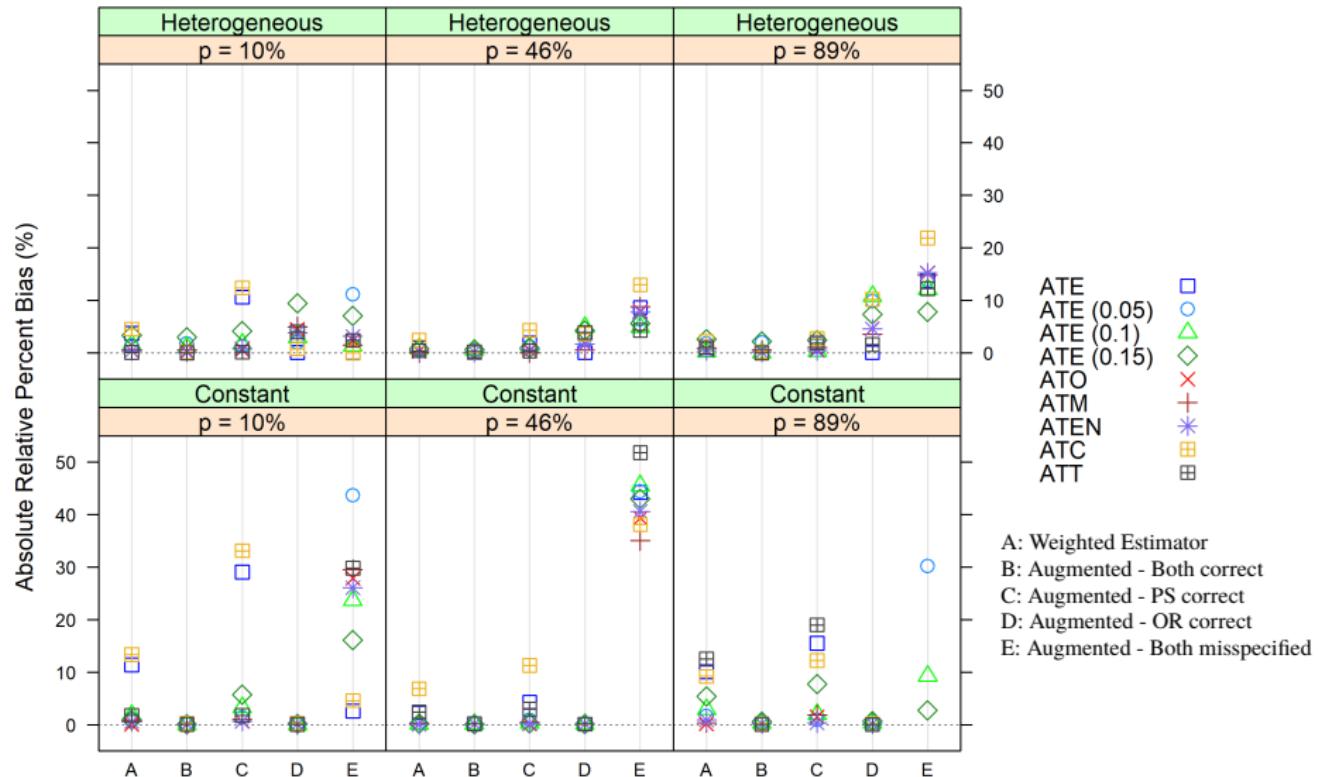
ATO  
ATM

ATEN  
ATT

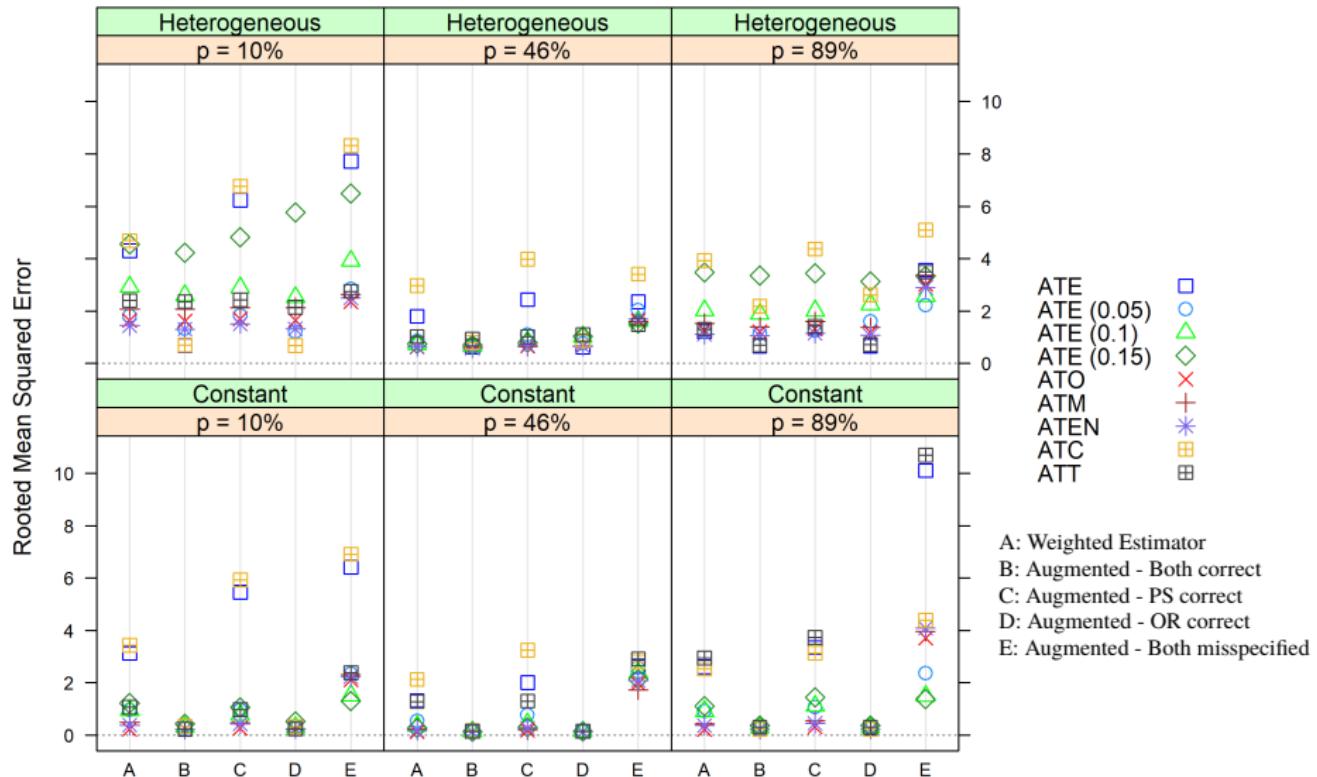
A: Weighted Estimator  
B: Augmented - Both correct  
C: Augmented - PS correct  
D: Augmented - OR correct  
E: Augmented - Both misspecified



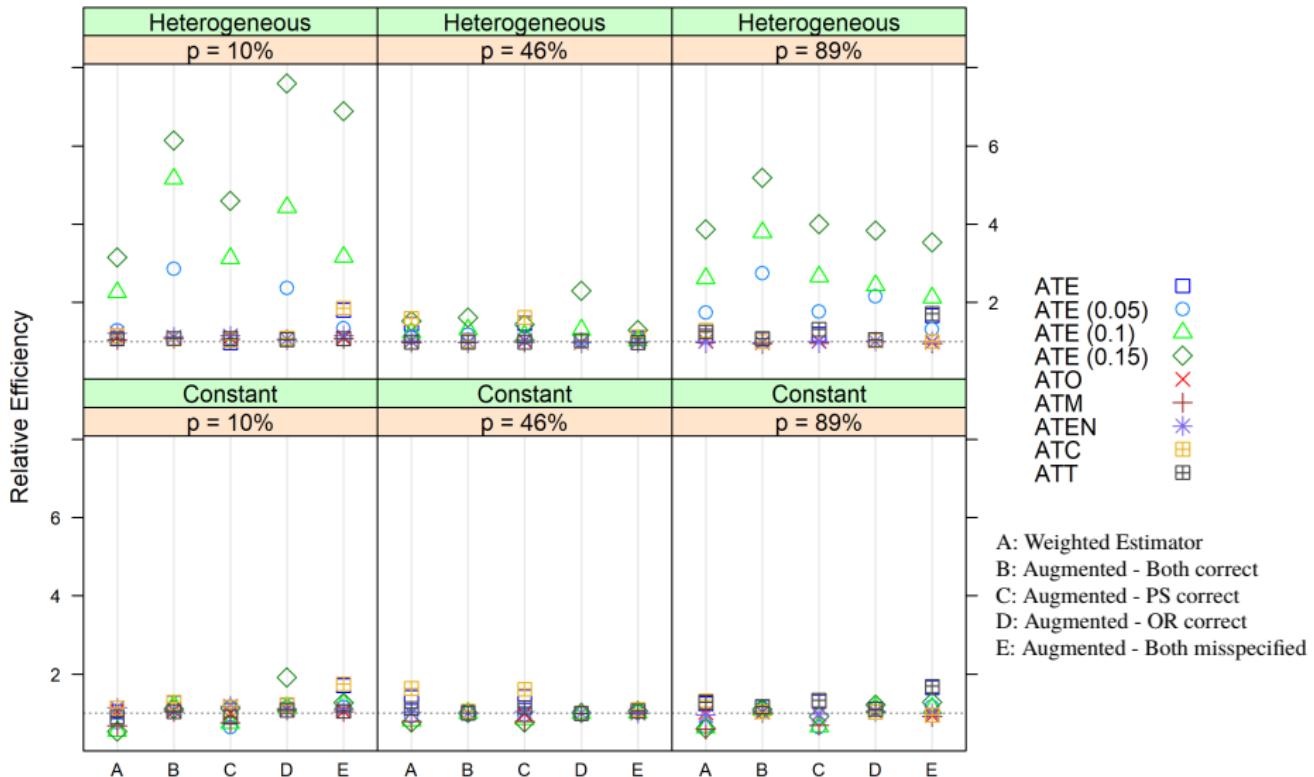
# Simulation - ARBias (%)



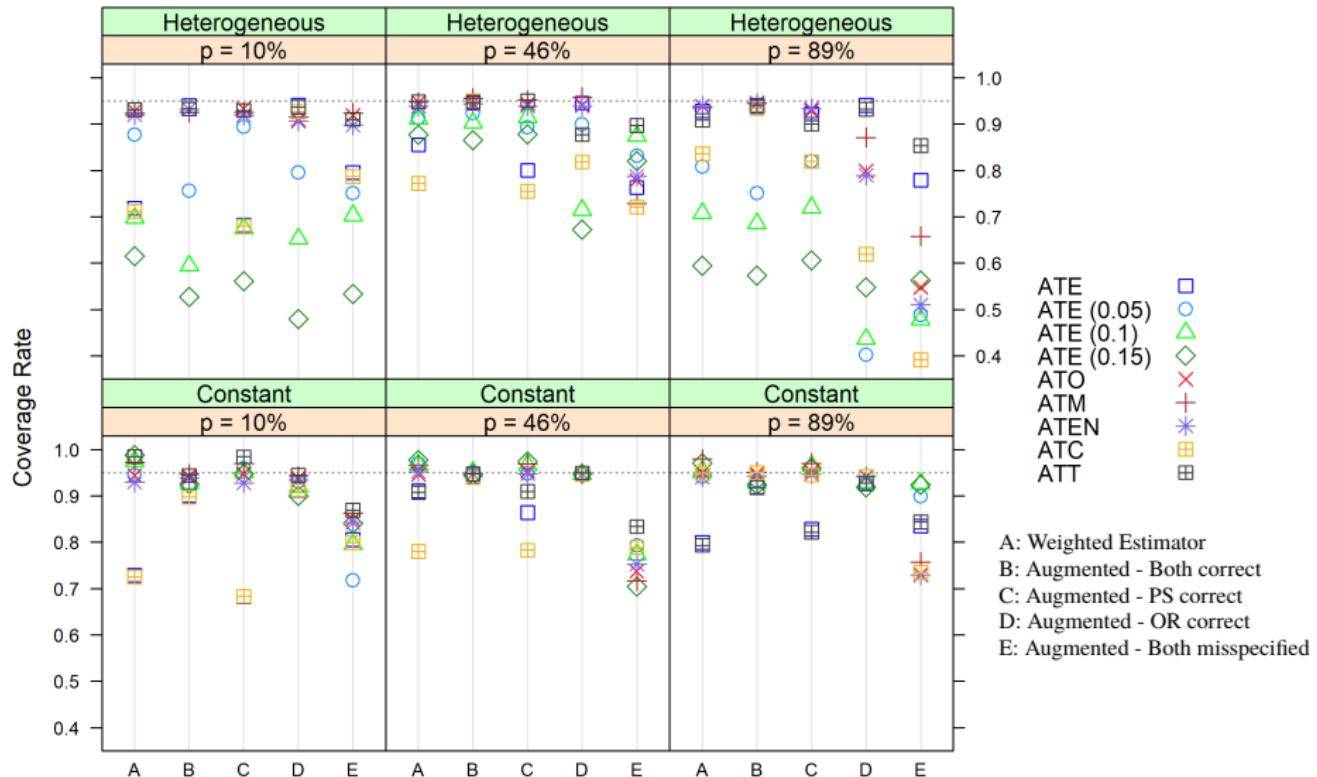
# Simulation - RMSE



# Simulation - Relative Efficiency



# Simulation - Coverage Rate



## Simulation - Summary

- ▶ When  $p = P(Z = 1)$  is high, ATE weights ATT more and overlap estimators (ATO, ATM and ATEN) weight ATC more, and vice versa
- ▶ When  $p \approx 0.5$ , about half of subjects receive the treatment and no lack of positivity, they all have similar estimates
- ▶ The ARBias and RMSE indicate that the IPW estimates (with/without trimming) in general have larger biases and errors in estimating corresponding estimands than overlap estimates in each case
- ▶ The differences among overlap augmented estimates are slight
- ▶ When both PS and OR models are misspecified, all augmented estimates have larger biases and errors than them in other cases
- ▶ The relative efficiency and coverage rate suggest the closed-form (asymptotic) sandwich variance estimations work better for overlap estimators

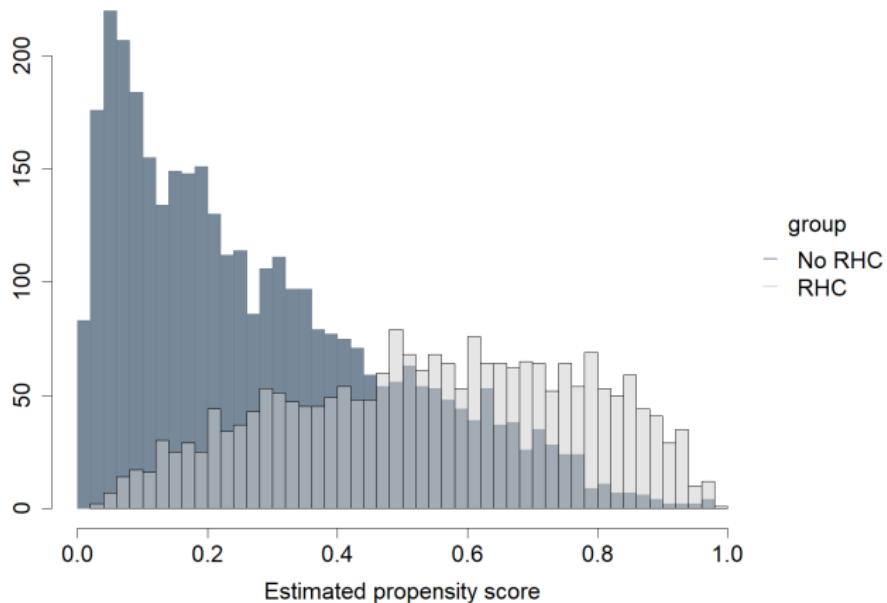
## Data Example

A right heart catheterization (RHC) data (Openly accessible at: <https://hbiostat.org/data/>). It is to investigate the effectiveness of the RHC diagnostic procedure during the initial care of hospitalized, critically ill patients.

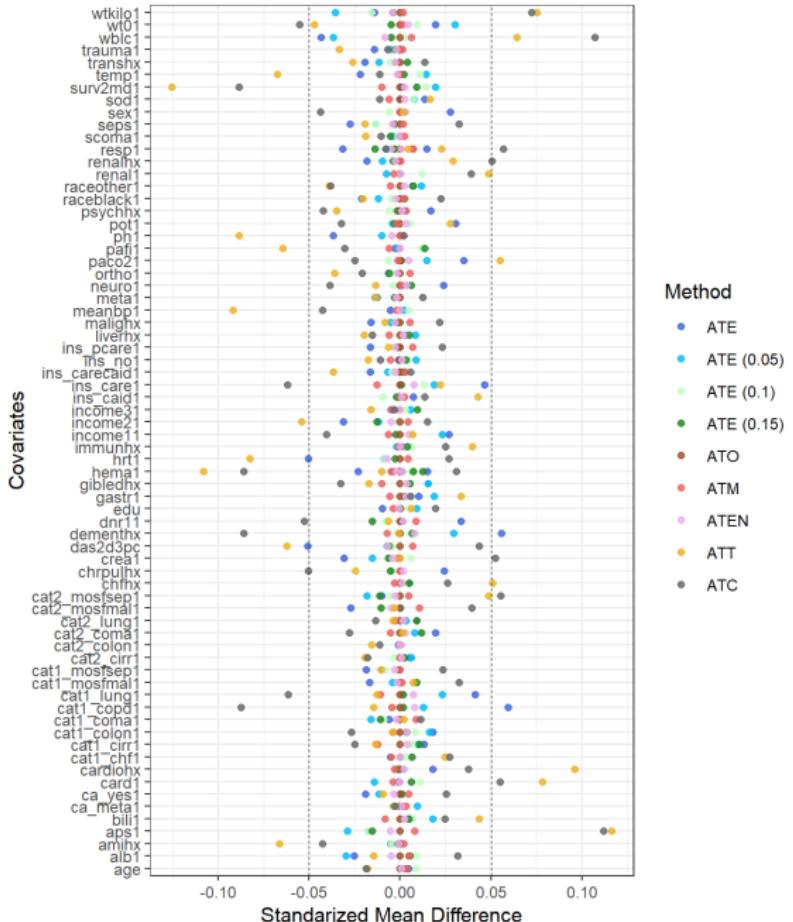
- ▶ Treatment: The RHC diagnostic procedure
  - ▶ 5735 hospitalized patients enrolled
  - ▶ **2184 (38%)** patients received the RHC treatment ( $Z = 1$ )
- ▶ Outcome: Patient's length of stay in the intensive care unit (ICU) during the first 24 hours
- ▶ 72 covariates are considered (continuous, categorical), such as age, race and some medical/biological indices, in both PS and OR models

## Data Example - Propensity Score

PS model:  $\text{logit}(e(X)) = \beta_0 + X\beta$



# Data Example - Covariates Balance



# Data Example - Causal Effects

Table: Treatment Effects of the RHC Procedure on Patients

Estimand	Prop.	Weighted Estimator			Augmented Estimator		
		Est.	SE	p-value	Est.	SE	p-value
ATE	100%	0.130	0.032	<0.001	0.129	0.033	<0.001
ATE (0.05)	93%	0.099	0.030	<0.001	0.102	0.030	<0.001
ATE (0.1)	82%	0.102	0.029	<0.001	0.100	0.029	<0.001
ATE (0.15)	73%	0.078	0.029	0.008	0.079	0.029	0.007
ATO	100%	0.095	0.028	<0.001	0.098	0.028	<0.001
ATM	100%	0.094	0.028	<0.001	0.095	0.028	<0.001
ATEN	100%	0.100	0.028	<0.001	0.102	0.028	<0.001
ATC	100%	0.156	0.035	<0.001	0.148	0.037	<0.001
ATT	100%	0.090	0.046	0.049	0.099	0.043	0.021

Prop.: proportion of sample used; Est.: point estimation; SE: standard error

## Conclusion

- ▶ Overlap estimands and ATE weight ATC and ATT on opposite directions based on simulations from different  $p$  cases
- ▶ Overlap (augmented) estimators are in general more robust to the model misspecifications
  - ▶ Lower ARBias and RMSE
  - ▶ Better relative efficiency and coverage rate
- ▶ The RHC data analysis further confirms our findings

## References

-  F. Li, K. L. Morgan, and A. M. Zaslavsky, “Balancing covariates via propensity score weighting,” *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 390–400, 2018.
-  D. Hajage, F. Tubach, P. G. Steg, D. L. Bhatt, and Y. De Rycke, “On the use of propensity scores in case of rare exposure,” *BMC medical research methodology*, vol. 16, no. 1, p. 38, 2016.
-  Y. Li and L. Li, “Propensity score analysis methods with balancing constraints: A monte carlo study,” *Statistical Methods in Medical Research*, vol. 30, no. 4, pp. 1119–1142, 2021.
-  Y. Zhou, R. A. Matsouaka, and L. Thomas, “Propensity score weighting under limited overlap and model misspecification,” *Statistical Methods in Medical Research*, vol. 29, no. 12, pp. 3721–3756, 2020.

# Thank you!

Yi Liu

[yl724@duke.edu](mailto:yl724@duke.edu)