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, 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: $\boldsymbol{X} = (X_1, ..., X_p)'$
- Propensity score: $e(\boldsymbol{X}) = P(Z = 1|\boldsymbol{X})$
  - Positivity assumption: $0 < e(\boldsymbol{X}) < 1$
- Outcome: Y
  - Potential outcome $Y(z), z = 0, 1$
  - Observed outcome $Y = ZY(1) + (1 - Z)Y(0)$
  - Unconfoundness assumption:
    $\{Y(0), Y(1)\} \perp Z|\boldsymbol{X} \Rightarrow E(Y(z)|\boldsymbol{X}, Z = z) = E(Y(z)|\boldsymbol{X}), z = 0, 1$

Background

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

A generalized class of *weighted* ATE (WATE)[1]:

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

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

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

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

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

Background

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

Target	$h(\boldsymbol{x})$	Estimand	Method
overall	1	ATE	IPW
treated	$e(\boldsymbol{x})$	ATT	IPWT
control	$1-e(\boldsymbol{x})$	ATC	IPWC
restricted	$\mathbf{1}\{\alpha \leq e(\boldsymbol{x}) \leq 1-\alpha\}$	ATE	IPW Trimming
overlap	$e(\boldsymbol{x})(1-e(\boldsymbol{x}))$	ATO	OW
overlap	$\min\{e(\boldsymbol{x}), 1-e(\boldsymbol{x})\}$	ATM	MW
overlap	$-[e(\boldsymbol{x})\ln(e(\boldsymbol{x}))$ $+(1-e(\boldsymbol{x}))\ln(1-e(\boldsymbol{x}))]$	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 $\{(\boldsymbol{X}_i, Y_i, Z_i), i = 1, \ldots, N\}$. τ_h can be estimated by the *weighted estimator*

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

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

Background

Observed data $\{(\boldsymbol{X}_i, Y_i, Z_i), i = 1, \ldots, N\}$. The *augmented estimator* of τ_h is given by

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

where $m_z(\boldsymbol{X}) = E(Y(z)|\boldsymbol{X}),\ z = 0, 1$, 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
  - Our hunch: ATE = pATT +$(1 - p)$ATC, while the 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 primary interest in different questions
- Want to know which estimand(s) is useful when we have different p in our study, i.e., which are closer to ATC/ATT
- Want to investigate the effect of model misspecifications to these estimands

Simulation

Data generating process (DGP). We follow Li and Li[2] 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 $X_7 = X_2^2$

[2] 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: $Z \sim \text{Bern}(\text{expit}(\boldsymbol{X}\beta))$, where $\beta = (\beta_0, 0.3, 0.4, 0.4, 0.4, -0.1, -0.1, 0.1)'$
- β_0 is varying for having different proportions of treatment in different simulated data
- Outcome regression: $Y(0) = 0.5 + X_1 + 0.6X_2 + 2.2X_3 - 1.2X_4 + (X_1 + X_2)^2 + \varepsilon$ and $Y(1) = Y(0) + \delta(\boldsymbol{X})$, for $\varepsilon \sim N(0, 4)$
  - Constant treatment effect: $\delta(\boldsymbol{X}) = 4$
  - Heterogeneous treatment effect: $\delta(\boldsymbol{X}) = 4 + 3(X_1 + X_2)^2 + X_1 X_3$
- $M = 2000$ 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 sepcified
- 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[3].

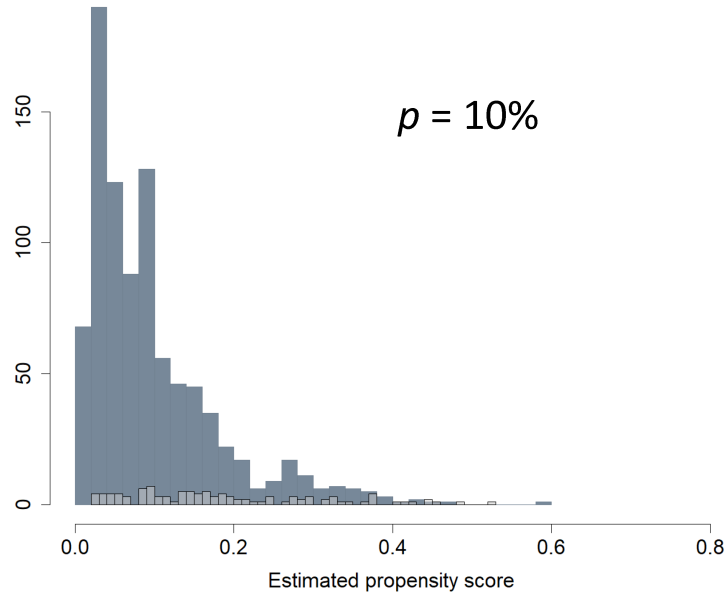
[3] Y. Zhou, R. A. Matsouaka, and L. Thomas, "Propensity score weighting under limited overlap and model misspecification," *Statistical Methods in Medical Research*, vol. 29, no. 12, pp. 3721–3756, 2020.

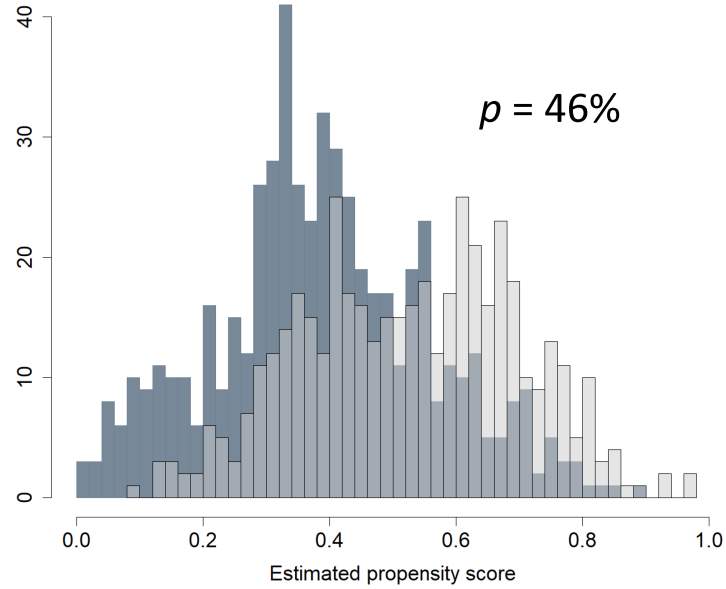
Simulation

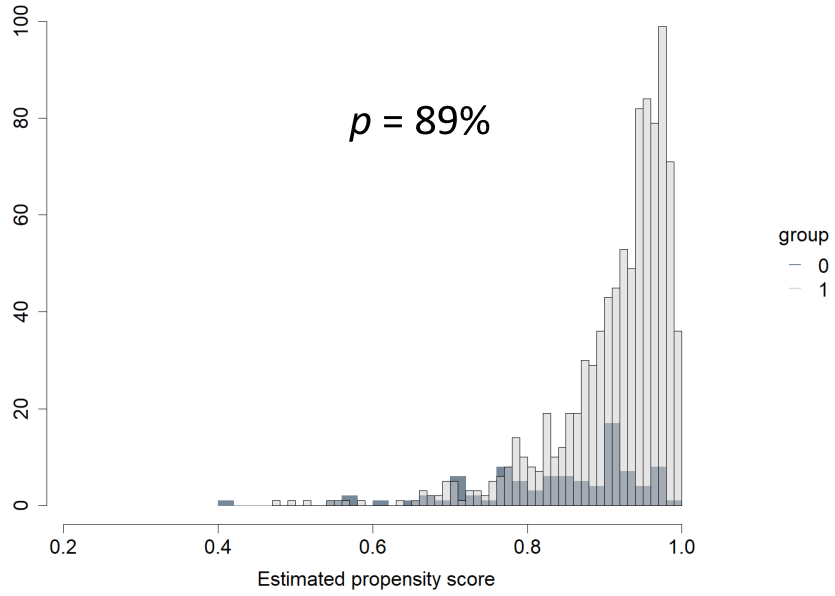
We present results from 3 PS models that simulate different proportions p of subjects in treatment group:

- 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





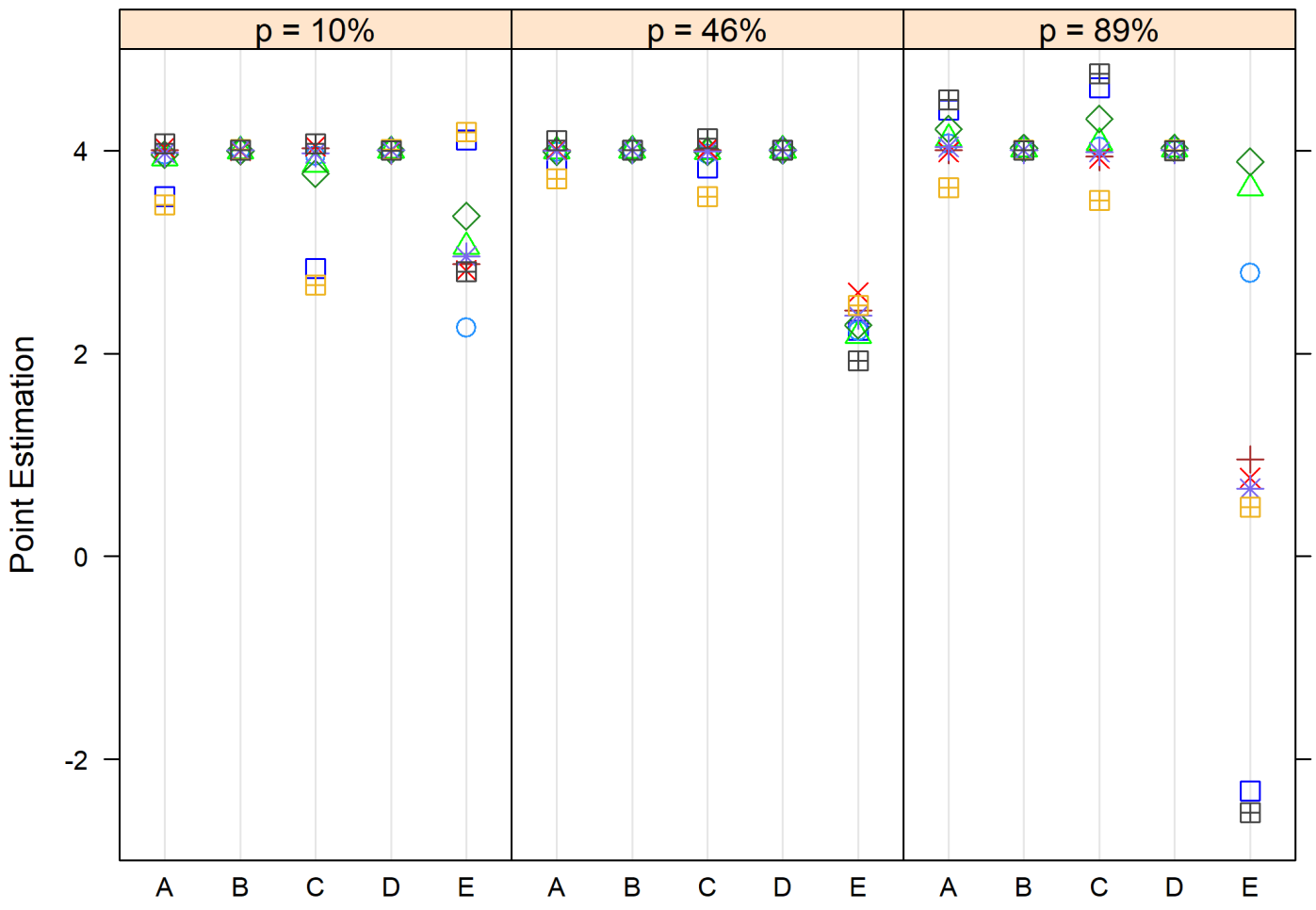


Simulation - Point Estimation

Constant treatment effect:

- Truth: $\tau_h = 4$

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

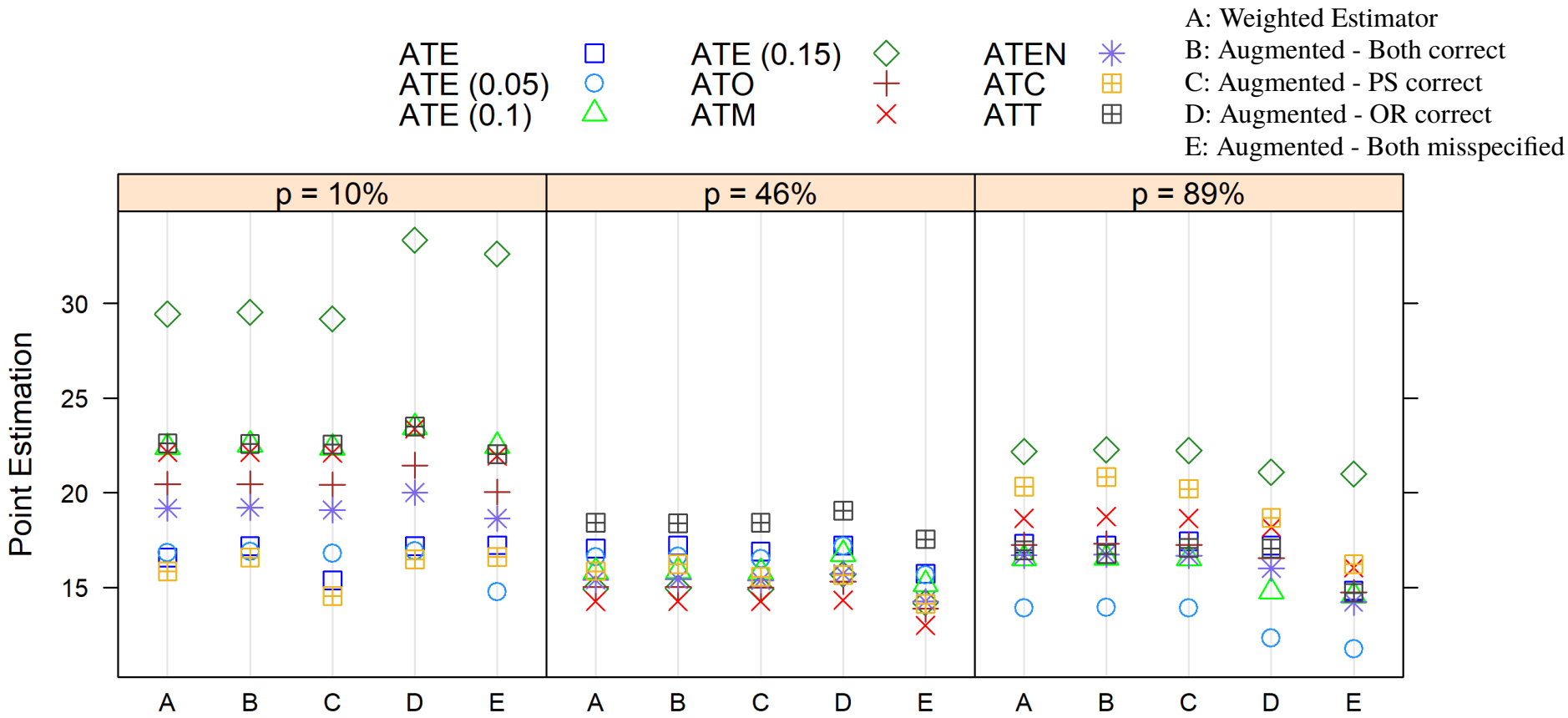


Simulation - Point Estimation

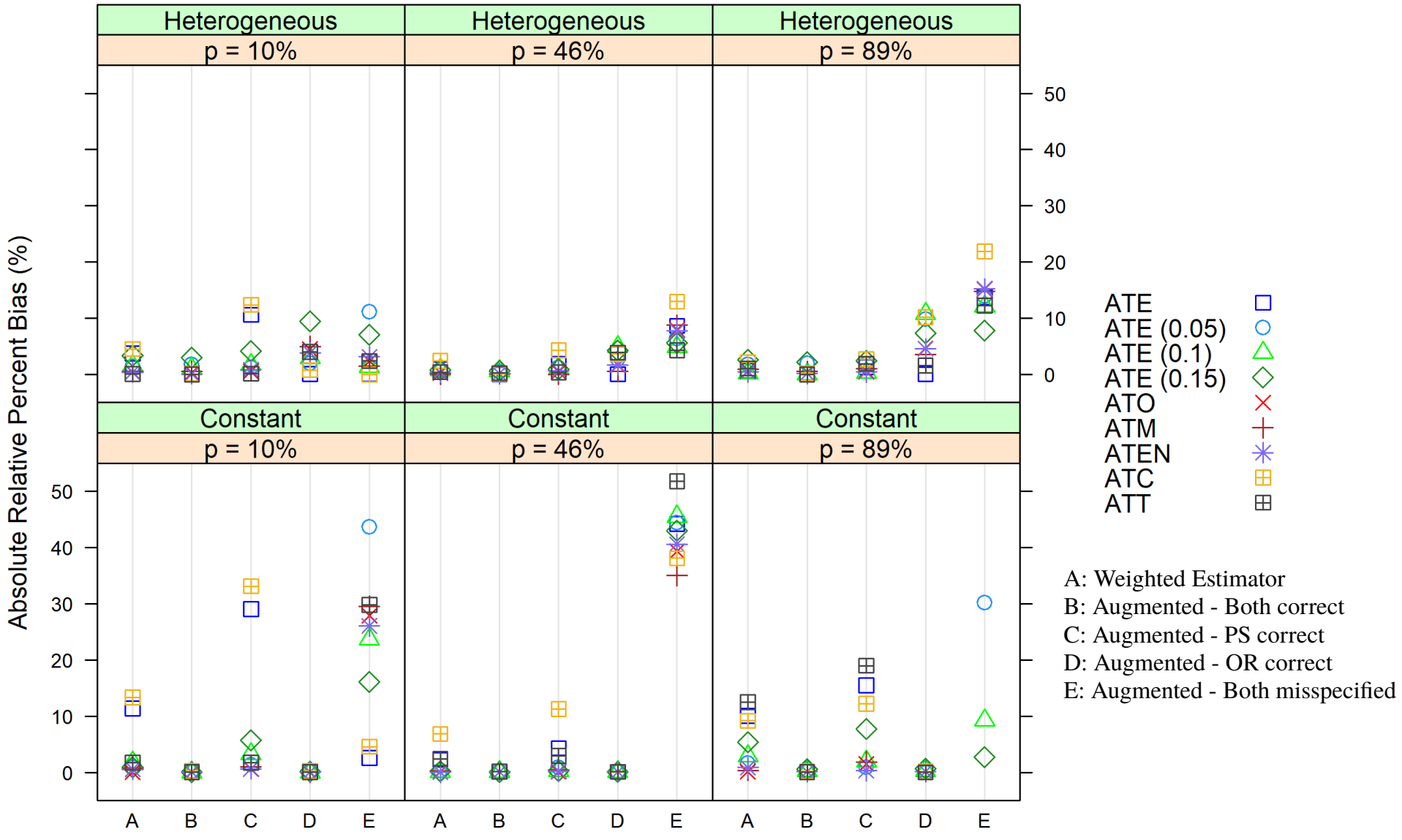
Heteorgeneous treatment effect:

- Truth:

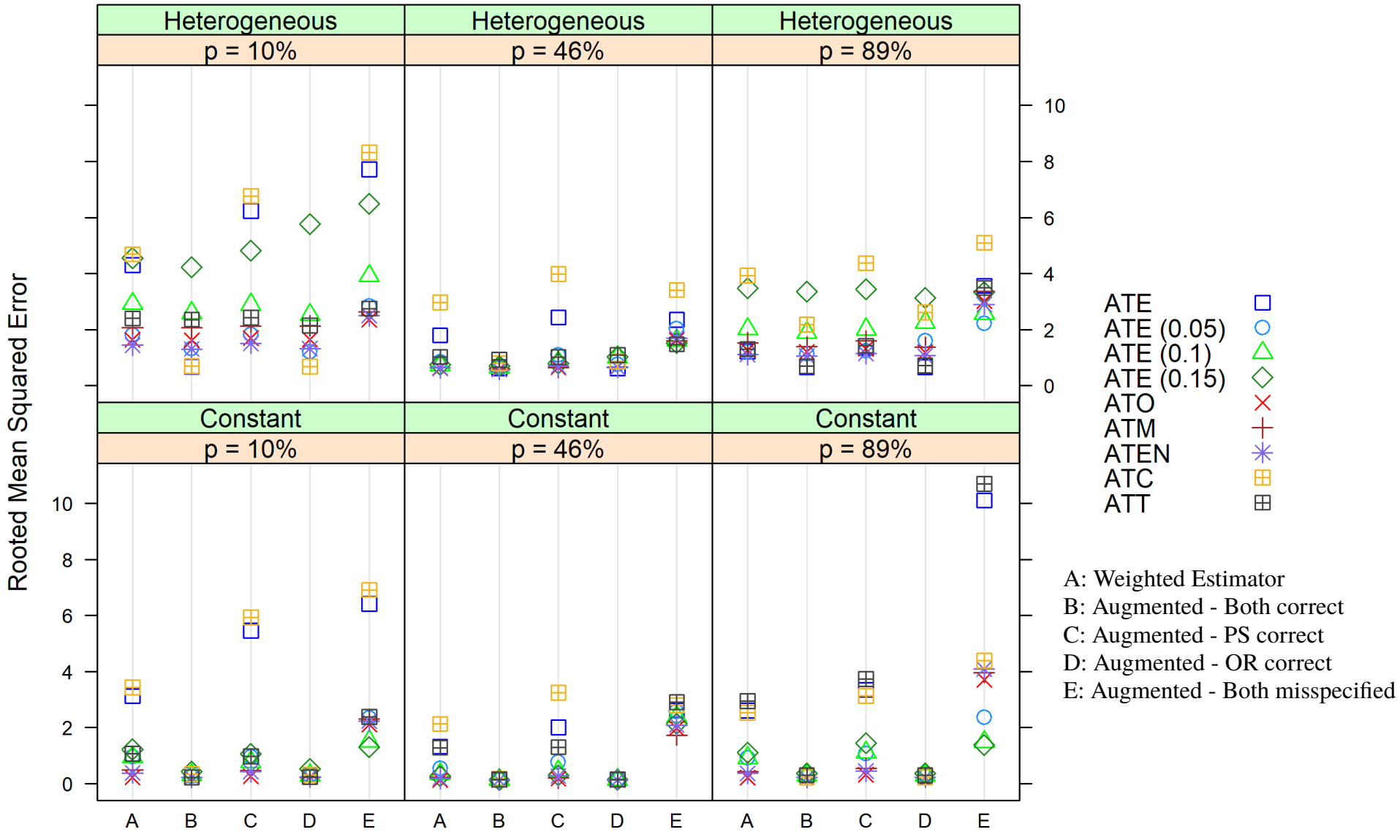
p	ATE	ATE (0.05)	ATE (0.1)	ATE (0.15)	ATO	ATM	ATEN	ATC	ATT
10%	17.22	16.61	22.79	30.46	20.53	22.28	19.26	16.62	22.59
46%	17.22	16.74	15.98	15.06	15.07	14.25	15.48	16.26	18.34
89%	17.22	13.67	16.59	22.78	17.36	18.84	16.78	20.79	16.78



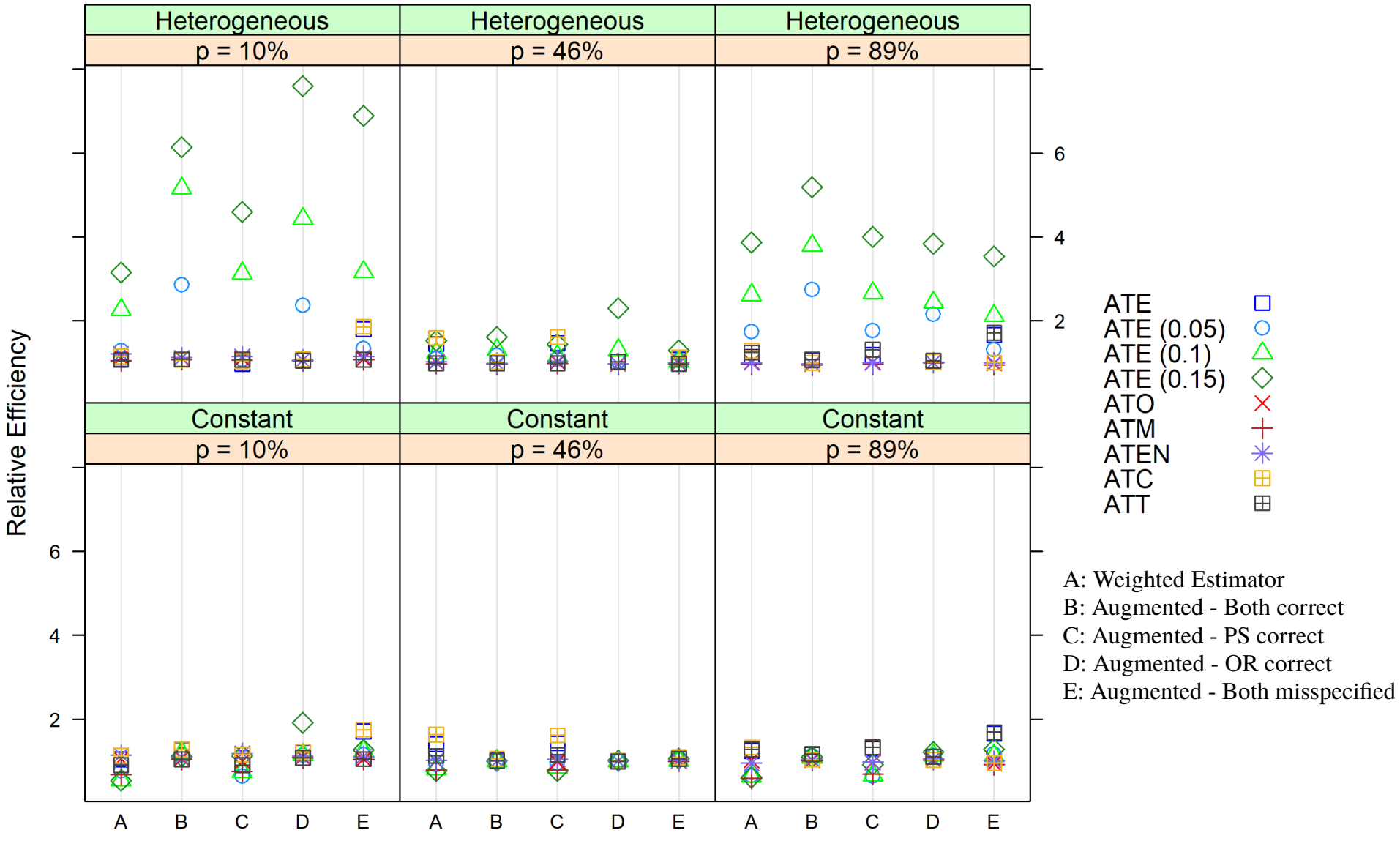
Simulation - ARBias (%)



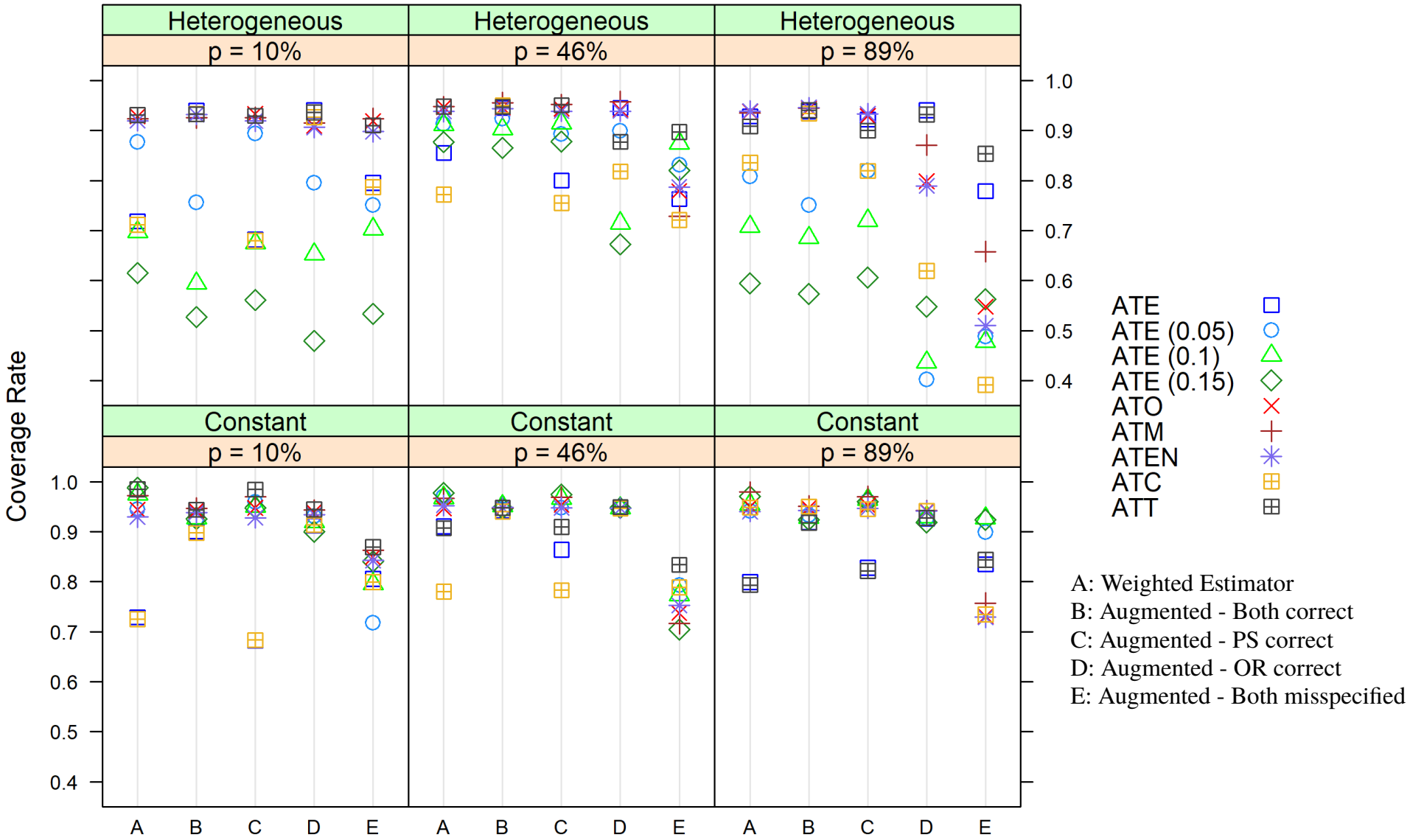
Simulation - RMSE



Simulation - Relative Efficiency



Simulation - Coverage Rate



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

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 receiving the treatment and no lack of positivity, they all have similar estimates
- When both PS and OR models are misspecified, all estimators have relatively large biases
- The relative efficiency and coverage rate for overlap estimators (ATO, ATM, ATEN) are better in each cases than IPW (with/without trimming) estimators

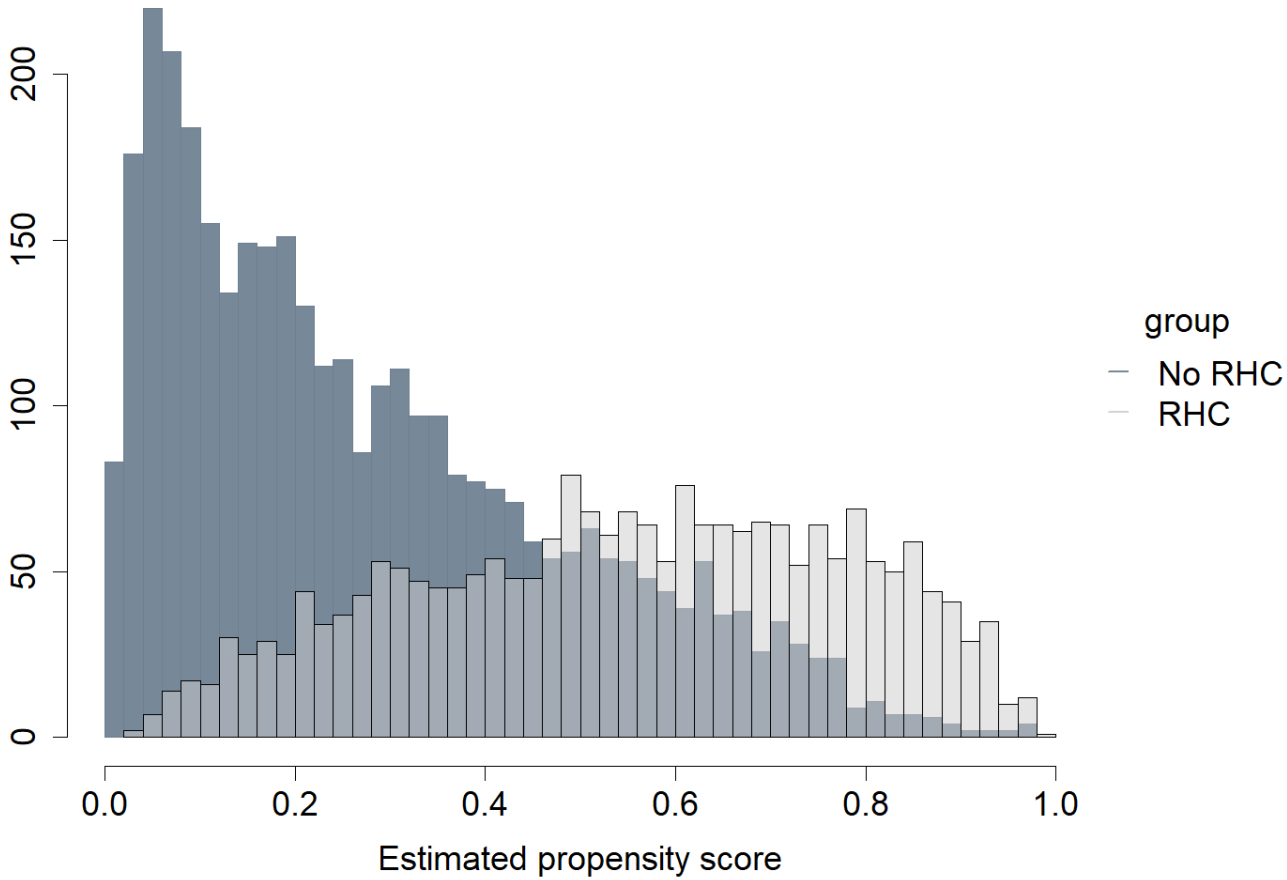
Data Example

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.

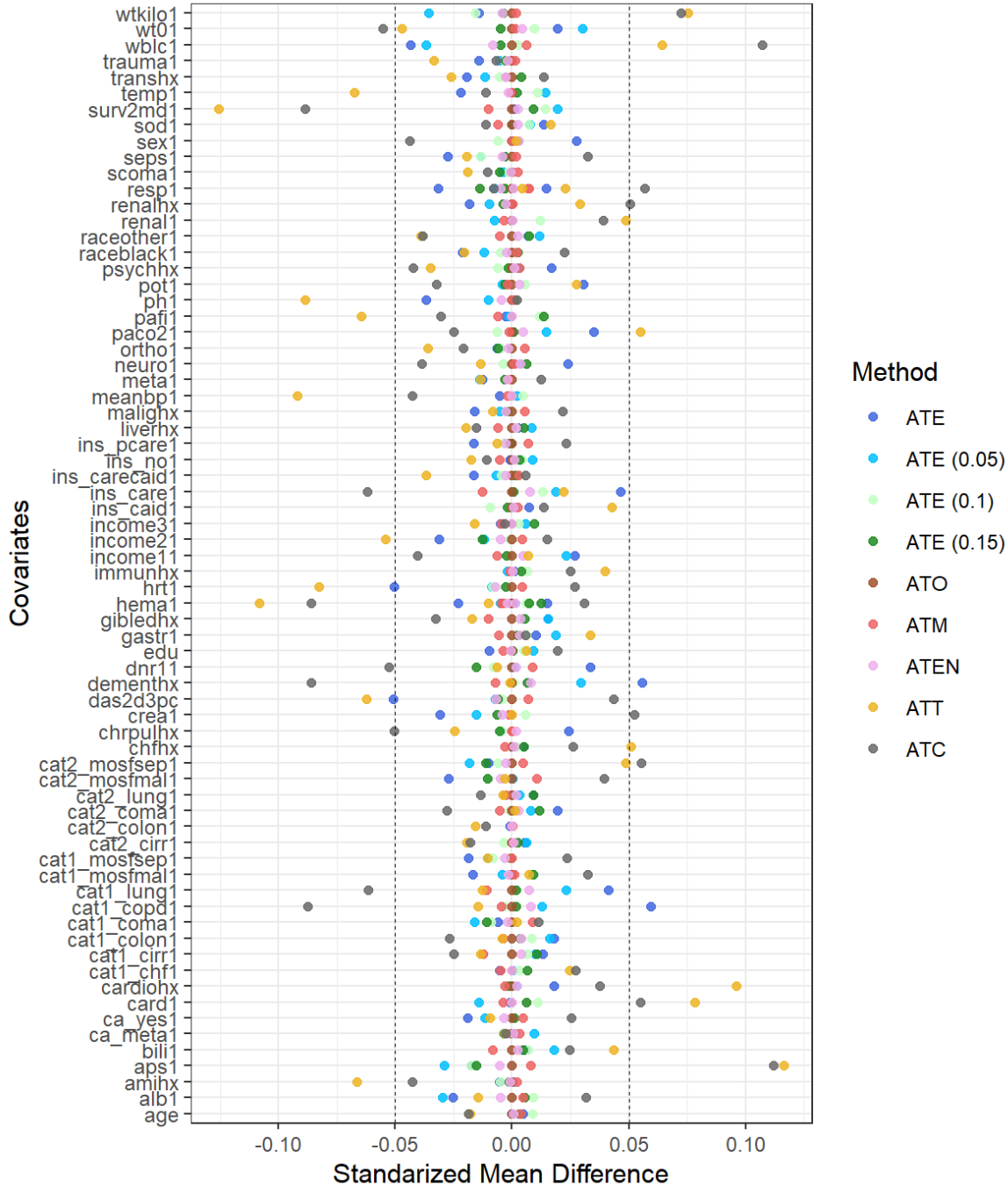
- Treatment: The RHC diagnostic procedure
  - 5735 hospitalized patients enrolled
  - **2184 (38%)** patients received the RHC treatment ($Z = 1$)
- Outcome: Patient's length of stay in the intensive care unit (ICU) during the first 24 hours
- 72 covariates are considered (continuous, categorical), such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

Data Example - Propensity Score

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



Data Example - Covariates Balance



Data Example - Causal Effects

Table: Treatment Effects of the RHC Procedure on Patients

		Weighted Estimator			**Augmented Estimator**		
Estimand	Prop.	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 the opposite directions
- Overlap (augmented) estimators are in general more robust to model misspcifications
  - The ARBias and RMSE indicate the IPW trimming weights always have larger errors when estimating the corresponding estimands, under model misspecifications
  - The relative effeciency and coverage rate suggest the closed-form (asymptotic) variance estimations work better for overlap estimators
- The RHC data application 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.


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