

Exploring finite-sample bias in propensity score weights

LUCY D'AGOSTINO MCGOWAN

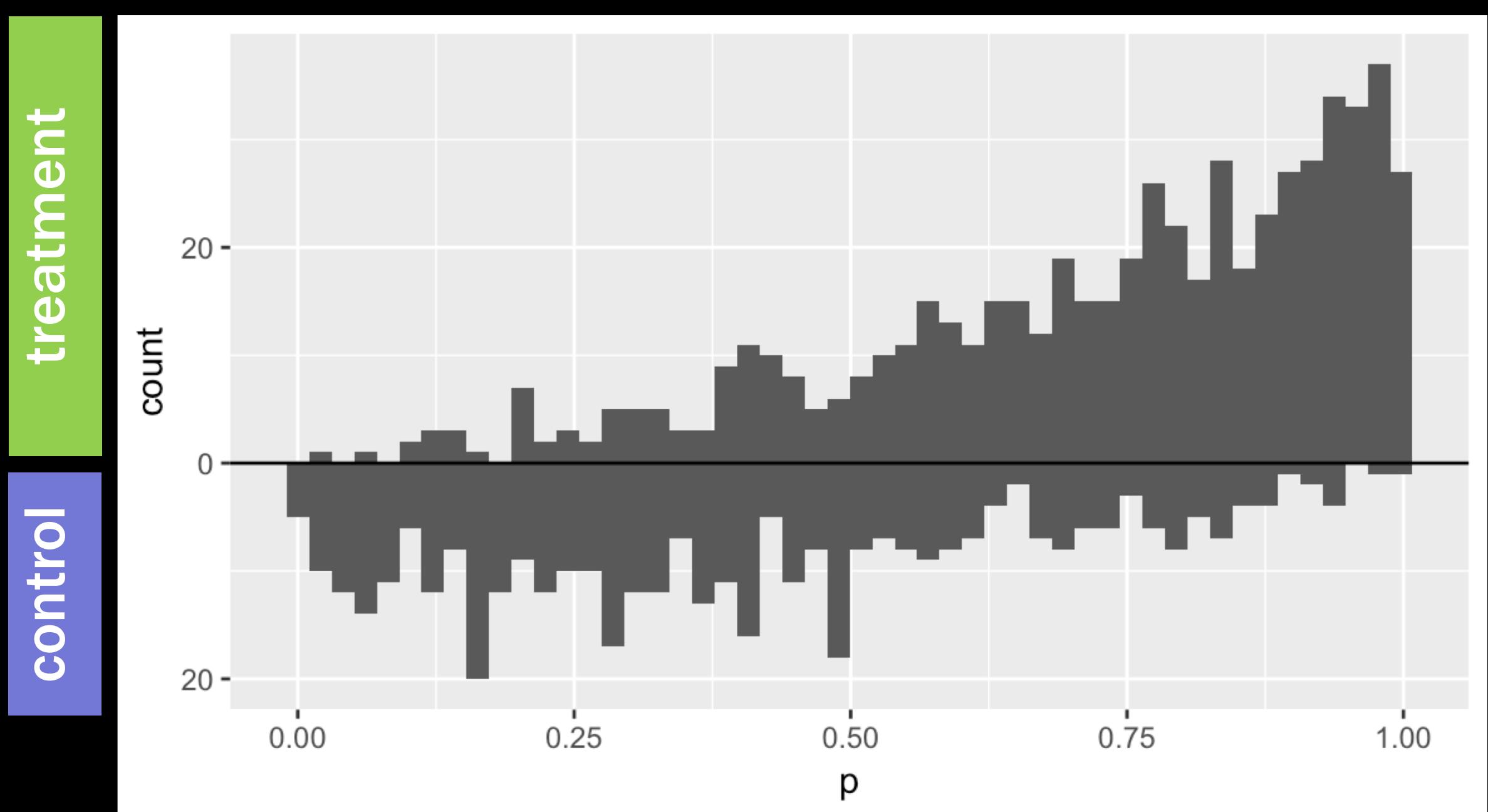
&

ROBERT GREEVY, JR.

Overview

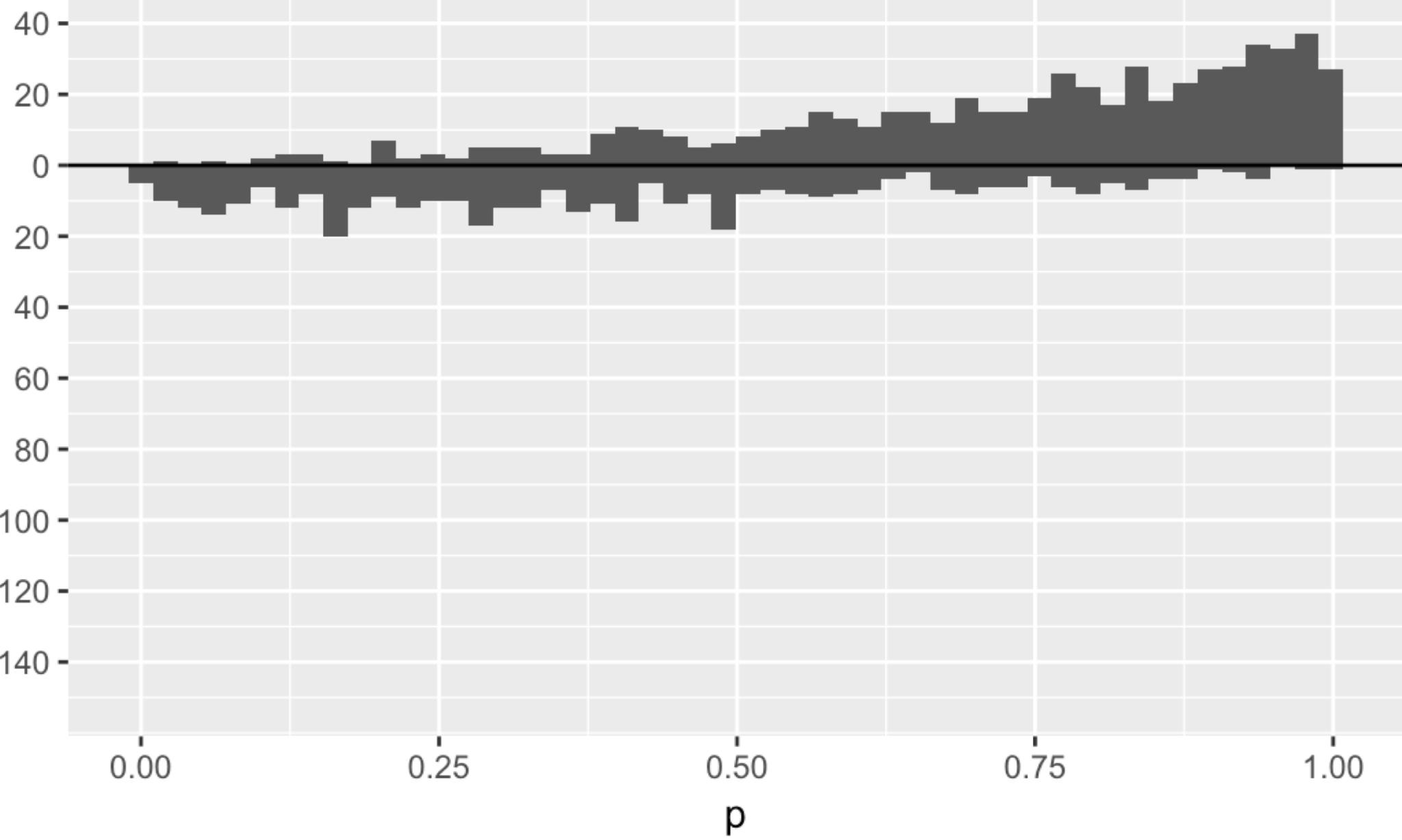
Overview

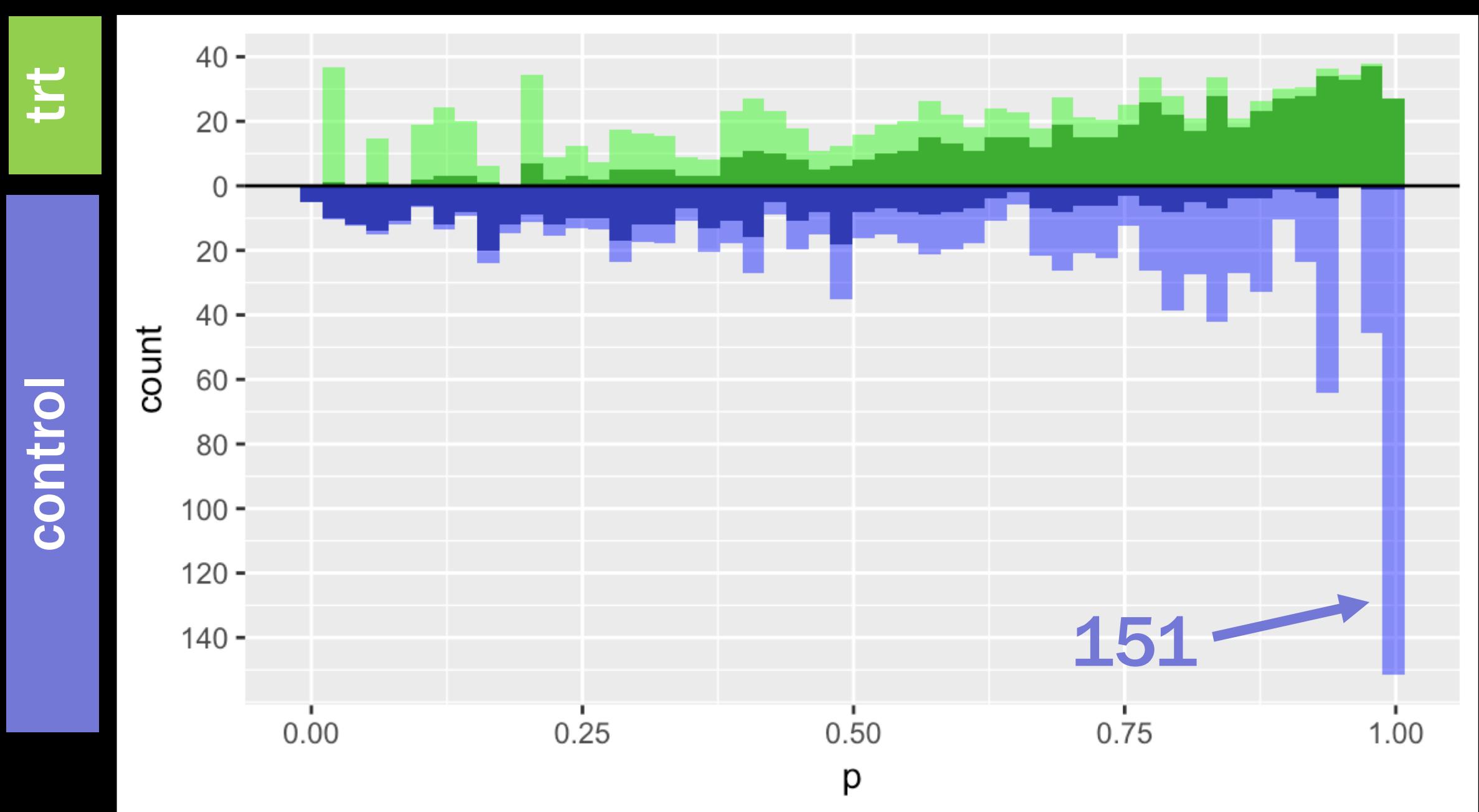
- 1 Finite sample bias**
- 2 Unmeasured confounding – the problem**
- 3 Unmeasured confounding – a solution**

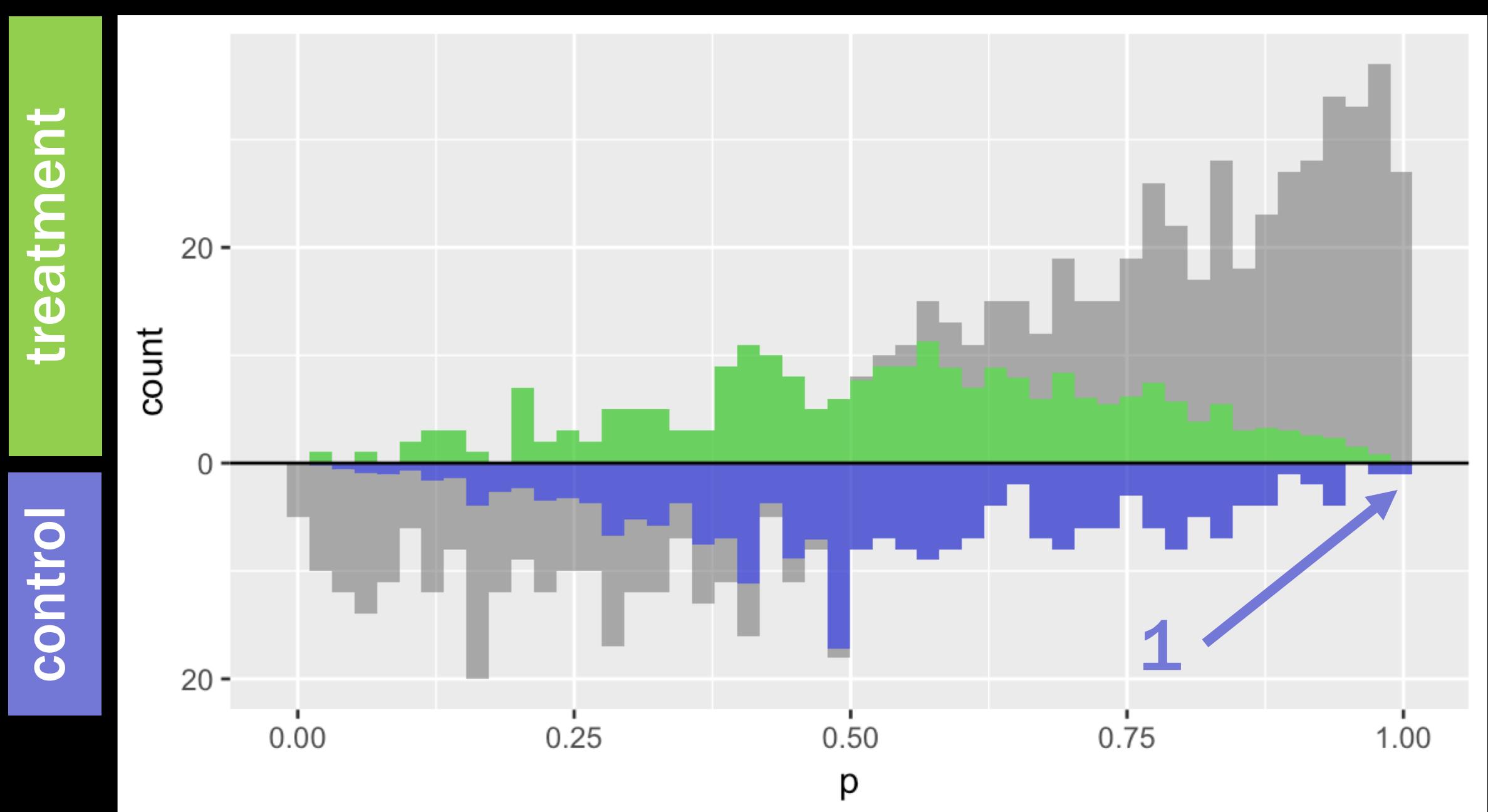


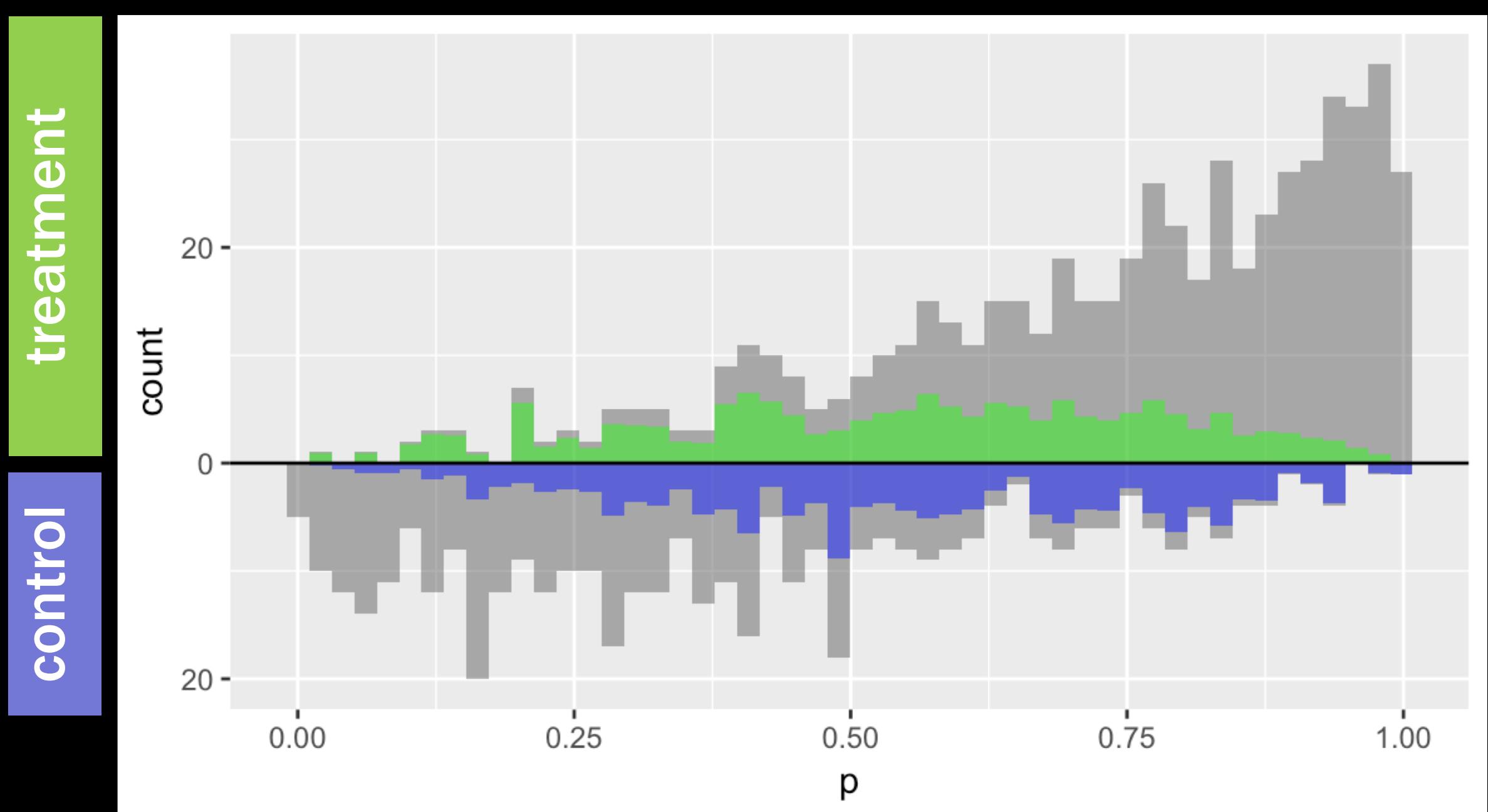
trt

control

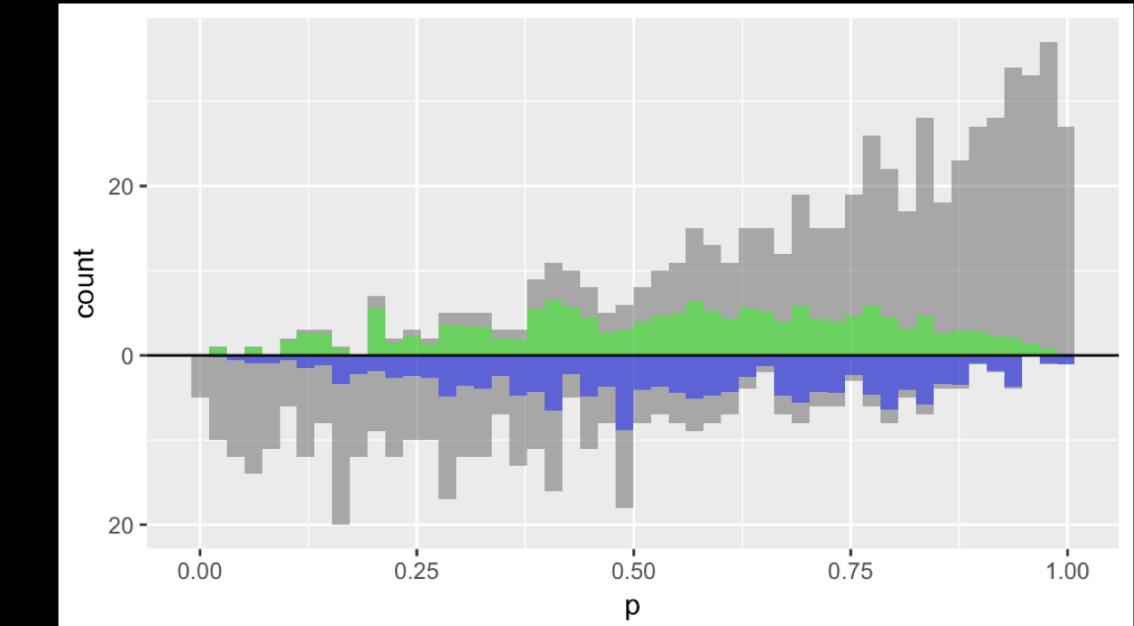
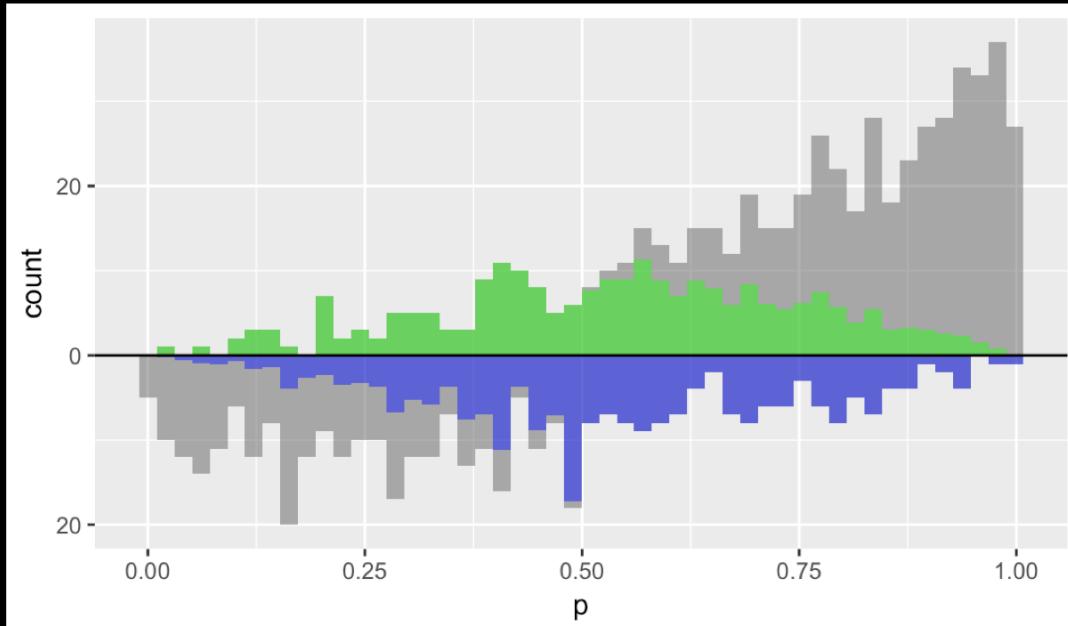
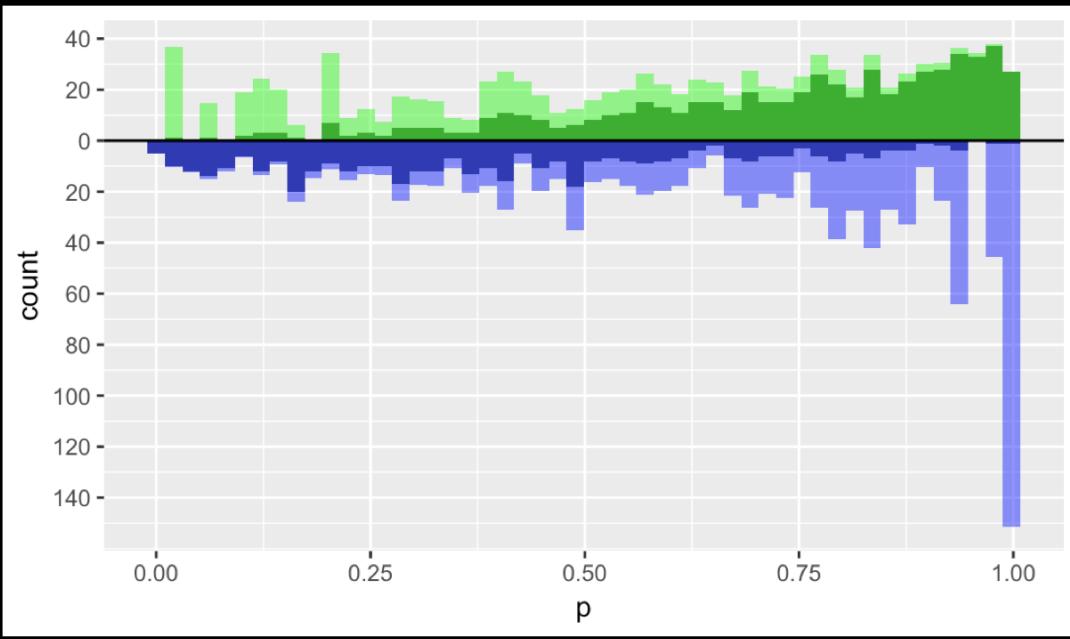








ATE

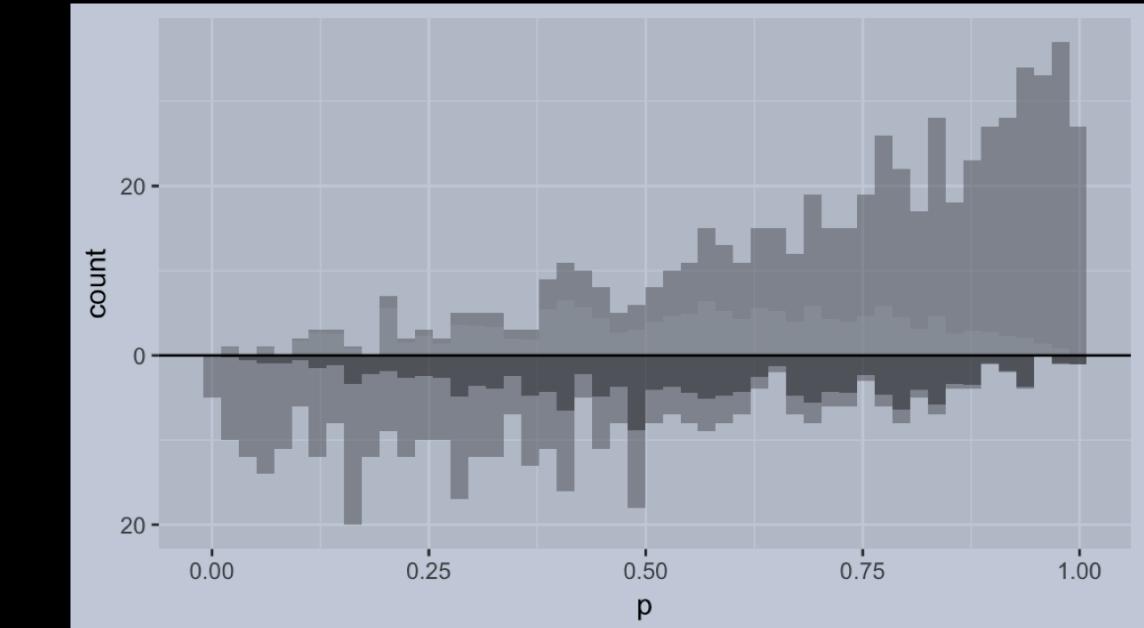
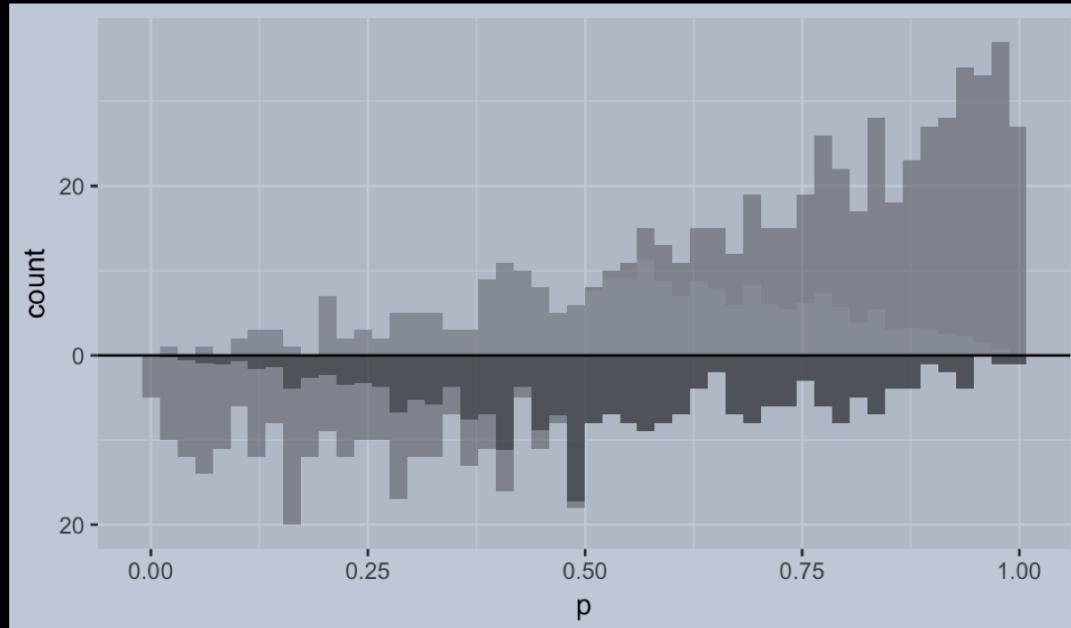
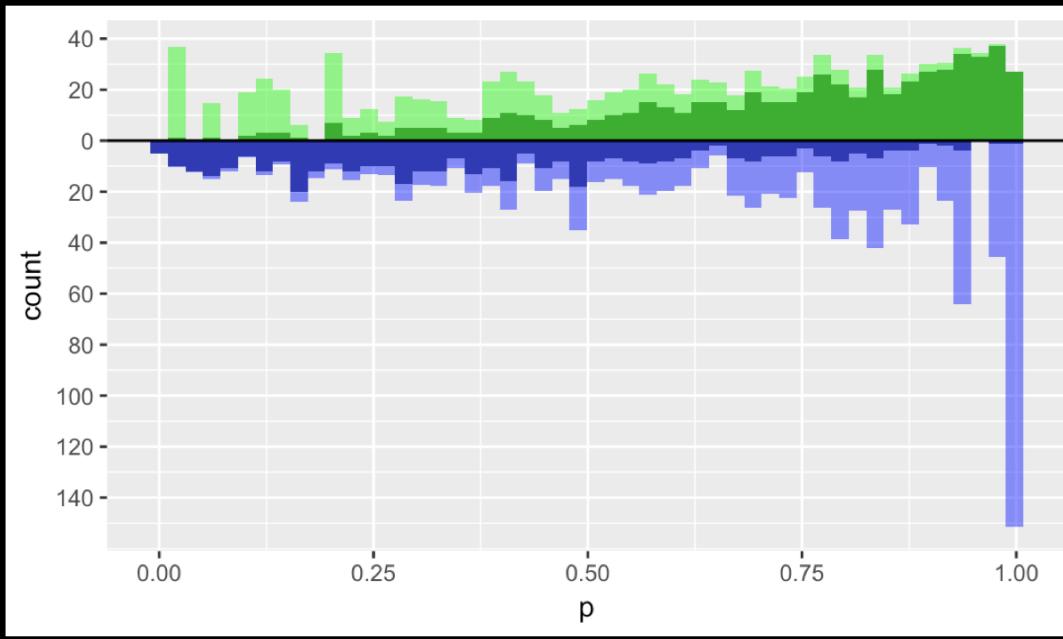


ATM

ENAR 2018

ATO

ATE

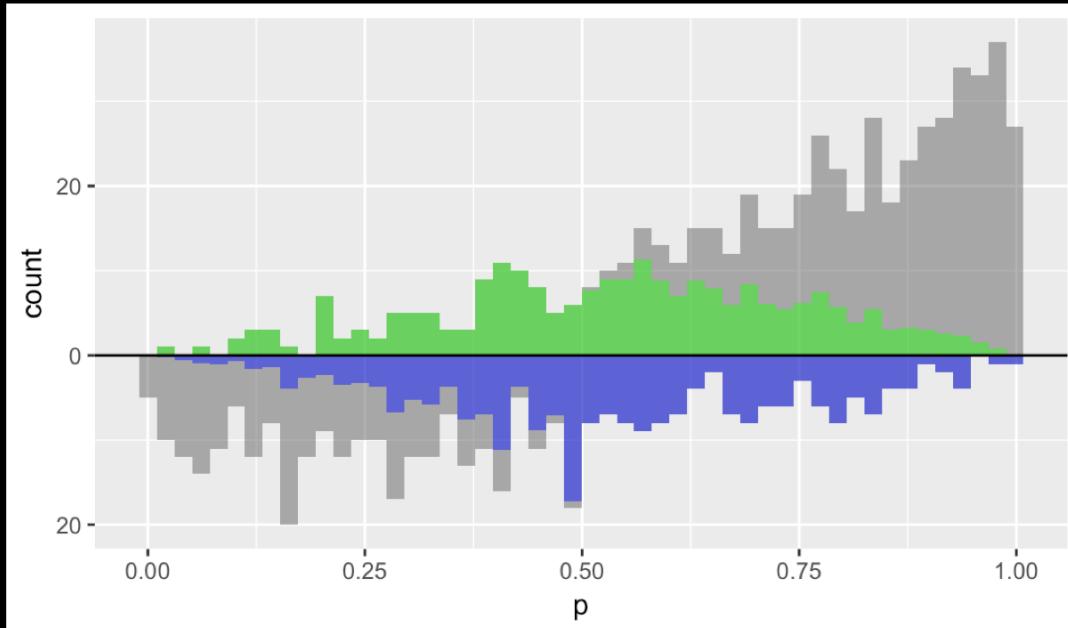
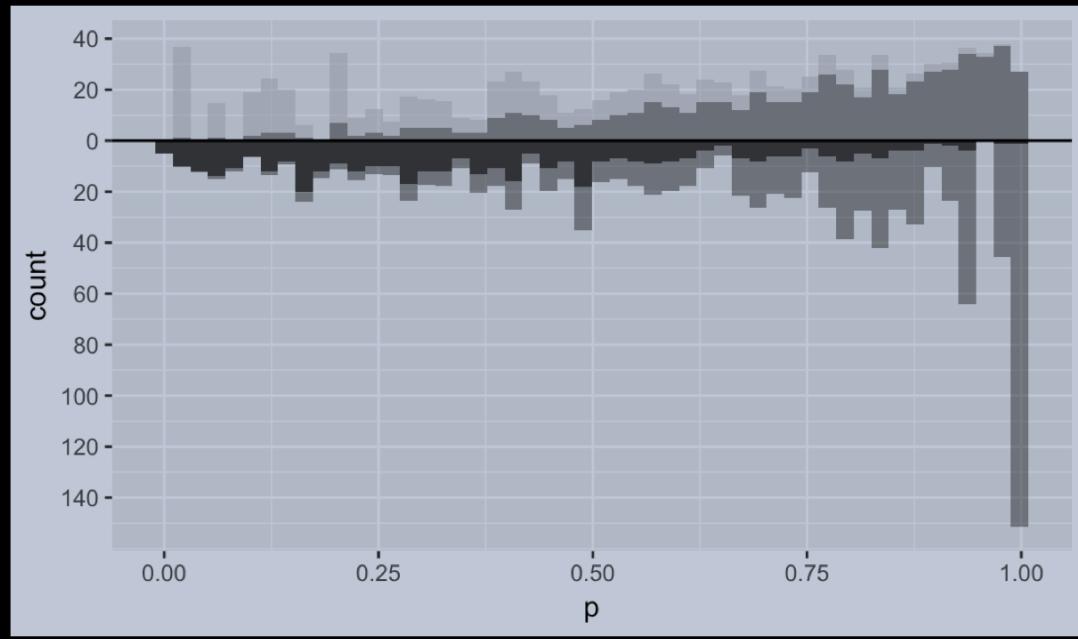


ATM

ENAR 2018

ATO

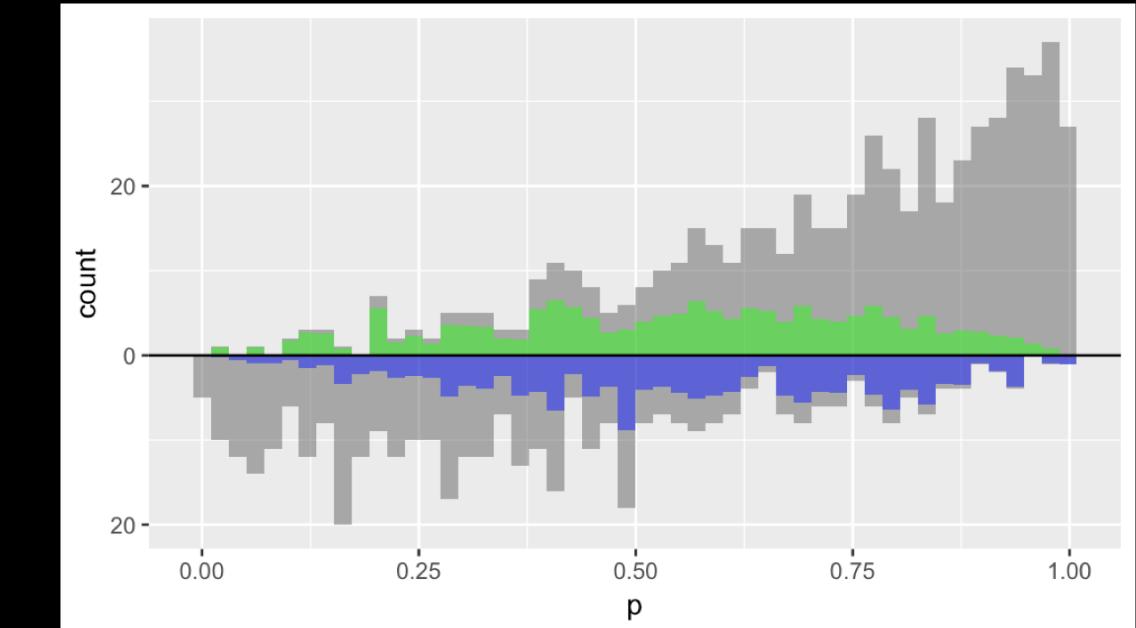
ATE



ATM

ENAR 2018

ATO



Finite-sample bias

Weighting Regressions by Propensity Scores

David A. Freedman

University of California, Berkeley

Richard A. Berk

University of Pennsylvania

Regressions can be weighted by propensity scores in order to reduce bias. However, weighting is likely to increase random error in the estimates, and to bias the estimated standard errors downward, even when selection mechanisms are well understood. Moreover, in some cases, weighting will increase the bias in estimated causal parameters. If investigators have a good causal model, it seems better just to fit the model without weights. If the causal model is improperly specified, there can be significant problems in retrieving the situation by weighting, although weighting may help under some circumstances.

Keywords: causation; selection; models; experiments; observational studies; regression; propensity scores

Estimating causal effects is often the key to evaluating social programs, but the interventions of interest are seldom assigned at random. Observational data are therefore frequently encountered. In order to

ENAR 2018

Doubly robust estimators

- Fit a **propensity score model**
- Fit two **outcome models**
- Use **these** to estimate an average treatment effect
- As long as **one is correctly specified**, your result will be **consistent**

Propensity score weighting method

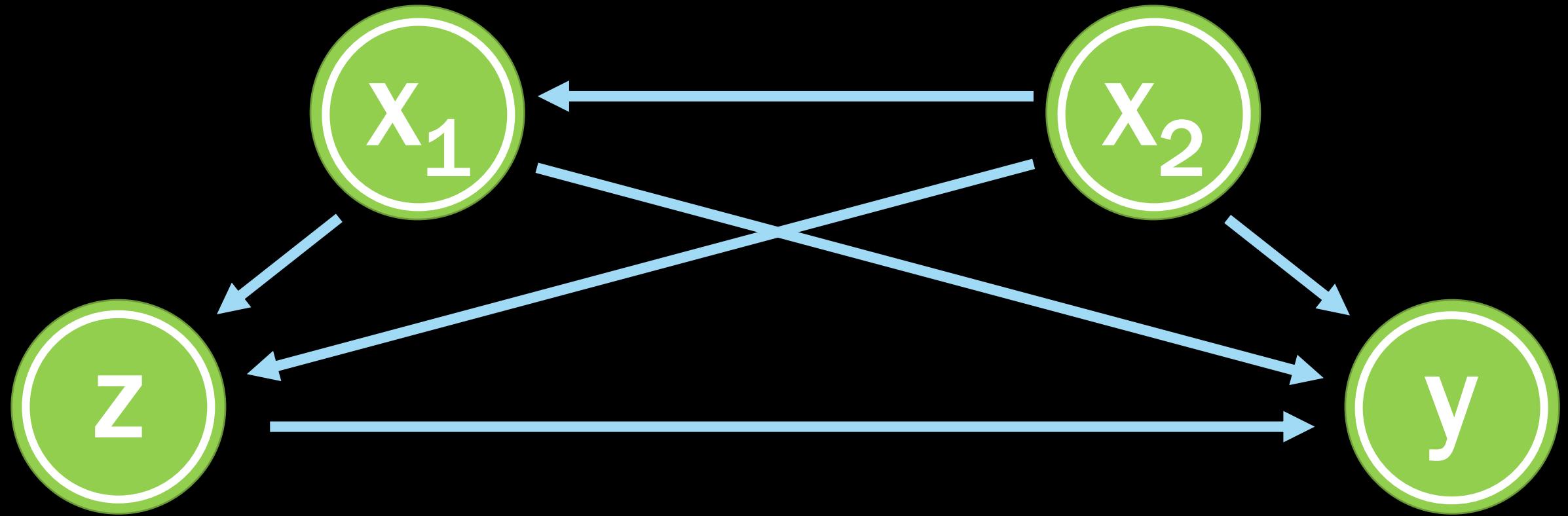
- The weights used in Freedman & Berk (2008) are estimating ATE
- *ATE weights can explode*
- Hypothesize that ATO and ATM weights will perform better

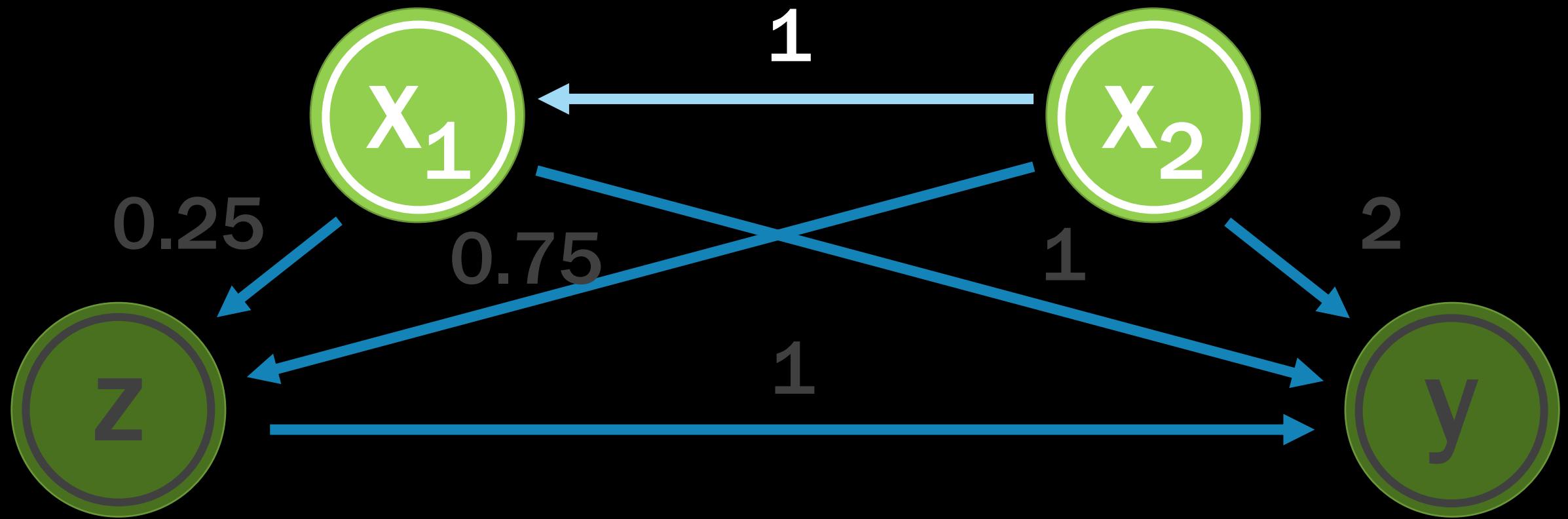
Fit a misspecified model

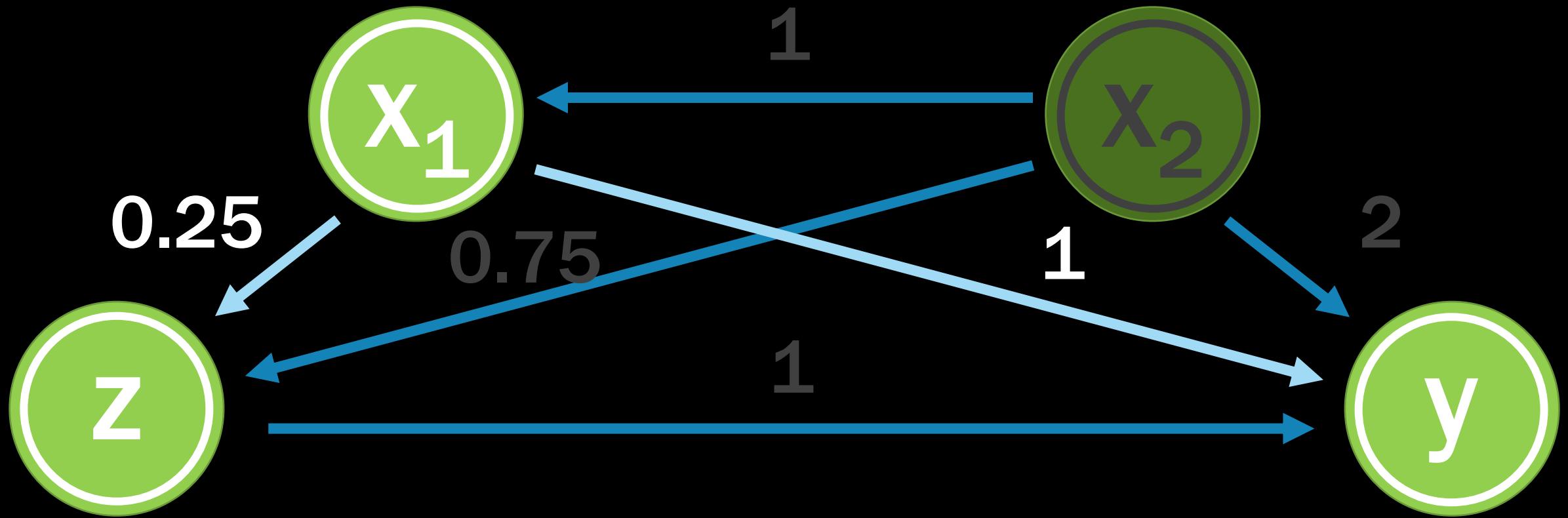
- Allow the propensity score to be **correct**
- The outcome model is **missing a covariate**

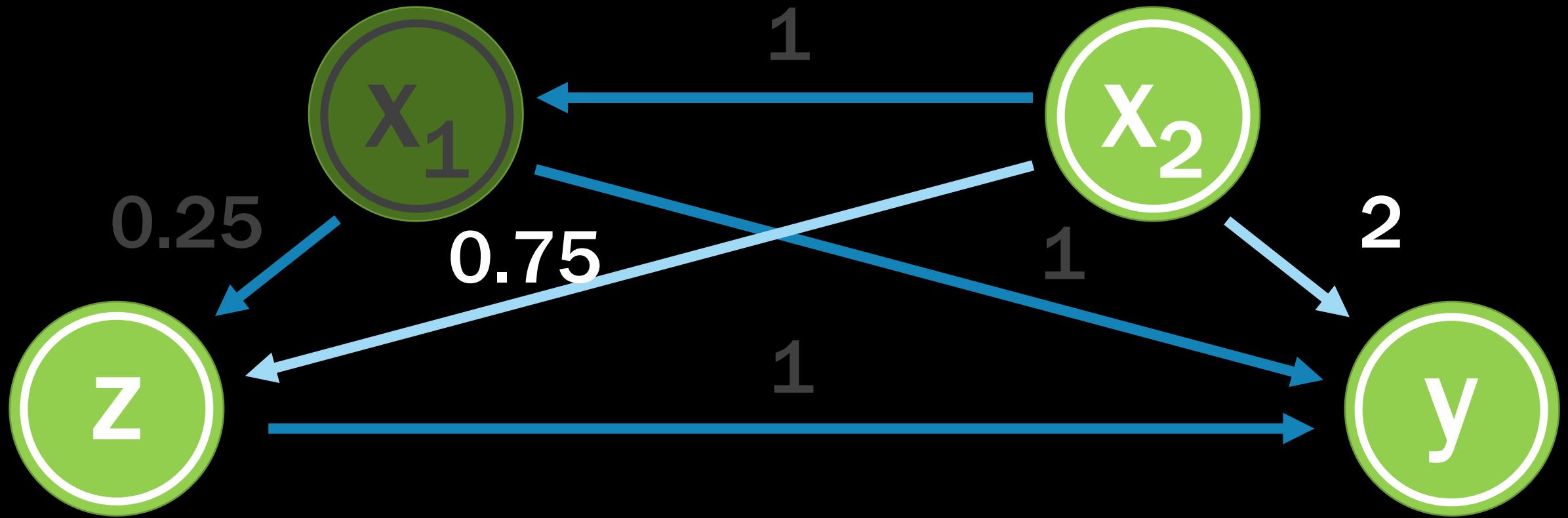
Freedman & Berk Simulation

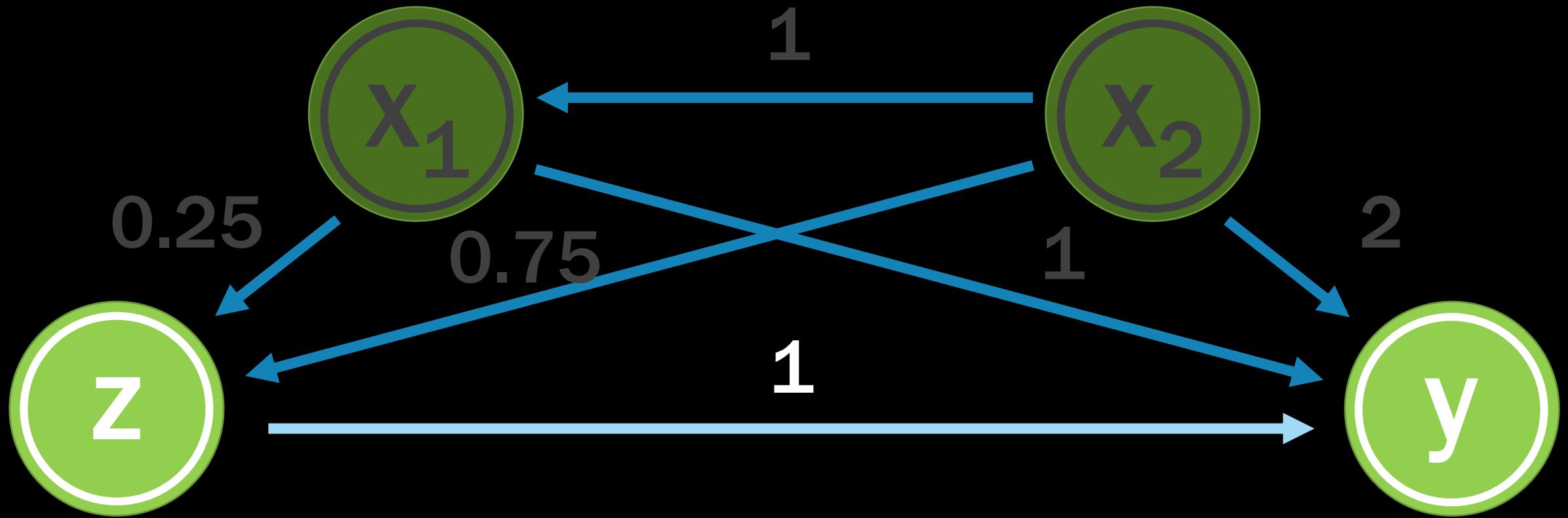
ENAR 2018

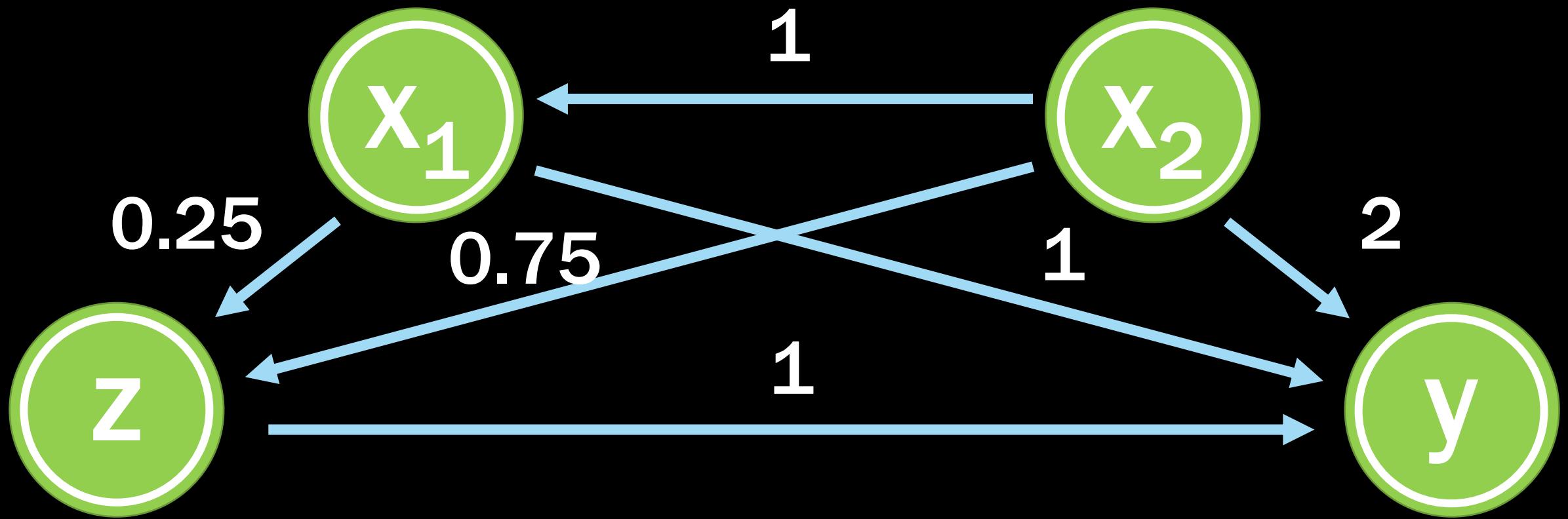




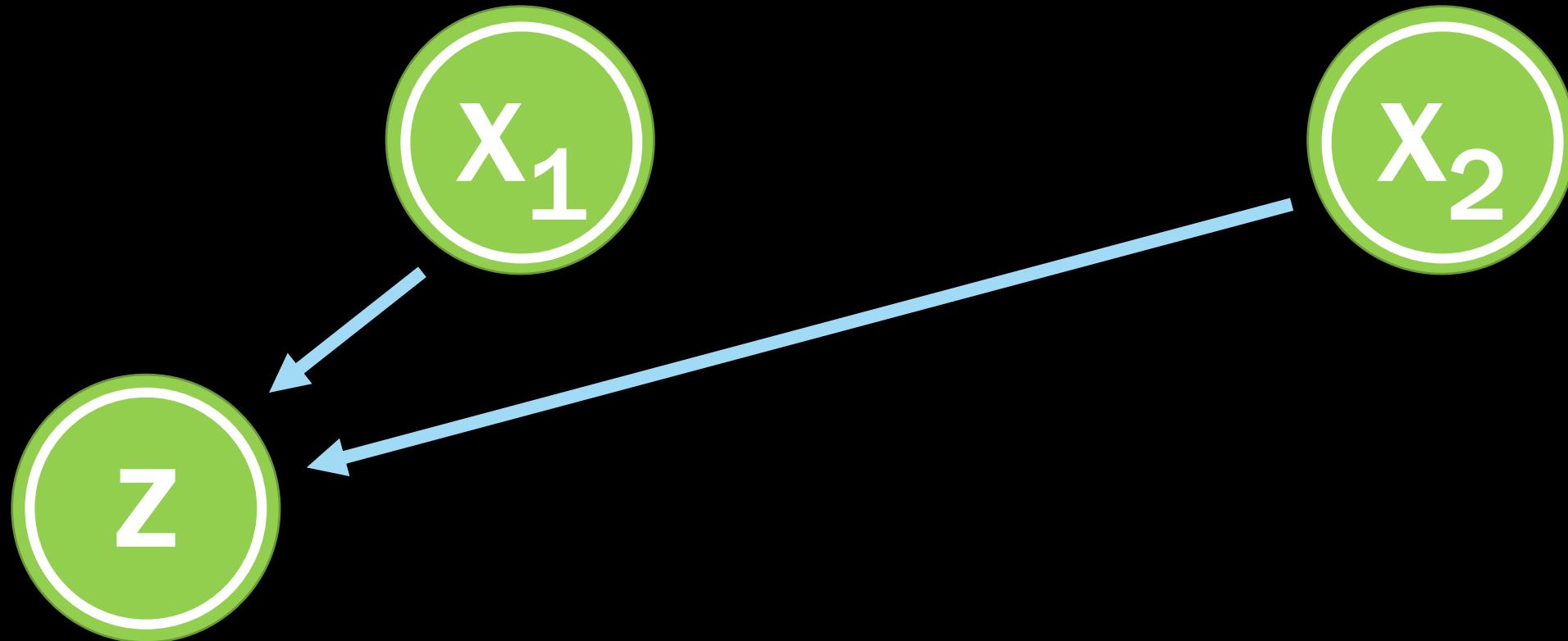




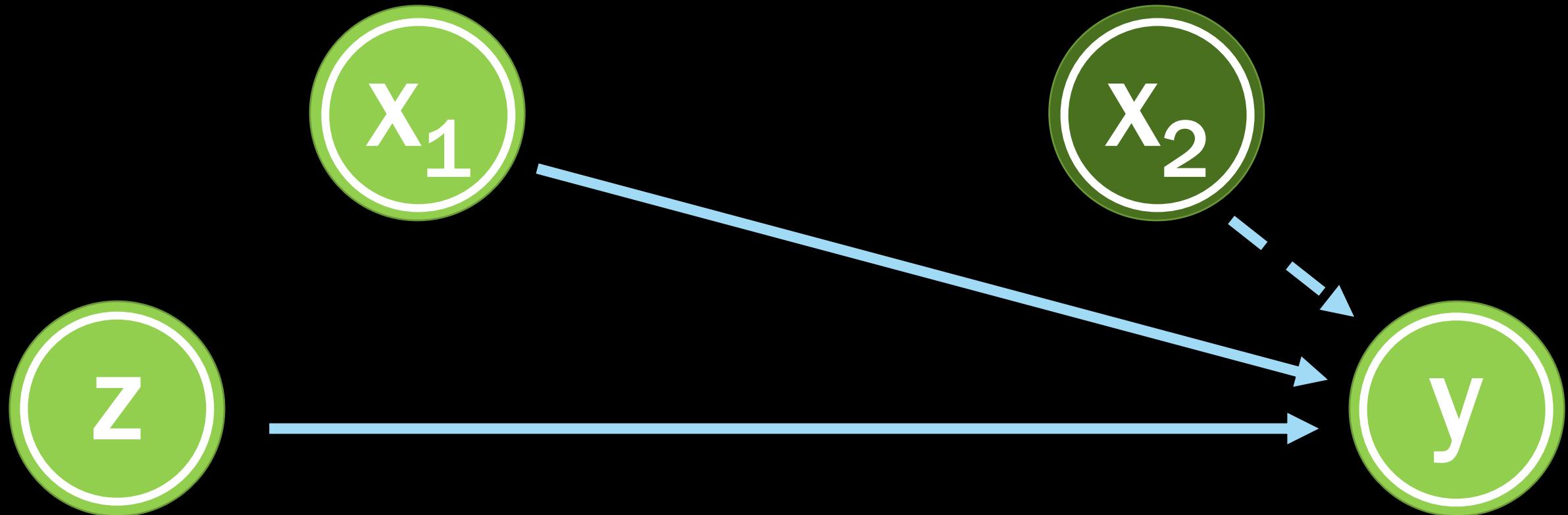


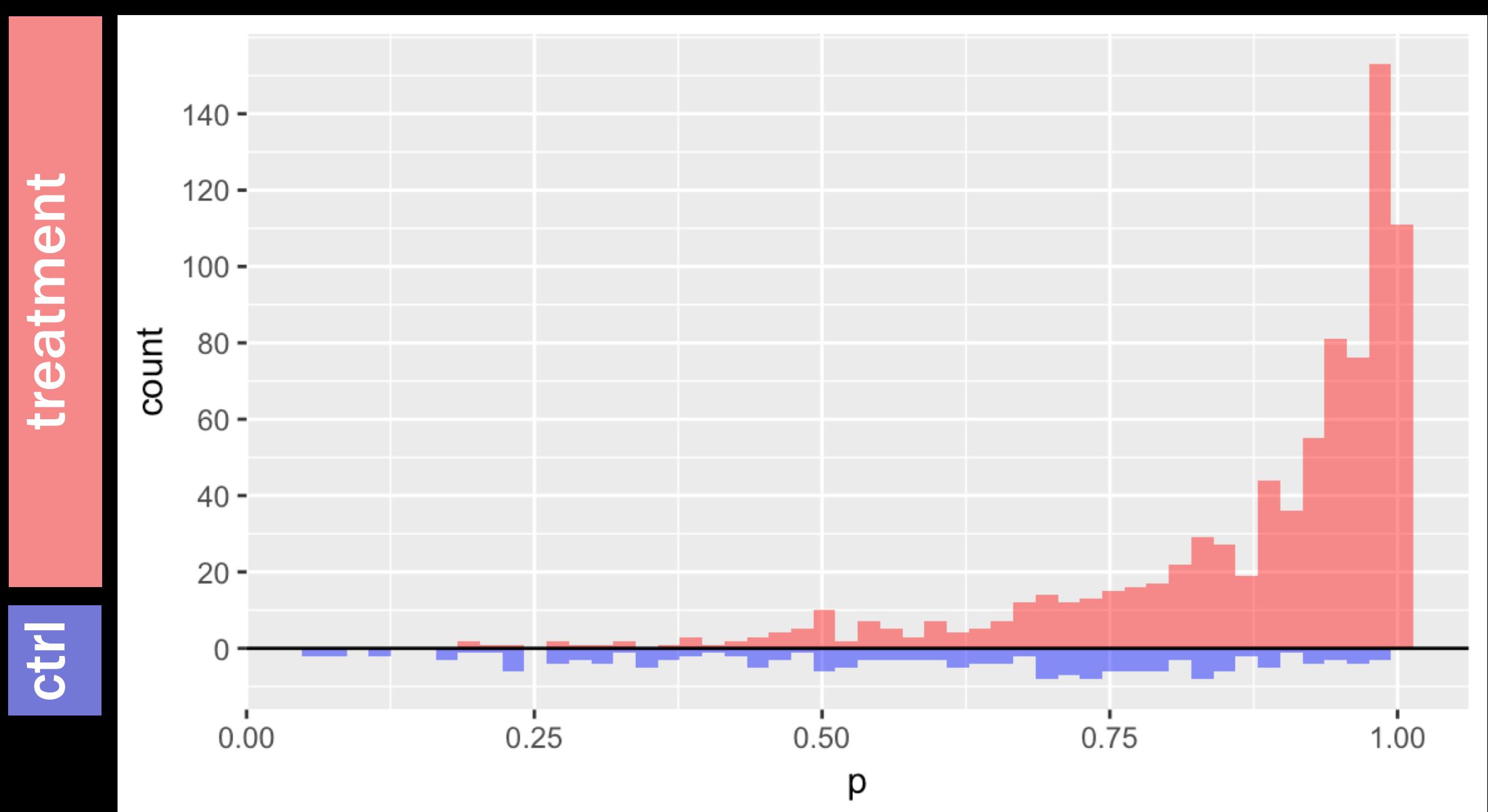


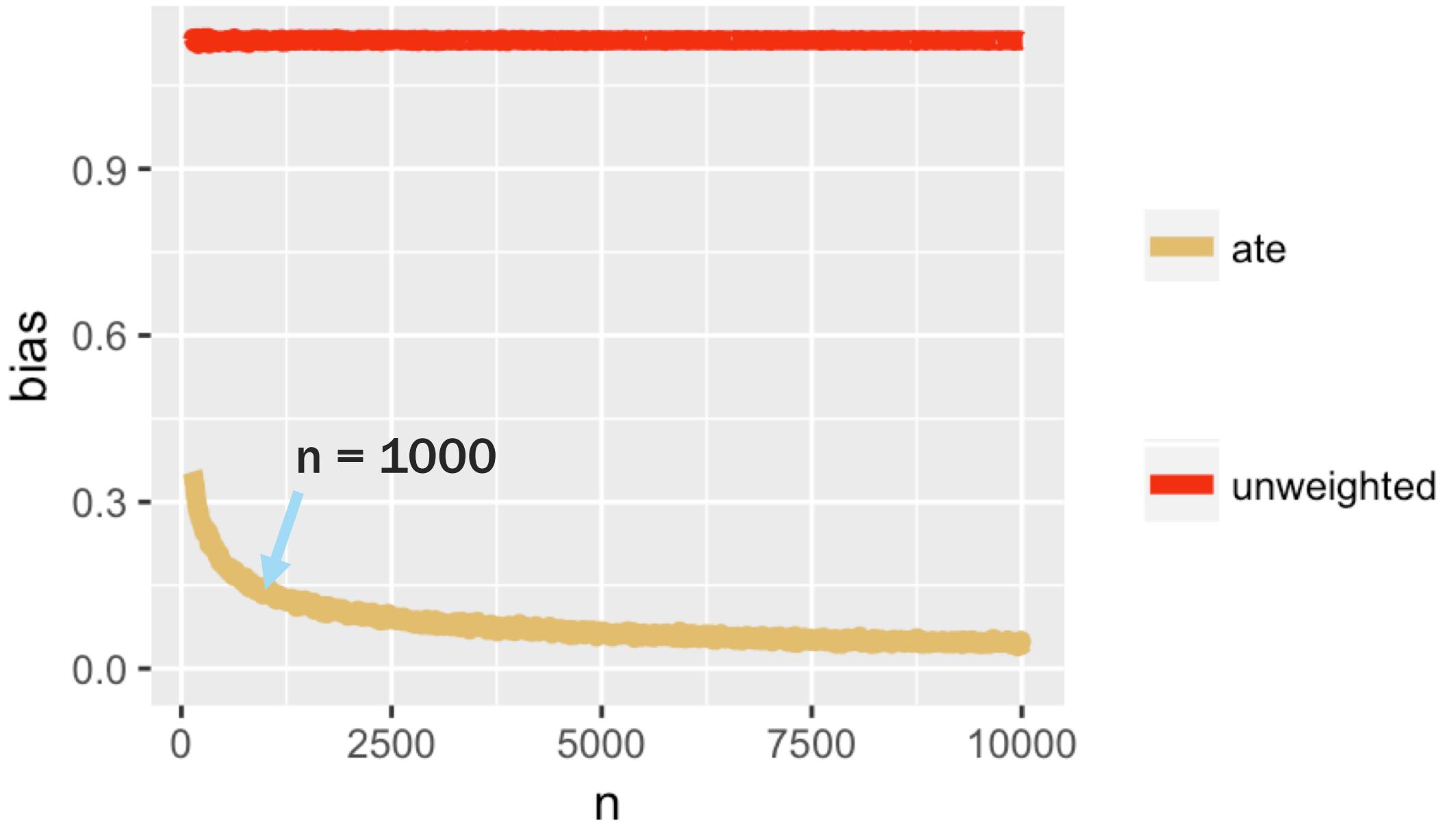
Propensity score model

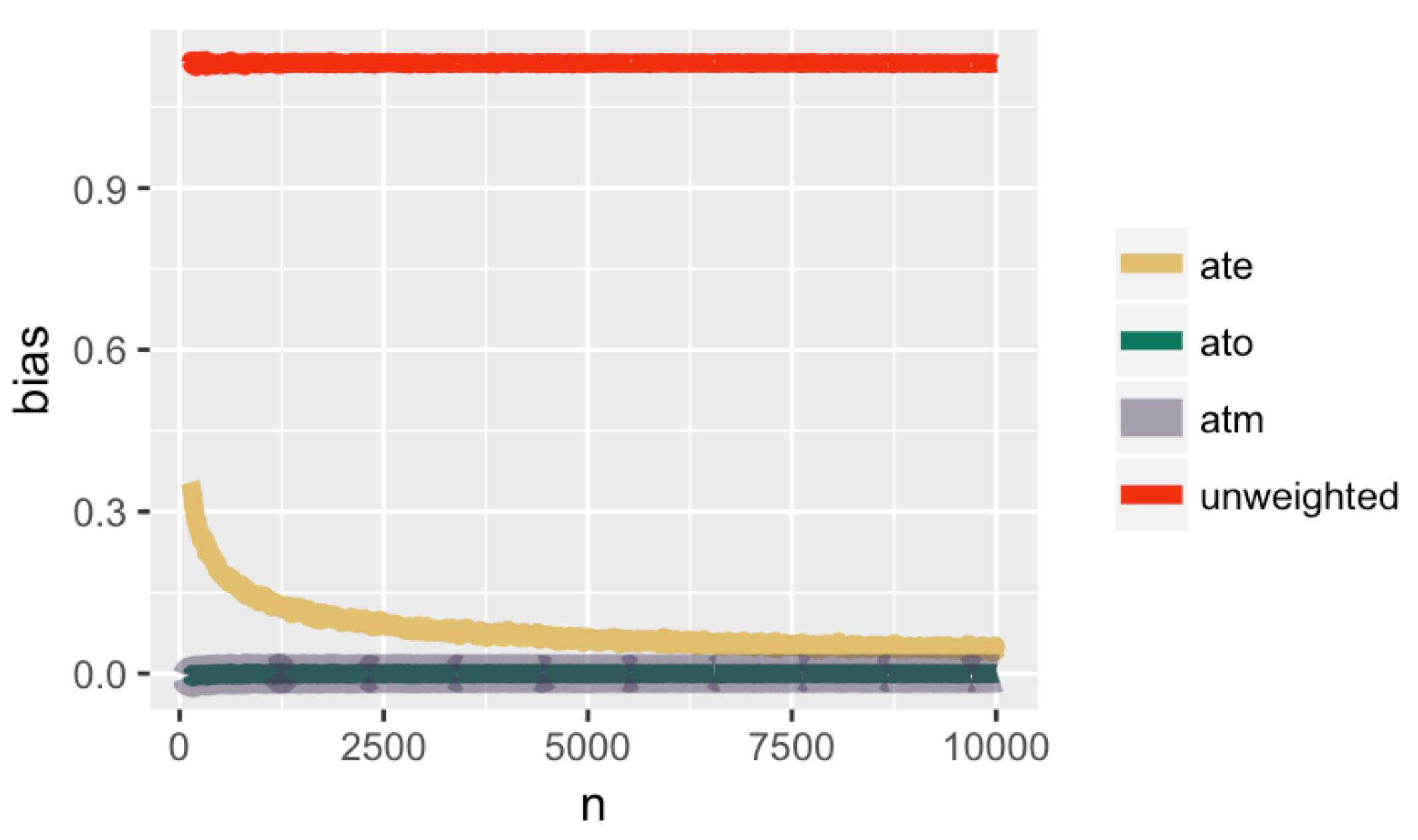


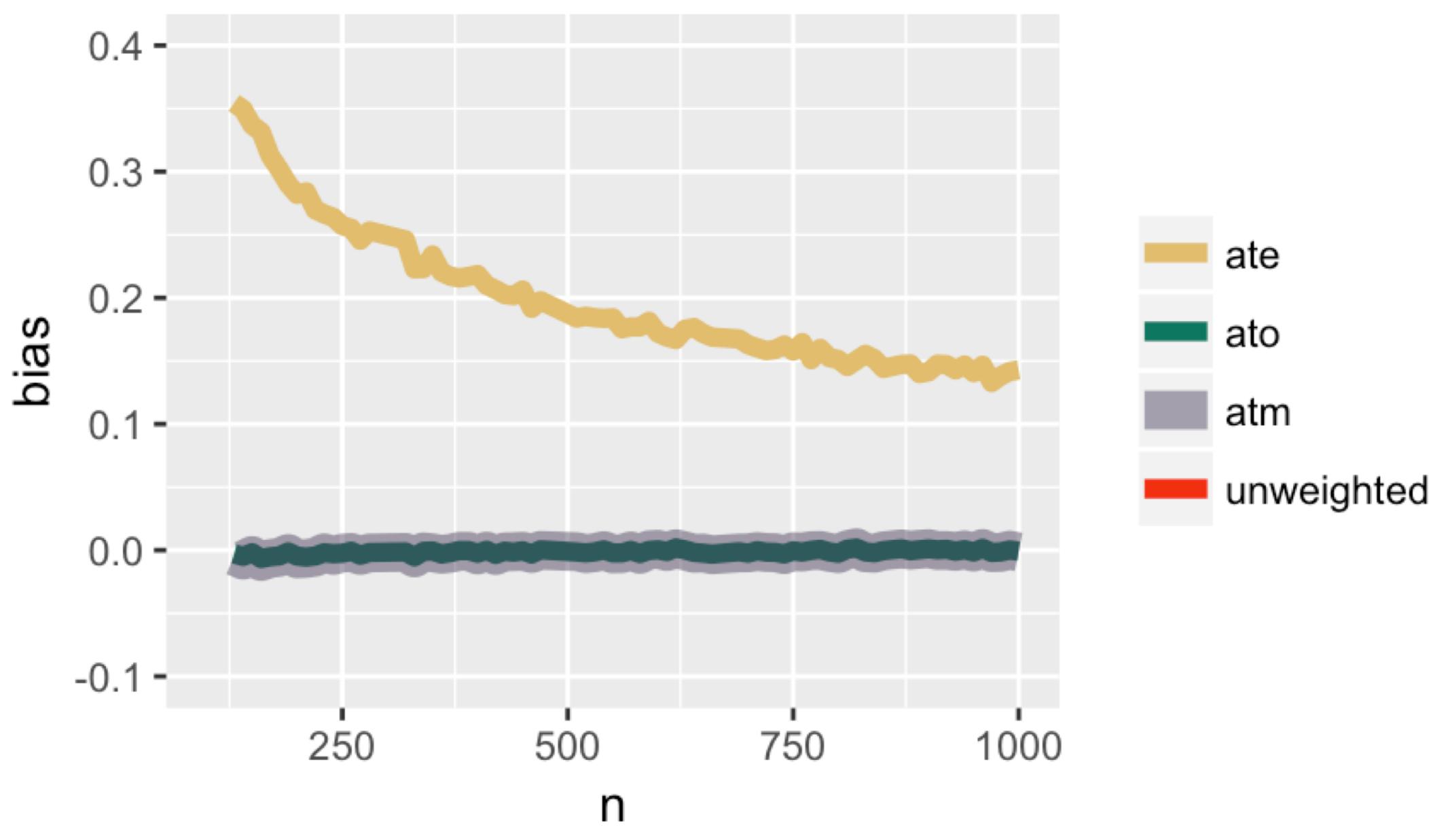
Outcome model





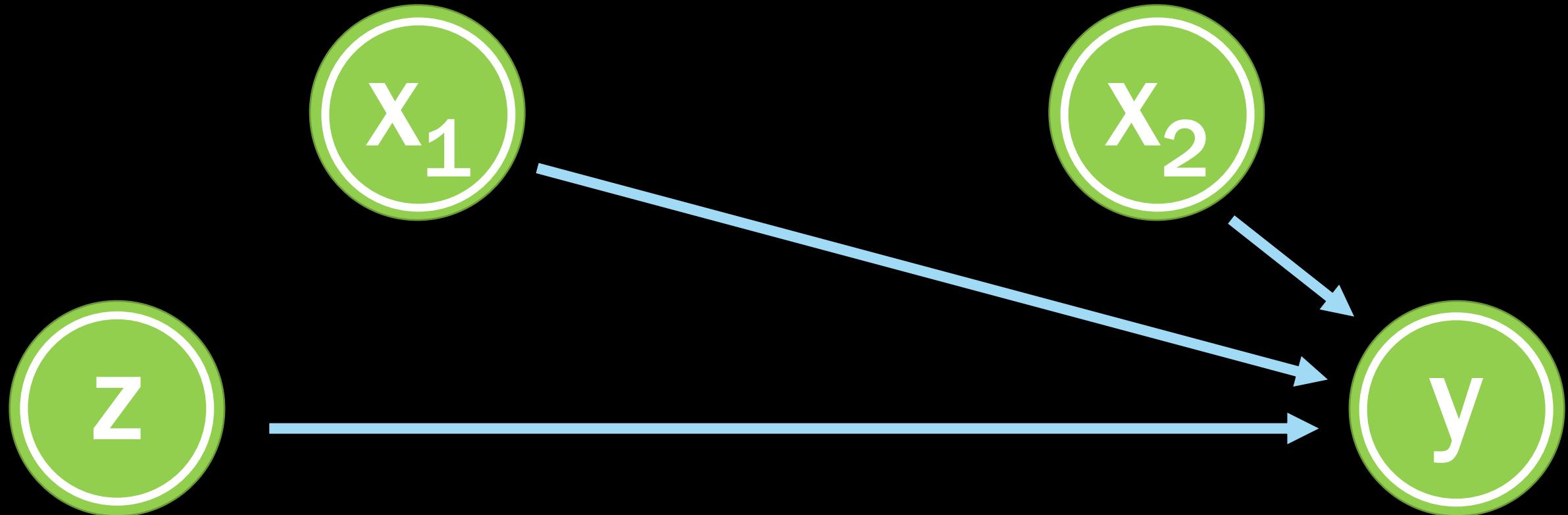


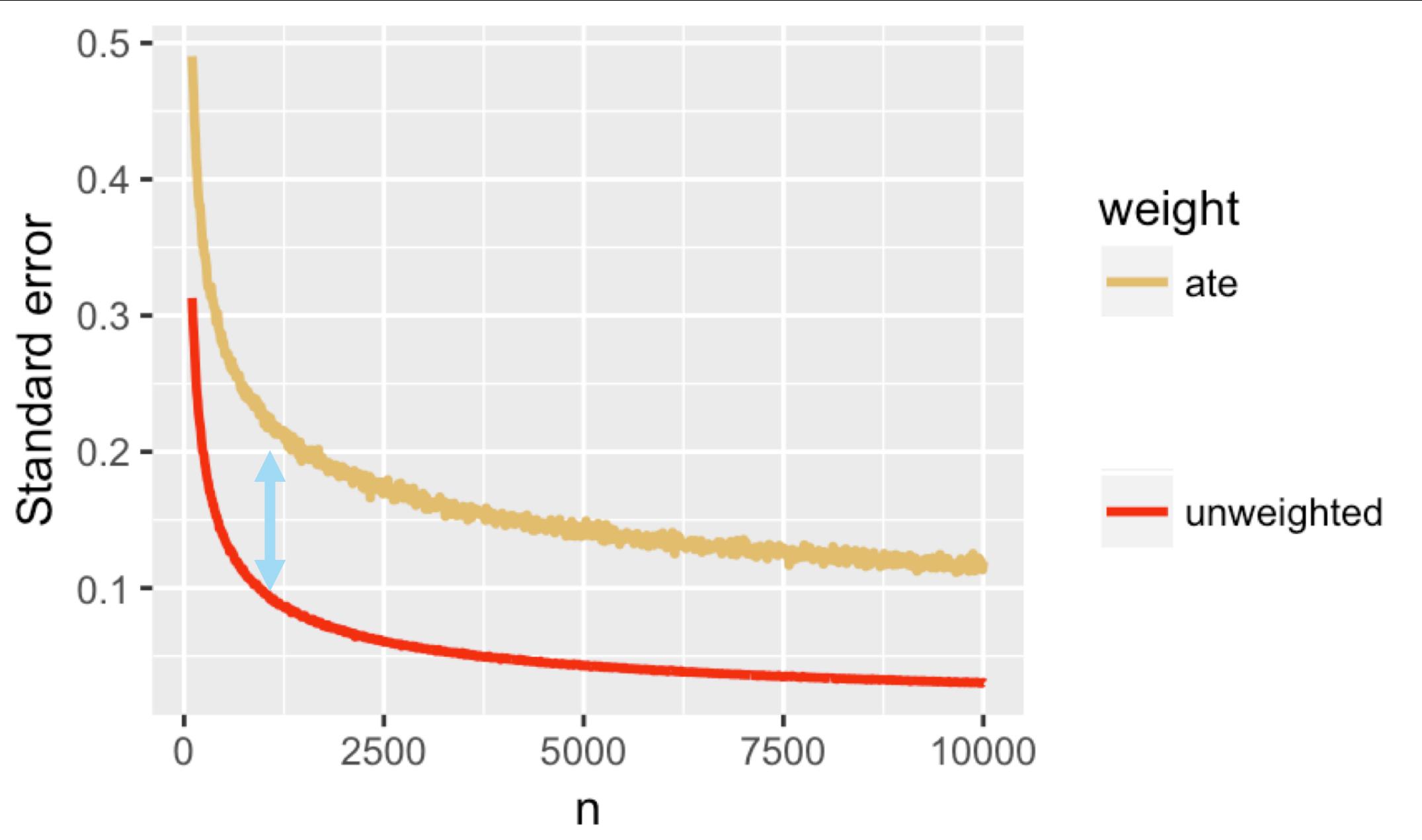


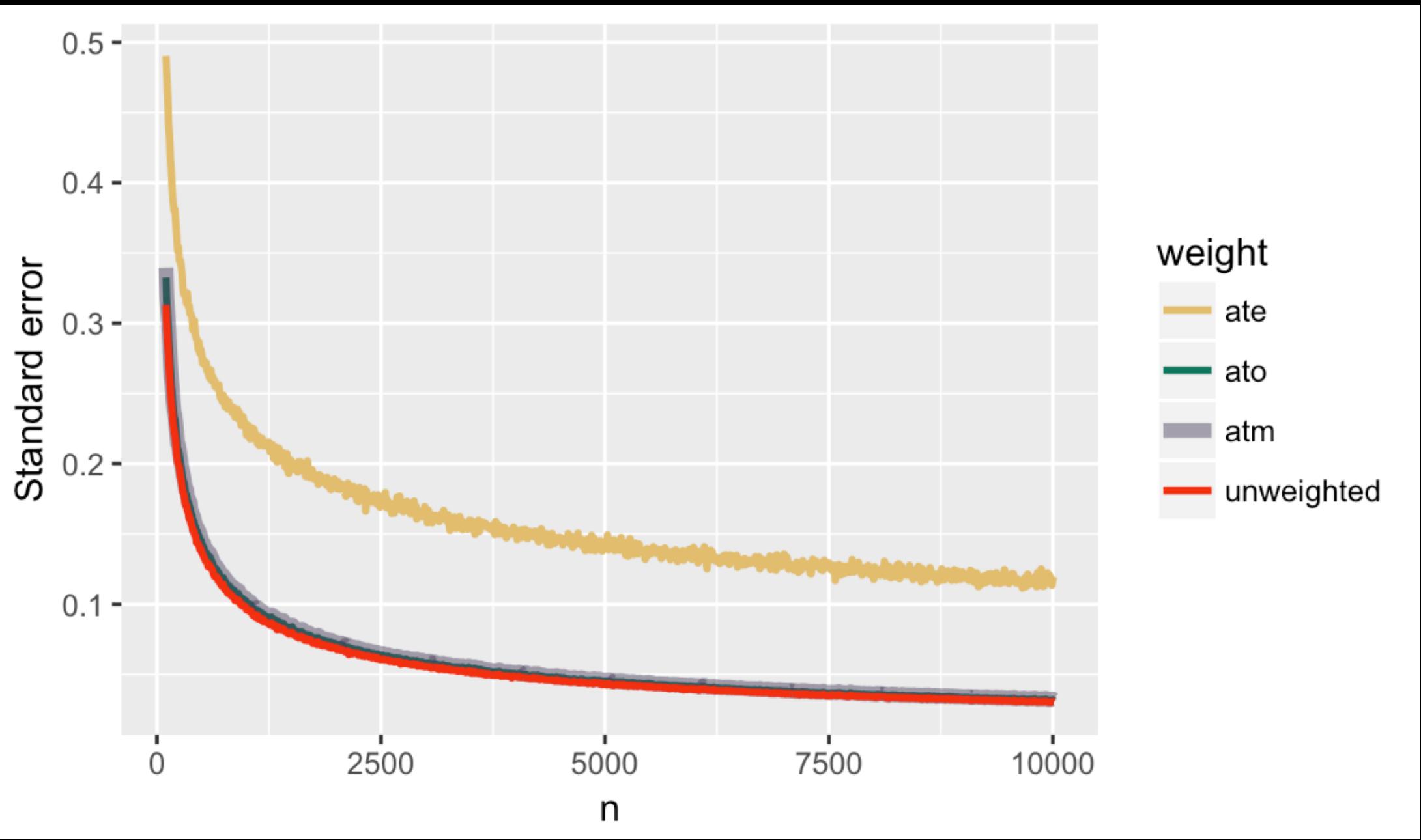


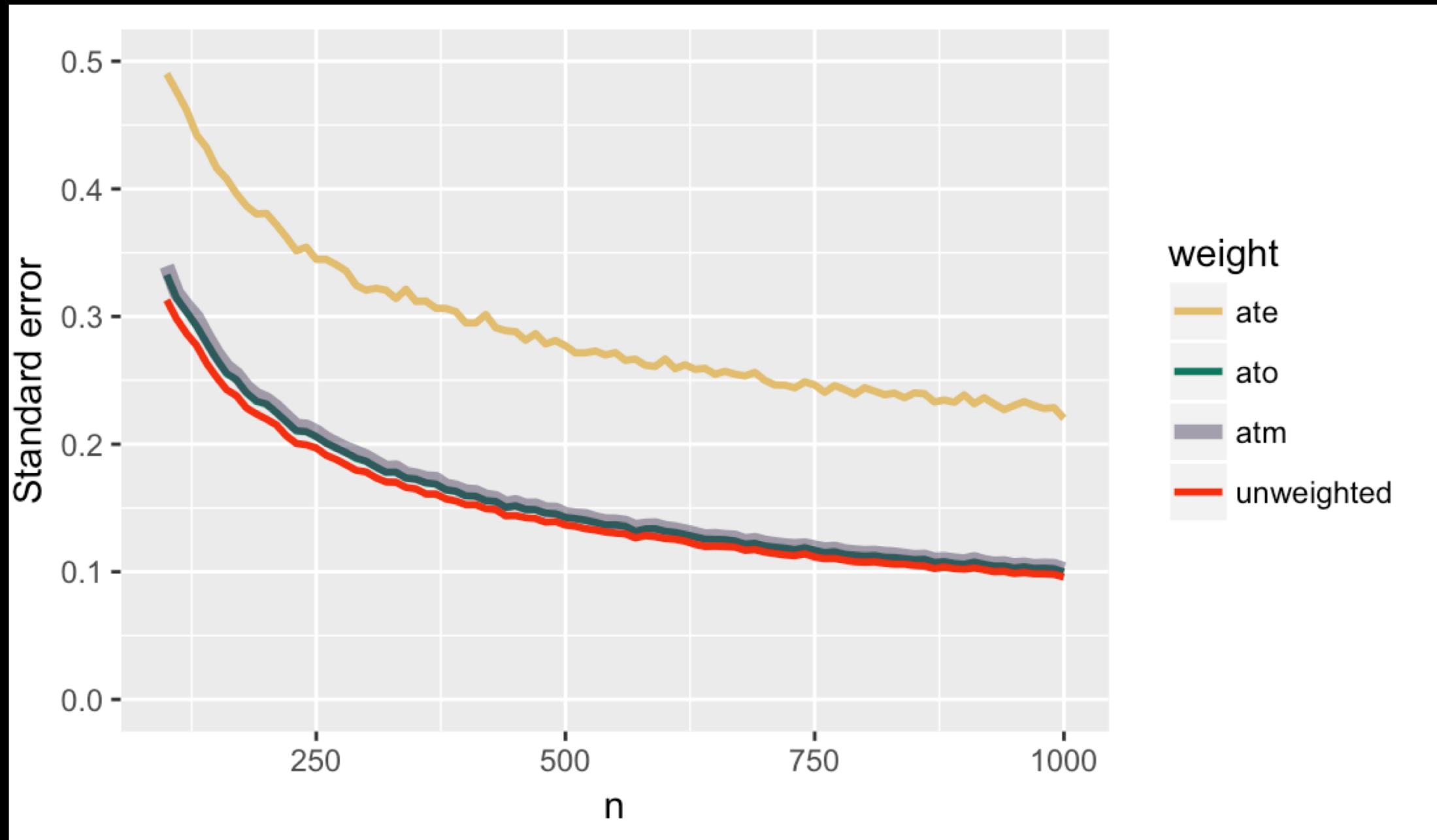
Standard errors

Outcome model









Recap

- Replicated the finite-sample bias seen by Freedman and Berk using the ATE weights
- ATM and ATO weights had improved finite-sample properties
- The variance for the ATO and ATM is preferable to that of the ATE

Unmeasured confounding the problem

ed: 1 January 2017 | Accepted: 3 January 2017
10.1111/jep.12714

WILEY Journal of Evaluation
International Journal of Public Health
Statistics, 2007

ORIGINAL ARTICLE

Improving causal inference with a doubly robust propensity score stratification and combines propensity score stratification and

Ariel Linden DrPH^{1,2}

¹ President, Linden Consulting Group, LLC Ann Arbor, MI, USA
² Research Scientist, Division of General Medicine, Medical School, University of Michigan, Ann Arbor, MI, USA

Correspondence
Ariel Linden, DrPH, Linden Consulting Group, LLC, 1301 North Bay Drive, Ann Arbor, MI, 48103, USA.
Email: alinden@lindenconsulting.org

Abstract

Rationale, aims and objectives When researchers typically use observational methods for causal inference, they often find themselves in a situation where they have to make assumptions about missing data. One common approach is to use propensity scores to adjust for confounding when estimating treatment effects. Another approach is to use inverse-probability weighting. A third approach is to use doubly robust (DR) estimators based on a model for the outcome and a model for the propensity score. DR estimators are "doubly robust" because they are consistent if either the outcome model or the propensity score model is correctly specified.

Method Monte Carlo simulations were used to compare the performance of various DR and non-DR estimators. The performance of the estimators was assessed by comparing their bias, variance, and mean squared error to those of the true population values.

Results The results show that DR estimators are more efficient than non-DR estimators when the propensity score model is misspecified. However, when the outcome model is misspecified, non-DR estimators are more efficient than DR estimators. The results also show that DR estimators are more robust to misspecification of the propensity score model than non-DR estimators.

Conclusions DR estimators are more efficient than non-DR estimators when the propensity score model is misspecified. However, when the outcome model is misspecified, non-DR estimators are more efficient than DR estimators. The results also show that DR estimators are more robust to misspecification of the propensity score model than non-DR estimators.

Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data¹

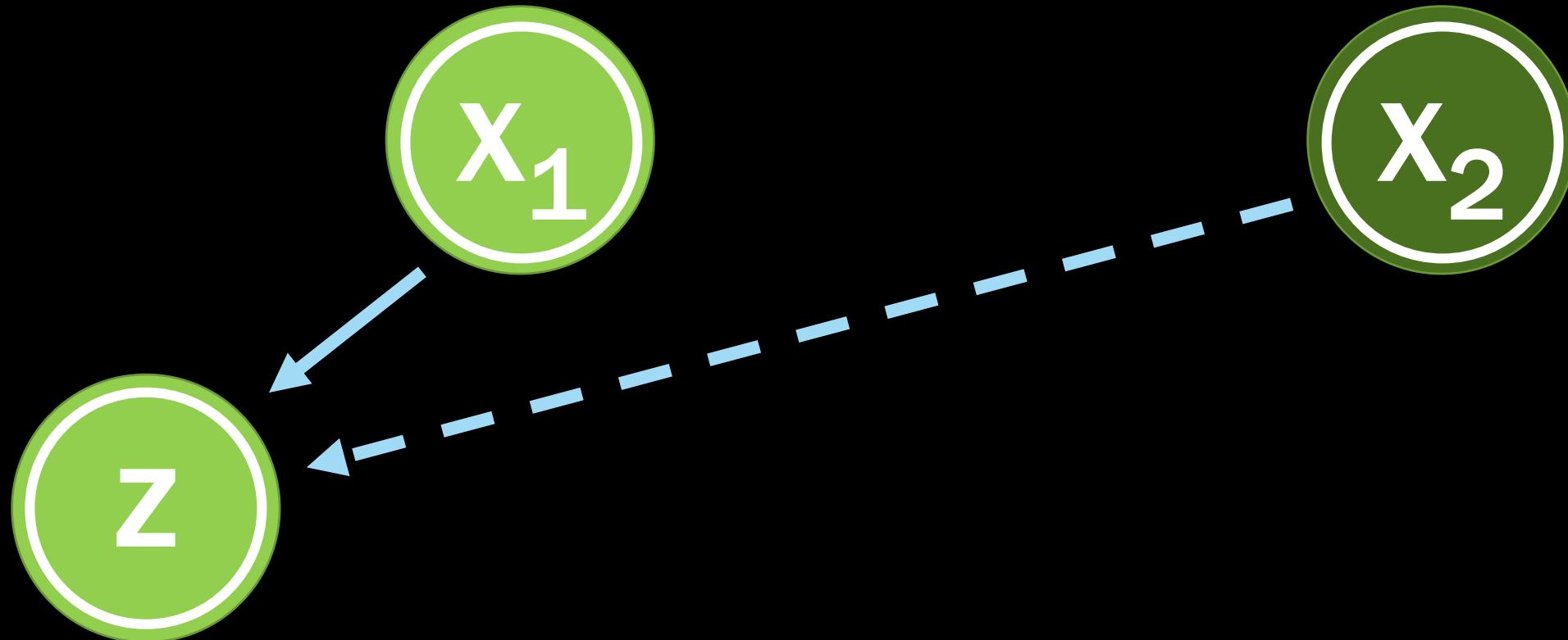
Joseph D. Y. Kang and Joseph L. Schafer

Abstract. When outcomes are missing for reasons beyond an investigator's control, there are two different ways to adjust a parameter estimate for covariates that may be related both to the outcome and to missingness. One approach is to model the relationships between the covariates and the outcome and use those relationships to predict the missing values. Another is to model the probabilities of missingness given the covariates and incorporate them into a weighted or stratified estimate. Doubly robust (DR) procedures apply both types of model simultaneously and produce a consistent estimate of the parameter if either of the two models has been correctly specified. In this article, we show that DR estimates can be constructed in many ways. We compare the performance of various DR and non-DR estimates of a population mean in a simulated example where both models are incorrect but neither is grossly misspecified. Methods that use inverse-probabilities as propensity model when some estimated propensities are small. Many DR methods perform better than simple inverse-probability weighting. None of the DR methods we tried, however, improved upon the performance of simple regression-based prediction of the missing values. This study does not represent every missing-data problem that will arise in practice. But it does demonstrate that, in at least some settings, two wrong models are not better than one.

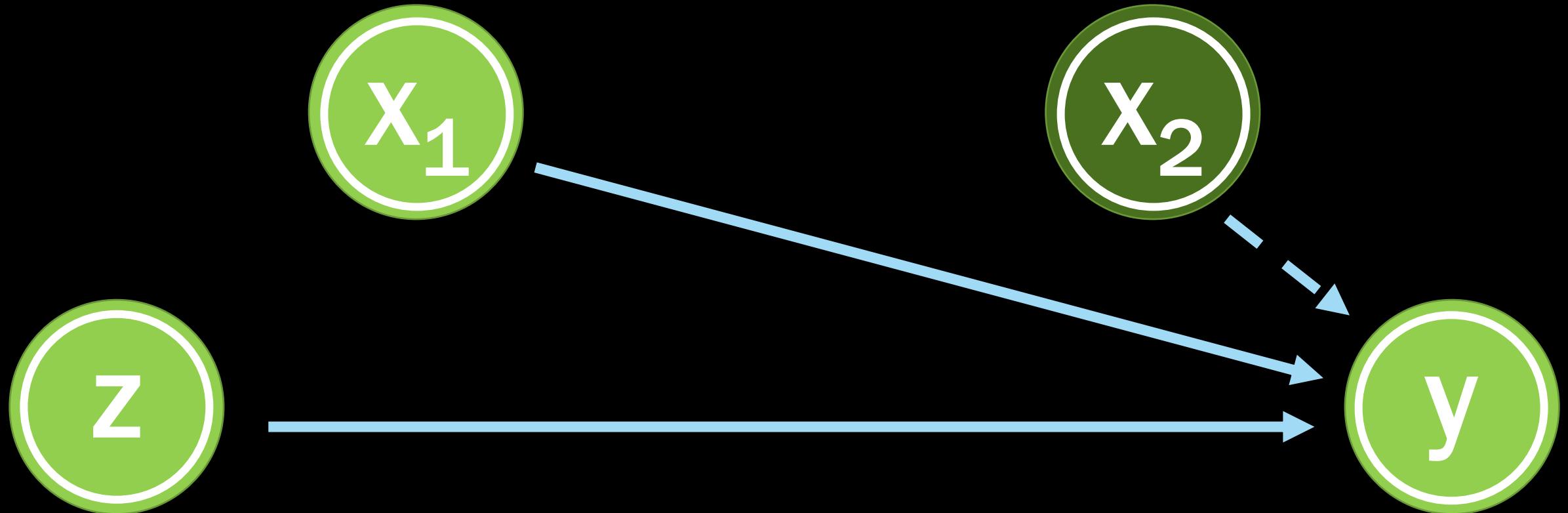
Key words and phrases: Causal inference, missing data, propensity score, model-assisted survey estimation, weighted estimating equations.

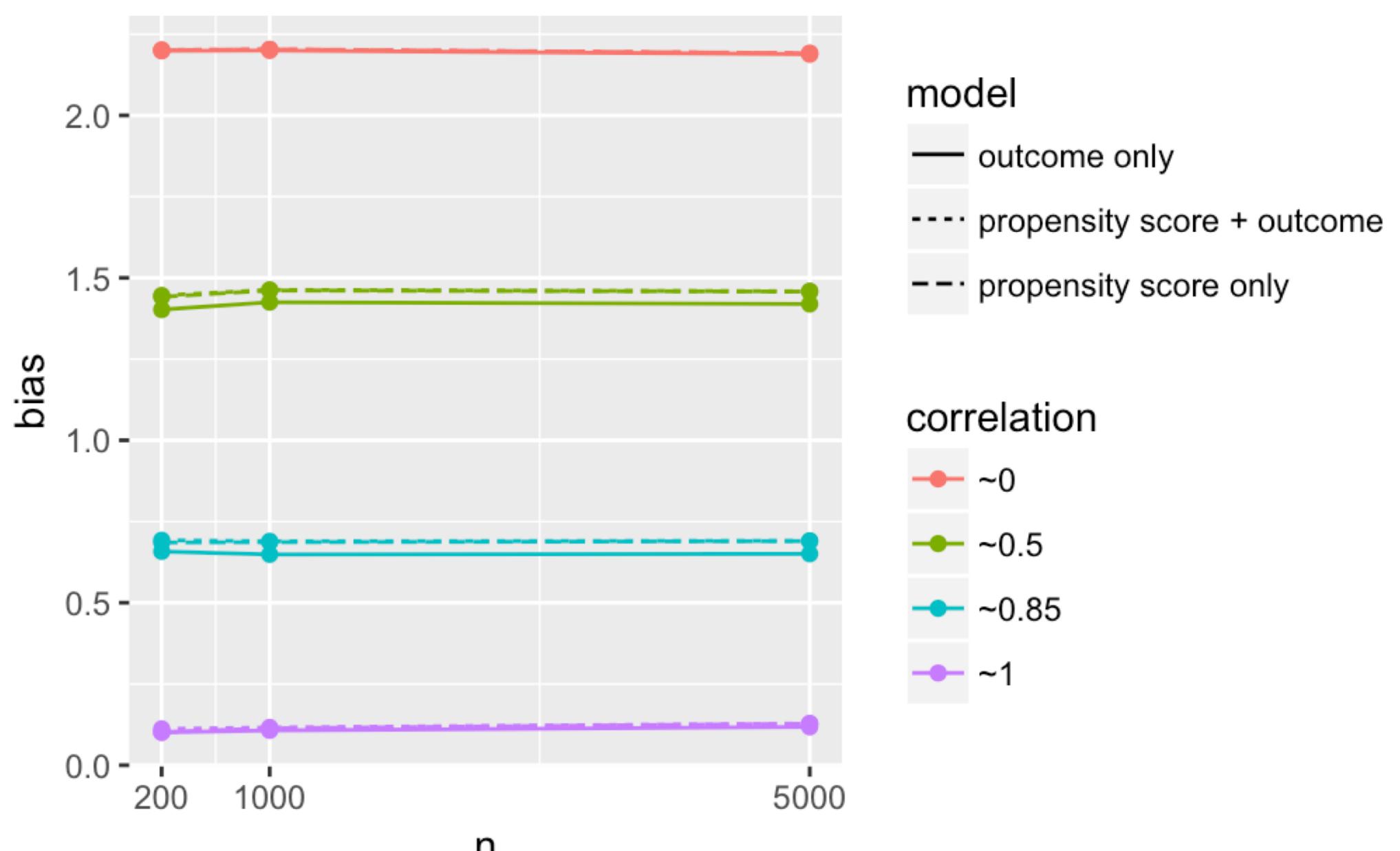
ENAR 2

Propensity score model



Outcome model





Unmeasured confounding a solution

E-value

$$\text{E - value} = LB_{obs} + \sqrt{LB_{obs} \times (LB_{obs} - 1)}$$

Ding and VanderWeele (2017)

Adjusted E-value

$$\text{E - value}_{adj} = \frac{LB_{obs}}{LB_{adj}} + \sqrt{\frac{LB_{obs}}{LB_{adj}} \times \left(\frac{LB_{obs}}{LB_{adj}} - 1 \right)}$$

Adjusted E-value

$$\text{E - value}_{adj} = \frac{LB_{obs}}{LB_{adj}} + \sqrt{\frac{LB_{obs}}{LB_{adj}} \times \left(\frac{LB_{obs}}{LB_{adj}} - 1 \right)}$$

Right Heart Catheterization Data

Connors et al (1996)

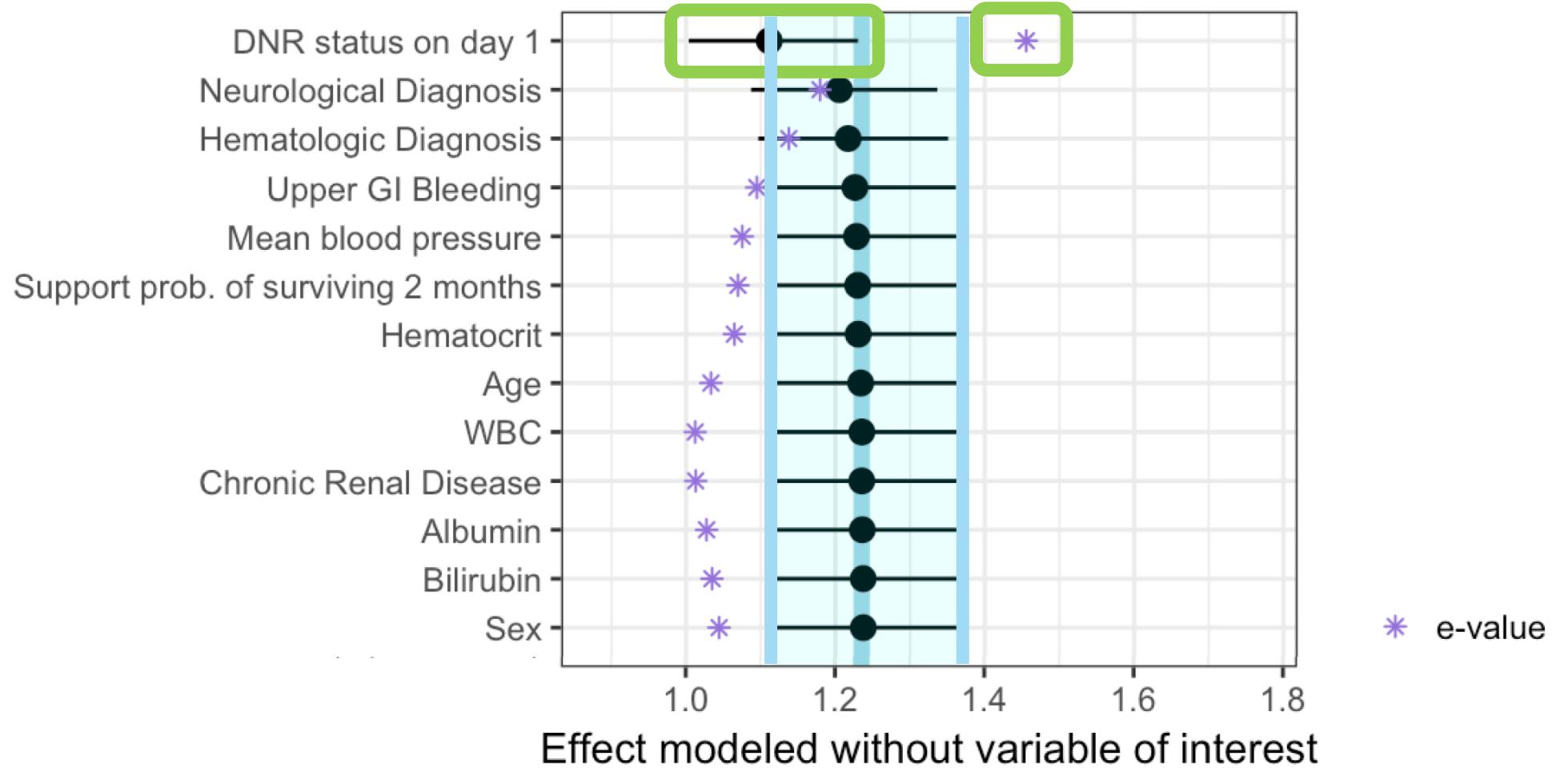
Right Heart Catheterization data

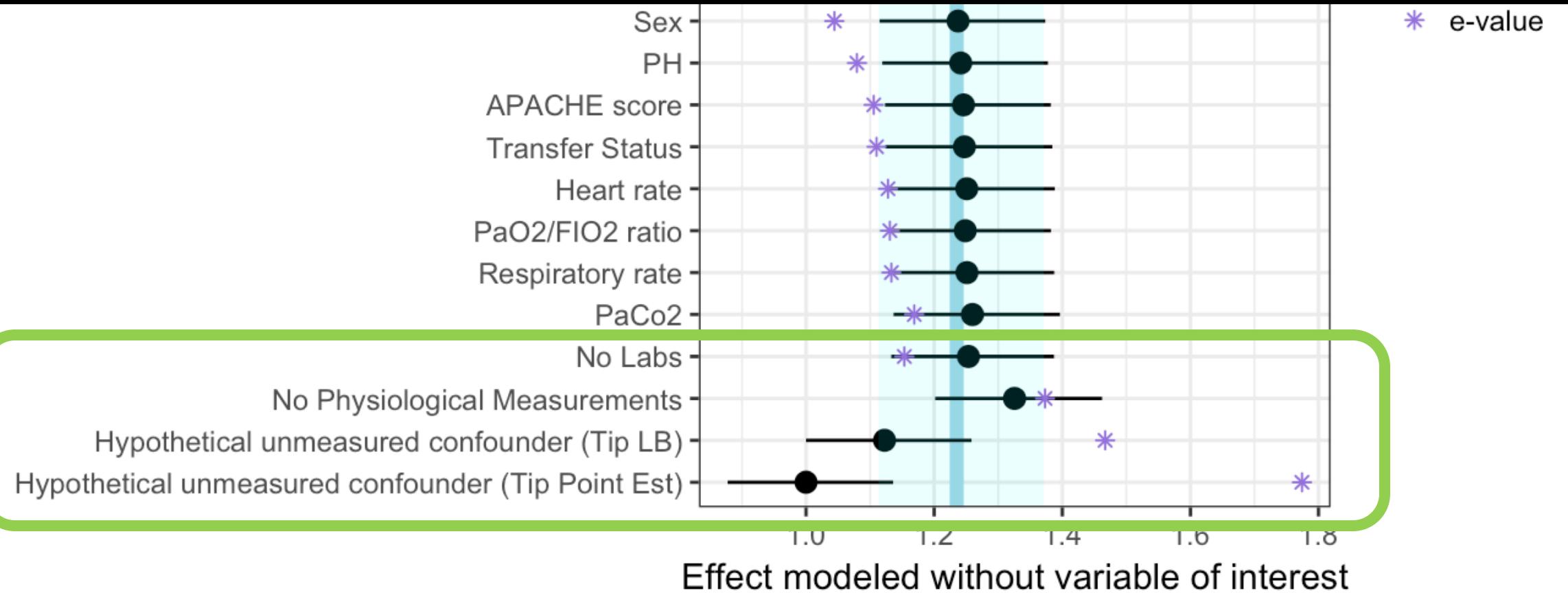
- We chose 20 covariates for demonstration purposes
 - demographics
 - comorbidities
 - physiological measurements
 - diagnosis categories
 - APACHE score
 - SUPPORT (probability of surviving 2 months)
 - DNR status on day 1

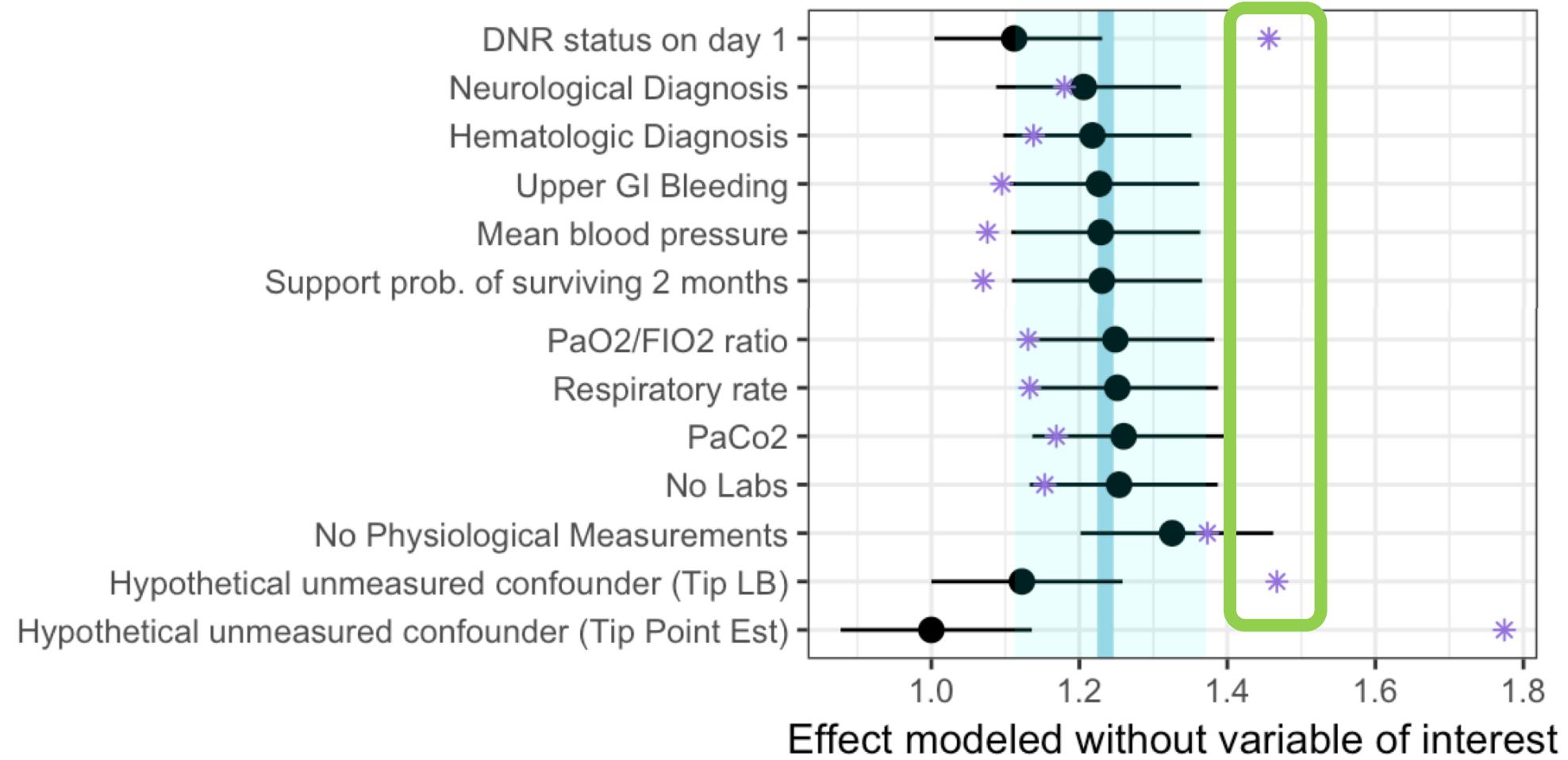
Connors et al (1996)

Right Heart Catheterization data

- Fit a propensity score model
- Use ATO weights
- Fit weighted cox model for 30 day survival







Thank you!



@LucyStats



<http://bit.ly/LucyStatsENAR2018>



lucymcgowan.com