

## Title in Progress

Amy Pitts, Hun Lee, Jimmy Kelliher,  
Tucker Morgan, and Waveley Qiu

2022-02-21

# Motivation

- ▶ Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians. . .

# Motivation

- ▶ Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians. . .
- ▶ . . . but outside of a randomized control trial (RCT), confounding variables can bias our estimates of treatment effects.

# Motivation

- ▶ Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians. . .
- ▶ . . . but outside of a randomized control trial (RCT), confounding variables can bias our estimates of treatment effects.
- ▶ Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders. . .

# Motivation

- ▶ Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians. . .
- ▶ . . . but outside of a randomized control trial (RCT), confounding variables can bias our estimates of treatment effects.
- ▶ Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders. . .
- ▶ . . . but there is no consensus on the best way to estimate standard errors when using the PSM algorithm.

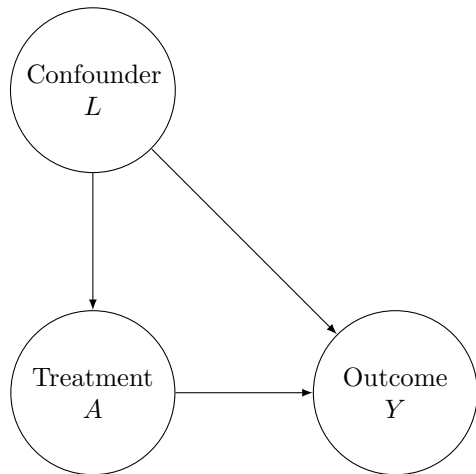
# Motivation

- ▶ Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians. . .
- ▶ . . . but outside of a randomized control trial (RCT), confounding variables can bias our estimates of treatment effects.
- ▶ Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders. . .
- ▶ . . . but there is no consensus on the best way to estimate standard errors when using the PSM algorithm.
- ▶ How can we assess which procedures reliably estimate standard errors?

# Motivation

A simulation study!

## A Quick Foray into Confounding





# Taking a Step Back, What is Propensity Score Matching?

- ▶ A *propensity score* is the probability that an individual receives a treatment  $A$ ; that is,  $P(A = 1)$ . In an RCT, treatments are randomized, and hence outcomes  $Y$  are independent of treatment  $A$ .

# Enter the Bootstrap

- ▶ Bootstrapping is one of the most common procedures for estimating standard errors.

# Enter the Bootstrap

- ▶ Bootstrapping is one of the most common procedures for estimating standard errors.
- ▶ The PSM algorithm intakes an unmatched dataset and outputs a matched one.

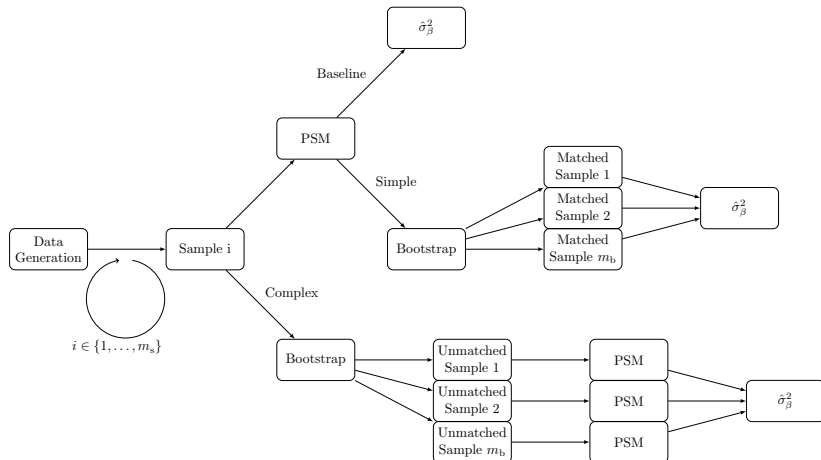
# Enter the Bootstrap

- ▶ Bootstrapping is one of the most common procedures for estimating standard errors.
- ▶ The PSM algorithm intakes an unmatched dataset and outputs a matched one.
- ▶ When do we execute the bootstrap - before the match or after it?

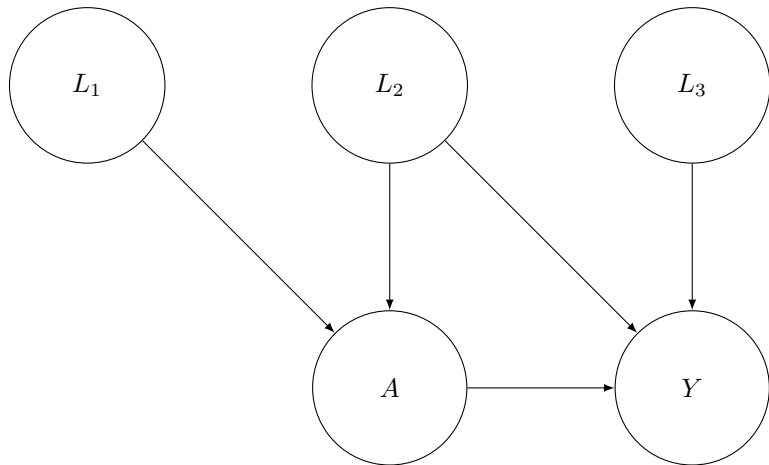
# Enter the Bootstrap

- ▶ Bootstrapping is one of the most common procedures for estimating standard errors.
- ▶ The PSM algorithm intakes an unmatched dataset and outputs a matched one.
- ▶ When do we execute the bootstrap - before the match or after it?
- ▶ Let's try both!

# Roadmap of the Simulation Study



## Data Generation



## Data Generation - Continuous Outcome

For each individual  $i \in \{1, \dots, n\}$ , we consider covariates  $L_{1i}, L_{2i}, L_{3i} \sim N(0, 1)$ . Treatments are distributed according to law  $A_i \sim B(\pi_i)$ , where  $\pi_i$  - the true propensity to be treated - is subject to the data-generating process

$$\log \left( \frac{\pi_i}{1 - \pi_i} \right) = \alpha_0 + \alpha_1 L_{1i} + \alpha_2 L_{2i}.$$

Given this, we further define the data-generating process of our continuous outcome via

$$Y_i = \beta_1 A_i + \beta_2 L_{2i} + \beta_3 L_{3i} + \varepsilon_i,$$

where  $\varepsilon_i$  denotes random error. Because  $L_{2i}$  effects both  $A_i$  and  $Y_i$ , it acts as a confounder in estimating the treatment effect.



## Data Generation - Binary Outcome

For each individual  $i \in \{1, \dots, n\}$ , we consider covariates  $L_{1i}, L_{2i}, L_{3i} \sim N(0, 1)$ . Treatments are distributed according to law  $A_i \sim B(\pi_i)$ , where  $\pi_i$  - the true propensity to be treated - is subject to the data-generating process

$$\log \left( \frac{\pi_i}{1 - \pi_i} \right) = \alpha_0 + \alpha_1 L_{1i} + \alpha_2 L_{2i}.$$

Given this, we further define the data-generating process of our binary outcome via  $Y_i \sim B(\tau_i)$  where

$$\log \left( \frac{\tau_i}{1 - \tau_i} \right) = \beta_1 A_i + \beta_2 L_{2i} + \beta_3 L_{3i}.$$

Observe that we have omitted a random error term, as realizations of  $Y_i$  are innately subject to noise.

# Parameters of Interest

- ▶ The sample size of each dataset  $n_{\text{sample}} \in \{100, 1000\}$
- ▶ The population proportion of treated individuals  $\pi \in \{0.113, 0.216, 0.313\}$
- ▶ The true average treatment effect  $\beta_1 \in \{0.15, 0.30\}$

## *Other Parameters*

- ▶ The number of datasets  $m_{\text{sample}} = 100$
- ▶ The number of bootstrap re-sample  $m_{\text{boot}} = 500$
- ▶ The sample size of bootstrap re-samples  $n_{\text{simple}} = n_{\text{complex}} = n_{\text{sample}} \times \pi$
- ▶ Strength of Covariate Correlation on Treatment Status  $\alpha_1, \alpha_2$  (continuous data (1,2), binary data ( $\log(1.25), \log(1.75)$ )) )
- ▶ Strength of Covariate Correlation on Outcome Variable  $\beta_2, \beta_3$  (continuous data (2,1), binary data ( $\log(1.75), \log(1.25)$ )) )

# Measures of Interest

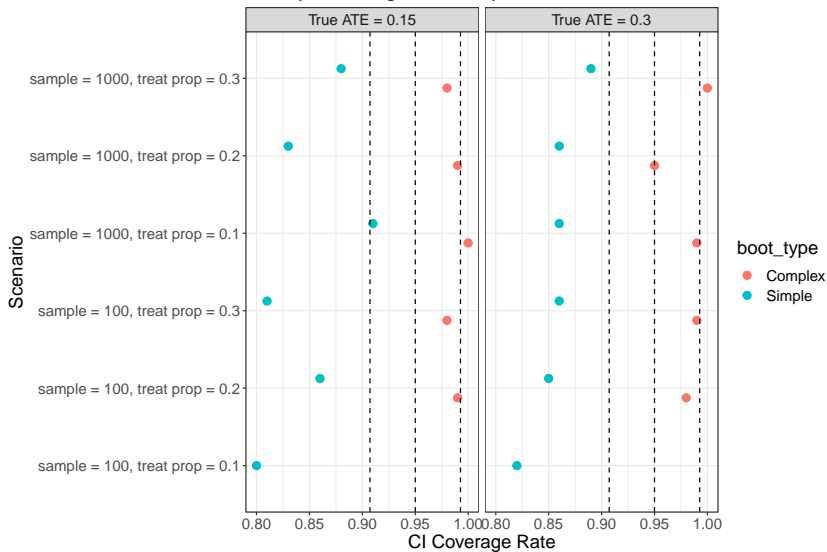
- ▶ **Coverage Rate:** Looks at the rate of the true average treatment effect falling in the 95% confidence intervals.  
 $\hat{ATE} \pm 1.96 \times SE$
- ▶ **Standard Error:** the variability of the average estimate.

## *Other Measures*

- ▶ **Bias:** This is mean of the average estimate subtract the true ATE
- ▶ **95% Confidence Intervals:**

# Results

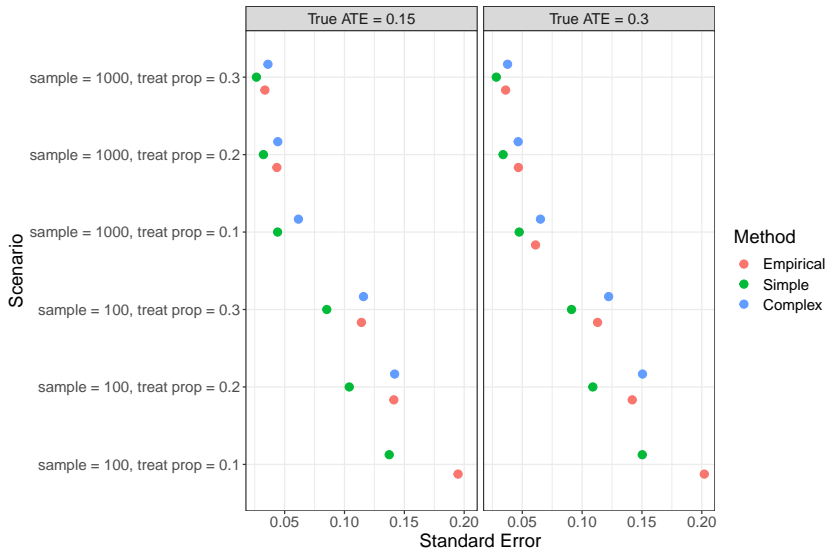
Binary Coverage Rates by Parameters of Interest



# Results

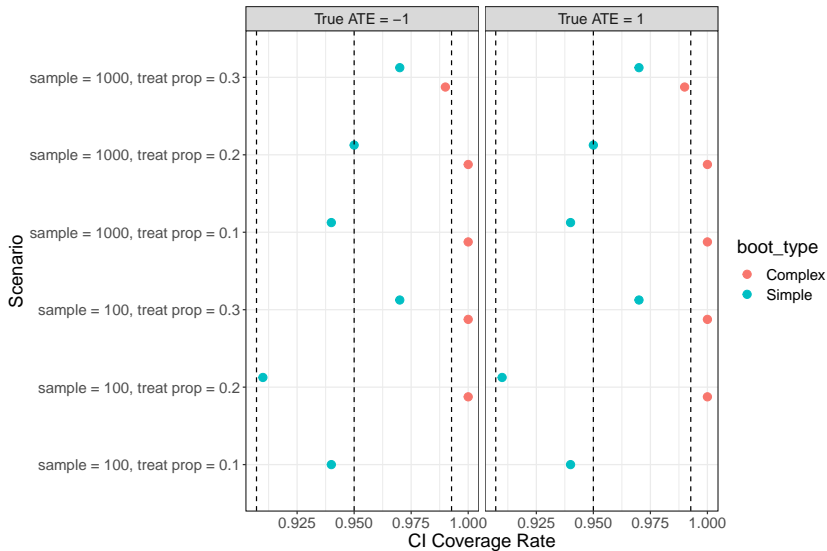
## Warning: Removed 1 rows containing missing values (geom\_

Binary Simulation Standard Error



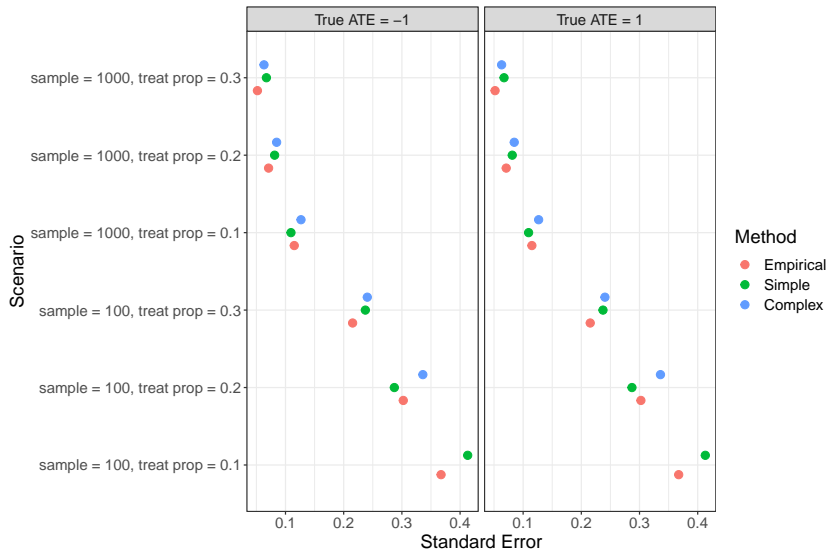
# Results

Continuous Coverage Rates by Parameters of Interest



# Results

Continuous Simulation Standard Error



## Summary of Results



# Limitations

- ▶ Sample size / treatment (or exposure) prevalence

# Limitations

- ▶ Sample size / treatment (or exposure) prevalence
- ▶ Small number of initial samples, limited in detecting differences in coverage rate

# Limitations

- ▶ Sample size / treatment (or exposure) prevalence
- ▶ Small number of initial samples, limited in detecting differences in coverage rate
- ▶

# Limitations

- ▶ Sample size / treatment (or exposure) prevalence
- ▶ Small number of initial samples, limited in detecting differences in coverage rate
- ▶
- ▶

## Future Work

- ▶ Larger number of initial samples, narrower coverage window

## Future Work

- ▶ Larger number of initial samples, narrower coverage window
- ▶ Increased sample size, changes in bootstrap performance?

## Future Work

- ▶ Larger number of initial samples, narrower coverage window
- ▶ Increased sample size, changes in bootstrap performance?
- ▶ Changes in treatment propensity model

## Future Work

- ▶ Larger number of initial samples, narrower coverage window
- ▶ Increased sample size, changes in bootstrap performance?
- ▶ Changes in treatment propensity model
- ▶ Non-normal distributions of covariates