Overcoming positivity violations in observational studies with non-overlap bounds

Herb Susmann, PhD
Division of Biostatistics
Department of Population Health
NYU Grossman School of Medicine

October 22, 2025
Emerging Leaders in Research Lecture Series

NYU Grossman School of Medicine

herbsusmann.com/dph2025

Alec McClean
alecmcclean.github.io

Iván Díaz
idiaz.xyz

arXiv > stat > arXiv:2509.20206

Statistics > Methodology

[Submitted on 24 Sep 2025]

Non-overlap Average Treatment Effect Bounds

Herbert P. Susmann, Alec McClean, Iván Díaz

The average treatment effect (ATE), the mean difference in potential outcomes under treatment and control, is a canonical causal effect. Overlap, which says that all subjects have non-zero probability of either treatment status, is necessary to identify and estimate the ATE. When overlap fails, the standard solution is to change the estimand, and target a trimmed effect in a subpopulation satisfying overlap; however, this no longer addresses the original goal of estimating the ATE. When the outcome is bounded, we demonstrate that this compromise is unnecessary. We derive non-overlap bounds: partial identification bounds on the ATE that do not require overlap. They are the sum of a trimmed effect within the overlap subpopulation and worst-case bounds on the ATE in the non-overlap subpopulation. Non-overlap bounds have width proportional to the size of the non-overlap subpopulation, making them informative when overlap violations are limited -- a common scenario in practice. Since the bounds are non-smooth functionals, we derive smooth approximations of them that contain the ATE but can be estimated using debiased estimators leveraging semiparametric efficiency theory. Specifically, we propose a Targeted Minimum Loss-Based estimator that is $\sqrt{n}$-consistent and asymptotically normal under nonparametric assumptions on the propensity score and outcome regression. We then show how to obtain a uniformly valid confidence set across all trimming and smoothing parameters with the multiplier bootstrap. This allows researchers to consider many parameters, choose the tightest confidence interval, and still attain valid coverage. We demonstrate via simulations that non-overlap bound estimators can detect non-zero ATEs with higher power than traditional doubly-robust point estimators. We illustrate our method by estimating the ATE of right heart catheterization on mortality.

Summary

We propose a **robust** method for estimating treatment effects from observational data.

We propose a method for learning about treatment effects without requiring **overlap**.

We propose **partial identification bounds** for the Average Treatment Effect and robust targeted estimators that can achieve **higher power** than traditional doubly-robust estimators when overlap fails.

Motivating example

- Scientific question: what is the effect of right heart catheterization (RHC) on survival ? (Murphy and Cluff, 1990; Connors et al., 1996)
- Study population: critically ill adult patients.
- Exposure: right heart catheterization within 24 hours of admittal.
- Outcome: 30-day mortality.
- Design: prospective cohort study.
- Setting: five US hospitals, 1989-1994.
- Data: 5,735 critically ill adult patients.
    - 72 measured covariates

Formalizing the scientific question

- What is the treatment effect of interest?
  - *"Estimand: A description of the exact treatment effect a study aims to quantify." (Kahan et al., 2024).*
- What is the average treatment effect over the study population of critically ill patients?
- We can formalize the estimand using the concept of *counterfactual outcomes* from *causal inference*.

Defining the estimand with counterfactuals

- Observed data:
  - X: set of patient characteristics.
  - A: whether the patient received right-heart catheterization within 24 hours of admittance.
  - Y: 30-day survival.
- Counterfactual outcomes:
  - Y^1: 30-day survival *if the patient had received RHC*.
  - Y^0: 30-day survival *if the patient had* **not** *received RHC*.
- Average Treatment Effect:

$$\text{ATE} = \mathbb{E}\left[Y^1 - Y^0\right].$$

- Interpretation: difference in mean 30-day survival if *every* patient had been given RHC versus if *no* patient had been given RHC.

Estimating the ATE

- Average Treatment Effect:

$$\text{ATE} = \mathbb{E}\left[Y^1 - Y^0\right].$$

- We can't observe *both* Y^1 **and** Y^0 for the same patient.
- **Instead**, we can make an *educated guess*:

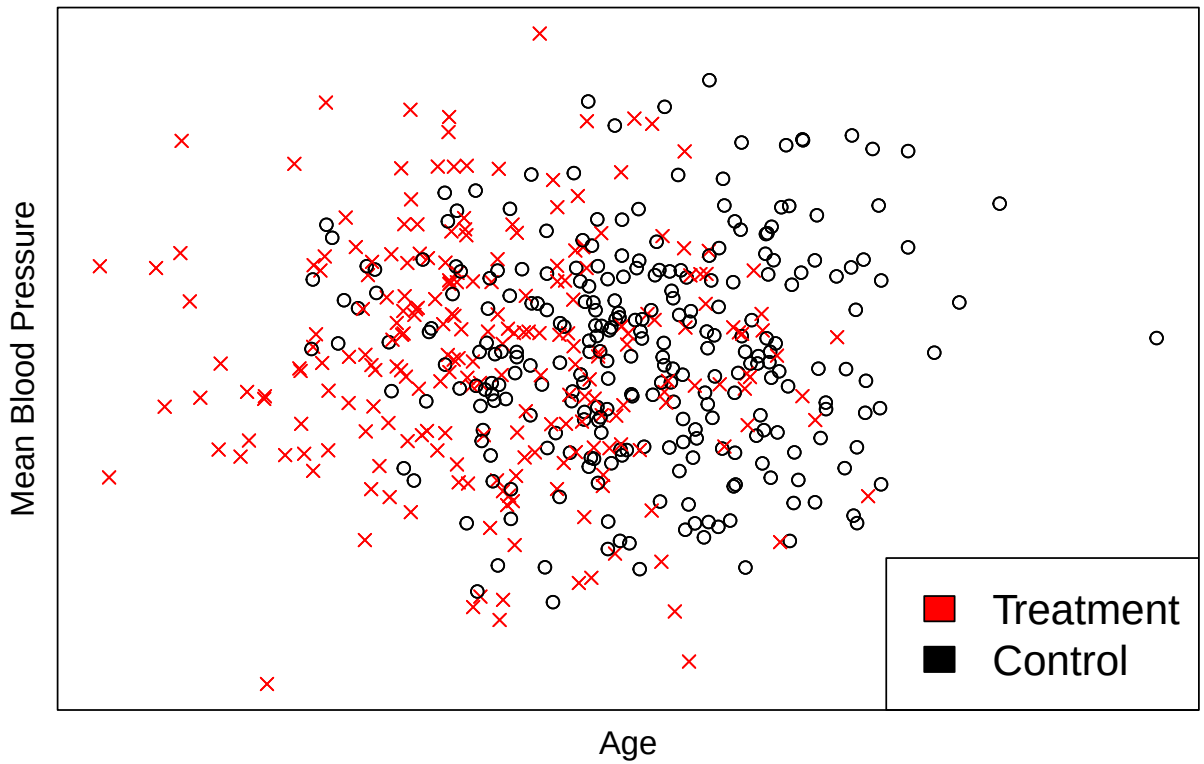
$$\text{ATE} = \mathbb{E}[\mathbb{E}[Y \mid A = 1, X]] - \mathbb{E}[\mathbb{E}[Y \mid A = 0, X]]$$

Identification Assumptions

Core assumptions:

- **No unmeasured confounding.**
  All common causes of RHC and 30-day survival are measured.
- **Overlap.**
  Every type of patient had a positive probability of receiving RHC, or not receiving RHC.

Overlap is violated when some strata rarely receive either the treatment or the control



$$\text{ATE} = \mathbb{E}[\mathbb{E}[Y \mid A = 1, X] - \mathbb{E}[Y \mid A = 0, X]]$$

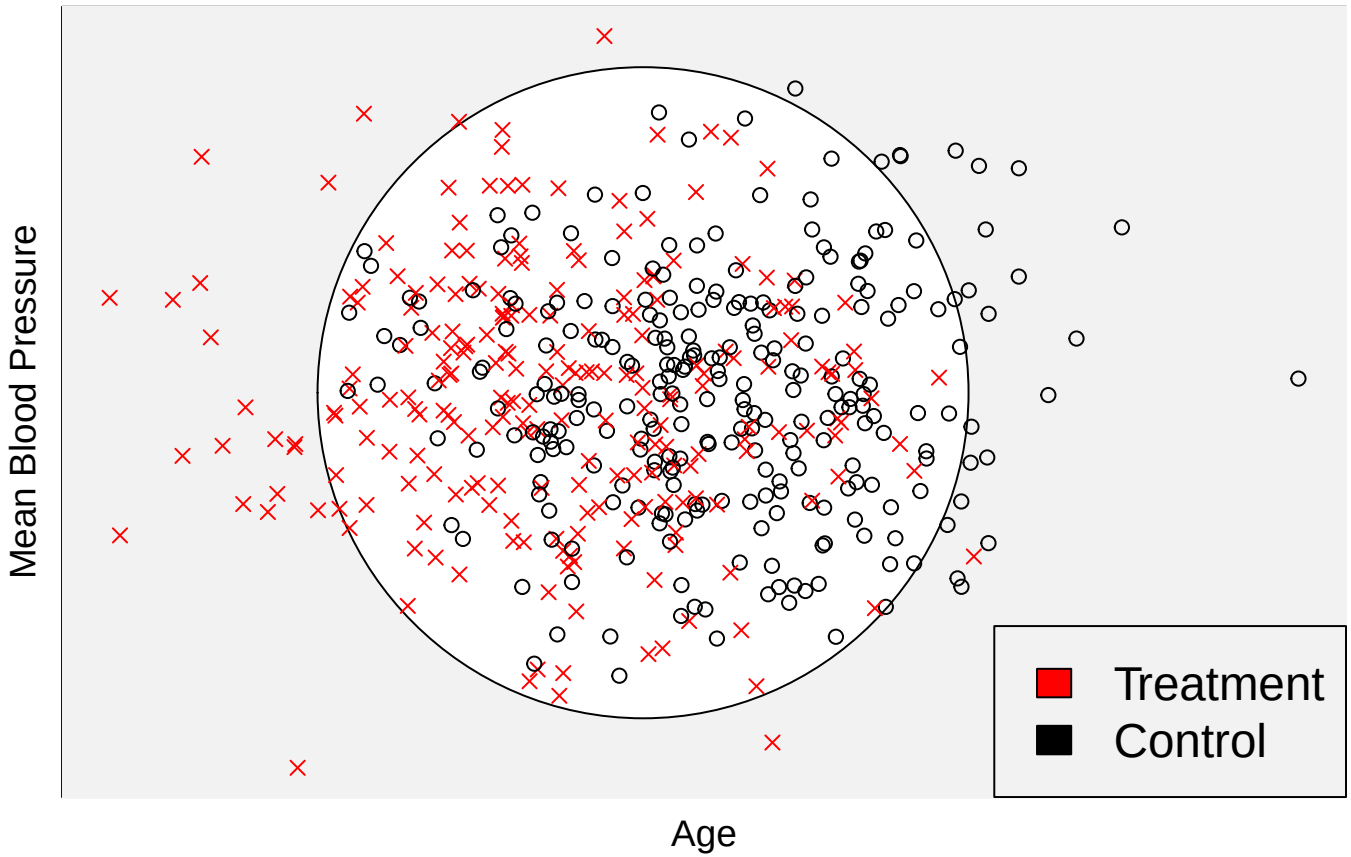
When overlap fails

- Practical implications of overlap violations:
  - Bias,
  - High uncertainty (wide confidence intervals).
- Using a traditional approach, the estimated Average Treatment Effect of RHC on 30-day survival has the 95% confidence interval $(-1, 1)$
- The uncertainty is so high we learn **nothing** about the treatment effect.

What to do when overlap fails

- Current best practice when positivity fails is to *change the estimand*.
- Common techniques target an *Overlap Treatment Effect*.
  - Propensity score trimming (Crump et al., 2009)
  - Overlap weighting (Li et al., 2018)
  - Incremental propensity score interventions (Kennedy, 2019)

Overlap Treatment Effect

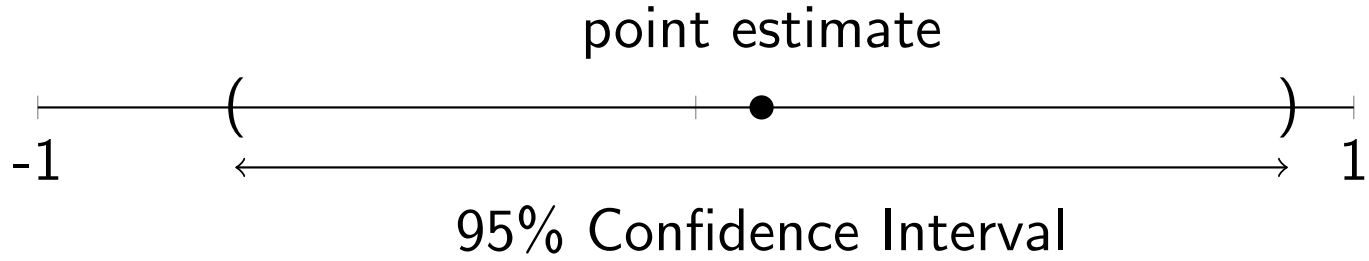


Changing the estimand changes what you learn!

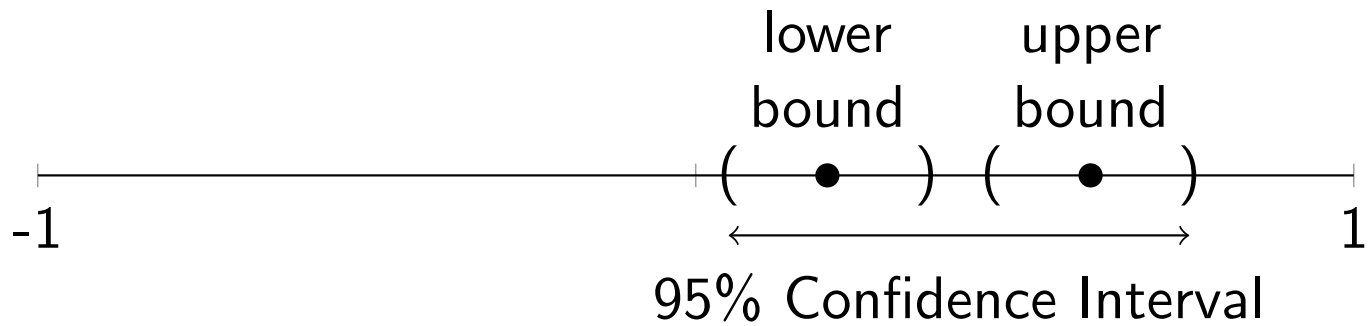
- The new estimand *might not answer the scientific question of interest* (Rizk 2025 Journal of Clinical Epidemiology).
- **Average Treatment Effect:** expected mean difference in survival *over population of critically ill adult patients.*
- **Overlap Treatment Effect:** expected mean difference in survival *over population of critically ill adult patients satisfying overlap.*

Our proposal: estimate *bounds* on the treatment effect

- Our approach: **estimate bounds for the ATE.**
- Traditional approach: estimate **point estimate** and **95% confidence interval**:



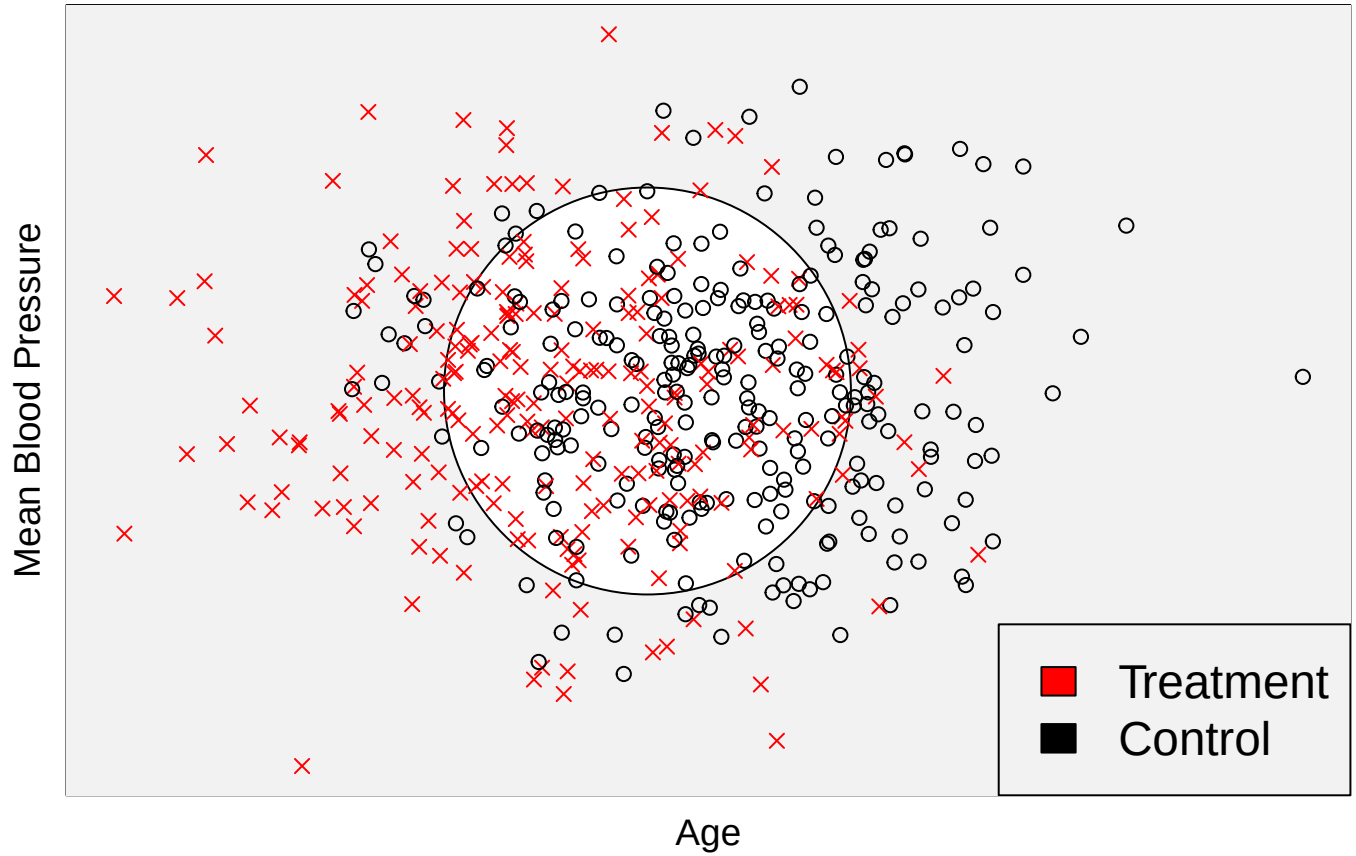
- Our approach: estimate **lower** and **upper bounds** and combine to form a **95% confidence interval**:

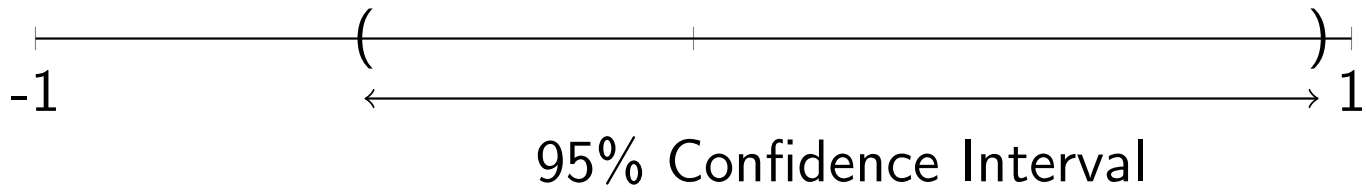


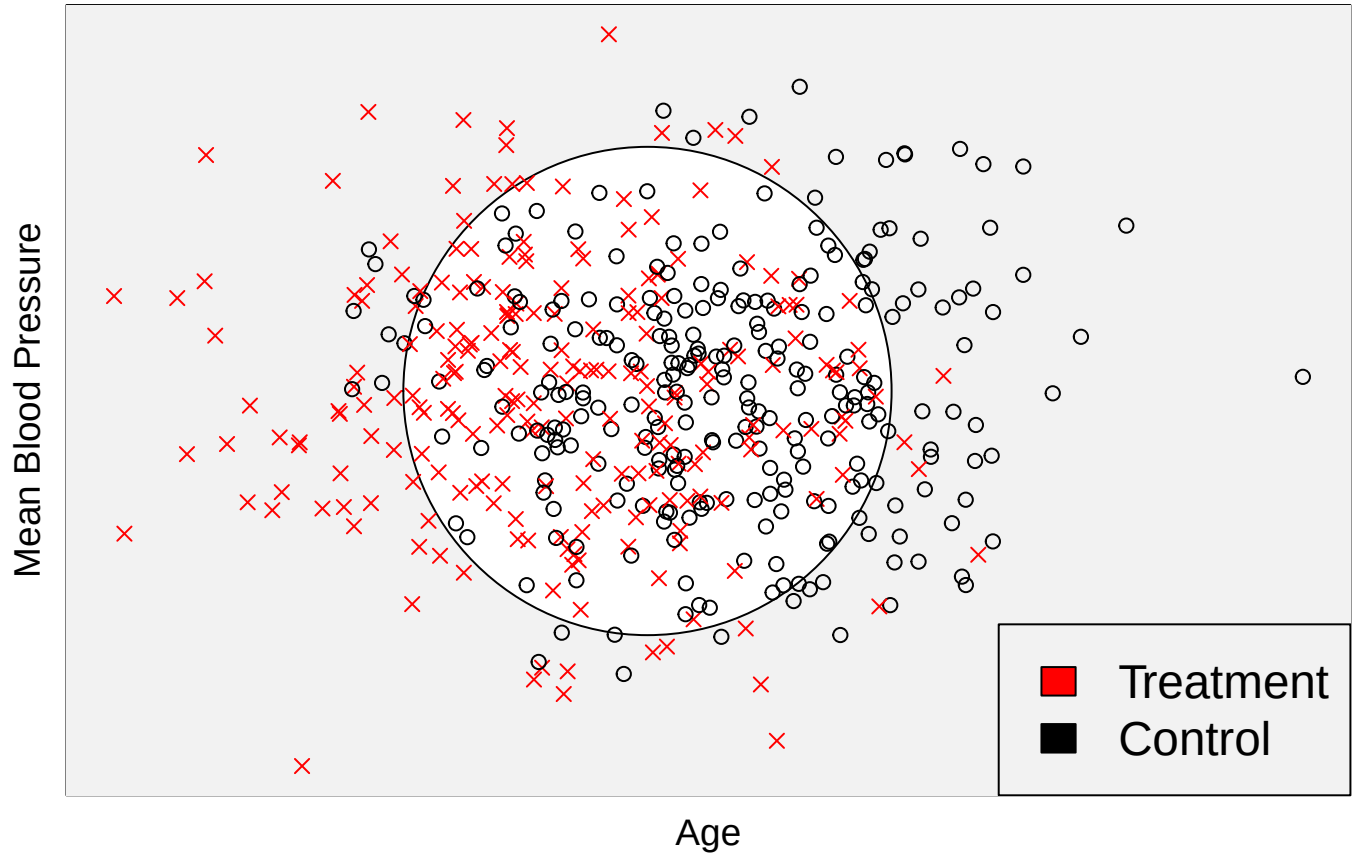
How the bounds are constructed

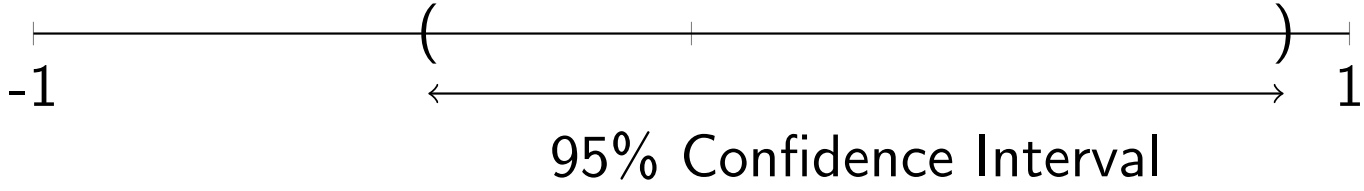
- Divide the population into overlap and non-overlap subpopulations.
- Set an *overlap threshold* c.
  - If $\mathbb{P}(\text{treatment} \mid \text{covariates}) \in [c, 1-c]$ then patient is in the *overlap* subpopulation.
  - If $\mathbb{P}(\text{treatment} \mid \text{covariates}) \notin [c, 1-c]$ then patient is in the *non-overlap* subpopulation.
- Assume that the outcome is *bounded*: $0 \leq Y \leq 1$.
- The ATE is bounded by:

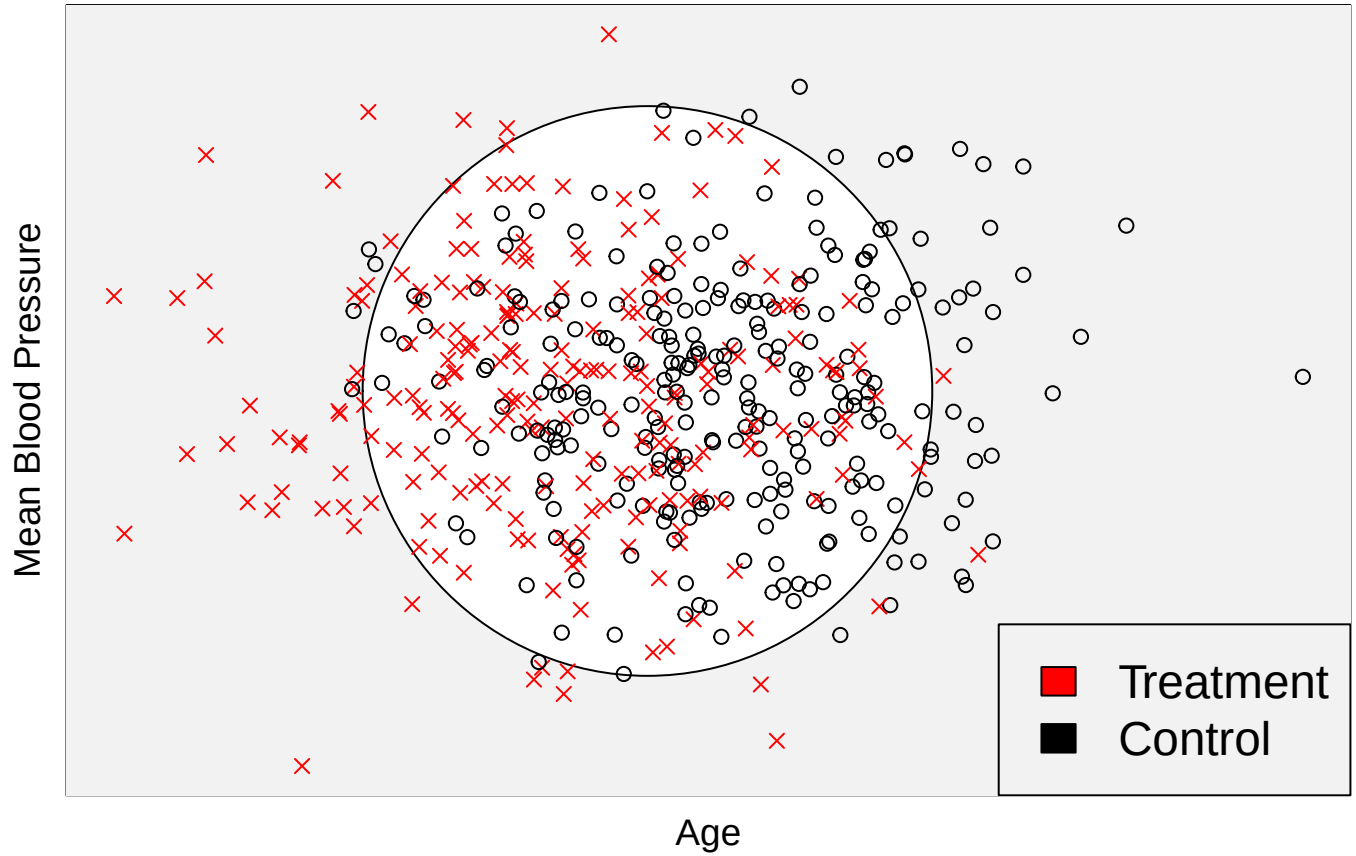
$$\text{ATE} \in \mathbb{E}(Y^1 - Y^0 \mid \text{overlap})\mathbb{P}(\text{overlap}) \pm \mathbb{P}(\text{non-overlap}).$$

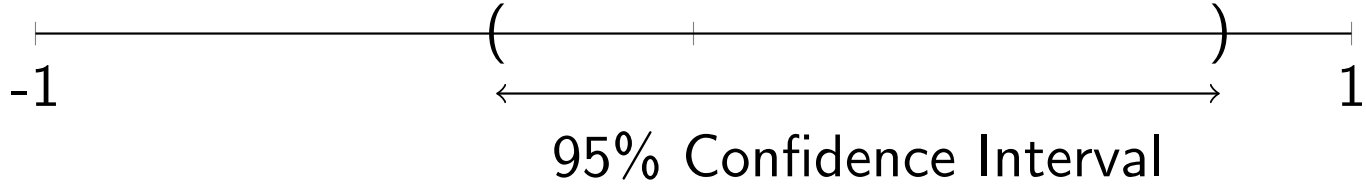

Mean Blood Pressure
Age
Treatment
Control

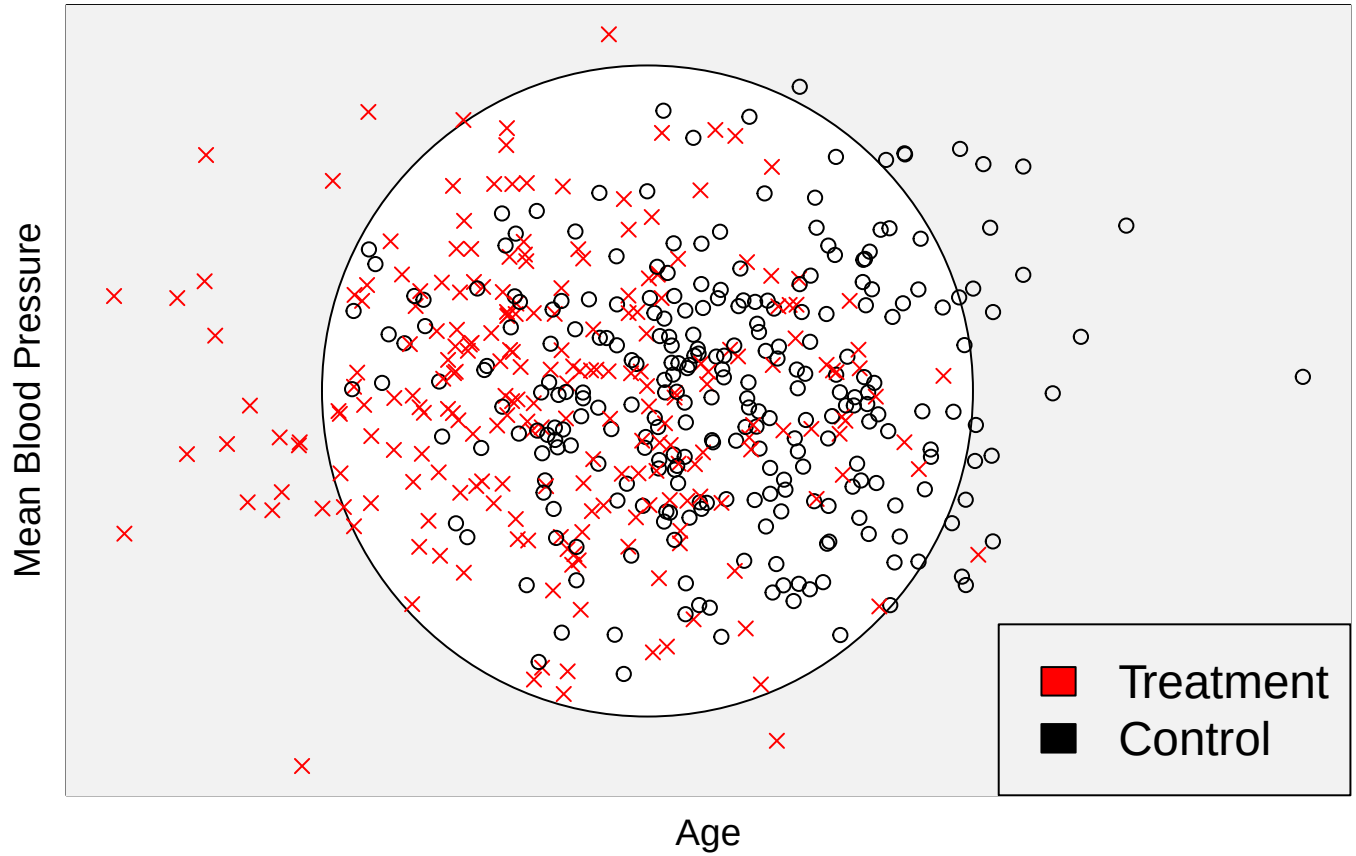

-1
1
95% Confidence Interval

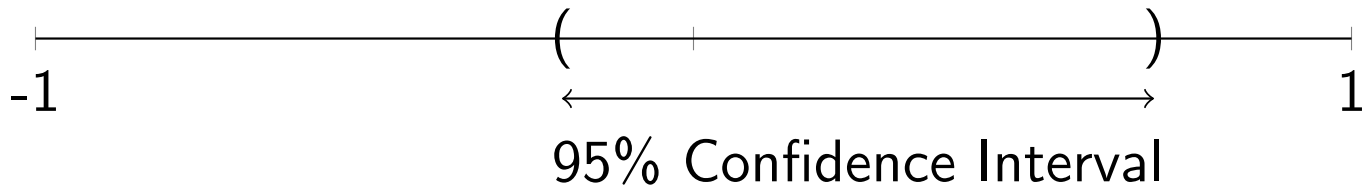

Mean Blood Pressure
Age
Treatment
Control

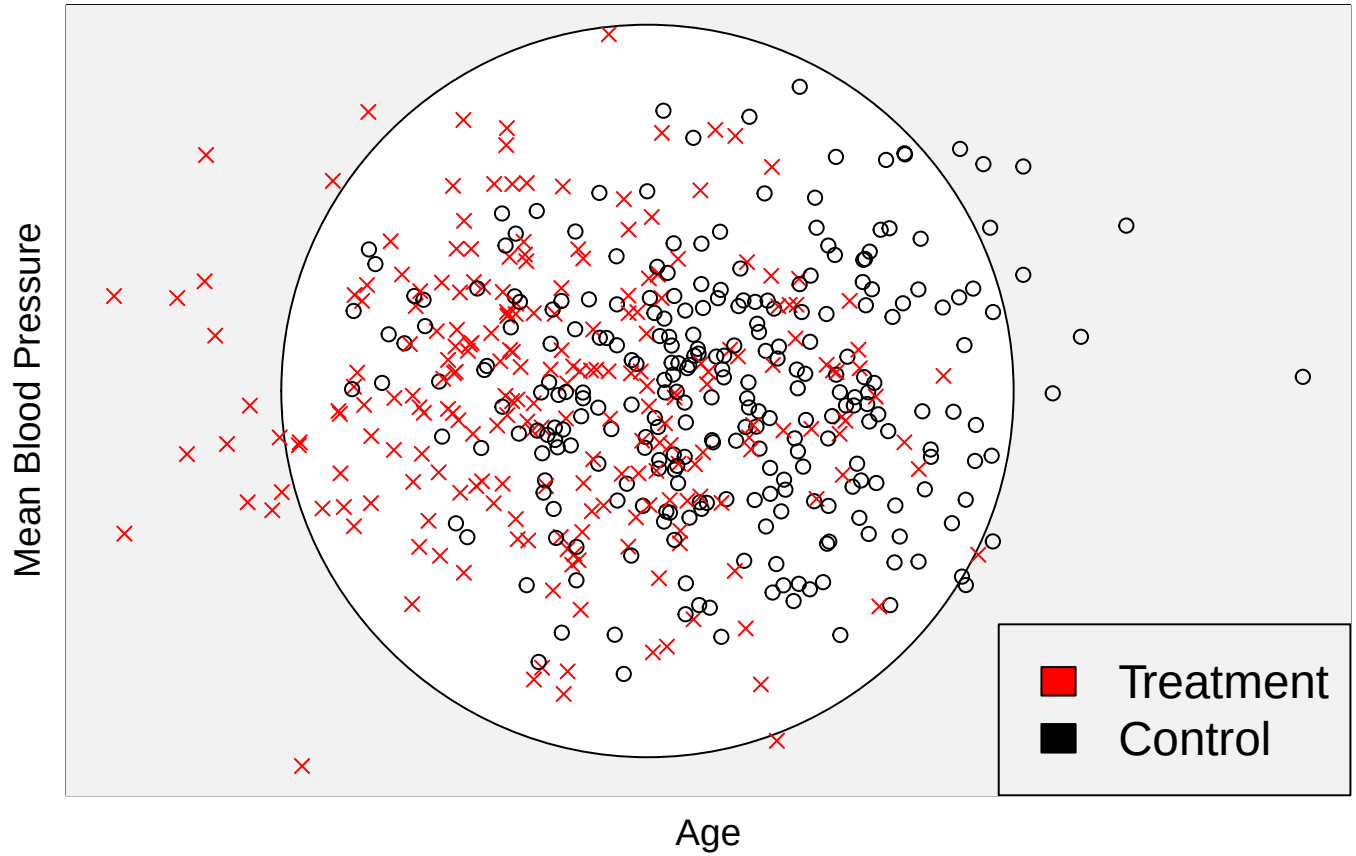

-1
1
95% Confidence Interval

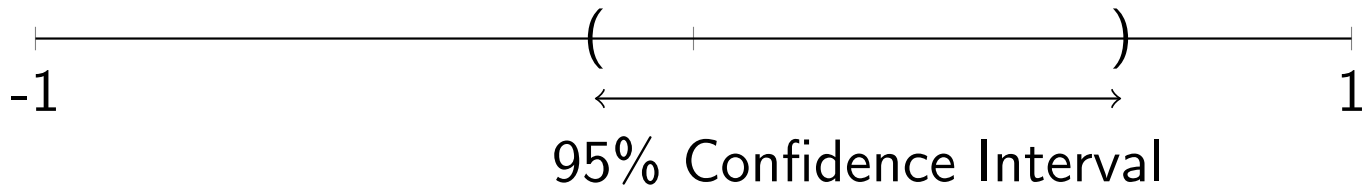

Mean Blood Pressure
Age
Treatment
Control

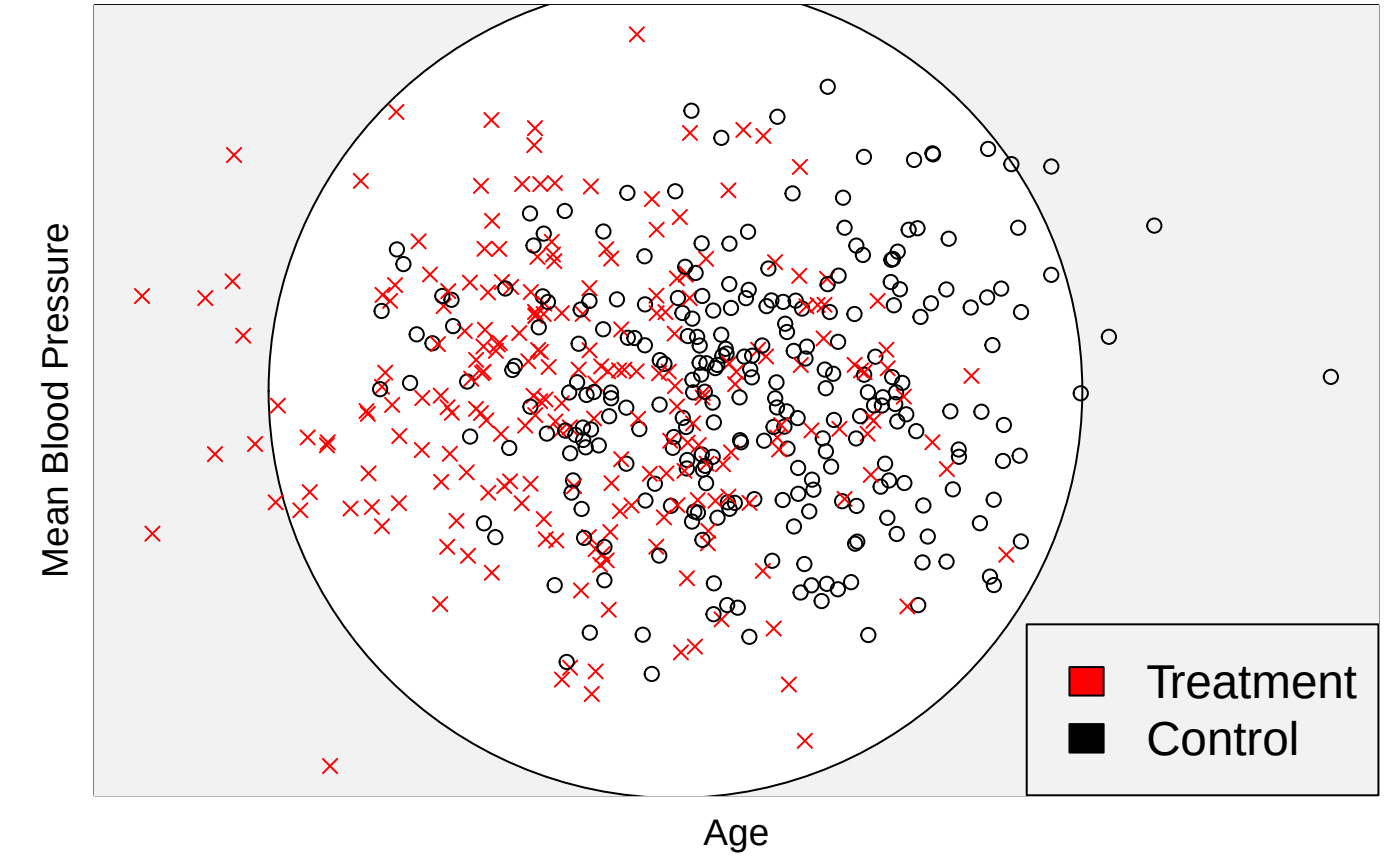

-1
1
95% Confidence Interval

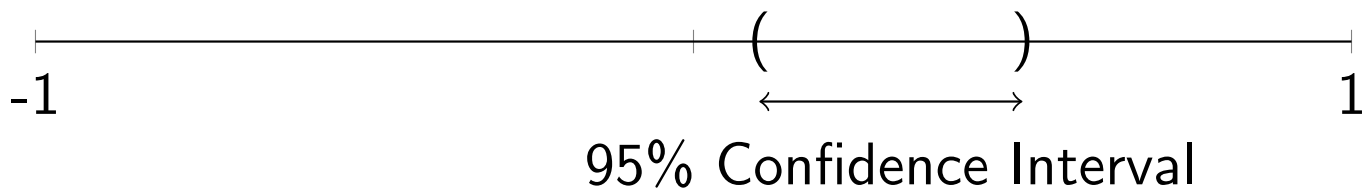

Mean Blood Pressure
Age
Treatment
Control

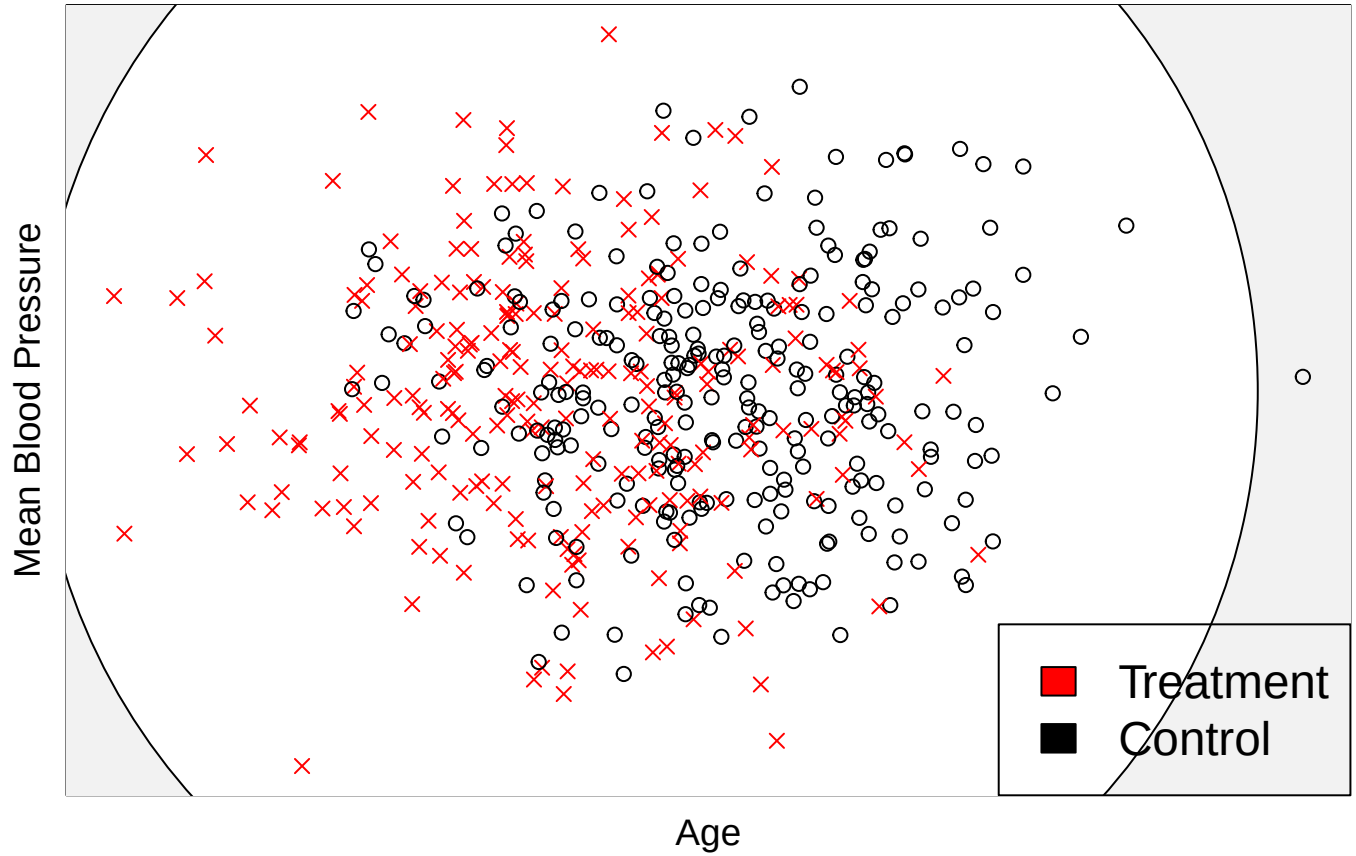

-1
1
95% Confidence Interval

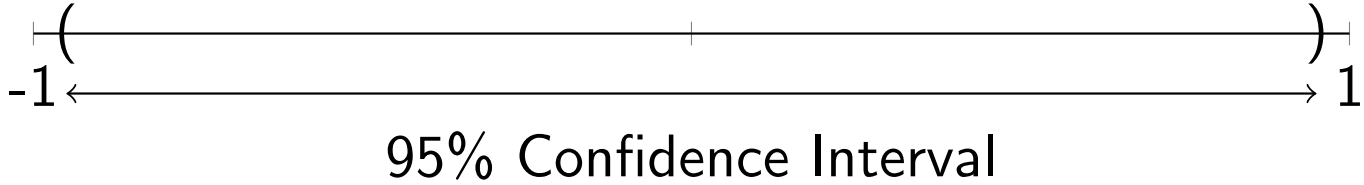

Mean Blood Pressure
Age
Treatment
Control


-1
1
95% Confidence Interval

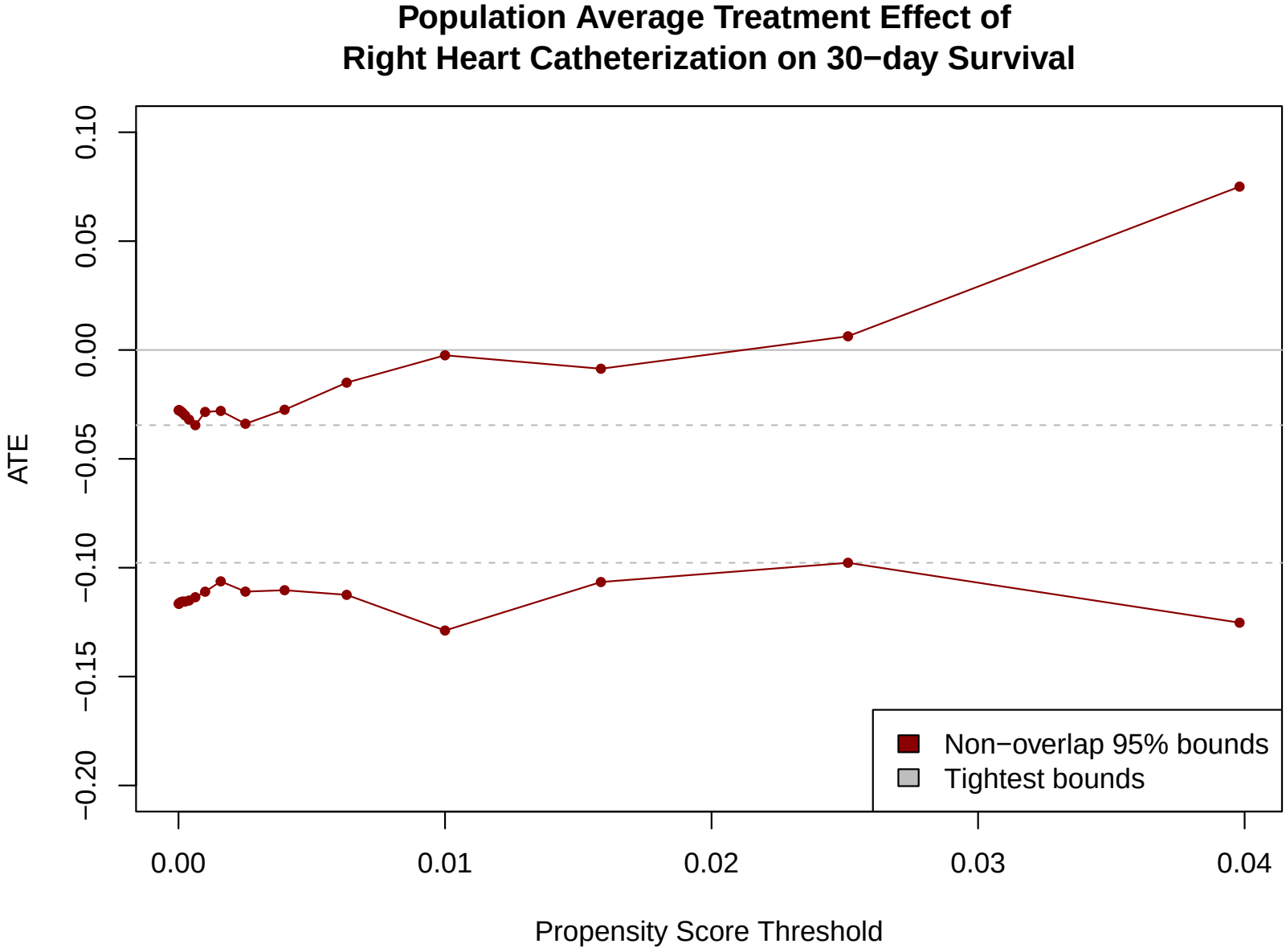

Mean Blood Pressure
Age
Treatment
Control


-1
1
95% Confidence Interval





95% Confidence Interval


Population Average Treatment Effect of
Right Heart Catheterization on 30−day Survival
ATE
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
0.00
0.01
0.02
0.03
0.04
Propensity Score Threshold
Non−overlap 95% bounds
Tightest bounds

Under the hood

- Estimation is based on Targeted Minimum Loss-based Estimation (TMLE).
  - Allows for the use of flexible machine-learning and AI to estimate propensity score and outcome regression models.
- When there are no overlap violations, bounds reduce to traditional confidence intervals around a doubly-robust point estimate.
- Uniform confidence sets via multiplier bootstrap allow for testing a range of thresholds without multiple testing problems.

arXiv > stat > arXiv:2509.20206

Statistics > Methodology

[Submitted on 24 Sep 2025]

Non-overlap Average Treatment Effect Bounds

Herbert P. Susmann, Alec McClean, Iván Díaz

The average treatment effect (ATE), the mean difference in potential outcomes under treatment and control, is a canonical causal effect. Overlap, which says that all subjects have non-zero probability of either treatment status, is necessary to identify and estimate the ATE. When overlap fails, the standard solution is to change the estimand, and target a trimmed effect in a subpopulation satisfying overlap; however, this no longer addresses the original goal of estimating the ATE. When the outcome is bounded, we demonstrate that this compromise is unnecessary. We derive non-overlap bounds: partial identification bounds on the ATE that do not require overlap. They are the sum of a trimmed effect within the overlap subpopulation and worst-case bounds on the ATE in the non-overlap subpopulation. Non-overlap bounds have width proportional to the size of the non-overlap subpopulation, making them informative when overlap violations are limited -- a common scenario in practice. Since the bounds are non-smooth functionals, we derive smooth approximations of them that contain the ATE but can be estimated using debiased estimators leveraging semiparametric efficiency theory. Specifically, we propose a Targeted Minimum Loss-Based estimator that is $\sqrt{n}$-consistent and asymptotically normal under nonparametric assumptions on the propensity score and outcome regression. We then show how to obtain a uniformly valid confidence set across all trimming and smoothing parameters with the multiplier bootstrap. This allows researchers to consider many parameters, choose the tightest confidence interval, and still attain valid coverage. We demonstrate via simulations that non-overlap bound estimators can det… …er power than traditional doubly-robust point estimators. We illustrate our method b… …t heart catheterization on mortality.

A … n: … or

2. For B bootstrap samples, draw i.… $_{i=1}^{n}$ with $\mathbb{E}(\xi)=0$, $\mathbb{E}(\xi^2)=1$, and form the studentized residuals for each …

$$T^{(b)} \quad \frac{1}{\;} \sum^{n} \xi^{(b)} \left\{ \varphi_{L,k}(Z_i, \quad \ldots \quad \frac{1}{\;} \sum^{n} \xi^{(b)} \left\{ \varphi_{U,k}(Z_i,\widehat{\eta}) - \widehat{U}_k \right\}$$

Practical benefits: higher precision, higher power

	Mean Width		Power	
N	Traditional	Bounds	Traditional	Bounds
(A) Simulations conditional on non-overlap				
100	1.91	**0.46**	0.02	**0.49**
250	1.83	**0.29**	0.07	**0.88**
500	1.65	**0.22**	0.16	**1.00**
1000	1.35	**0.17**	0.31	**1.00**

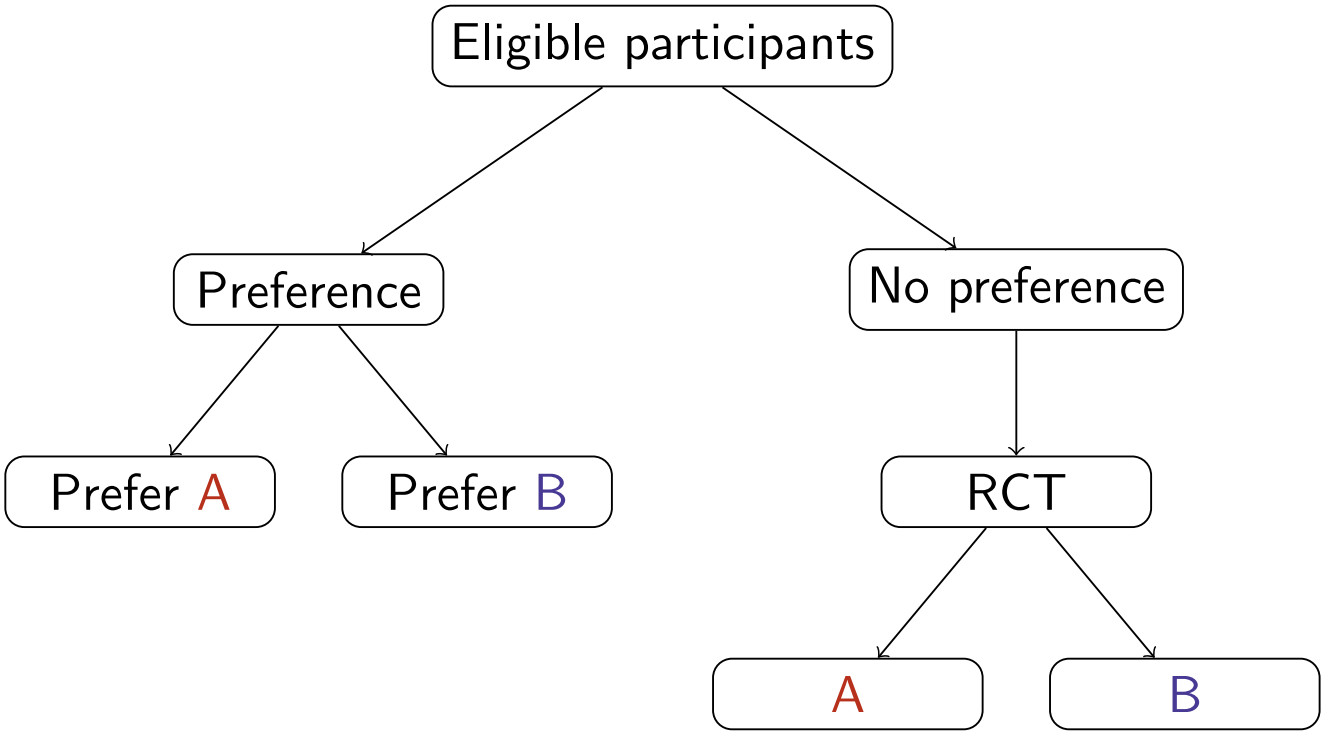
No free lunch: tradeoffs

- No need for overlap to hold, but your outcome must be *bounded*.
- Trade point identification for bounds.
- Only performs better than traditional approaches when the size of the non-overlap subpopulation is small relative to the effect size.

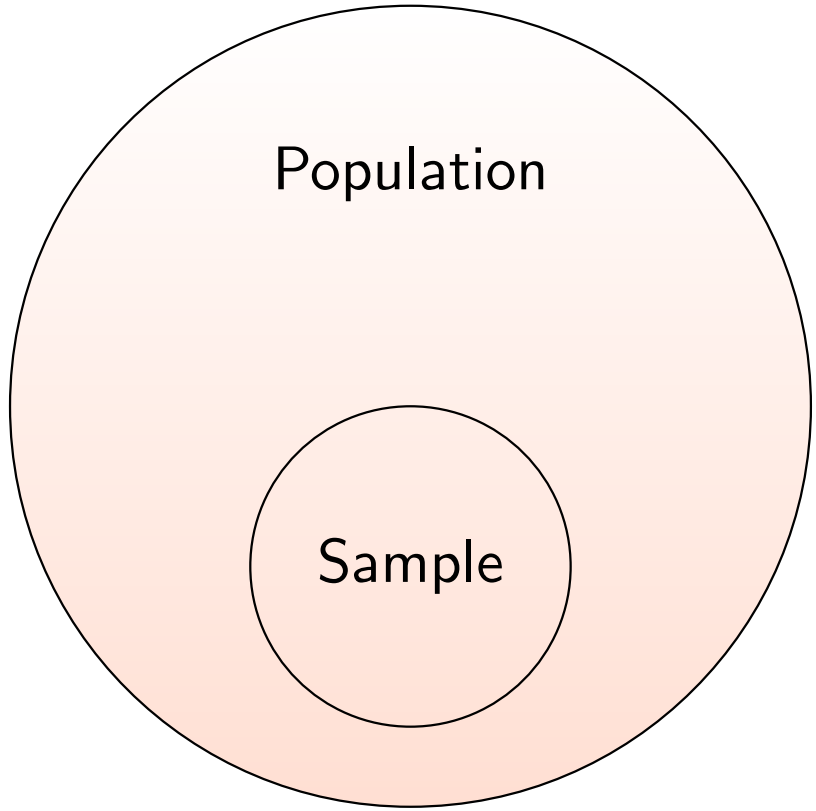
Target the estimand that answers the scientific question

- What is the estimand that answers the scientific question of interest?
- If the scientific estimand is the Average Treatment Effect, *then target the Average Treatment Effect*.
- If there are overlap violations that make estimating a point estimate difficult, *non-overlap bounds may nevertheless yield useful information*.
- As a *secondary* analysis, an alternative estimand can be estimated and reported.

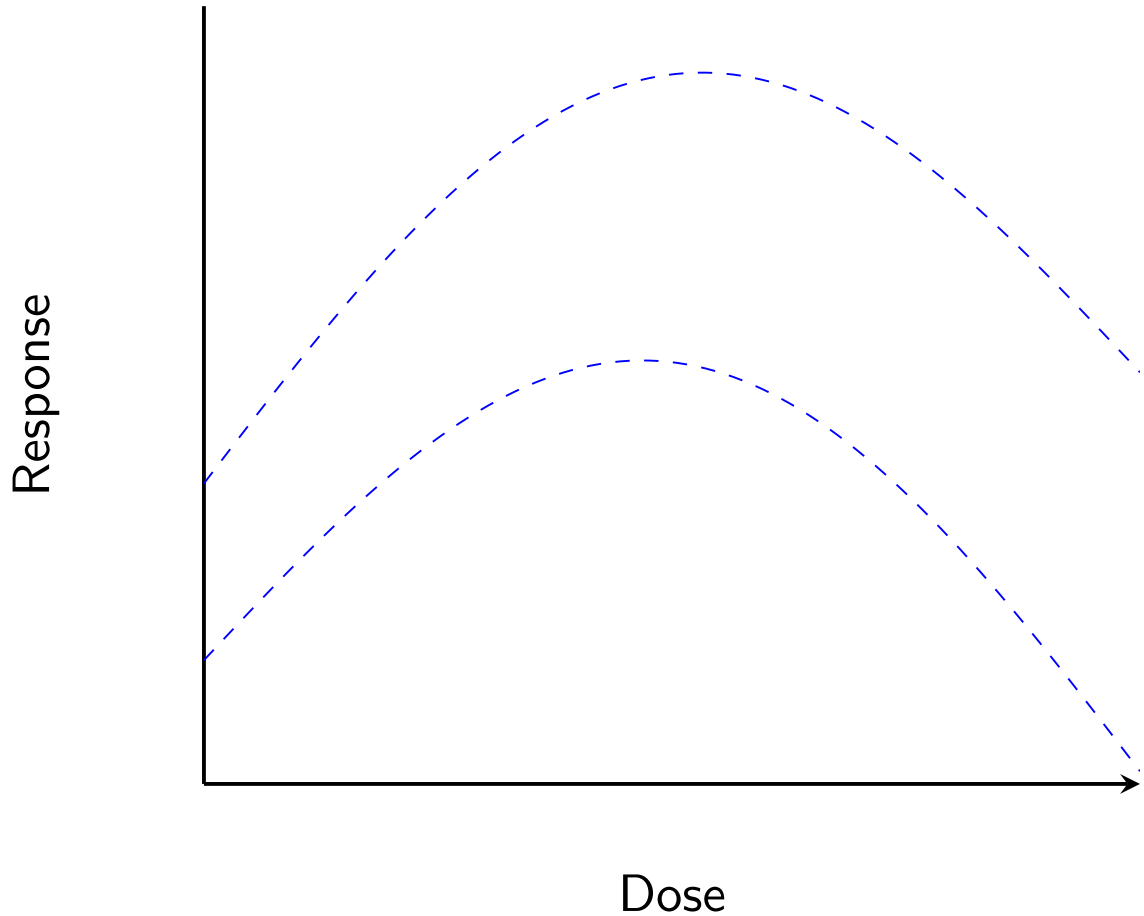
Next steps: partially randomized patient preference trials



Next steps: generalizing trial results



Next steps: dose-response curves



Summary

Setting: estimating treatment effects in observational studies.

Problem: when there is not **overlap** between treatment and control groups, traditional estimators can fail.

Solution: focus on estimating **non-overlap bounds** for the treatment effect.

Summary

`effectbounds` R package
`github.com/herbps10/effectbounds`

README | Apache-2.0 license

effectbounds

R-CMD-check.yaml passing

Overview

The `effectbounds` package provides tools for estimating non-overlap bounds for causal effects.

The identification of causal effects typically relies on the *overlap assumption* (also known as *positivity*), which requires that all units have a positive probability of being in either the treatment or control group.

When overlap fails in finite-samples, with some units having very small estimated probability of receiving the treatment (or control), then estimators of the causal effect can perform poorly.

Non-overlap bounds are an approach for estimating causal effects even when non-overlap is violated, by focusing on estimating *bounds* on the effect.

Installation

You can install the development version of effectbounds from GitHub:

```
# install.packages("devtools")
devtools::install_github("herbps10/effectbounds")
```

Connors, Alfred F., J., Speroff, T., Dawson, N. V., Thomas, C., Harrell, Frank E., J., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., Fulkerson, William J., J., Vidaillet, H., Broste, S., Bellamy, P., Lynn, J., and Knaus, W. A. (1996). The effectiveness of right heart catheterization in the initial care of critically iii patients. *JAMA*, 276(11):889–897.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199.

Kahan, B. C., Hindley, J., Edwards, M., Cro, S., and Morris, T. P. (2024). The estimands framework: a primer on the ich e9(r1) addendum. *BMJ*, 384.

Kennedy, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400.

Murphy, D. J. and Cluff, L. E. (1990). The SUPPORT study. *Journal of Clinical Epidemiology*, 43:V–X.