

Generalizing Observational Study Results

Applying Propensity Score Methods to Complex Surveys

National Conference on Health Statistics
August 8, 2012



JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Megan Schuler
Eva DuGoff
Elizabeth Stuart

Objectives

- Provide a tutorial for using propensity score methods with complex survey data
- Present results from a simulation study investigating the performance of various propensity score methods with survey weights
- Original motivation: Effectiveness study of type of primary healthcare provider on healthcare spending in MEPS dataset

Background

- Nationally representative survey data represent important data sources for effectiveness studies
 - Challenge = potential confounding
- Lack of clear guidelines on how to use propensity score methods in this context
 - Wide variability in methods and inferences in current literature

Propensity Score Overview

- Propensity score = probability of receiving the treatment, conditional on covariates

$$p(x_i) = \Pr[T_i=1 | X_i = \mathbf{x}_i]$$

- Conditioning on propensity score will reduce confounding (Rosenbaum & Rubin, 1983)

Propensity Score Methods

- Multiple techniques to condition on propensity score:
 1. Matching: match individuals on propensity score
 2. Subclassification: create classes of individuals with similar propensity scores
 3. Weighting: weight individuals using propensity scores

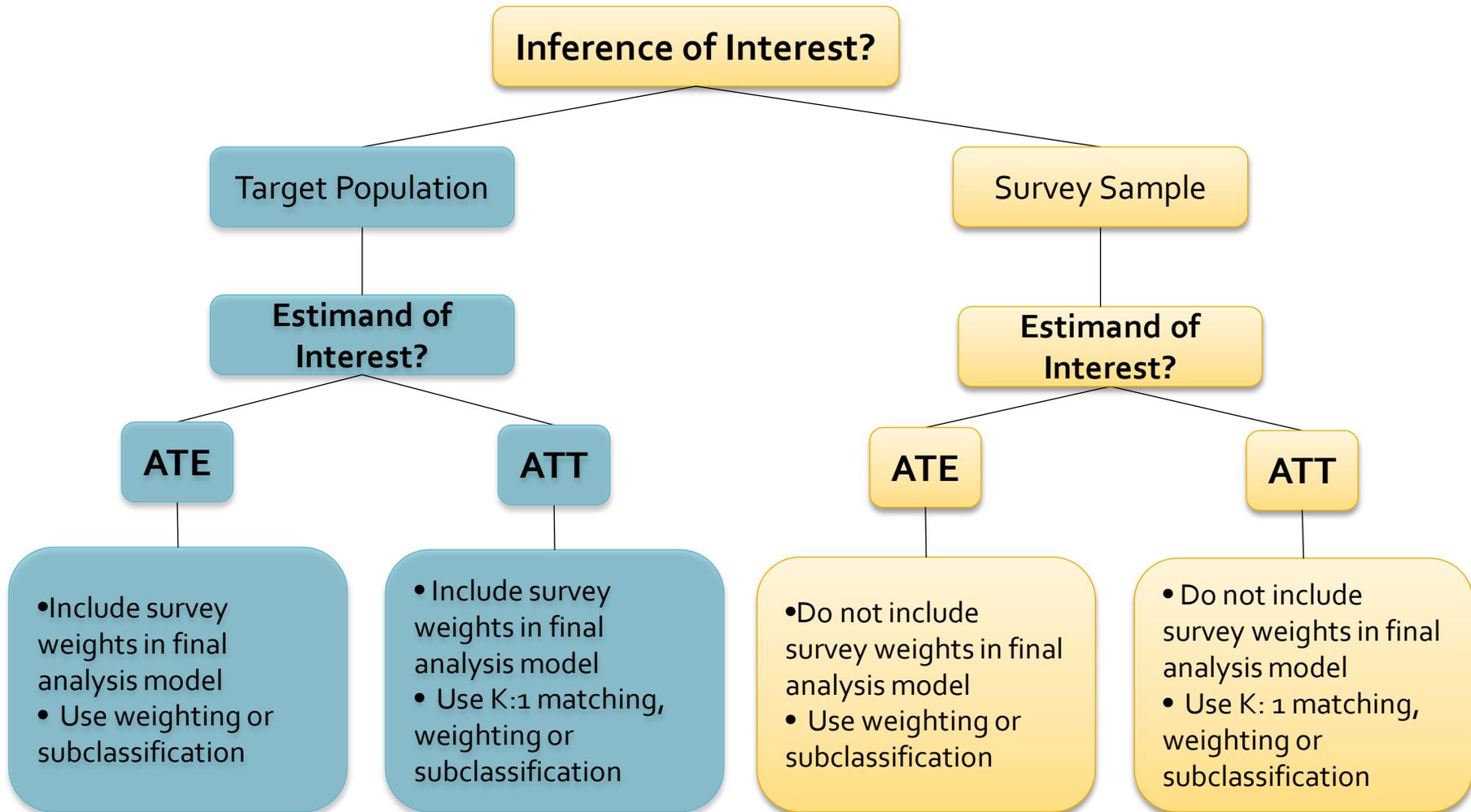
Causal Estimands

- **Average Treatment Effect (ATE)**
- ATE = compares mean outcome if *entire population* had received Treatment to mean outcome if *entire population* had received Control

Causal Estimands

- **Average Treatment Effect (ATE)**
- ATE = compares mean outcome if *entire population* had received Treatment to mean outcome if *entire population* had received Control
- **Average Treatment Effect on Treated (ATT)**
- ATT = compares mean outcomes for *individuals who in reality received Treatment* to the mean outcomes if *these same* individuals had instead received Control

Conceptual Flowchart



Simulation Study Overview

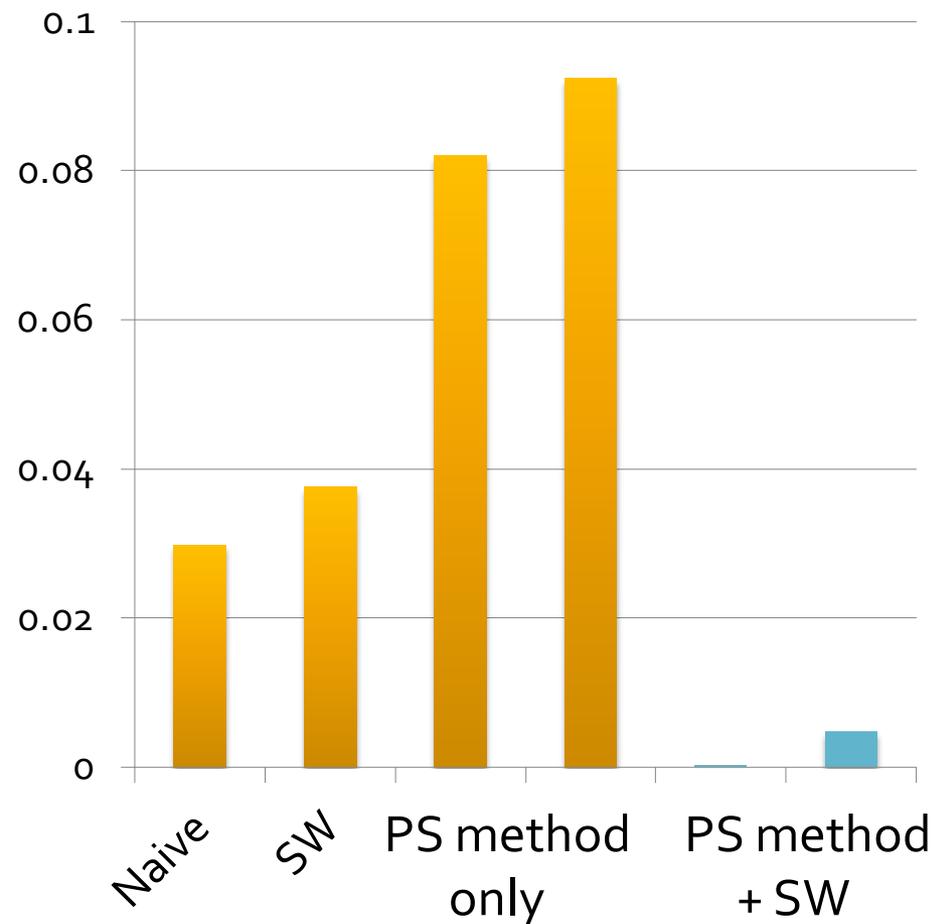
- Performed simulation study to compare propensity score methods when generalizing results to original study population
- Setup
 - 100,000 individuals
 - single covariate
 - survey weight (no clustering)
 - 2,000 simulations

Estimating the ATE

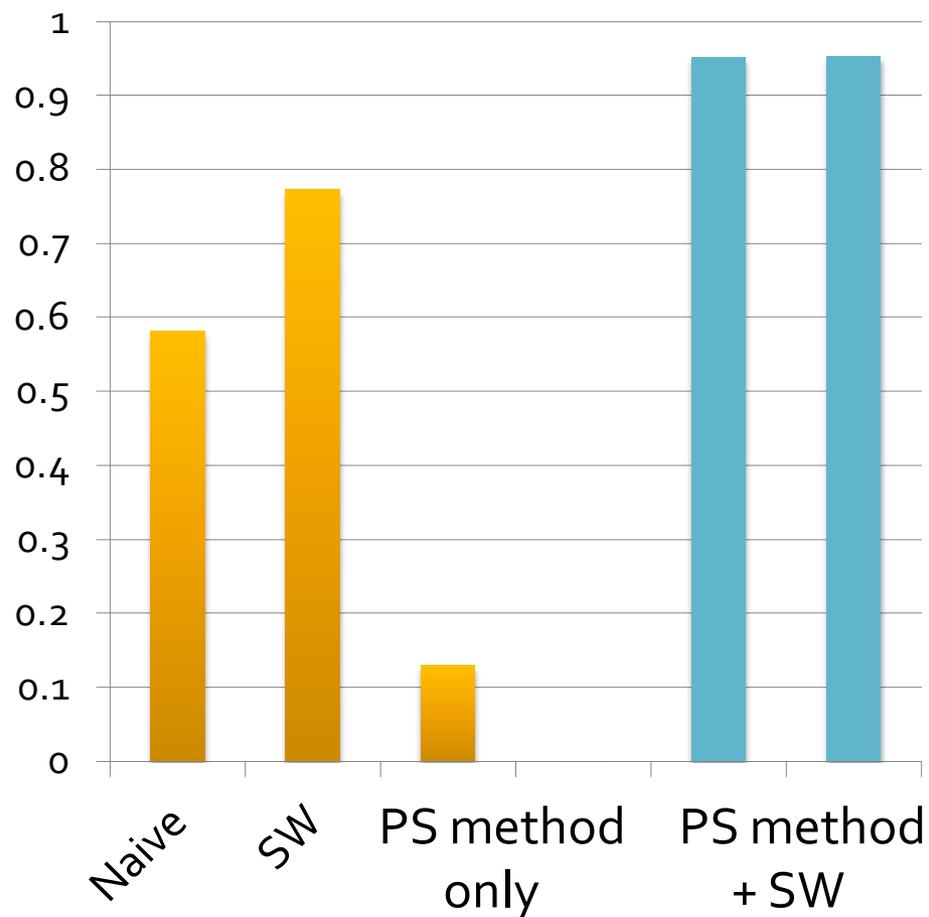
- Reference methods
 - Naive (no propensity scores, no survey weights)
 - Survey weights only
- Appropriate propensity score methods
 - Weighting
 - Subclassification
- Evaluated each approach, with and without survey weights

ATE Results

Absolute Bias



95% CI Coverage

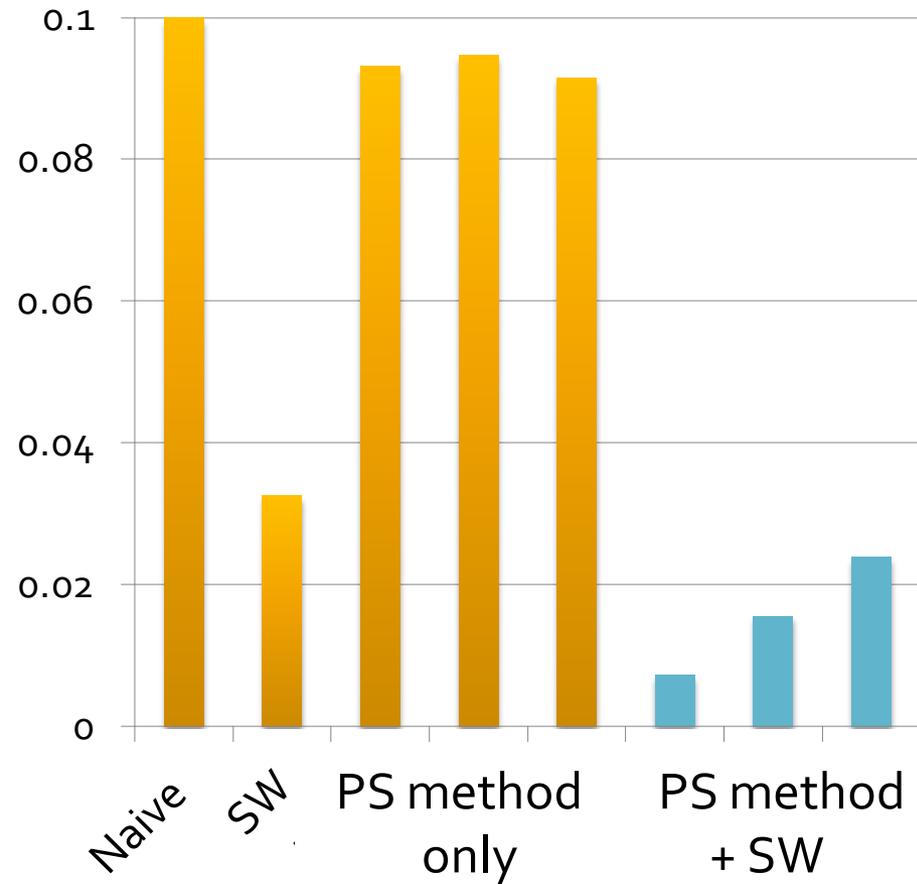


Estimating the ATT

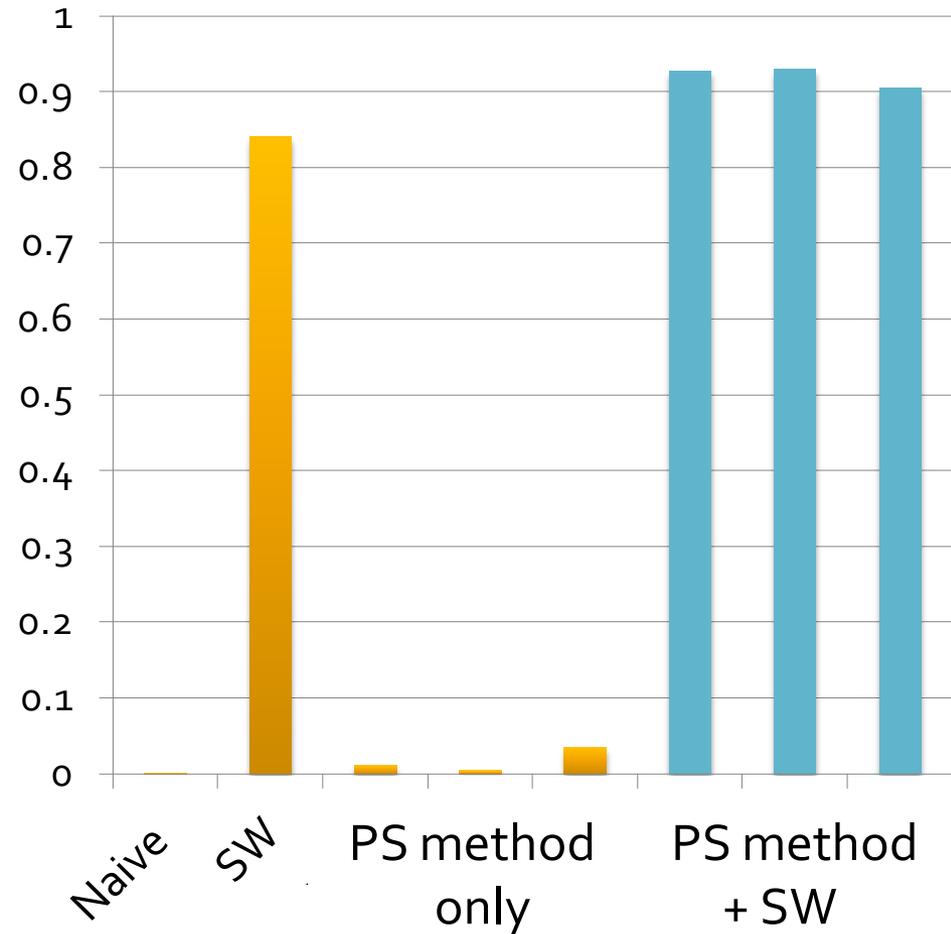
- Reference methods
 - Naive (no propensity scores, no survey weights)
 - Survey weights only
- Appropriate propensity score methods
 - Weighting
 - Subclassification
 - Nearest Neighbor matching (1:1)
- Evaluated each approach, with and without survey weights

ATT Results

Absolute Bias



95% CI Coverage



Simulation Summary

- In general, combination of propensity score method and survey weighting is necessary to achieve unbiased treatment estimates
 - PS Weighting: multiply PS weights by SW
 - Subclassification: use SW to combine across subclasses
 - Nearest Neighbor Matching: SW regression within matched sample
- Propensity score methods perform similarly
 - ATE: weighting, subclassification
 - ATT: weighting, subclassification, nearest neighbor matching

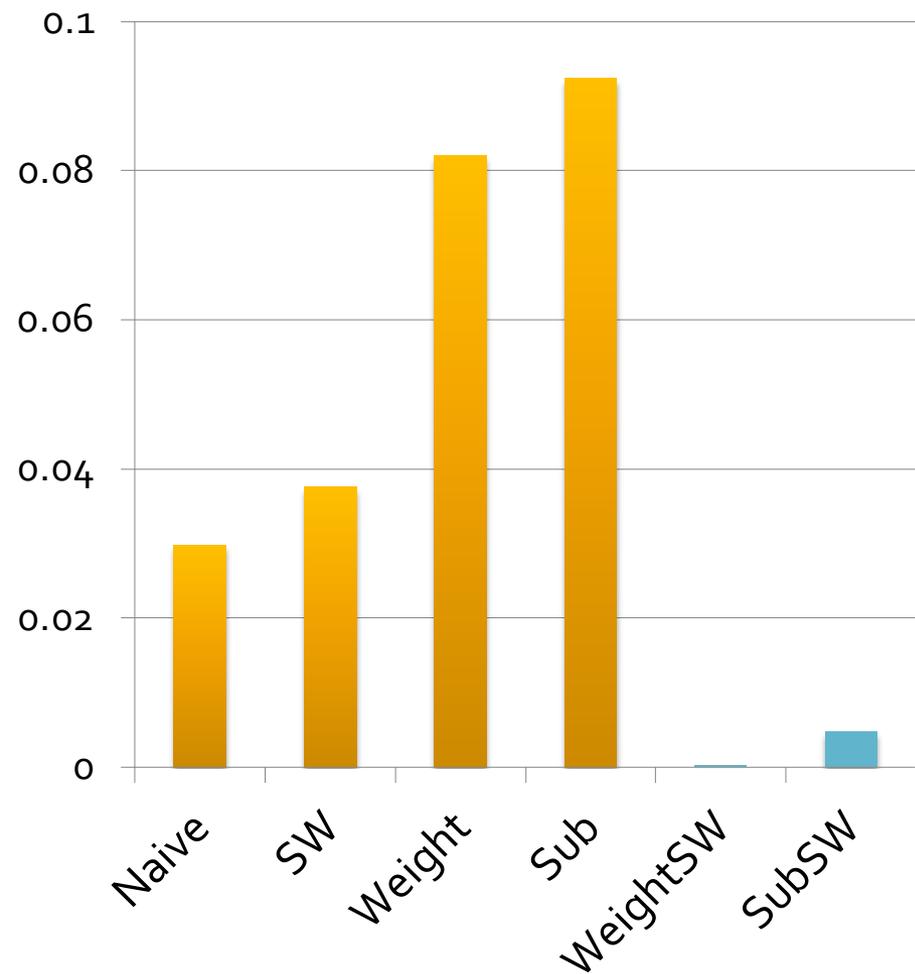
Discussion

- First quantitative investigation of methods for combining propensity score methods and survey weights
- Future work could explore:
 - Further differentiating between performance of various PS methods
 - More complex survey designs
 - Effects of PS model misspecification

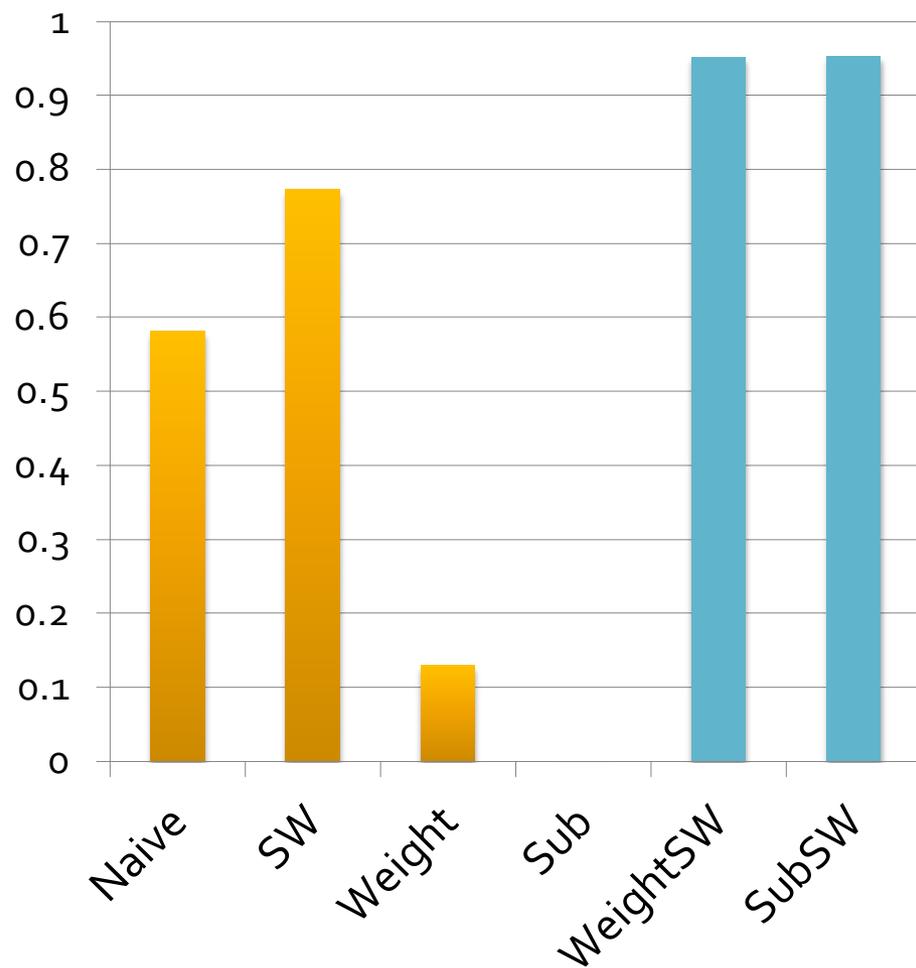
Thanks!

ATE Results

Absolute Bias

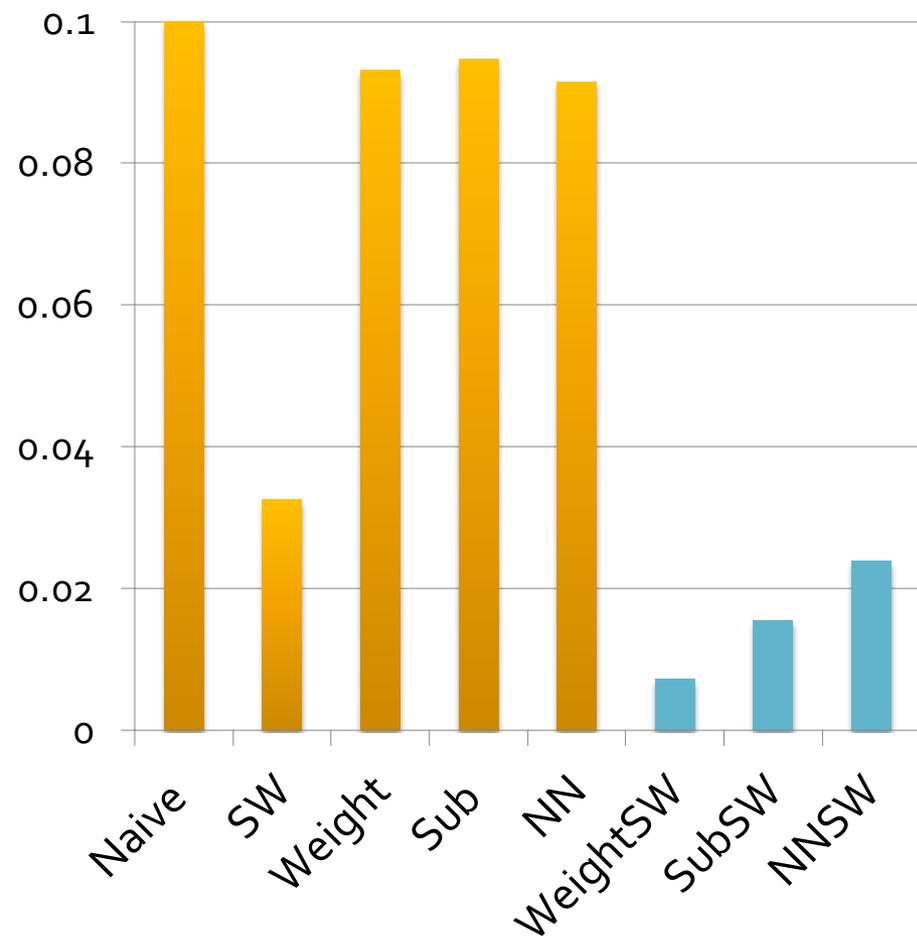


95% CI Coverage



ATT Results

Absolute Bias



95% CI Coverage

