State-of-the-art Strategies for Addressing Selection Bias When Comparing Two or More Treatment Groups

Beth Ann Griffin

Acknowledgements

- This work has been generously supported by NIDA grants 1R01DA034065 and 1R01DA045049

- My colleagues

Lane Burgette
Craig Martin
Joe Pane
Brian Vegetabile
Shuangshuang Liu
Daniel McCaffrey
Matt Cefalu
Donna Coffman
Ricky Sanchez
Marika Suttorp Booth

Goals

- To increase your understanding of
  - How to define and estimate causal effects
  - How to use propensity scores when estimating causal effects
  - How to assess the validity of key assumptions
- To provide you with step-by-step instructions for using propensity score weights when you have:
  - 2 groups (binary treatment)
  - More than 2 groups (multiple treatments)
  - Time-varying treatments

BACKGROUND ON CAUSAL EFFECTS ESTIMATION

Why causal effects?

- Causal effects are of great interest in many fields
  - Interested in **causal effects**, not merely **correlations**
- Causal effects describe **how individuals will change their behaviors** (substance use, mental health, criminal involvement) in response to
  - A new treatment
  - A prevention program
  - A new policy
  - A new exposure

Motivating example

- **Case study objective:** To estimate the causal effect of MET/CBT5 versus "usual care" among adolescent clients
  - Data from 2 SAMHSA CSAT discretionary grants

MET/CBT5	"Usual Care"
• Longitudinal, observational data	• Longitudinal, observational data
• MET/CBT5 at 37 sites from EAT	• "Usual care" at 4 sites from ATM

Motivating example

- **Case study objective:** To estimate the causal effect of MET/CBT5 versus "usual care" among adolescent clients
  - Data from 2 SAMSHA CSAT discretionary grants

MET/CBT5

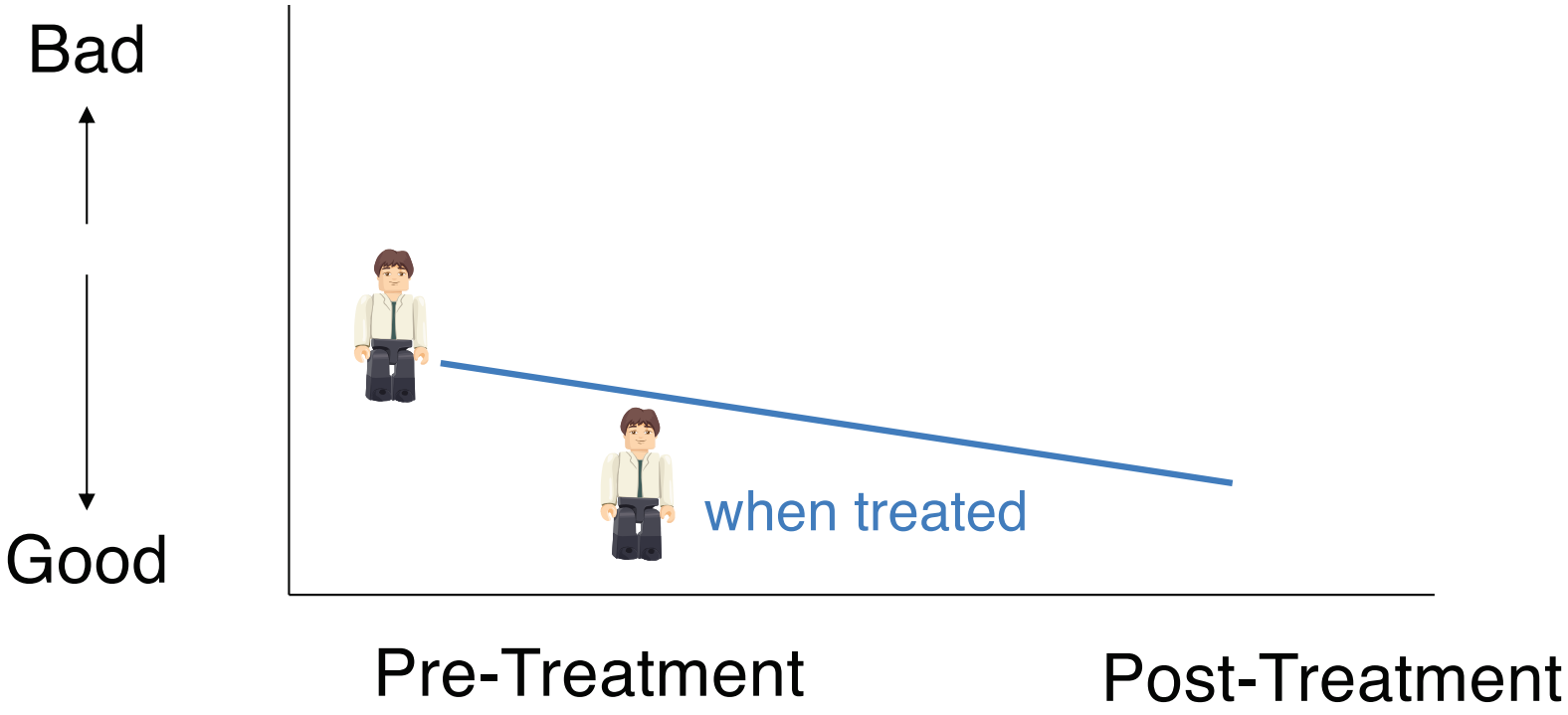
"Usual Care"

Various situations can benefit from use of causal inference methods

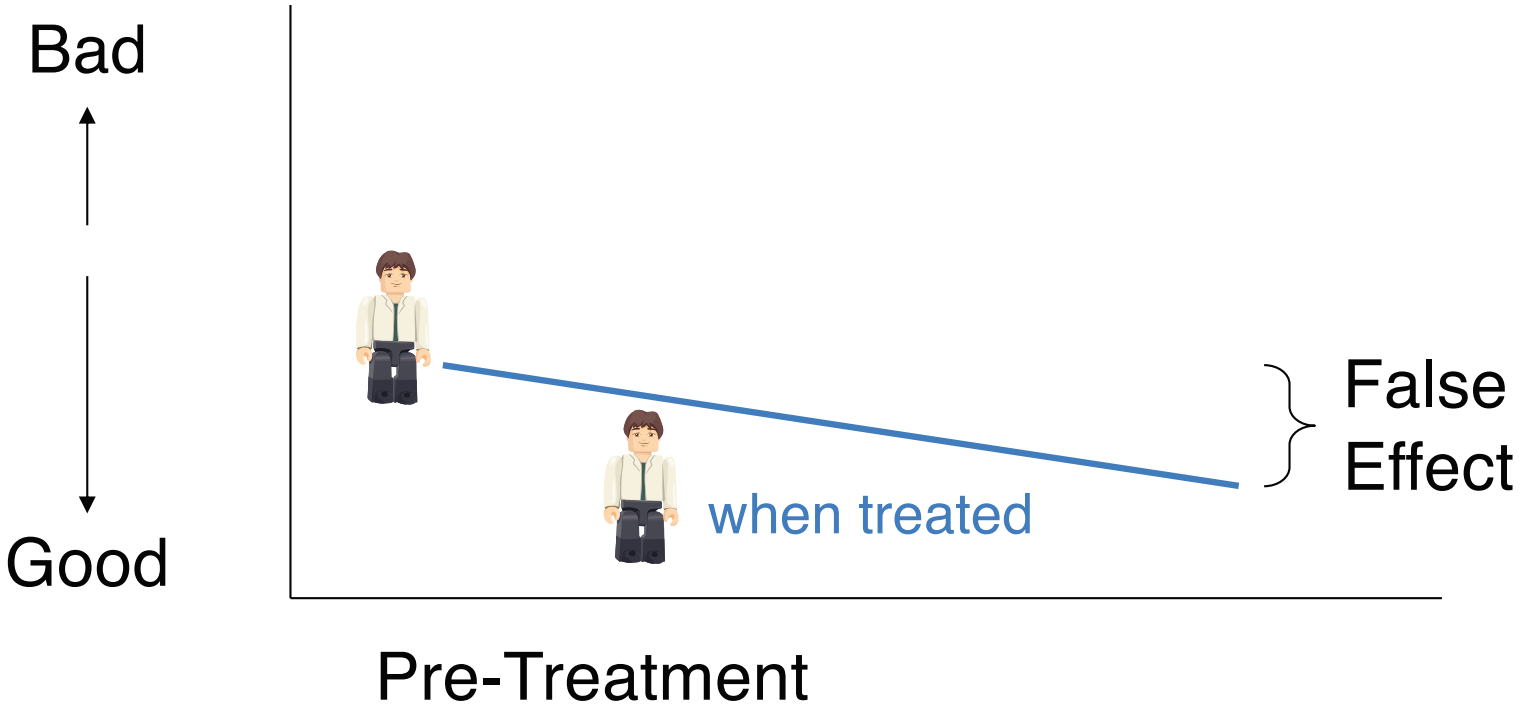
- Longitudinal observational studies (like our case study)
- Administrative databases with outcomes (i.e., in a health plan, who gets what treatment and what is their subsequent service use)
- Aggregating across studies to maximize power and look at moderation or mediation
  - E.g., different RCTs done over time that compared A to B then A+C to B and you want to compare A to A+C

How do we get a causal effect estimate?

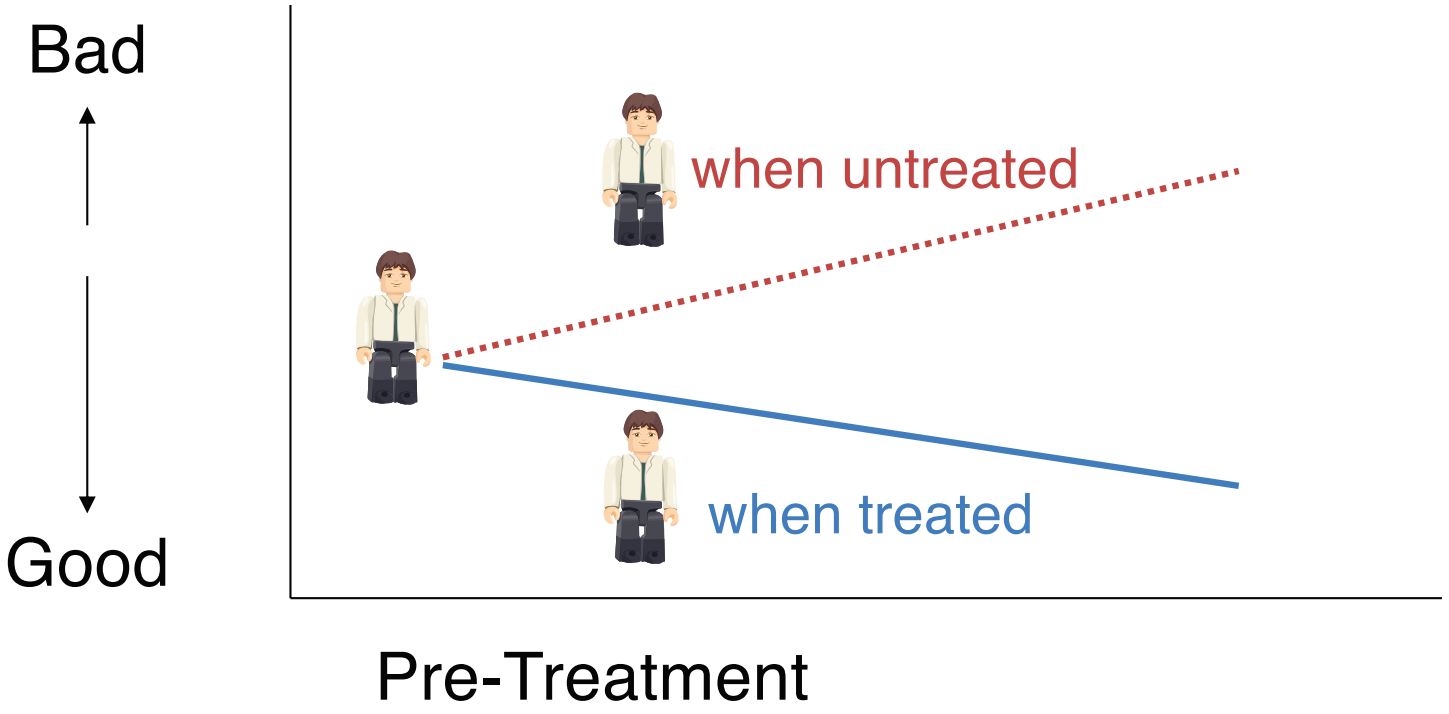
How do we get a causal effect estimate?



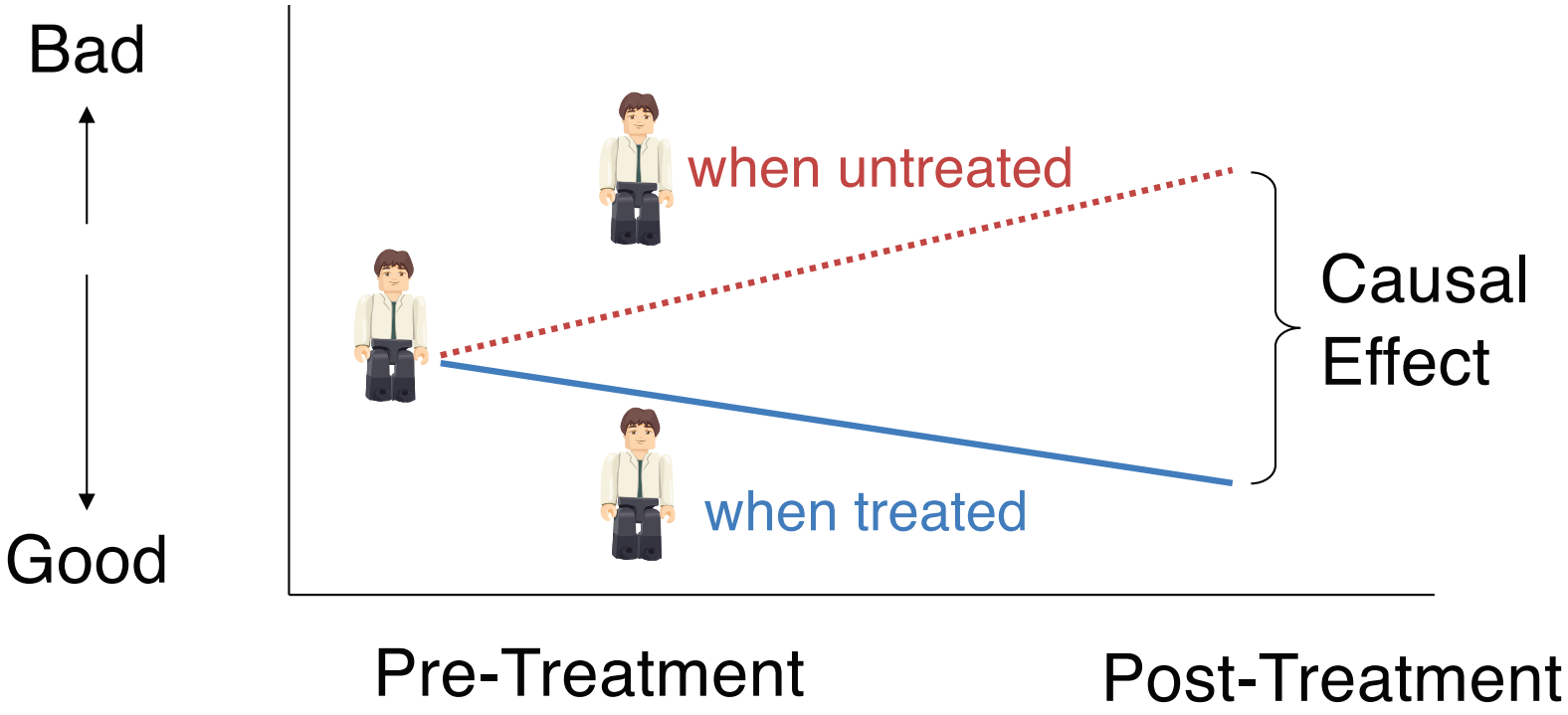
How do we get a causal effect estimate?



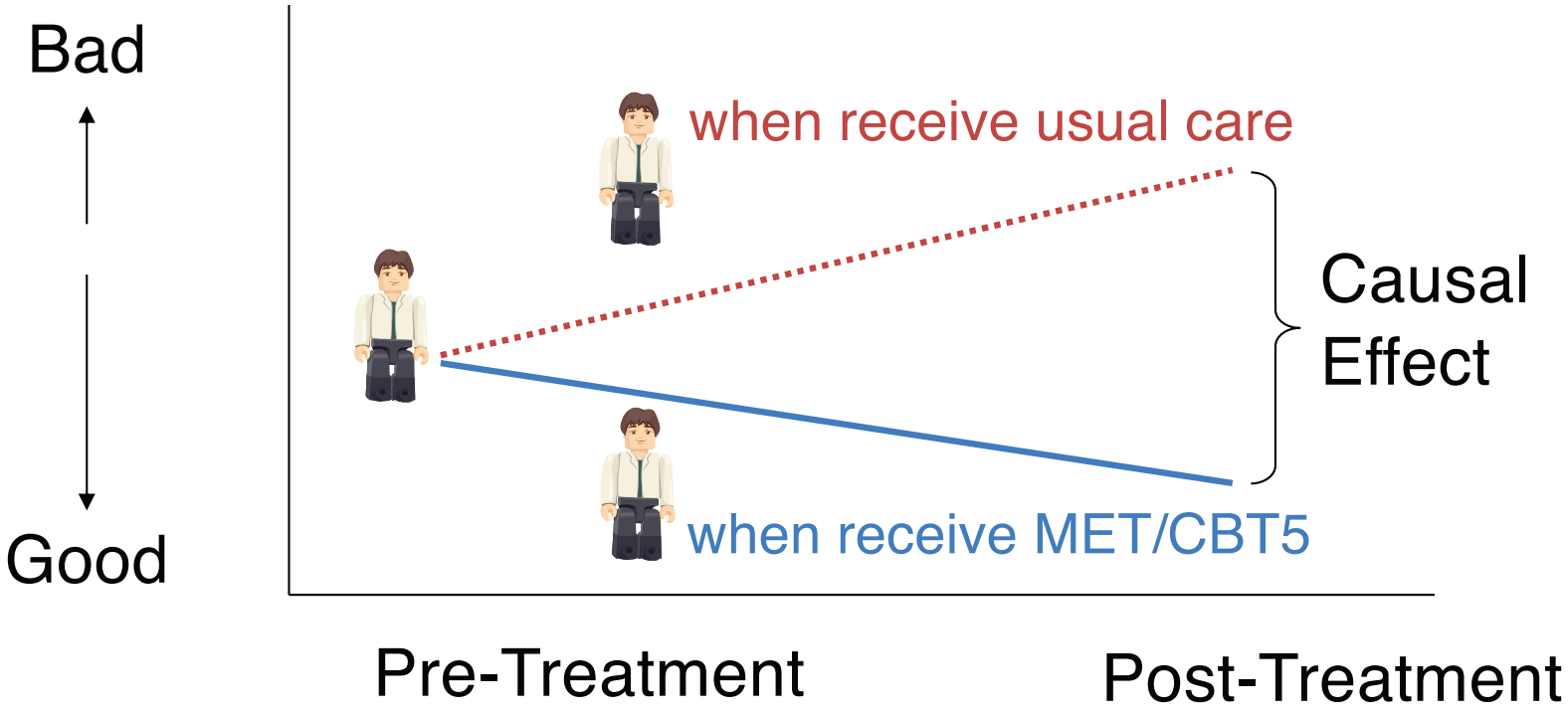
How do we get a causal effect estimate?



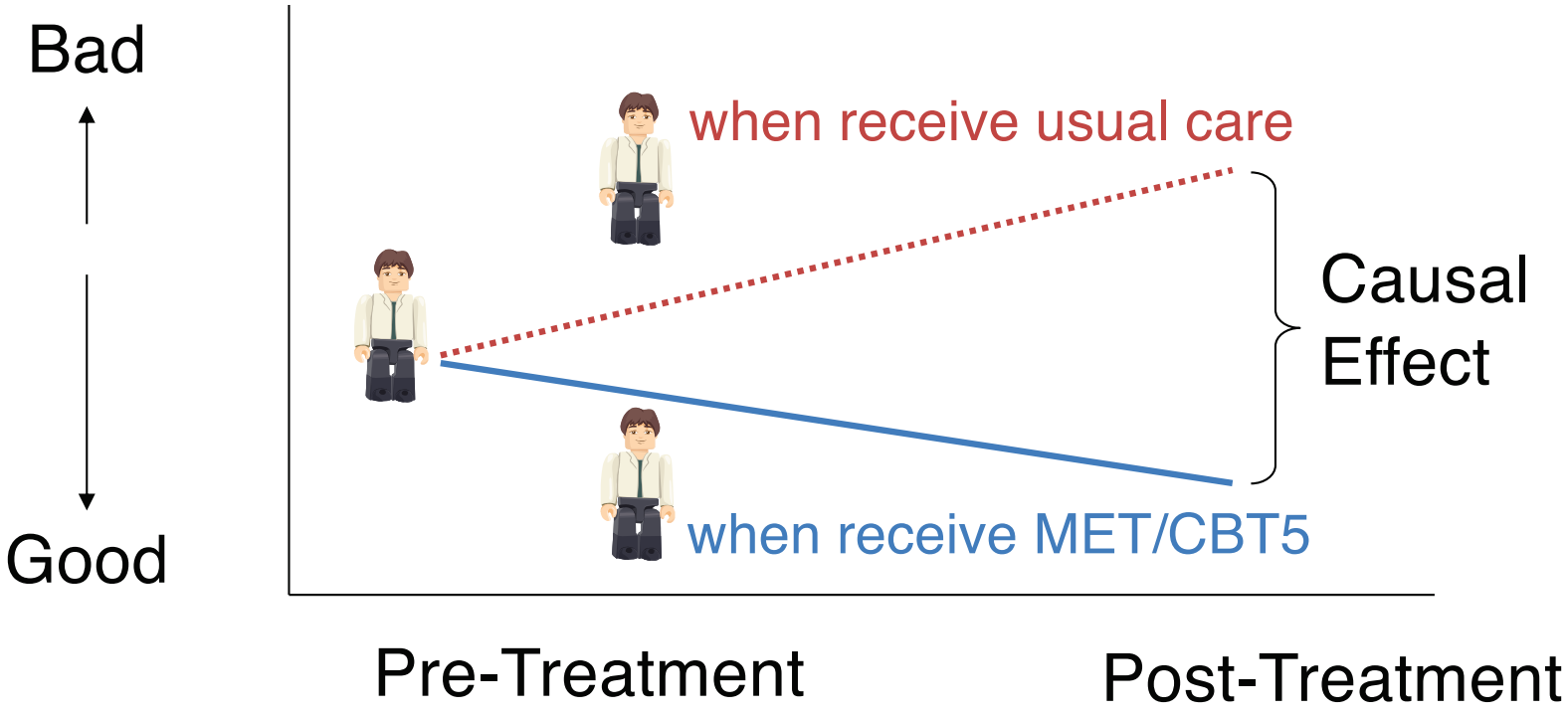
Causal effect = difference between two potential outcomes



Causal effect of MET/CBT5 vs. Usual Care



Causal effect of MET/CBT5 vs. Usual Care

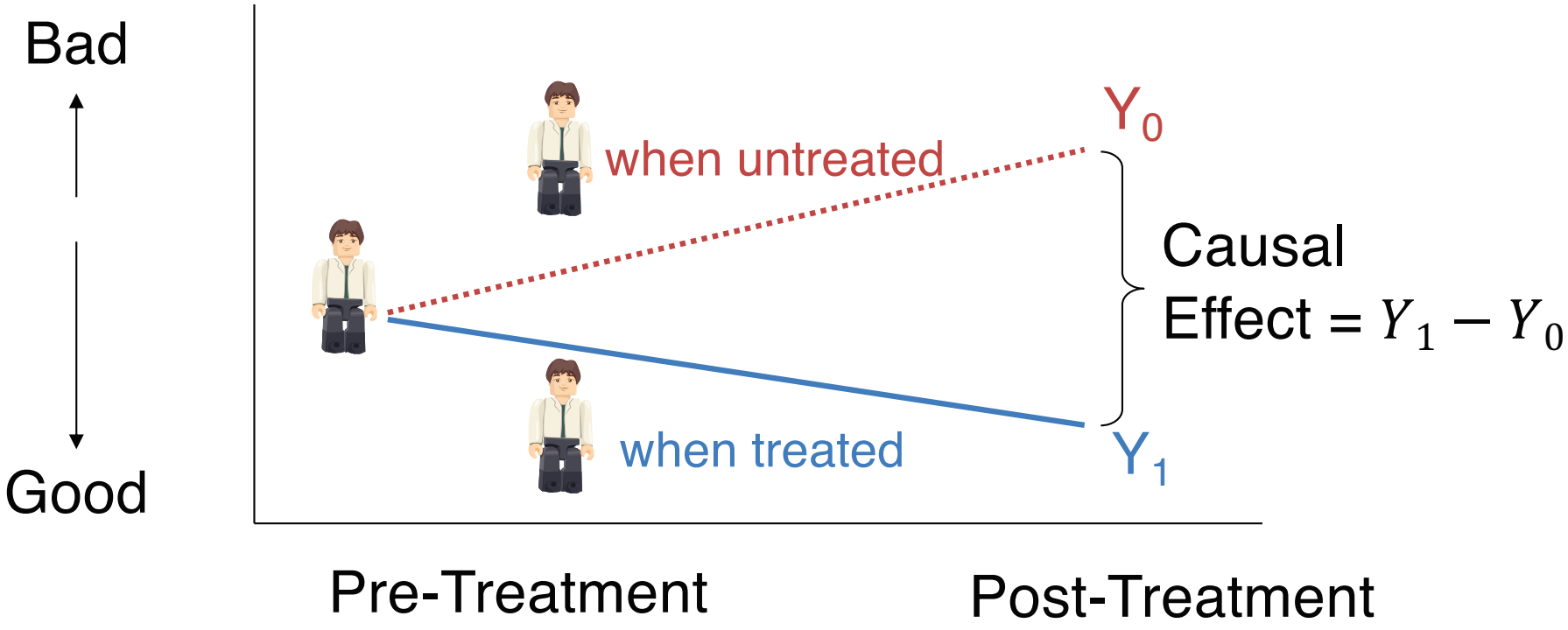


Fundamental problem for causal inference: We only observe ONE of these potential outcomes/counterfactuals

Potential outcomes framework

- Two potential outcomes for each study participant:
  - Outcome after receiving treatment = Y_1
  - Outcome after receiving comparison condition = Y_0 (can be either another treatment or no treatment control)
- Y_1 and Y_0 exist for all individuals in the population regardless of whether the individual received treatment ($Z = 1$) or comparison condition ($Z = 0$)
- Observe only one of these outcomes for each participant

Causal effect = difference between two potential outcomes



Potential outcomes and assigned treatment

- Since Y_1 and Y_0 exist for all individuals in the population, 3 sets of distributions are important:

1. Distributions for the entire populations (denote the densities by f(Y_1), f(Y_0), and f(Y_0,Y_1))

2. Distributions for the subsample of where Z is observed to be = 1 (denote the densities by f(Y_1|Z =1), f(Y_0|Z =1), and f(Y_0,Y_1 |Z =1))

3. Distributions for the subsample of where Z is observed to be = 0 (denote the densities by f(Y_1|Z =0), f(Y_0|Z =0), and f(Y_0,Y_1 |Z =0))

Potential outcomes, assigned treatment, and selection

- Selection occurs when individuals with Z = 1 look different from individuals with Z = 0
- Formally, we define this as

$$f(Y_0, Y_1 \mid Z = 1) \neq f(Y_0, Y_1 \mid Z = 0)$$

The treatment effect

- $Y_1 - Y_0$ is the treatment effect for the individual person
  - Every individual has a potential treatment effect
  - Treatments can (and do typically) vary across individuals
- So average treatment effects are typically estimated

Primary types of causal effects

- Average treatment effect in the population (ATE)
  - Answers the question:
    - How effective is the treatment in the population? (if you have a *control condition*)
    - What is the relative effectiveness of two treatments on average in the population? (if you have *2 treatments*)
  - $E(Y_1 - Y_0)$
- Average treatment effect in the treated population (ATT)
  - Answers the question:
    - How would those who received treatment have done had they received the comparison condition?
  - $E(Y_1 - Y_0|Z = 1)$

Primary types of causal effects: case study

- Average treatment effect in the population (ATE)
  - Answers the question:
    - What is the relative effectiveness of MET/CBT5 versus usual care on average in the population?
  - $E(Y_1 - Y_0)$
- Average treatment effect in the treated population (ATT)
  - Answers the question:
    - How would those who had received MET/CBT5 have done had they received usual care?
  - $E(Y_1 - Y_0|Z = 1)$

Primary types of causal effects: case study

- Average treatment effect in the control population (ATC?)
  - Answers the question:
    - How would those who received control have done had they received the treatment condition?
    - How would those who had received usual care have done had they received MET/CBT5?
  - $E(Y_1 - Y_0|Z = 0)$

Population vs Sample Average Treatment Effects

- Restrict inferences to observed sample
  - Sample Average Treatment Effect (SATE or SATT)
  - Average treatment effect for member of the sample only
  - Sample of Y_0's and Y_1's and covariates are treated are treated as fixed
  - Random error is due to the chance assignment of the treatment indicators
- Infer to a larger population from which the sample arose
  - Population Average Treatment Effect (PATE or PATT)
  - Extrapolates beyond the observed sample
  - Sample of Y_0's and Y_1's and covariates are treated are treated as random
  - Treatment assignment is also random

Challenge of Causal Effect Estimation

- **The treatment effect is the average of the individual differences in potential outcomes across the entire population or sample**
  - **ATE=**$E(Y_1 - Y_0)$
- **Simple calculations show that ATE equals the average of all individuals' potential outcomes under treatment minus the average of all individuals' potential outcomes under control**
  - $E(Y_1 - Y_0)$ **=** $E(Y_1)$ **-** $E(Y_0)$
- **These averages include the potential outcomes under treatment for individuals who receive treatment (**$Z = 1$**) and those who do not (**$Z = 0$**)**
  - $E(Y_1) = E(Y_1|Z = 1)\pi + E(Y_1|Z = 0)(1 - \pi)$**,** $\pi = Pr(Z = 1)$
  - $E(Y_0) = E(Y_0|Z = 1)\pi + E(Y_0|Z = 0)(1 - \pi)$
- **Estimating the treatment effect requires estimates of** $E(Y_1|Z = 1)$**,** $E(Y_1|Z = 0)$**,** $E(Y_0|Z = 1)$**, and** $E(Y_0|Z = 0)$

Challenge of Causal Effect Estimation

- **We observe potential outcomes under treatment only for individuals who receive treatment and we observe potential outcomes under control only for individuals who do not receive treatment**
  - **Y_1 when $Z = 1$ and Y_0 when $Z = 0$**
  - **We do not observe Y_0 when $Z = 1$ or Y_1 when $Z = 0$**
- **We can directly estimate $E(Y_1|Z = 1)$ and $E(Y_0|Z = 0)$**
- **Need to extrapolate to estimate $E(Y_1|Z = 0)$ and $E(Y_0|Z = 1)$**
- **But if pretreatment differences exist then the average potential outcome under treatment for individuals who receive the treatment will not equal the average potential outcome under treatment for individuals who do not receive treatment**
  - **$E(Y_1|Z = 1) \neq E(Y_1|Z = 0)$ and $E(Y_0|Z = 1) \neq E(Y_0|Z = 0)$**
  - **Similar results hold for potential outcomes under control**

Solving the Fundamental Problem of Causal Inference

- **Key challenge: Estimating the means for the unobserved counterfactuals, $E(Y_1|Z=0)$ and $E(Y_0|Z=1)$**
- **Solution: Find or create groups where the distributions or the expected values of Y_1 are the same for individuals who receive treatment ($Z=1$) and individuals who receive control ($Z=0$) and the distributions or expected values of Y_0 are the same for individuals who receive treatment and individuals who receive control**
- **Within these groups potential outcomes are unrelated to treatment assignment**
  - **This is known as *strong ignorability***

Strong Ignorability

- **Condition 1: The distribution of potential outcomes, (Y_0, Y_1), is the same for individuals with $Z = 1$ and $Z = 0$ and at each value of X** $(Y_0, Y_1) \perp\!\!\!\perp Z \mid X$
- **Condition 2: $0 < p(X) < 1$ for all X**

Methods to Achieve Strong Ignorability

- Random experiments
- Define groups on the basis of observed covariates and *assume* strong ignorability holds within these groups
  - Modeling assignment or selection (propensity scores)
- Regression discontinuity
- Instrumental variables (IV)

Random Experiments

- Collect group of participants and randomly assign to treatment or comparison
  - Researchers control assignment
- By design and control, assignment, Z, unrelated to any characteristics of the participant
- Provides a random sample of Y_1s and a random sample of Y_0s
- Because of randomization
  - $E(Y_1|Z=1)$ = $E(Y_1|Z=0)$ and $E(Y_0|Z=1)$ = $E(Y_0|Z=0)$
  - (Y_0, Y_1) is independent of treatment assignment, Z
- Any realized assignment might result in groups with differences *but* on average the procedure will provide an unbiased estimate of $E(Y_0)$ and $E(Y_1)$
  - The behavior of differences can be characterized through probability distributions

Experiments vs Observational Studies

- Random experiments = gold standard for estimating causal effects
  - Randomization (if it works) makes groups being compared balanced on baseline characteristics
  - Treatment assignment is unrelated to potential outcomes (**strong ignorability**)
- Randomization is not always feasible
- Observational studies provide another way to get at causal effects
  - Treatment assignment is not controlled by the researcher
  - Groups being compared are usually imbalanced
  - Can use causal inference methods to try to replicate what a randomized study does

Traditional Methods for Observational Studies

- Change scores (post score less pre score in pre-post designs)
  - Assumes that potential change is unrelated to treatment for all individuals
- Regression modeling
  - Assumes strong ignorability conditional on the values of the covariates in the model
  - Also assumes form for the expected value of the outcomes as a function of covariates

Biggest challenge to causal estimation is selection effects

- Selection occurs when the people getting the treatments being compared differ

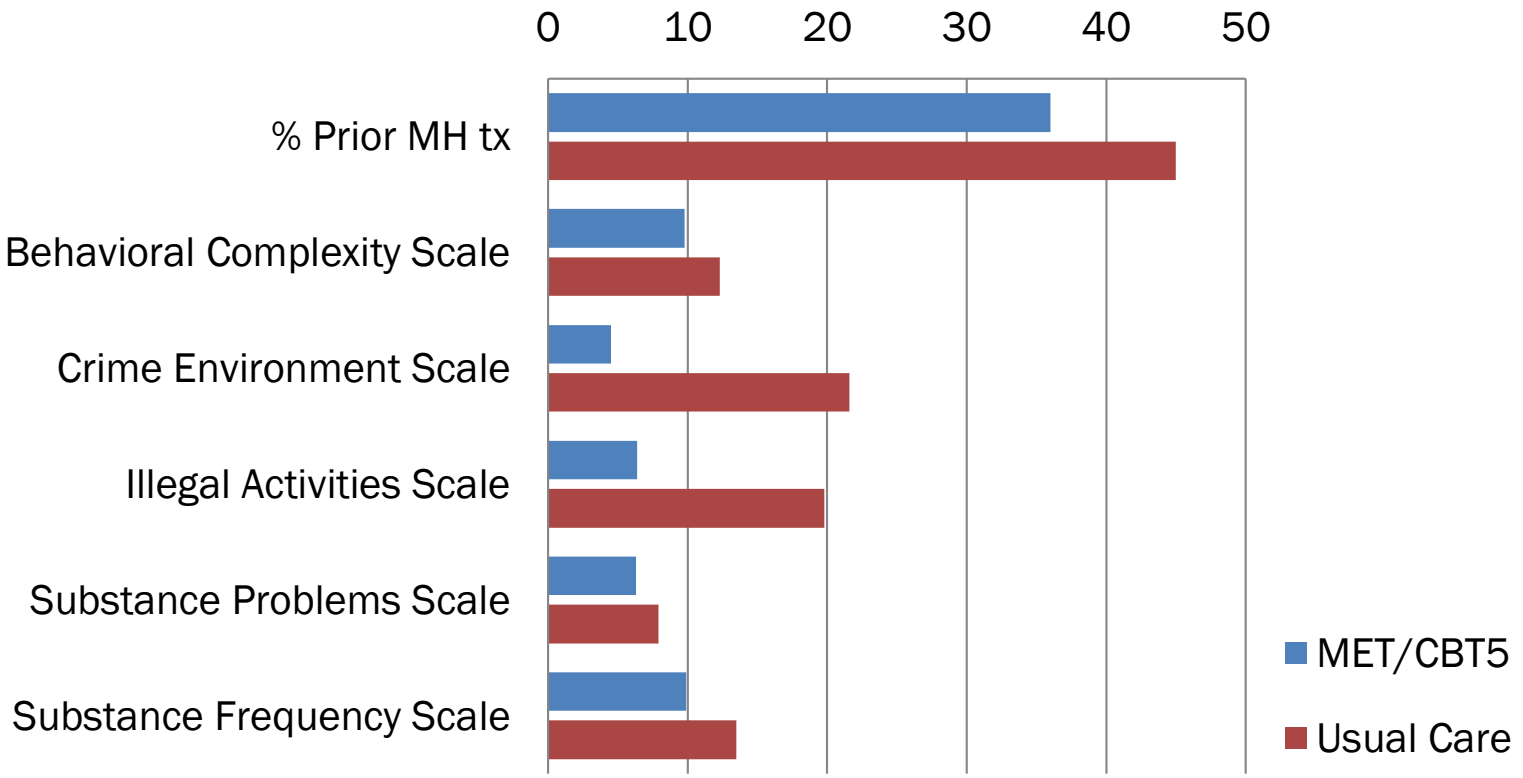


Illustration: The potential impact of selection on our case study

- Mean substance frequency scale (SFS) at 12-months =

  0.11 (~10 days of use per month) for usual care

  0.07 (~ 6 days of use per month) for MET/CBT5

  → Effect size (ES) difference of 0.41

Illustration: The potential impact of selection on our case study

- Mean substance frequency scale (SFS) at 12-months =

  0.11 (~10 days of use per month) for usual care

  0.07 (~ 6 days of use per month) for MET/CBT5

  → Effect size (ES) difference of 0.41

- I am going to report that usual care results in 0.41 standard deviations greater SFS than MET/CBT5
  - Is this a good idea?

Illustration: The potential impact of selection on our case study

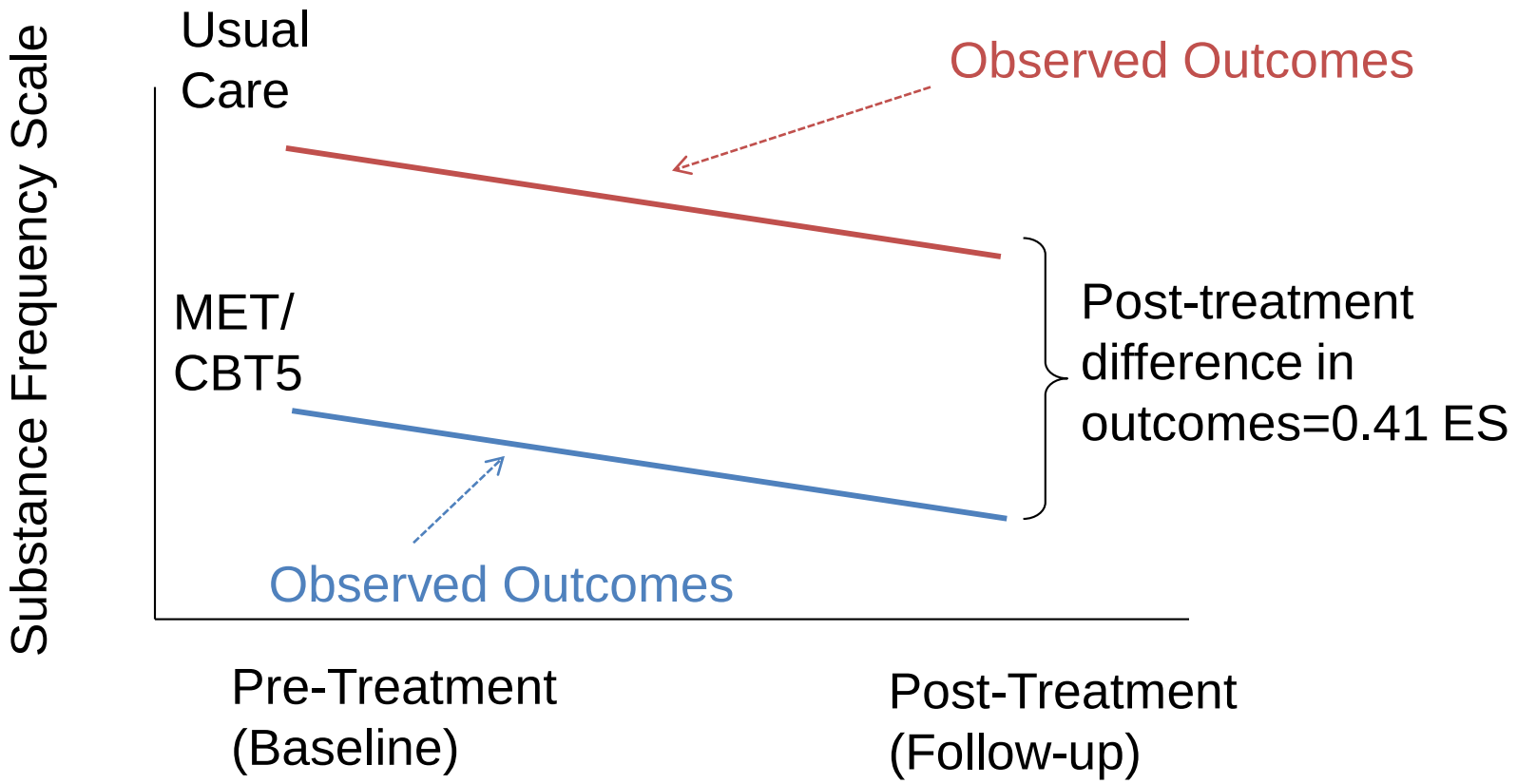
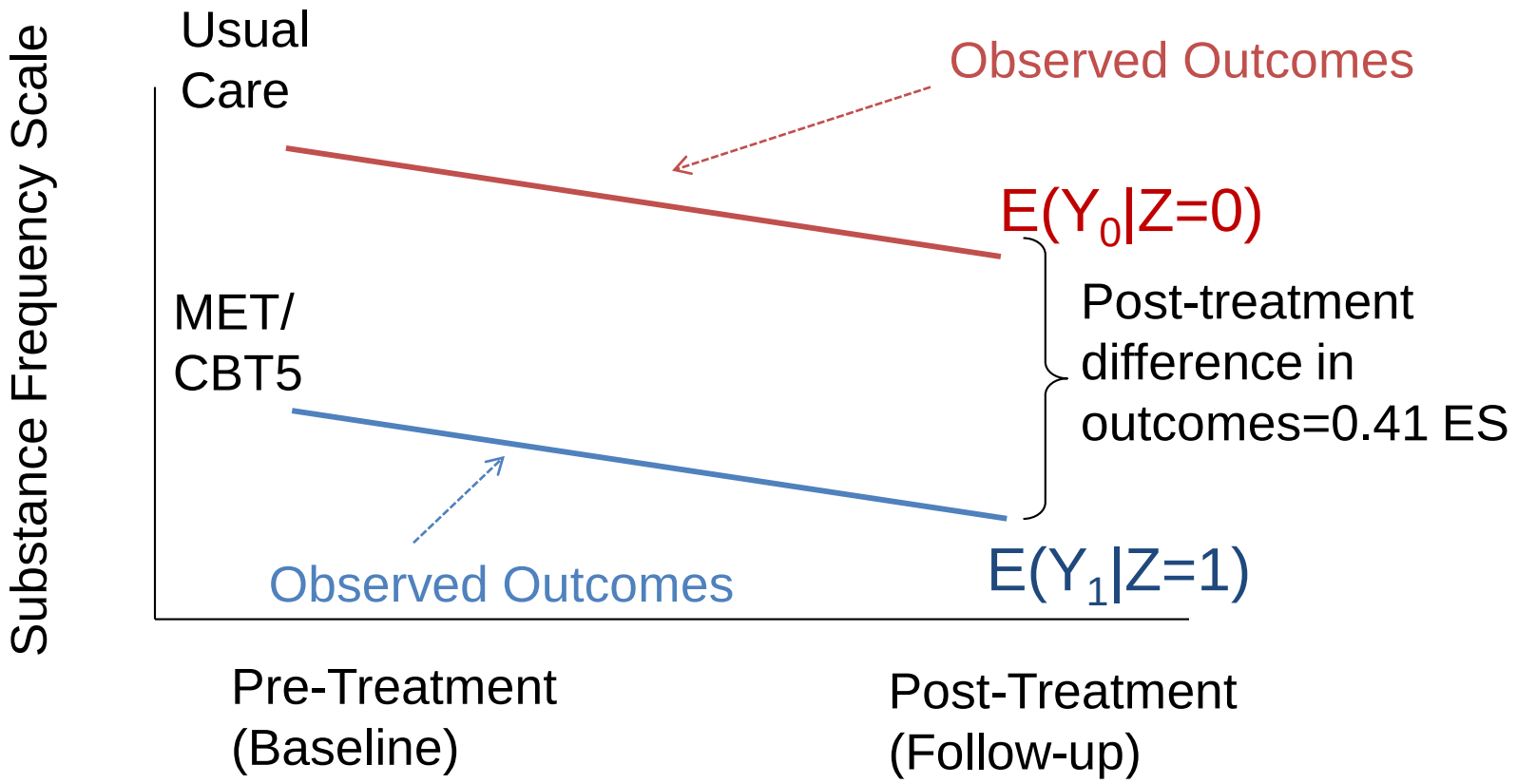


Illustration: The potential impact of selection on our case study



Clearly a case where $E(Y_1|Z=0) \neq E(Y_1|Z=1)$ and $E(Y_0|Z=1) \neq E(Y_0|Z=0)$

Illustration: The potential impact of selection on our case study

- Mean substance frequency scale (SFS) at 12-months =

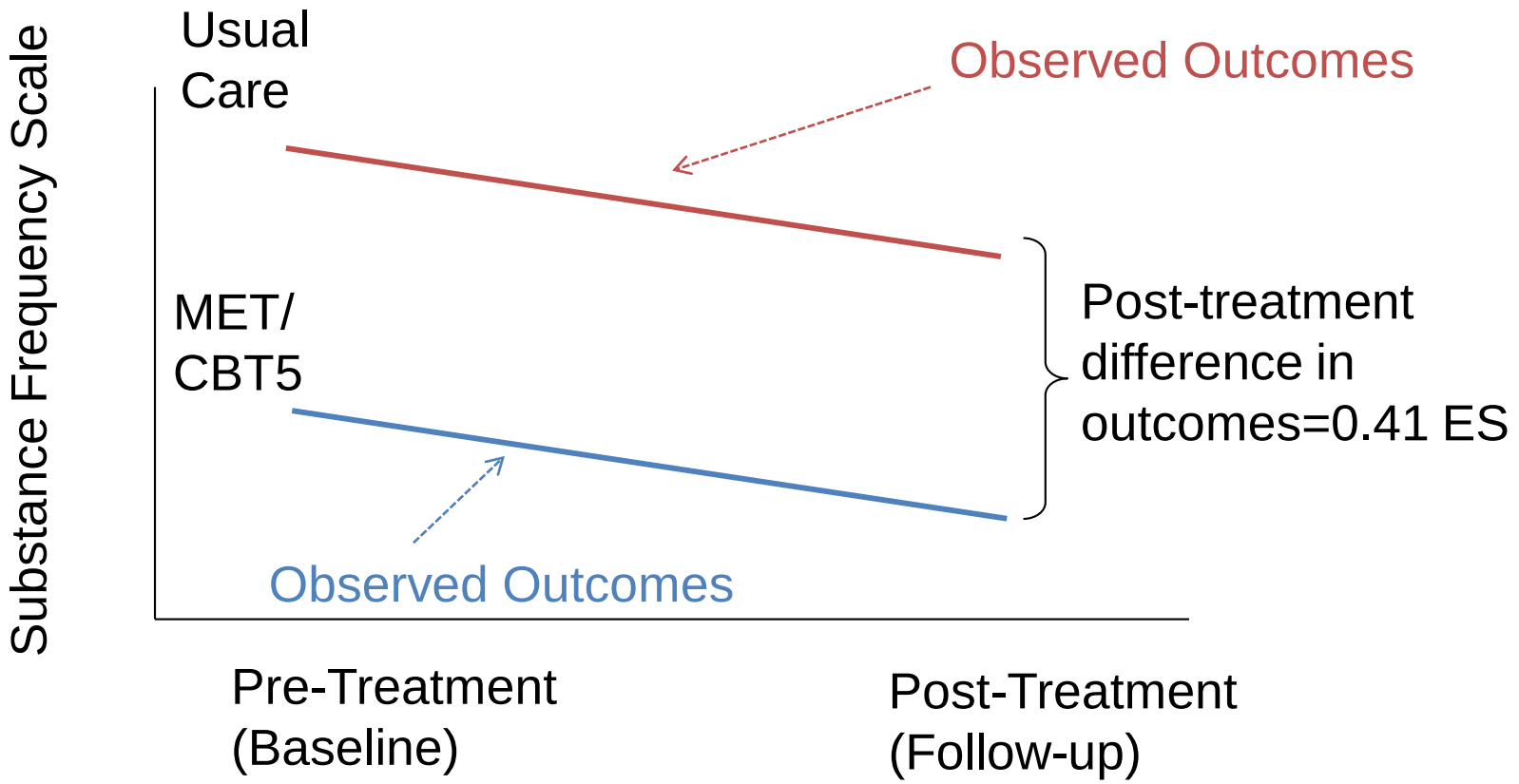
  0.11 (~10 days of use per month) for usual care

  0.07 (~ 6 days of use per month) for MET/CBT5

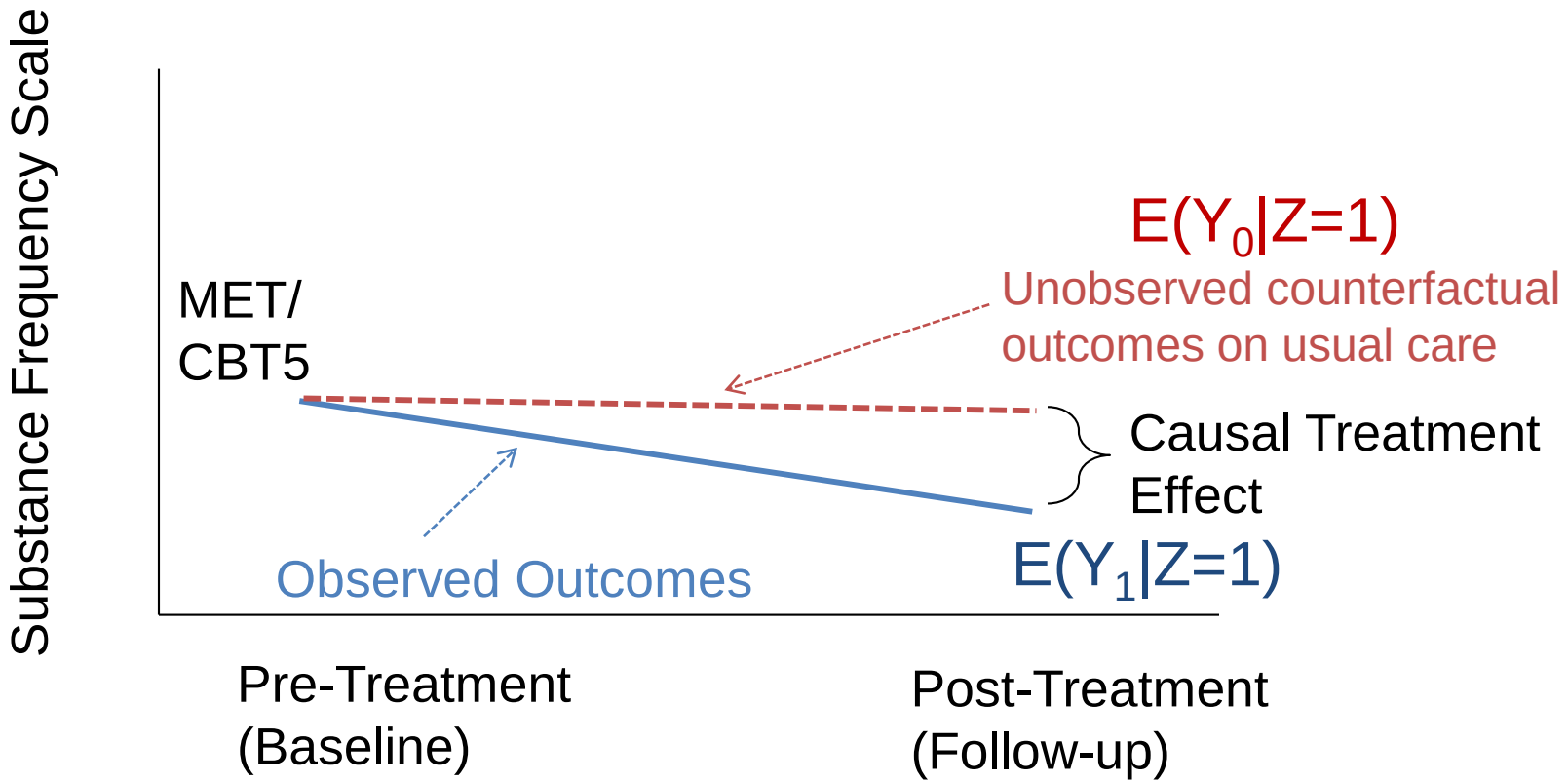
  → Effect size (ES) difference of 0.41

- I am going to report that usual care results in 0.41 standard deviations greater SFS than MET/CBT5
  - Is this a good idea?
  - No!!! Groups might differ on pretreatment characteristics that influence SFS at 12-months such that even in the absence of the treatment the groups might differ on outcomes at 12-months

Different kids, different treatments



WANT: Same kids, different treatments



Propensity scores help us deal with the challenge of selection

- The propensity score is an individual's probability of receiving treatment given pretreatment characteristics

$$p(X) = \Pr(Z = 1|X)$$

  - Used to create **balance** between treatment and comparison conditions
  - Balancing on $p(X)$ can balance distribution of X between treatment and comparison groups

Propensity scores help us deal with the challenge of selection

- The propensity score is an individual's probability of receiving treatment given pretreatment characteristics

$$p(X) = \Pr(Z = 1|X)$$

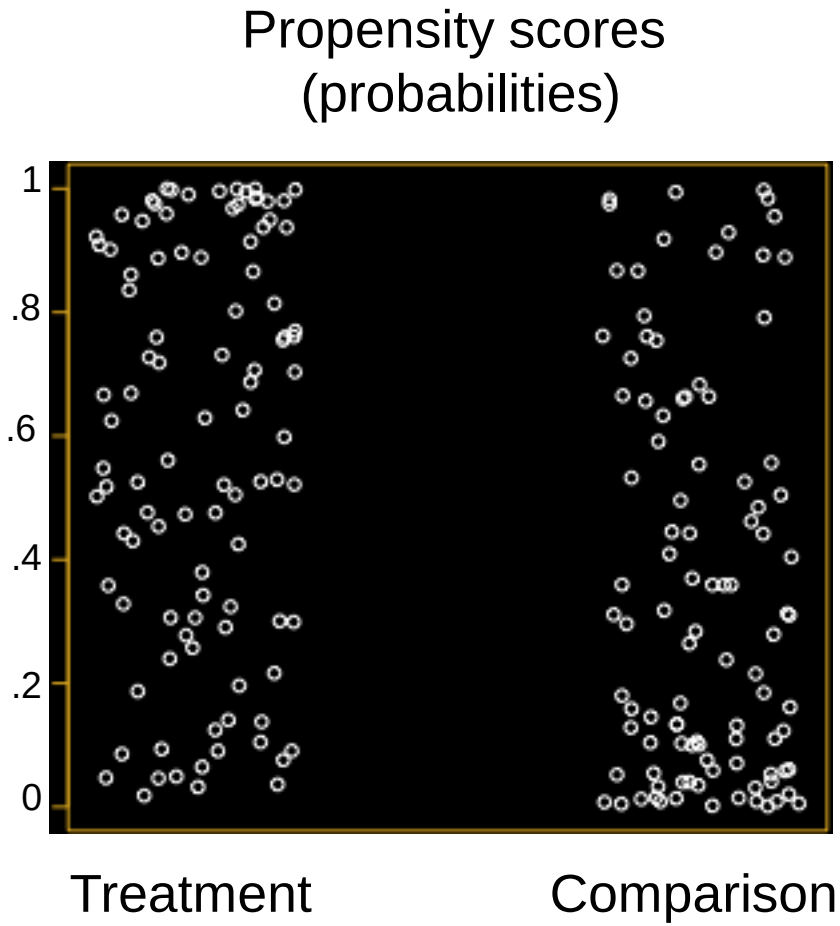
  - Used to create **balance** between treatment and comparison conditions
  - Balancing on $p(X)$ can balance distribution of X between treatment and comparison groups

- If **strong ignorability** holds, propensity score is all that is required to control for observed, pretreatment differences between groups
  - Treatment assignment is **strongly ignorable** if
    (1) $(Y_1, Y_0) \perp Z|X$ – e.g, no unobserved confounders
    (2) $0 < p(X) < 1$ – e.g, overlap between groups

The Propensity Score and Strong Ignorability

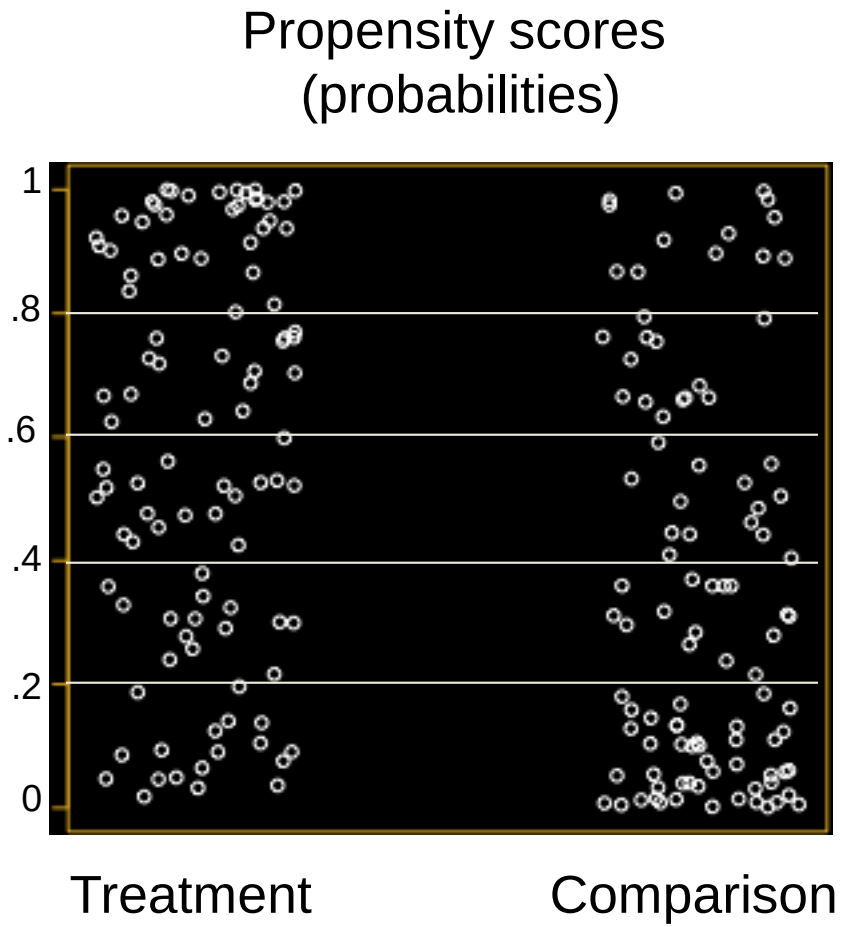
- **If distribution of (Y_0, Y_1) is the same for individuals with $Z = 1$ and $Z = 0$ at each value of X, then then it is the same for individuals with $Z = 1$ and $Z = 0$ at each value of $p(X)$**
  - **If $(Y_0, Y_1) \perp Z|X$ then $(Y_0, Y_1) \perp Z|p(X)$**
- **If treatment is strongly ignorable with respect to X (i.e., the distribution of (Y_0, Y_1) is the same for individuals with $Z = 1$ and $Z = 0$ and at each value and $0 < p(X) < 1$ for all X), then for every value of $p(X)$, the expected value of Y_1 for individuals with $p(X)$ and $Z = 1$ is equal to the expected value of Y_1 for individuals with $p(X)$ and $Z = 0$ and the same for Y_0**

Various ways to use propensity scores to estimate causal effects



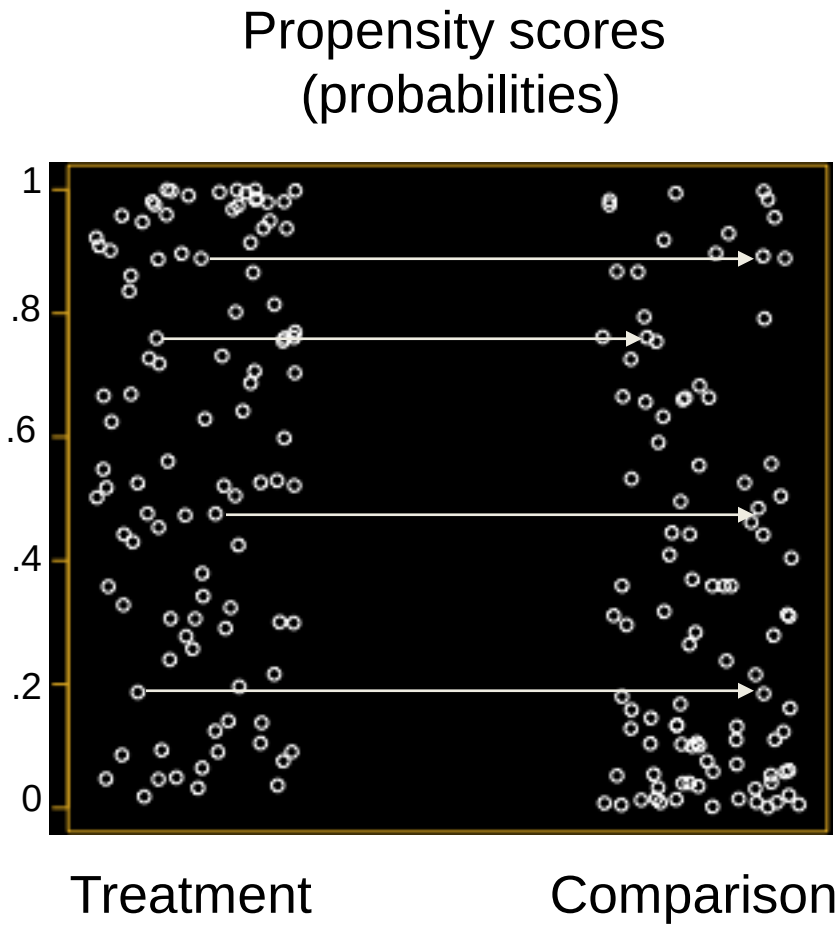
Various ways to use propensity scores to estimate causal effects

- Stratification



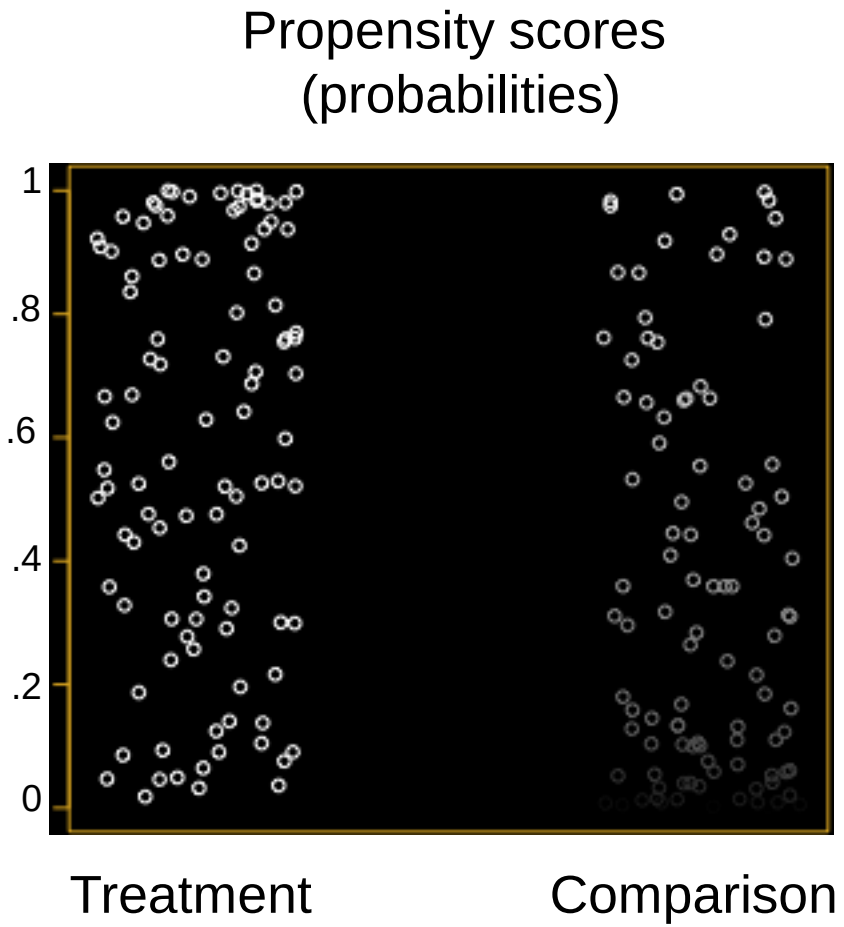
Various ways to use propensity scores to estimate causal effects

- Stratification
- Matching



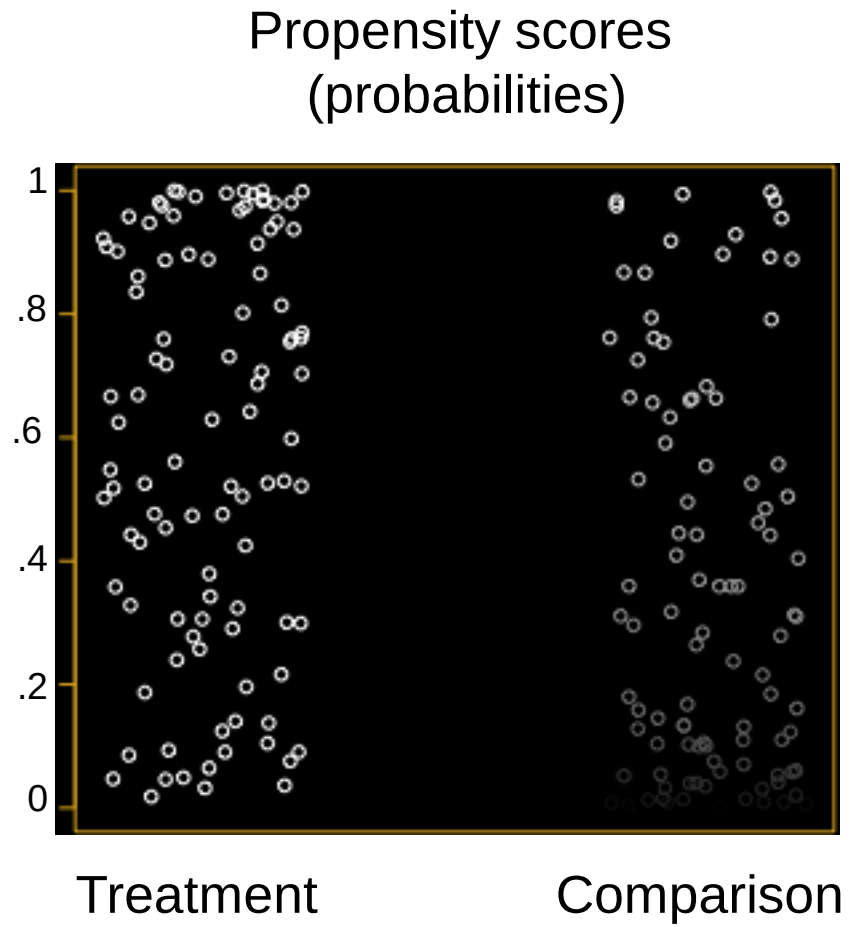
Various ways to use propensity scores to estimate causal effects

- Stratification
- Matching
- Weighting



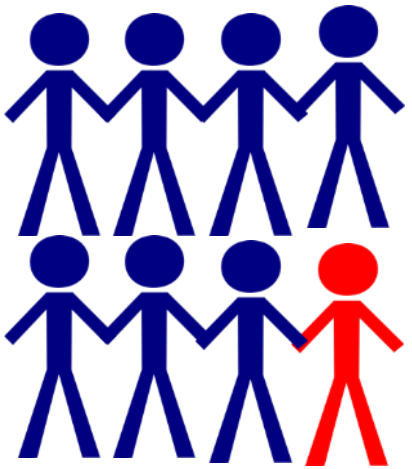
Various ways to use propensity scores to estimate causal effects

- Stratification
- Matching
- Weighting
  - Today, we will focus on how to use propensity score weights to correct for selection effects on observed pretreatment characteristics

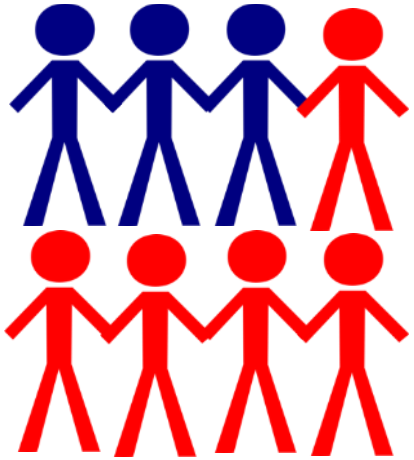


Additional illustration of weighting

Treated

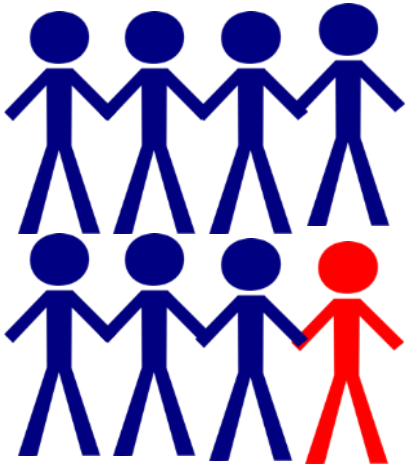


Control

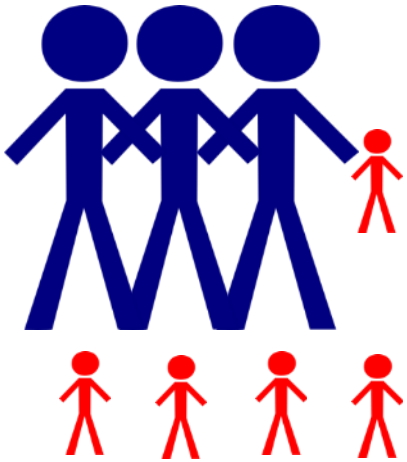


Additional illustration of weighting

Treated



Control



Propensity score weights for different causal effects

- For ATE:
  - weight treatment group by $1/\hat{p}(X_i)$
  - weight comparison group by $1/(1-\hat{p}(X_i))$
- For ATT:
  - weight comparison group by $\hat{p}(X_i)/(1-\hat{p}(X_i))$
- Both can provide unbiased estimates of the causal treatment effect in question assuming
  1. covariates are balanced in treatment and control groups after weighting
  2. there are no unobserved confounders and overlap (strong ignorability)

Several Important Assumptions to Consider

Inter-Dependence Among Participant Outcomes

- **Participant's potential outcomes depend on treatment assignment of other participants**
  - **Outcomes are better if best friend is also in treatment**
  - $U = 1$ **if best friend in treatment; 0 otherwise**
  - Y_{00}, Y_{10}, Y_{01}, Y_{11}
  - $T_{U=0} = Y_{10} - Y_{00}$
  - $T_{U=1} = Y_{11} - Y_{01}$
- **Typically assume no inter-dependence or interference**
- **Stable Unit Treatment Value Assumption (SUTVA)**

Key Assumption Assignment Depends Only on Observed Covariates

- Key assumption for both ANCOVA and propensity score methods
- If assignment depends on unobservables, then method will produce biased results
- Assumption cannot be tested directly with data

Overlap

- All individuals have a positive probability of receiving treatment and control:

$$0 < p(X) < 1$$

  - E.g, you don't just have women in your control group

- This is also difficult to test in practice

- Simple summary statistics can help you assess for obvious areas of concern

  - Check for 0/empty cells with binary and categorical covariates in the treatment and control groups

  - Compare the minimums and maximums for continuous covariates between the treatment and control groups