Inverse Probability Weighting

STSCI/INFO/ILRST 3900: Causal Inference

10 Sep 2026

Learning goals for today

At the end of class, you will be able to:

1. Estimate the average causal effect using data from a conditionally randomized experiment using inverse probability weighting
2. Explain why conditional exchangeability might be reasonable in some observational data

Logistics

- Ch 1.3 and 2.4 in Hernan and Robins 2023
- Quiz 1 will be in class on Sep 17
- Review in discussion on Sep 16

Conditional Exchangeability

Exchangeability may not hold when considering everyone, but may hold when considering each sub-group separately

Covid vaccine example:

-
  - Old: randomly assigned vaccine with probability 2/3
  - Young: randomly assigned vaccine with probability 1/4

Conditional Exchangeability

Exchangeability may not hold when considering everyone, but may hold when considering each sub-group separately

Covid vaccine example:

-
  - Old: randomly assigned vaccine with probability 2/3
  - Young: randomly assigned vaccine with probability 1/4
  - Vaccinated group has higher proportion of old people than unvaccinated group
  - If old/young have different distribution of potential outcomes, vaccinated group has different distribution of potential outcomes compared to unvaccinated group
  - Exchangeability does not hold!

Conditional Exchangeability

Consider the data as coming from two different experiments

Conditional Exchangeability

Consider the data as coming from two different experiments

- Consider just old group: exchangeability holds
- Average potential outcomes for treated old are same as average potential outcomes for all old

Conditional Exchangeability

Consider the data as coming from two different experiments

- Consider just old group: exchangeability holds
- Average potential outcomes for treated old are same as average potential outcomes for all old
- Conditional ATE (the ATE for old people) is difference in observed conditional means

$$
\begin{aligned}
\text{ATE for old} &= \mathrm{E}(Y^{a=1} \mid L = \text{old}) - \mathrm{E}(Y^{a=0} \mid L = \text{old}) \\
&= \mathrm{E}(Y \mid A = 1, L = \text{old}) - \mathrm{E}(Y \mid A = 0, L = \text{old})
\end{aligned}
$$

Standardization

- **Standardization** allows us to estimate the ACE by combining estimates from each sub-population

$$\mathrm{E}(Y^{a=1}) = \sum_{\ell} \underbrace{\mathrm{E}(Y^{a=1} \mid L = \ell)}_{\text{Avg for group } \ell} \times \underbrace{Pr(L = \ell)}_{\text{Size of group } \ell}$$

Standardization

- **Standardization** allows us to estimate the ACE by combining estimates from each sub-population

$$E(Y^{a=1}) = \sum_{\ell} \underbrace{E(Y^{a=1} \mid L = \ell)}_{\text{Avg for group } \ell} \times \underbrace{Pr(L = \ell)}_{\text{Size of group } \ell}$$

- Under conditional exchangeability

$$\begin{aligned} E(Y^{a=1}) &= \underbrace{E(Y^{a=1} \mid L = \text{old})}_{\text{Avg for old}} \times \underbrace{Pr(L = \text{old})}_{\text{Size of old}} \\ &\quad + \underbrace{E(Y^{a=1} \mid L = \text{young})}_{\text{Avg for young}} \times \underbrace{Pr(L = 0)}_{\text{Size of young}} \\ &= \underbrace{E(Y \mid A = 1, L = \text{old})}_{\text{Avg for treated old}} \times \underbrace{Pr(L = \text{old})}_{\text{Size of old}} \\ &\quad + \underbrace{E(Y \mid A = 1, L = \text{young})}_{\text{Avg for treated young}} \times \underbrace{Pr(L = \text{young})}_{\text{Size of young}} \end{aligned}$$

Generalization (transportability)

- If Conditional ATE differs between sub-groups, we say that the treatment effect is **heterogeneous**

$$\text{ATE old} \neq \text{ATE young}$$

Generalization (transportability)

- If Conditional ATE differs between sub-groups, we say that the treatment effect is **heterogeneous**

$$\text{ATE old} \neq \text{ATE young}$$

- If treatment effect is heterogeneous, ATE is defined with respect to a population of interest:
  May need to use different weights when calculating ATE for population
- Example: most psychology experiments consist primarily of undergraduate students. Results may be accurate for undergraduate students, but may not say much about broader population

Generalization (transportability)

- Drawing representative sample for RCTs yields more reliable conclusions
- Transporting effect to different population
  - Assume conditional ATE is same for each sub-group across populations
  - Standardization should use population weights in target population

Generalization (transportability)

- Suppose the vaccine experiment was conducted at Cornell where $Pr(L = \text{young}) = .95$ and $Pr(L = \text{old}) = .05$
- Suppose

$$ATE \text{ young} = .2 \qquad\qquad ATE \text{ old} = -.4$$

- ATE for Cornell population is

$$ATE = \underbrace{.2}_{ATE \text{ young}} \times \underbrace{.95}_{\text{Size young}} + \underbrace{-.4}_{ATE \text{ old}} \times \underbrace{.05}_{\text{Size old}} = 0.17$$

- In New York, $Pr(L = \text{young}) = .6$ and $Pr(L = \text{old}) = .4$
- Assuming Conditional ATE in New York and Cornell are same, ATE for New York is

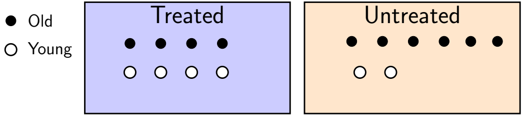
$$.2 \times .6 + (-.4) \times .4 = -0.04$$

Inverse probability weighting

- Standardization: constructs an estimate of $\mathrm{E}(Y^a)$ through a weighted average
- Inverse probability weighted (IPW) estimator also gets ATE and is equivalent to standardization

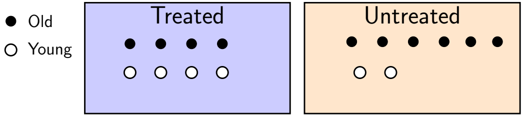
Inverse probability weighting

Observe: A conditionally randomized experiment:



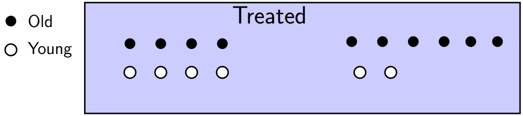
Inverse probability weighting

Observe: A conditionally randomized experiment:



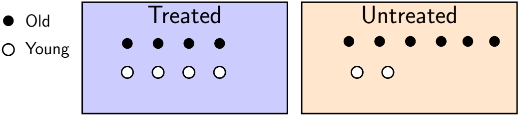
In order to get the ATE $= E(Y^{a=1}) - E(Y^{a=0})$

Want: a world where everyone is treated



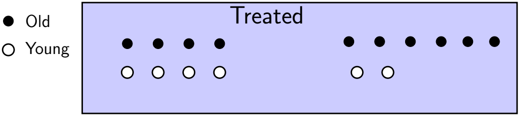
Inverse probability weighting

Observe: A conditionally randomized experiment:

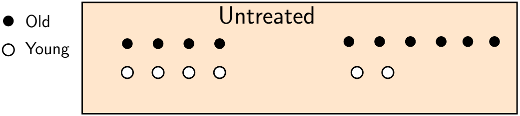


In order to get the ATE $= E(Y^{a=1}) - E(Y^{a=0})$

Want: a world where everyone is treated

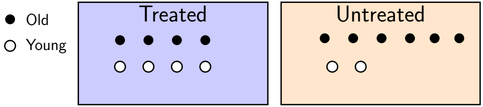


Want: a world where everyone is not treated

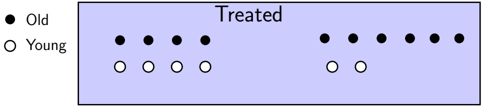


Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:

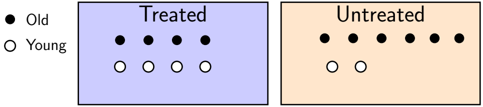


Want: a world where everyone is treated

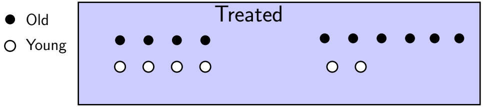


Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



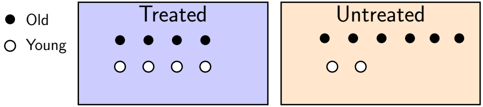
Want: a world where everyone is treated



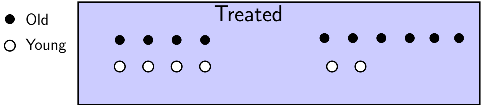
Main idea: Reweight observed units to match the hypothetical world we want

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated

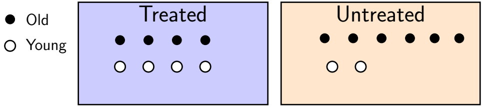


Main idea: Reweight observed units to match the hypothetical world we want

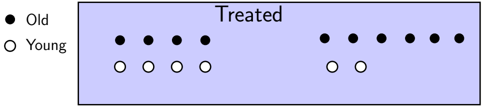
- **Observe**: 4 Treated Young people
- **Want**: 6 Treated Young people

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated

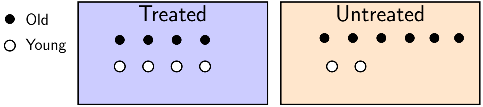


Main idea: Reweight observed units to match the hypothetical world we want

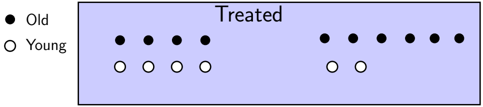
- **Observe**: 4 Treated Young people
- **Want**: 6 Treated Young people
- Count each observed Treated Young Person as 6/4 hypothetical persons

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated

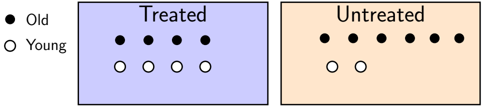


Main idea: Reweight observed units to match the hypothetical world we want

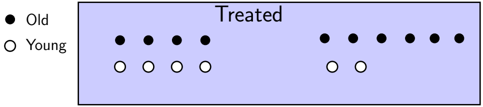
- **Observe**: 4 Treated Young people
- **Want**: 6 Treated Young people
- Count each observed Treated Young Person as 6/4 hypothetical persons
- Note: Probability of being treated conditional on being young: 4/6; i.e., $Pr(A_i = 1 \mid L = \text{young}) = 4/6$ (propensity score)

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



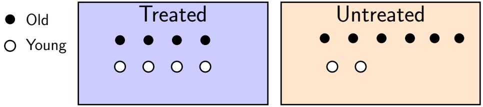
Want: a world where everyone is treated



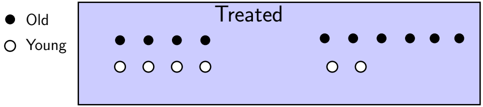
Main idea: Reweight observed units to match the hypothetical world we want

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated

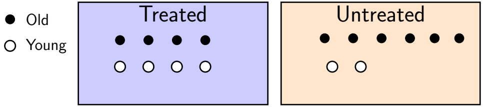


Main idea: Reweight observed units to match the hypothetical world we want

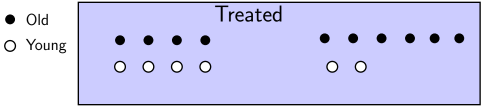
- **Observe**: 4 Treated Old people
- **Want**: 10 Treated Old people

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated

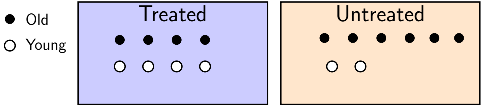


Main idea: Reweight observed units to match the hypothetical world we want

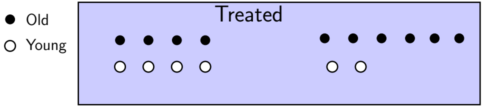
- **Observe**: 4 Treated Old people
- **Want**: 10 Treated Old people
- Count each observed Treated Old Person as 10/4 hypothetical persons

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is treated



Main idea: Reweight observed units to match the hypothetical world we want

- **Observe**: 4 Treated Old people
- **Want**: 10 Treated Old people
- Count each observed Treated Old Person as 10/4 hypothetical persons
- Note: Probability of being treated conditional on being old: 4/10; i.e., $Pr(A_i = 1 \mid L = \text{old}) = 4/10$ (propensity score)

Inverse probability weighting: Main Idea

Reweight observed units to match the hypothetical treated world

- Reweigh each treated **old** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being old}}$$

- Reweigh each treated **young** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being young}}$$

Inverse probability weighting: Main Idea

Reweight observed units to match the hypothetical treated world

- Reweigh each treated **old** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being old}}$$

- Reweigh each treated **young** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being young}}$$

- **General rule**: reweigh each treated individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given age}}$$

Inverse probability weighting: Main Idea

Reweight observed units to match the hypothetical treated world

- Reweigh each treated **old** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being old}}$$

- Reweigh each treated **young** individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given being young}}$$

- **General rule**: reweigh each treated individual by

$$\text{Weight} = \frac{1}{\text{Prob of being treated given age}}$$

- Probability of being treated given covariates is called **propensity score**

$$\pi_i = Pr(A_i = 1 \mid L = \text{Age}_i)$$

Inverse probability weighting: Math

- To get $\mathrm{E}(Y^{a=1})$, we want the hypothetical world all treated individuals

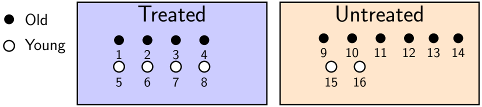
$$\mathrm{E}(Y^{a=1}) = \frac{1}{\#\text{ obs in hyp}} \sum_{i \text{ is in hyp}} Y_i^{a=1}$$

- To approximate the hypothetical world we use the reweighted actual treated observations

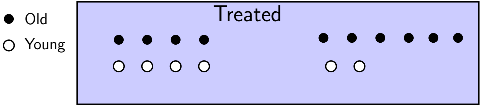
$$\begin{aligned}
\mathrm{E}(Y^{a=1}) &= \frac{1}{\#\text{ obs in hyp}} \sum_{i:A_i=1} (\text{Weight for i}) \times Y_i^{a=1} \\
&= \frac{1}{\#\text{ obs in hyp}} \sum_{i:A_i=1} \frac{1}{Pr(A_i = 1 \mid L = \text{Age}_i)} \times Y_i
\end{aligned}$$

Inverse probability weighting: Calculation

Observe: A conditionally randomized experiment:



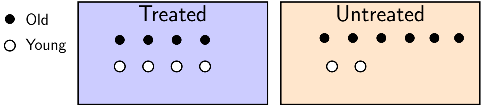
Want: a world where everyone is treated



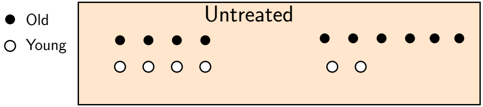
$$\text{E}(Y^{a=1}) = \frac{1}{\#\text{ obs in hyp}} \sum_{i:A_i=1} \frac{1}{Pr(A_i = 1 \mid L = \text{Age}_i)} \times Y_i$$

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



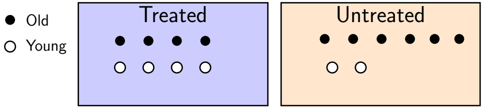
Want: a world where everyone is untreated



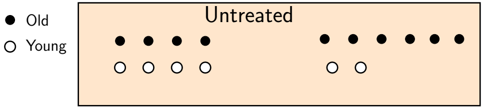
Main idea: Reweigh observed units to match the hypothetical world we want

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated

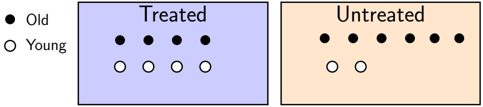


Main idea: Reweigh observed units to match the hypothetical world we want

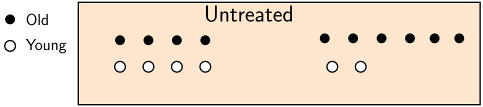
- **Observe**: 2 untreated young people
- **Want**: 6 untreated young people

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated

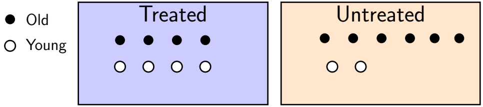


Main idea: Reweigh observed units to match the hypothetical world we want

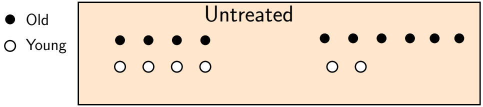
- **Observe**: 2 untreated young people
- **Want**: 6 untreated young people
- Count each observed untreated young Person as 6/2 hypothetical persons

Inverse probability weighting: Intuition

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated

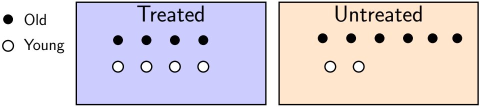


Main idea: Reweigh observed units to match the hypothetical world we want

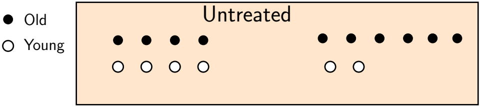
- **Observe**: 2 untreated young people
- **Want**: 6 untreated young people
- Count each observed untreated young Person as 6/2 hypothetical persons
- Note: Probability of **untreated** conditional on being young: 2/6; i.e., $Pr(A_i = 0 \mid L = \text{young}) = 2/6 = 1 - Pr(A_i = 0 \mid L = \text{young})$

Inverse probability weighting: Practice

Observe: A conditionally randomized experiment:



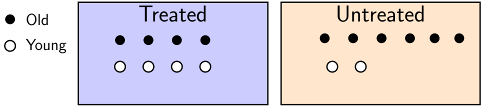
Want: a world where everyone is untreated



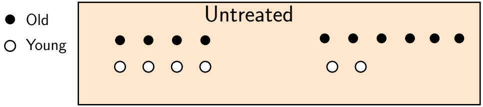
Main idea: Reweight observed units to match the hypothetical world we want

Inverse probability weighting: Practice

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated

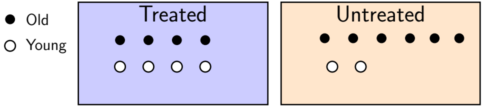


Main idea: Reweight observed units to match the hypothetical world we want

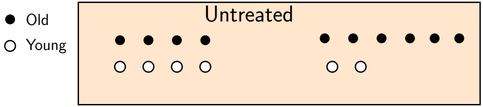
- **Observe**: ? untreated old people
- **Want**: ? untreated old people

Inverse probability weighting: Practice

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated

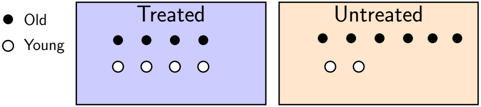


Main idea: Reweight observed units to match the hypothetical world we want

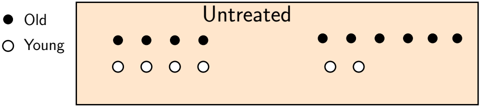
- **Observe**: ? untreated old people
- **Want**: ? untreated old people
- Count each observed untreated old person as ? hypothetical persons

Inverse probability weighting: Practice

Observe: A conditionally randomized experiment:



Want: a world where everyone is untreated



Main idea: Reweight observed units to match the hypothetical world we want

- **Observe**: ? untreated old people
- **Want**: ? untreated old people
- Count each observed untreated old person as ? hypothetical persons
- Note: Probability of untreated conditional on being old: ?; i.e., $Pr(A_i = 0 \mid L = \text{old}) = ?$

IPW: Everything together

- To approximate the hypothetical **treated** world each **treated** individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of being } \textbf{treated} \text{ given age}}$$

IPW: Everything together

- To approximate the hypothetical **treated** world each **treated** individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of being } \textbf{treated} \text{ given age}}$$

- To approximate the hypothetical **untreated** world each **untreated**individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of being } \textbf{untreated} \text{ given age}}$$

- **General rule**: Each individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of their obs treatment status given their age}}$$

IPW: Everything together

General rule: Each individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of their obs treatment status given their age}}$$

- Propensity Score: $\pi_i = Pr(A_i = 1 \mid L_i)$
- Let N be the number of observations in the hypothetical world (which is the same as the number of actual observations)

IPW: Everything together

General rule: Each individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of their obs treatment status given their age}}$$

- Propensity Score: $\pi_i = Pr(A_i = 1 \mid L_i)$
- Let N be the number of observations in the hypothetical world (which is the same as the number of actual observations)
- Hypothetical world where everyone is treated

$$\mathrm{E}(Y^{a=1}) = \frac{1}{N} \sum_{i:A_i=1} \frac{1}{\pi_i} \times Y_i$$

IPW: Everything together

General rule: Each individual should be reweighted by

$$\text{Weight} = \frac{1}{\text{Prob of their obs treatment status given their age}}$$

- Propensity Score: $\pi_i = Pr(A_i = 1 \mid L_i)$
- Let N be the number of observations in the hypothetical world (which is the same as the number of actual observations)
- Hypothetical world where everyone is treated

$$\mathrm{E}(Y^{a=1}) = \frac{1}{N} \sum_{i:A_i=1} \frac{1}{\pi_i} \times Y_i$$

- Hypothetical world where everyone is not treated

$$\mathrm{E}(Y^{a=0}) = \frac{1}{N} \sum_{i:A_i=0} \frac{1}{1-\pi_i} \times Y_i$$

- $ATE = \quad \frac{1}{N} \sum_{i:A_i=1} \frac{1}{\pi_i} \times Y_i \quad - \quad \frac{1}{N} \sum_{i:A_i=0} \frac{1}{1-\pi_i} \times Y_i$

Conditional exchangeability in observational data

- When conditional exchangeability holds, we can estimate causal effects from the observed data
- Use either standardization or inverse probability weighting
- By design, conditional exchangeability holds in conditionally randomized experiments
- Marginal exchangeability is very unlikely in observational data
- Conditional exchangeability is may be more reasonable in observational data

Excercise

Suppose we have data gathered by surveying individuals in Fall of 2021

- Whether the individual was vaccinated for Covid
  $A_i = 1$ if vaccinated, $A_i = 0$ if not vaccinated
- Whether the individual tested positive for Covid in 2021
  $Y_i = 1$ if positive test, $Y_i = 0$ if no positive test
- What additional information could you gather about each individual to make conditional exchangeability might be plausible?

$$Y^{a=1}, Y^{a=0} \perp\!\!\!\perp A \mid L$$

Conditional exchangeability in observational data

- Even if gathering data was possible for every covariate we want, when do we stop?

Conditional exchangeability in observational data

- Even if gathering data was possible for every covariate we want, when do we stop?
- Never 100% sure that conditional exchangeability holds
- Is it reasonable?

Conditional exchangeability in observational data

- Even if gathering data was possible for every covariate we want, when do we stop?
- Never 100% sure that conditional exchangeability holds
- Is it reasonable?
- In observational data, conditional exchangeability is an assumption we make (but can't typically verify)
- Requires expert knowledge

Conditional exchangeability in observational data

- Even if gathering data was possible for every covariate we want, when do we stop?
- Never 100% sure that conditional exchangeability holds
- Is it reasonable?
- In observational data, conditional exchangeability is an assumption we make (but can't typically verify)
- Requires expert knowledge
- Causal claims are data + outside knowledge

Learning goals for today

At the end of class, you will be able to:

1. Estimate the average causal effect using data from a conditionally randomized experiment using inverse probability weighting
2. Explain why conditional exchangeability might be reasonable in some observational data