Causal Inference & Causal Learning

Observational Study

Xiaolei Lin

School of Data Science
Fudan University

Mar 22, 2024

Recap

- Neyman's Repeated Sampling Approach to Completely Randomized Experiments
- Comparing Neyman's and Fisher's perspectives to (causal) inference

Today's plan

- Confounding in Observational studies
- g-formula in Randomized Studies
- g-formula in Observational Studies

ACE

- PACE & SACE

$$SACE = \frac{1}{N}\sum_{i=1}^{N}(Y_i^1 - Y_i^0)$$

$$PACE = E(Y^1 - Y^0)$$

- These quantities are referred to as "the estimands" of interest

Identifiability

- To recover the population or sample mean of the potential outcomes, we would like it to be the case that those who had treatment 1 and those who had treatment 0 are comparable (in their counterfactual outcomes)
- If that were the case, then the outcomes of those who had treatment 1 would be similar to the outcomes if the whole population had been given treatment 1
- And the outcomes of those who had treatment 0 would be similar to the outcomes if the whole population had been given treatment 0
- However, this will often not be the case

Confounding

- Randomized experiments can identify and quantify ACE because the randomized assignment of treatment leads to exchangeability.
- Observational studies are much less convincing due to lack of randomized treatment assignment
- Heart transplantation
  - who received the heart transplant were more likely to have a severe heart condition
  - if those who received a transplant had not received it, they would have been expected to have a greater death risk than those who did not actually receive the heart transplant

Confounding

Possible solutions:

- We have thus far considered that by study design we can make treatment assignment independent of outcome risk factors
- However, even if the groups who received treatment 1 and those who received treatment 0 are not comparable, it is possible that within **strata of other variables** those who received treatment 1 and those who received treatment 0 are comparable
- Then analyze data as if treatment had been randomly assigned conditional on measured covariates

Identifiability conditions

Informally, an observational study can be conceptualized as a conditionally randomized experiment if the following conditions hold:

- the values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data
- the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates C
- the probability of receiving every value of treatment conditional on C is greater than zero, i.e., positive

Marginal Exchangeability

$$Y^a \perp A$$

- The treated and the untreated are exchangeable because the treated, had they remained untreated, would have experienced the same average outcome as the untreated did.
- Does this hold for the heart transplantation example?
- More treated (69%) than untreated (43%) in critical condition

Exchangeability

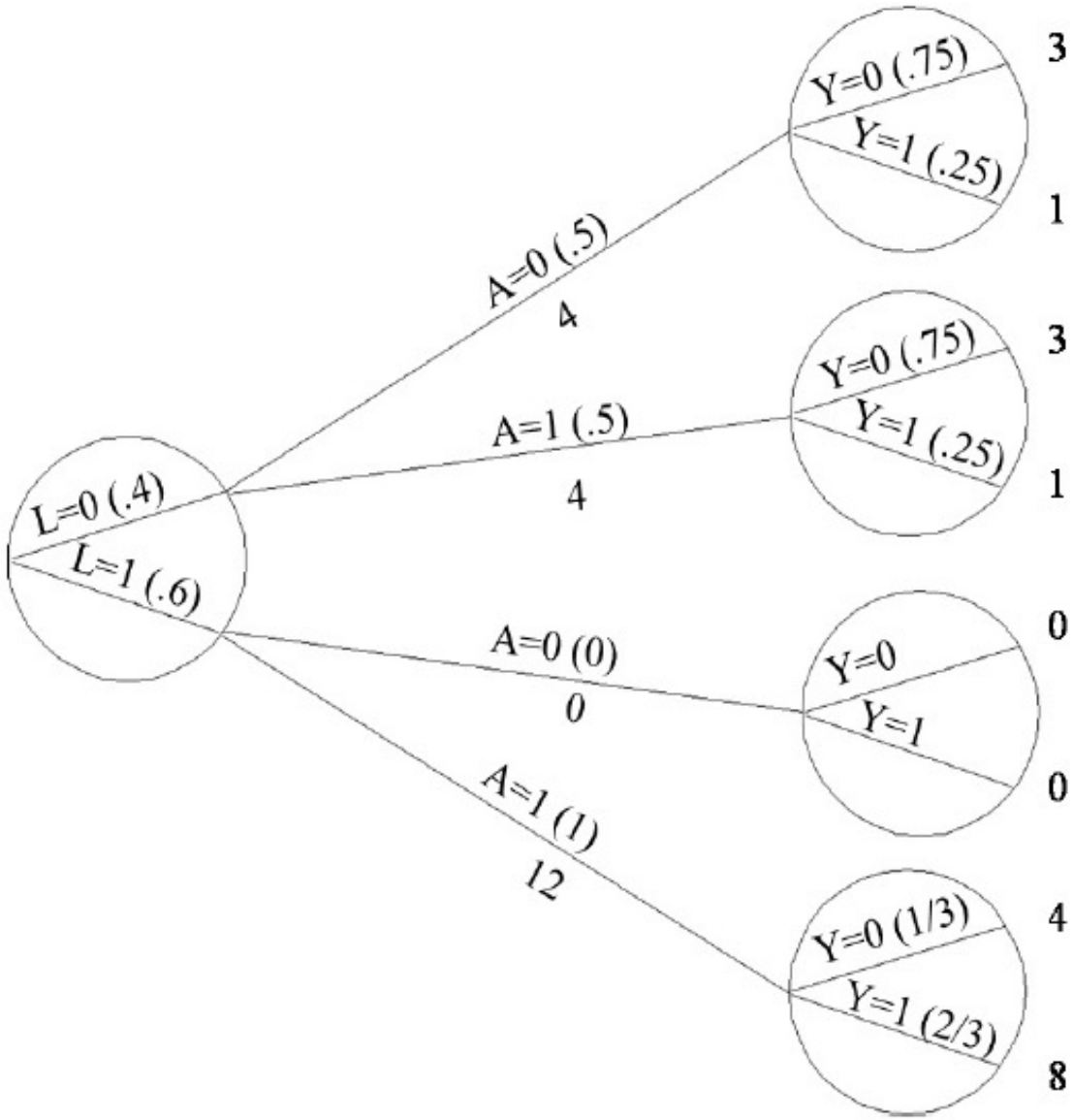
	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Conditional Exchangeability

$$Y^a \perp A \mid C$$

- The treated and the untreated are conditionally exchangeable within levels of the pre-treatment covariate C
- In the subset of $C = 1$ (critical condition), the treated and the untreated are exchangeable because the treated, had they remained untreated, would have experienced the same average outcome as the untreated did
- Similarly for $C = 0$
- Violations of conditional exchangeability?

Positivity


L=0 (.4)
L=1 (.6)
A=0 (.5)
4
A=1 (.5)
4
A=0 (0)
0
A=1 (1)
12
Y=0 (.75)
Y=1 (.25)
3
1
Y=0 (.75)
Y=1 (.25)
3
1
Y=0
Y=1
0
0
Y=0 (1/3)
Y=1 (2/3)
4
8

Consistency

$$Y^a = Y$$

- The observed outcome for every treated individual equals his/he/her outcome if she had received treatment
- the observed outcome for every untreated individual equals his/he her outcome if she had remained untreated
- The two main components of consistency
  1. a precise definition of the counterfactual outcomes Y^a via a detailed specification of the superscript a
  2. the linkage of the counterfactual outcomes to the observed outcomes

Consistency

Causal question of interest: causal effect of heart transplant on 5-year mortality

- Intervention: heart transplant + pre-operative procedures + anesthesia + surgical technique + post-operative care + immunosuppressive therapy
- Specify all these details in the protocol and define the version of intervention
- For interventions in randomized experiment, a precise specification of the version of interest may include its type (e.g., surgical technique), amount and timing (e.g., duration, frequency)

Consistency

Define precise version of treatment is difficult for observational studies with causal questions involving biological (blood pressure, LDL-cholesterol, body weight) or social (e.g., socioeconomic status) treatments

- example 1: causal effect of exercise on blood pressure; the duration, frequency, intensity, and type of exercise (swimming, running, playing basketball...)? how the time devoted to exercise would otherwise be spent (playing with your children, rehearsing with your band, watching television...)
- example 2: causal effect of obesity at age 40 on the risk of mortality by age 50; many different ways to become obese, genetic, physical inactivity, calorie intake, smoking cessation, lack of bariatric surgery; each of these mechanisms may have different effects on mortality even if they all led to the same body weight

Consistency

absolute precision in the definition of the treatment is not needed for useful causal inference

- benefits of running clockwise around your neighborhood's park are the same as those of running counterclockwise
- we only need sufficiently well-defined interventions a for which no meaningful vagueness remains
- largely rely on scientific consensus and may vary over time

Target trial

If all identifiability conditions are met, we can analyze the causal effects of observational studies as in conditional randomized experiment (target trial)

Randomization

- Identification through randomization: Suppose we randomize our population of patients with a headache to either aspirin or no aspirin with equal probability $\frac{1}{2}$
- The crude association between A and Y will coincide with the average causal effect of A on Y
- This is not a coincidence but by design

Baseline variables

- Baseline covariates always exist, whether they are used to design the trial (as inclusion or exclusion criteria), for balancing, or for effciency
- We have distinguished between two types of baseline pre-randomization characteristics:
  1. Observed baseline characteristics C that correlate with the observed outcome
  2. Unobserved baseline characteristics U that correlate with the observed outcome

The g-formula in randomized trials

- Recall that under randomization, we have A independent of C, U
- Consequently:

$$
\begin{aligned}
E(Y^a) &= \sum_c E(Y^a \mid C = c) Pr(C = c) \\
&= \sum_c E(Y^a \mid A = a, C = c) Pr(C = c) \\
&= \sum_{c,u} E(Y \mid A = a, C = c, U = u) Pr(C = c, U = u) \\
&= E(Y \mid A = a)
\end{aligned}
$$

- The first equality follows from the "law of iterated expectation" , whereby an average can be written as an average of averages. The second and third equalities follow from randomization (RA) and consistency (CA)

The g-formula in randomized trials

- The following equation is particularly of interest

$$E(Y^a) = \sum_c E(Y \mid A = a, C = c) Pr(C = c)$$

- This is a the so-called "g-formula" due to Robins, or "intervention formula" due to Pearl
- In Epidemiology it is also referred to as "direct standardization"
- "Direct standardization" averages the conditional mean over the distribution of baseline covariates in the population from which a randomized sample is obtained

The g-formula in randomized trials

- The g-formula representation of the causal effect estimated in a randomized trial makes clear that it depends on the inclusion and exclusion criteria used to define the study population
- Therefore, two studies with different inclusion criteria will in general yield two different causal effects
- e.g. Consider an HIV drug trial that excludes participants with CD4 count $>$500 v.s. one that excludes participants with CD4 count $>$ 350 (CD4: white blood cells that fight infection)

The g-formula in randomized trials

- We re-state the following key simplification of the g-formula in randomized experiments as the average outcome in the subgroup of individuals with $A = a$

$$E(Y^a) = \sum_c E(Y \mid A = a, C = c) Pr(C = c) = E(Y \mid A = a)$$

- As we will see, this simplification is specific to the marginal randomized design, and does not extend to the observational study setting

The g-formula in conditional randomized trials

Heart transplantation example (conditional randomization)

- 8 individuals with $C = 0$: $A = 1$ with probability 50%
  - $Pr(Y = 1 \mid C = 0, A = 1) = \frac{1}{4}$ & $Pr(Y = 1 \mid C = 0, A = 0) = \frac{1}{4}$
- 12 individuals with $C = 1$: $A = 1$ with probability 75%
  - $Pr(Y = 1 \mid C = 1, A = 1) = \frac{2}{3}$ & $Pr(Y = 1 \mid C = 1, A = 0) = \frac{2}{3}$
- Calculate causal risk ratio $\frac{Pr(Y^1=1)}{Pr(Y^0=1)} = \frac{\frac{1}{4}\times 0.4 + \frac{2}{3}\times 0.6}{\frac{1}{4}\times 0.4 + \frac{2}{3}\times 0.6} = \frac{0.5}{0.5} = 1$
- Marginal counterfactual risk = weighted average of the stratum specific risk

Observational studies

- Suppose that, as typically the case in observational studies, the exposure/treatment/intervention is not randomized
- Then, it will typically be the case that individuals select or are selected to take the active treatment based on their underlying health condition
- Say goodbye to unbiased estimates and exact p-value by design...
- But if we only have observational data, we want to do the best we can!

GOAL: Make data from observational studies look as much as possible like data from a randomized experiment

Timing of Treatment

- In a randomized experiment, the timing of the treatment assignment is clear (covariates are clear)
- In an observational study, this may not be clear

Solution? Clearly define the timing of the treatment

Case study: anti-retroviral therapy (ART)

- Suppose that in the MACS study, a multi-city cohort study of HIV positive men in the US, physicians were historically more likely to prescribe anti-retroviral therapy (ART) to patients with low CD4 count (< 250) than patients with higher CD4 count (>250)
- In this setting, immune-compromised patients were generally more likely to receive the active drug than healthier patients
- A crude comparison of treated and untreated participants could lead to inference of a harmful effect of ART, especially during time periods when therapy was less durably effective than it is now

Design vs Analysis

- In a randomized experiment, the design phase (collecting data, balancing covariates, specifying plan) is done before access to outcomes and analysis
- In an observational study, you typically get all data together (covariates, treatment, outcomes): design and analysis mingled

Solution? do the design part before having access to/ looking at outcomes

No Outcomes!

- In a randomized experiment, outcomes are not available in the design phase
- When analyzing observational studies, the design should NOT include outcomes
- Anything done to try to make the data look balanced between treatment groups should IGNORE THE OUTCOMES!
- This is the only way to be objective
- As long as outcomes are hidden, you can do whatever you want to achieve covariate balance

Analysis

- In randomized experiments, there is usually a pre-specified protocol for analysis
- In observational studies, people often try many different models and analyses - introduces subjectivity and bias

Solution? specify protocol with outcomes in advance, and do most of your work in the design phase to make analysis easy

Assignment Mechanisms

- In a randomized experiment, the assignment mechanism is regular (unconfounded, individualistic, probabilistic, controlled) by design
- In an observational study these are only assumptions, and may not hold

Solution? Do what we can to make these assumptions more plausible

Case study: anti-retroviral therapy (ART)

- Consider an analysis that "adjusts" for CD4 count measured at time of a decision about whether or not to initiate therapy is likely to provide a "better" answer, since it should account, at least in part, for differential treatment assignment due to a measurement of CD4
- This is the main distinction between comparisons drawn from randomized trials vs observational studies
- In the former, treatment assignment is under our control at the design stage; whereas in the latter, we must attempt to understand what risk factors of the outcome influenced treatment decision
- Such factors are typically also predictive of the outcome and could create a spurious association between treatment and outcome when in truth there is none. We shall formalize these ideas...

No unobserved confounding assumption (NUCS)

- Suppose that the data generating mechanism involves variables $\{A, C, Y, U\}$
- As before A is the exposure, Y the outcome, $\{C, U\}$ are pre-exposure variables
- As in a randomized trial, we wish to identify and estimate ACE, i.e $E(Y^a)$, so that we may compute the population average causal effect $E(Y^1 - Y^0)$
- C are observed correlates of A and Y , i.e. observed confounders.

No unobserved confounding assumption (NUCS)

Let $X \perp Y \mid Z$ to denote that X is independent of Y conditional on Z

Confounding: the effect of treatment A on outcome Y is unconfounded given covariates C if for all values a: $Y^a \perp A \mid C$

- i.e. within strata of the confounding variables, the treatment groups are comparable (i.e. they have similar potential outcomes) and we can draw causal conclusions
- Implication (for conditional average causal effects given C)

$$E(Y^1 \mid C = c) = E(Y^1 \mid A = 1, C = c) = E(Y \mid A = 1, C = c)$$

$$E(Y^0 \mid C = c) = E(Y^0 \mid A = 0, C = c) = E(Y \mid A = 0, C = c)$$

- We can compute causal effects from the data

$$E(Y^1 \mid C = c) - E(Y^0 \mid C = c) = E(Y^0 \mid A = 1, C = c) - E(Y \mid A$$

No unobserved confounding assumption (NUCS)

- We collect data on as many "pre-treatment" variables as possible that affect both the treatment/exposure under consideration and the outcome
- The assumption of no unobserved confounding essentially states that the observed C suffices to account for confounding, and therefore within levels of C, it is as if A were randomized (by nature)
- Thus, conditional on C, A should be independent of all unobserved covariates U.
- Recall that Y^0 and Y^1 are the quintessential unobserved pretreatment risk factors for the outcome Y and therefore the assumption of no unobserved confounding states that

$$(NUCA) A \perp \{Y^0, Y^1\} \mid C$$

No unobserved confounding assumption (NUCS)

- As stated before, the intuition behind NUCA is similar to that of RA; however, the randomization probability is now allowed to depend on C, in a manner not under our control
- Conceptually, NUCA can be achieved only if we are able to measure all common causes of A and Y (that is all risk factors for Y that also determine A)
- Thus, we are implicitly assuming that all variables in U can either:
  1. predict A conditional on C, OR
  2. predict Y conditional on C and A,
- but not both. If both conditions occur simultaneously, NUCA is violated

The g-formula in observational studies

- Next, we show that the no unmeasured confounding assumption suffices to again identify $E(Y^a)$ as follow

$$\begin{aligned} E(Y^a) &= E\{E(Y^a \mid C)\} \\ &= \sum_c E(Y^a \mid C = c) Pr(C = c) \\ &= \sum_c E(Y^a \mid A = a, C = c) Pr(C = c) \\ &= \sum_c E(Y \mid A = a, C = c) Pr(C = c) \end{aligned}$$

- Note that this is exactly the same g-formula previously derived in the context of a randomized trial

- This shows that the g-formula unifies randomized trials and observational studies, however, the g-formula no longer has a simple form as a conditional mean, i.e. $E(Y \mid A = a)$ is no longer equal to $E(Y^a)$

Example

- Consider a hypothetical study with one confounder

N	A	C	$E(Y \mid A = a, C = c)$
4000	1	0	24
3000	1	1	36
8000	0	0	10
9000	0	1	22

- $Pr(A = 1 \mid C = 1) = \frac{1}{4}$ and $Pr(A = 1 \mid C = 0) = \frac{1}{3}$
- Moreover, $E(Y \mid A = 1, C = c) = 24 + 12c$ so that C predicts Y given A

Example

- The crude mean

$$E(Y \mid A = 1) = \sum_c E(Y \mid A = 1, C = c) Pr(C = c \mid A = 1)$$
$$= 24 \times \frac{4}{7} + 36 \times \frac{3}{7} = \frac{204}{7}$$

- Whereas

$$E(Y^1) = \sum_c E(Y \mid A = 1, C = c) Pr(C = c)$$
$$= 24 \times \frac{1}{2} + 36 \times \frac{1}{2} = 30$$

- The difference between the two quantities reflects the effect of adjustment for confounding bias

Conditional Average Causal Effects

- Suppose that we wish to make inferences about a conditional causal effect given a collection of baseline covariates V included in $C = \{V, W\}$, i.e.

$$E(Y^1 \mid V) - E(Y^0 \mid V)$$

- Then one can show that under NUCA

$$E(Y^a \mid V) = \sum_w E(Y \mid W = w, V = v, A = a) Pr(W = w \mid V = v$$

- A special case often of interest takes $V = C$ and the g-formula becomes

$$E(Y^a \mid V) = E(Y^a \mid C) = E(Y^a \mid C, A = a) = E(Y \mid C, A = a)$$

The g-formula for Conditional Average Causal Effects

- A special case often of interest takes $V = C$ and the g-formula becomes

$$E(Y^a \mid V) = E(Y^a \mid C) = E(Y^a \mid C, A = a) = E(Y \mid C, A = a)$$

- Therefore, the standard approach of fitting a regression model for Y given$\{A, C\}$ recovers the conditional g-formula given all confounders

- Here adjustment is achieved by CONDITIONING on the confounders in the regression model.

- We will see alternative approaches later in this course

Probabilistic assignment

- In addition to uncounfounded assumption, the assignment mechanism has to be probabilistic in observational studies as well.
- Probabilistic: Every unit has some chance of being assigned to either treatment group, conditional on covariates

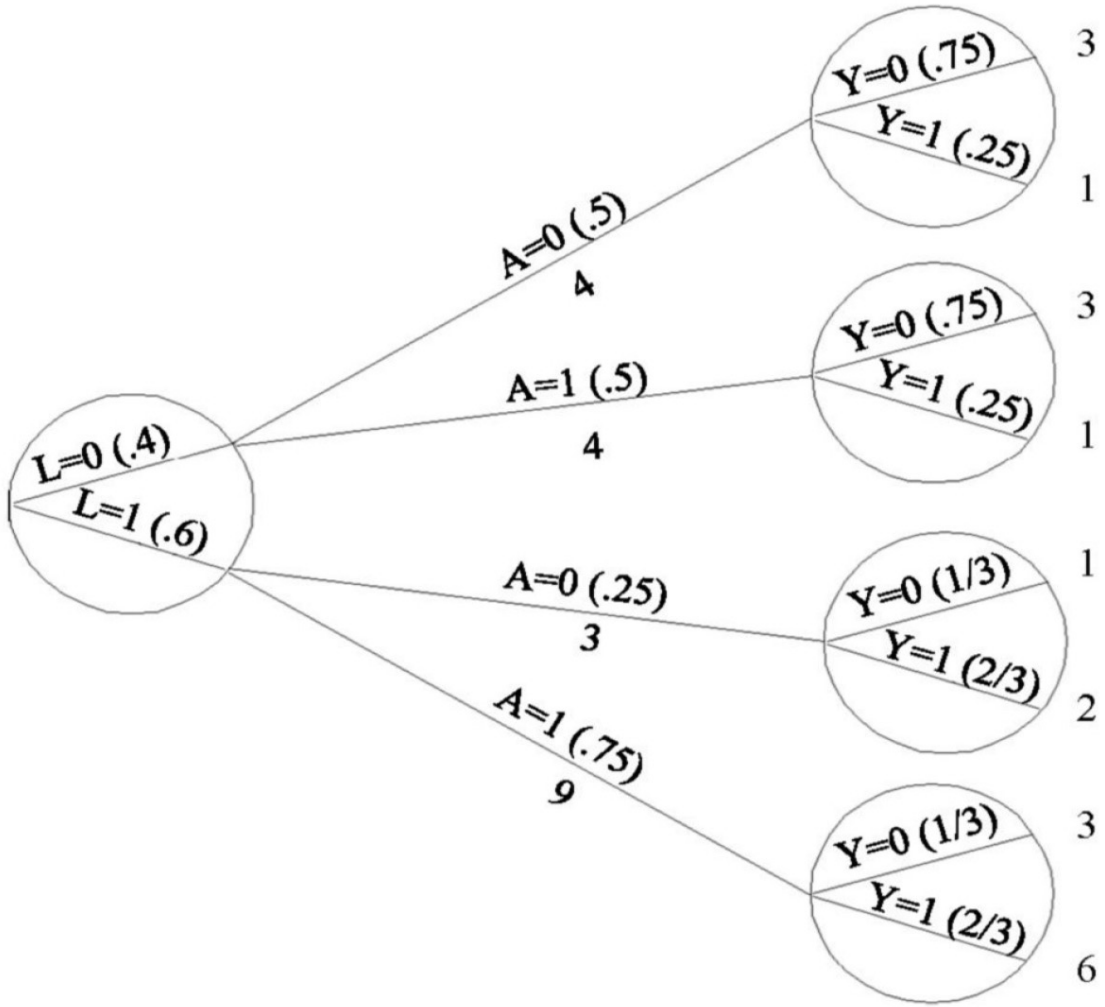
Solution? If certain types of units are only observed in one group, eliminate these units (restrict causal inferences to units who might get either treatment).

Or remove units not similar to any units in the opposite group (how can we measure this similarity?)

Observational studies vs randomized trials

- The g-formula for observational and randomized studies yield the same causal effect, provided the covariates C that confound A have the same distribution in both types of studies
- Underlying populations for which studies are representative must be comparable and inclusion and exclusion criterion must also match
- Randomization must have been proper i.e. missing data and censoring rates must be "comparable" across assignment arms and NUCA must hold.

Inverse probability weighting



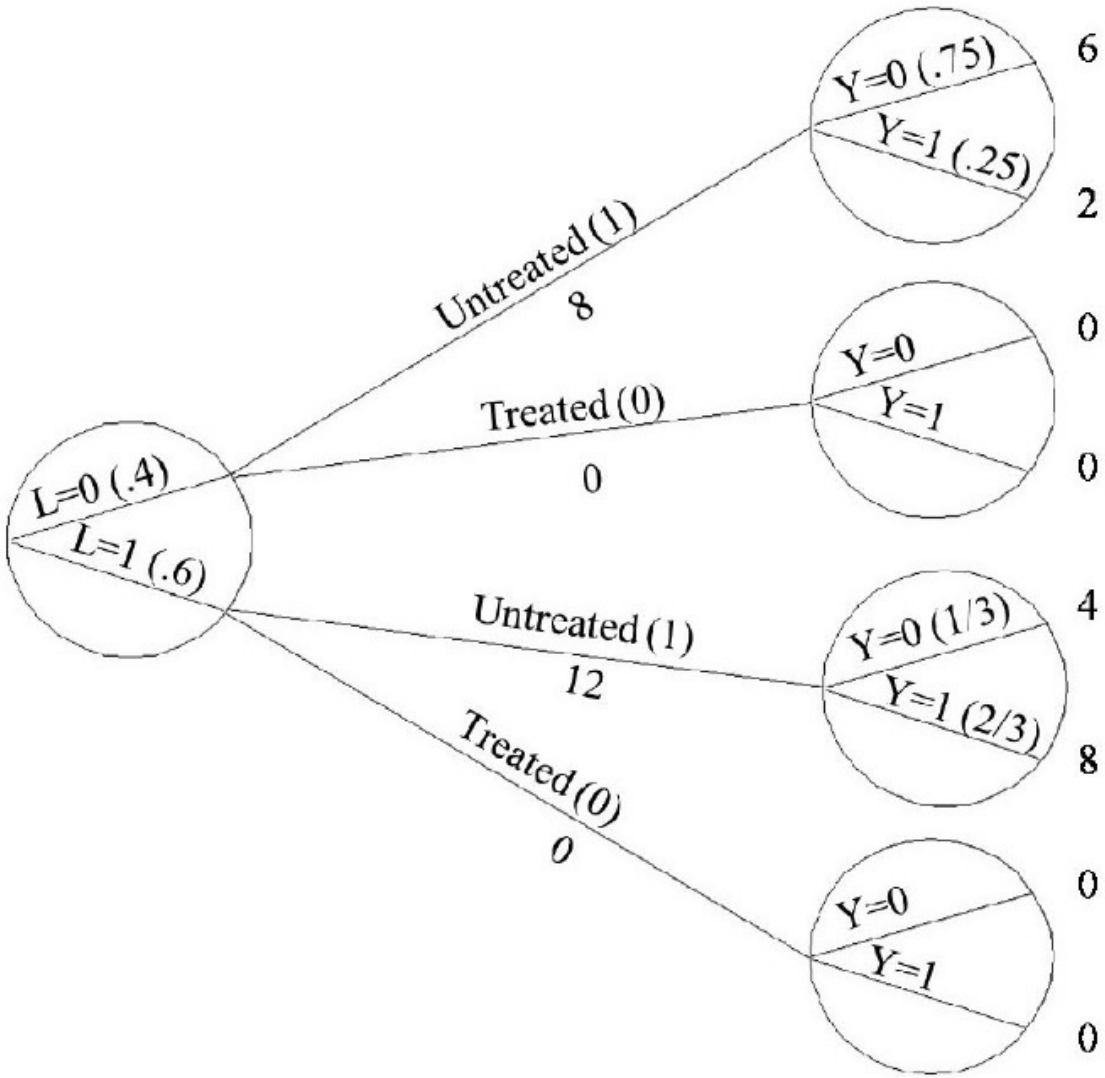
Let's calculate the causal risk ratio $\frac{Pr(Y^1=1)}{Pr(Y^0=1)}$

Inverse probability weighting

First, focus on the denominator $Pr(Y^0 = 1)$

- the counterfactual risk of death had everybody in the population remained untreated
- for $C = 0$: if all 8 individuals were untreated, 2 would have died
- for $C = 1$: if all 12 individuals were untreated, 8 would have died
- therefore, if all 20 individuals were untreated, 10 would have died

Inverse probability weighting

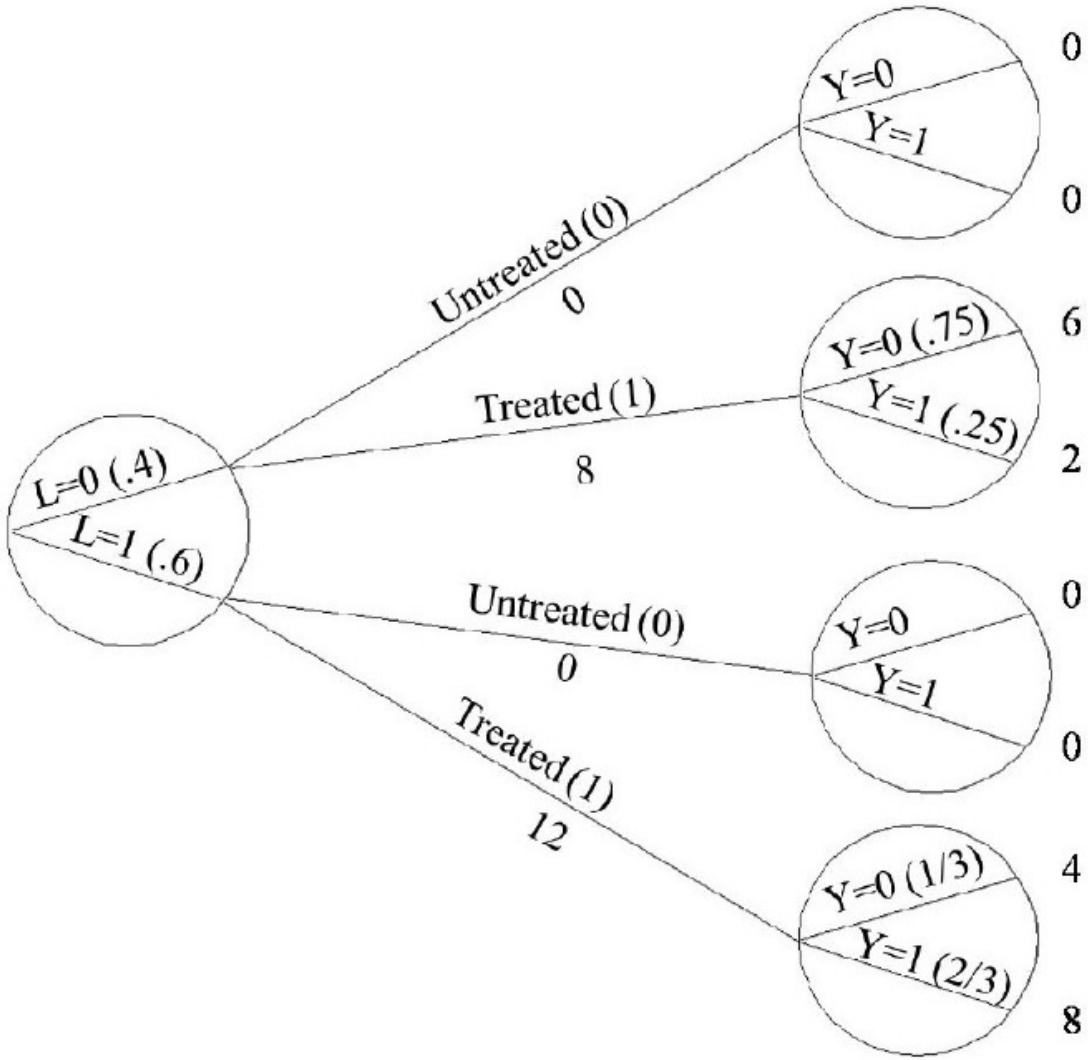


Inverse probability weighting

Next, focus on the numerator $Pr(Y^1 = 1)$

- the counterfactual risk of death had everybody in the population remained treated
- for $C = 0$: if all 8 individuals were treated, 2 would have died
- for $C = 1$: if all 12 individuals were treated, 8 would have died
- therefore, if all 20 individuals were treated, 10 would have died

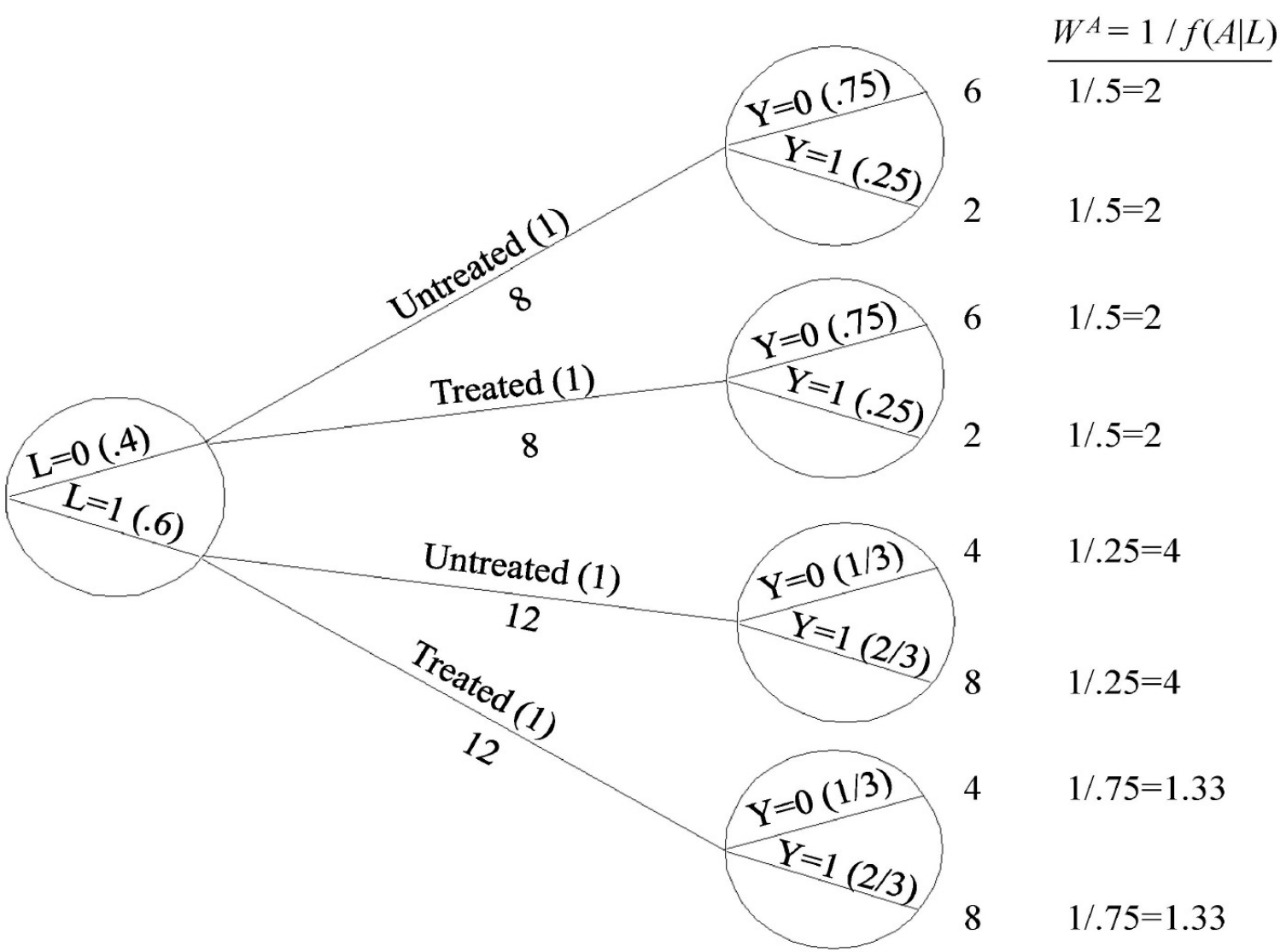
Inverse probability weighting



How IPW works

- The two trees are both simulations of what would have happened had all individuals in the population been untreated and treated
- These simulations are correct under conditional exchangeability
- Both simulations can be pooled to create a hypothetical population in which every individual appears as a treated and as an untreated individual
- This hypothetical population, twice as large as the original population, is known as the pseudo-population

Pseudo-population



Inverse probability weighting

- the associational risk ratio in the pseudo-population is equal to the causal risk ratio in the original population
- IP weighting yielded the same result as g-formula / standardization in this example (no causal effect of heart transplantation on mortality)
- IPW and g-formula / standardization are mathematically equivalent, but conceptually different