Week 3: Rubin's Causal Model & Controlling for Confounders

Causal Inference & Structural Equation Modeling

Noémi K. Schuurman
based on slides by Ellen Hamaker

February 2022

Overview

- **Potential Outcomes & Causal Effects**
- Estimating Causal Effects: Assumptions
- Estimating Causal Effects: Controlling for Confounders

Causal Inference Frameworks

In this course, we discuss two frameworks of causality:

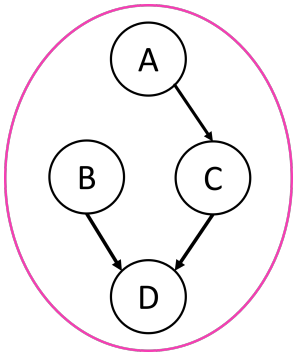
- **Structural causal model & DAGS** by Pearl: Previous weeks
- **Potential outcomes framework** by Rubin (also Imbens): Now
- After: Apply both

Judea Pearl and Don Rubin

Causal Inference Goals

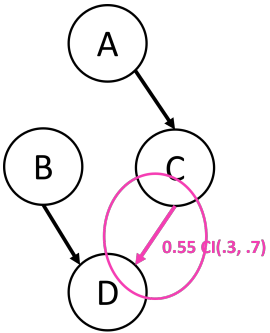
Causal Discovery or Causal 'Learning'

- Figuring out the causal structure among a bunch of variables - what are confounders, colliders, mediators?



Causal Identification

- Answering a specific, well-defined, causal question
- Estimating a specific, well-defined, causal effect



Defining Causal Effects: Because Statements vs. Causal Statements

Examples of **because statements**:

- My headache went away, because I took an aspirin.
- He is late, because he overslept.
- Hillary Clinton lost the election, because she is a woman.

Defining Causal Effects: Because Statements vs. Causal Statements

Examples of **because statements**:

- My headache went away, because I took an aspirin.
- He is late, because he overslept.
- Hillary Clinton lost the election, because she is a woman.

Causal statements should be based on:

- the outcome of some **action** (e.g.,intervention, manipulation, treatment), applied to a unit at a particular point in time
- **relative to** the outcome of **another action**

Defining Causal Effects: Because Statements vs. Causal Statements

Examples of **because statements**:

- My headache went away, because I took an aspirin.
- He is late, because he overslept.
- Hillary Clinton lost the election, because she is a woman.

Causal statements should be based on:

- the outcome of some **action** (e.g.,intervention, manipulation, treatment), applied to a unit at a particular point in time
- **relative to** the outcome of **another action**

When the alternative action is not well described, the causal question is not well defined.

Defining Causal Effects: Potential Outcomes

Suppose Don has a headache. A well-defined causal question requires us to define:

- **treatment levels** (aka exposure): Aspirin ($X = 1$) and No Aspirin ($X = 0$)
- **outcome**: Headache under both actions one hour later ($Y^{X=1}$ and $Y^{X=0}$)

Defining Causal Effects: Potential Outcomes

Suppose Don has a headache. A well-defined causal question requires us to define:

- **treatment levels** (aka exposure): Aspirin ($X = 1$) and No Aspirin ($X = 0$)
- **outcome**: Headache under both actions one hour later ($Y^{X=1}$ and $Y^{X=0}$)

Suppose Don has these **two potential outcomes**:

- Potential outcome $Y^{X=1} = 0$ (i.e., no headache 1h after aspirin)
- Potential outcome $Y^{X=0} = 1$ (i.e., headache 1h after no aspirin)
- Causal effect: $Y^1 - Y^0 = 0 - 1$: reduction (i.e., improvement) due to Aspirin

Defining Causal Effects: Potential Outcomes

Suppose Don has a headache. A well-defined causal question requires us to define:

- **treatment levels** (aka exposure): Aspirin ($X = 1$) and No Aspirin ($X = 0$)
- **outcome**: Headache under both actions one hour later ($Y^{X=1}$ and $Y^{X=0}$)

Suppose Don has these **two potential outcomes**:

- Potential outcome $Y^{X=1} = 0$ (i.e., no headache 1h after aspirin)
- Potential outcome $Y^{X=0} = 1$ (i.e., headache 1h after no aspirin)
- Causal effect: $Y^1 - Y^0 = 0 - 1$: reduction (i.e., improvement) due to Aspirin

Either we give Don the aspirin, **or** we don't:

- the potential outcome we observe is the **fact**
- the potential outcome we do NOT observe is the **counterfact**

Defining Causal Effects: Potential Outcomes

Two key features to notice here are:

- the causal effect is defined at the **level of the unit**
- we can only **observe one potential outcome per unit**

Defining Causal Effects: Potential Outcomes

Two key features to notice here are:

- the causal effect is defined at the **level of the unit**
- we can only **observe one potential outcome per unit**

The latter is known as: "**The fundamental problem of causal inference**" (p.947, Holland, 1986).

Defining Causal Effects: Potential Outcomes

Two key features to notice here are:

- the causal effect is defined at the **level of the unit**
- we can only **observe one potential outcome per unit**

The latter is known as: "**The fundamental problem of causal inference**" (p.947, Holland, 1986).

Note: measuring the **same person at different occasions** (under different treatments), is NOT necessarily the solution; that **requires additional assumptions** (cf. Holland, 1986):

- temporal stability: effect of treatment does not depend on time
- causal transience: there is no lingering effect from the earlier treatment

Defining Causal Effects: Individual vs. Average Causal Effect

Causality is defined as the **difference in potential outcomes of an individual**:

Individual causal effect:

$$ICE_i = Y_i^1 - Y_i^0$$

Defining Causal Effects: Individual vs. Average Causal Effect

Causality is defined as the **difference in potential outcomes of an individual**:

Individual causal effect:

$$ICE_i = Y_i^1 - Y_i^0$$

The ICE **may differ** for different individuals.

Defining Causal Effects: Individual vs. Average Causal Effect

Causality is defined as the **difference in potential outcomes of an individual**:

Individual causal effect:

$$ICE_i = Y_i^1 - Y_i^0$$

The ICE **may differ** for different individuals. We can't observe both potential outcomes, so we often focus on the **average causal effect** instead:

Average causal effect:

$$ACE = E[Y_i^1 - Y_i^0] = E[Y_i^1] - E[Y_i^0]$$

Defining Causal Effects: Individual vs. Average Causal Effect

Causality is defined as the **difference in potential outcomes of an individual**:

Individual causal effect:

$$ICE_i = Y_i^1 - Y_i^0$$

The ICE **may differ** for different individuals. We can't observe both potential outcomes, so we often focus on the **average causal effect** instead:

Average causal effect:

$$ACE = E[Y_i^1 - Y_i^0] = E[Y_i^1] - E[Y_i^0]$$

(*Also referred to as individual treatment effect (ITE) and average treatment effect (ATE).)

Example: Individual vs. Average Causal Effect

Treatment: aspirin ($X = 1$) or no aspirin ($X = 0$)
Outcome: headache ($Y = 1$) or no headache ($Y = 0$)

Example: Individual vs. Average Causal Effect

Treatment: aspirin ($X = 1$) or no aspirin ($X = 0$)
Outcome: headache ($Y = 1$) or no headache ($Y = 0$)

	Potential outcomes		ICE
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Charles	1	1	0
Susan	0	1	-1
Tracy	1	1	0
Ken	1	1	0
Pete	0	0	0
Helen	0	1	-1
Kate	0	0	0
George	0	1	-1

Example: Individual vs. Average Causal Effect

Treatment: aspirin ($X = 1$) or no aspirin ($X = 0$)
Outcome: headache ($Y = 1$) or no headache ($Y = 0$)

	Potential outcomes		ICE
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Charles	1	1	0
Susan	0	1	-1
Tracy	1	1	0
Ken	1	1	0
Pete	0	0	0
Helen	0	1	-1
Kate	0	0	0
George	0	1	-1

ACE $= E[Y^1] - E[Y^0]$

Example: Individual vs. Average Causal Effect

Treatment: aspirin ($X = 1$) or no aspirin ($X = 0$)
Outcome: headache ($Y = 1$) or no headache ($Y = 0$)

	Potential outcomes		ICE
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Charles	1	1	0
Susan	0	1	-1
Tracy	1	1	0
Ken	1	1	0
Pete	0	0	0
Helen	0	1	-1
Kate	0	0	0
George	0	1	-1

$\text{ACE} = E[Y^1] - E[Y^0] \quad = \frac{3}{8} - \frac{6}{8} = -0.375$

Example: Naive Estimate Based on Observational Data

We only observe one of each individual's potential outcome!
Note: Observational, not experimental data...

	Unobserved			Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Charles	**1**	1	0	1	1
Susan	**0**	1	-1	1	0
Tracy	**1**	1	0	1	1
Ken	**1**	1	-1	1	1
Pete	0	**0**	0	0	0
Helen	0	**1**	-1	0	1
Kate	0	**0**	0	0	0
George	0	**1**	-1	0	1

Example: Naive Estimate Based on Observational Data

We only observe one of each individual's potential outcome!
Note: Observational, not experimental data...

	Unobserved			Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Charles	**1**	1	0	1	1
Susan	**0**	1	-1	1	0
Tracy	**1**	1	0	1	1
Ken	**1**	1	-1	1	1
Pete	0	**0**	0	0	0
Helen	0	**1**	-1	0	1
Kate	0	**0**	0	0	0
George	0	**1**	-1	0	1

Probability of headache after aspirin: $E[Y|X=1] = \frac{1+0+1+1}{4} = 0.75$

Example: Naive Estimate Based on Observational Data

We only observe one of each individual's potential outcome!
Note: Observational, not experimental data...

	Unobserved			Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Charles	**1**	1	0	1	1
Susan	**0**	1	-1	1	0
Tracy	**1**	1	0	1	1
Ken	**1**	1	-1	1	1
Pete	0	**0**	0	0	0
Helen	0	**1**	-1	0	1
Kate	0	**0**	0	0	0
George	0	**1**	-1	0	1

Probability of headache after aspirin: $E[Y|X=1] = \frac{1+0+1+1}{4} = 0.75$
Probability of headache after no aspirin: $E[Y|X=0] = \frac{0+1+0+1}{4} = 0.5$

Example: Naive Estimate Based on Observational Data

We only observe one of each individual's potential outcome!
Note: Observational, not experimental data...

	Unobserved			Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Charles	**1**	1	0	1	1
Susan	**0**	1	-1	1	0
Tracy	**1**	1	0	1	1
Ken	**1**	1	-1	1	1
Pete	0	**0**	0	0	0
Helen	0	**1**	-1	0	1
Kate	0	**0**	0	0	0
George	0	**1**	-1	0	1

Probability of headache after aspirin: $E[Y|X=1] = \frac{1+0+1+1}{4} = 0.75$
Probability of headache after no aspirin: $E[Y|X=0] = \frac{0+1+0+1}{4} = 0.5$

Hence: $E[Y|X=1] - E[Y|X=0] = 0.75 - 0.5 = 0.25$

Example: Naive Estimate Based on Observational Data

We only observe one of each individual's potential outcome!
Note: Observational, not experimental data...

	Unobserved			Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Charles	**1**	1	0	1	1
Susan	**0**	1	-1	1	0
Tracy	**1**	1	0	1	1
Ken	**1**	1	-1	1	1
Pete	0	**0**	0	0	0
Helen	0	**1**	-1	0	1
Kate	0	**0**	0	0	0
George	0	**1**	-1	0	1

Probability of headache after aspirin: $E[Y|X=1] = \frac{1+0+1+1}{4} = 0.75$
Probability of headache after no aspirin: $E[Y|X=0] = \frac{0+1+0+1}{4} = 0.5$

Hence: $E[Y|X=1] - E[Y|X=0] = 0.75 - 0.5 = 0.25$

Naive conclusion: Aspirin increases one's chances of still having a headache 1 hour later.

Observing vs. Intervening

Observing $\neq$ intervening

$E(Y|X=1) - E(Y|X=0)$ is **not the same** as $E(Y^1) - E(Y^0)$

Observing that $E(Y|X=1) \neq E(Y|X=0)$
does **not imply a causal effect** of X on Y.

Fancy way of saying correlation $=/=$ causation
there may be confounding, or colliders we conditioned on unknowingly!

Observing vs. Intervening

Observing $\neq$ intervening

$E(Y|X=1) - E(Y|X=0)$ is **not the same** as $E(Y^1) - E(Y^0)$

Observing that $E(Y|X=1) \neq E(Y|X=0)$
does **not imply a causal effect** of X on Y.

Fancy way of saying correlation $=/=$ causation
there may be confounding, or colliders we conditioned on unknowingly!

Pearl calls observing vs intervening the difference between **seeing** and **doing** (more next week).

Recap

Causality is defined as the **difference in potential outcomes of an individual**:

Individual causal effect:

$$ICE_i = Y_i^1 - Y_i^0$$

As we cannot observe both potential outcomes, we typically focus on the **average causal effect** instead:

Average causal effect:

$$ACE = E[Y_i^1 - Y_i^0] = E[Y_i^1] - E[Y_i^0]$$

There is a **difference between intervening and observing**:

Observing $\neq$ intervening

$E(Y|X = 1) - E(Y|X = 0)$ is **not necessarily the same** as $E(Y^1) - E(Y^0)$

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)
4. Formalize the **potential outcomes**, one for each level of treatment

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)
4. Formalize the **potential outcomes**, one for each level of treatment
5. Specify the causal effect in terms of a parameter to estimate: the **estimand** (e.g., the difference in means between potential outcome distributions)

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)
4. Formalize the **potential outcomes**, one for each level of treatment
5. Specify the causal effect in terms of a parameter to estimate: the **estimand** (e.g., the difference in means between potential outcome distributions)
6. State the **assumptions** needed to estimate the causal effect (e.g., no unmeasured confounding)

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)
4. Formalize the **potential outcomes**, one for each level of treatment
5. Specify the causal effect in terms of a parameter to estimate: the **estimand** (e.g., the difference in means between potential outcome distributions)
6. State the **assumptions** needed to estimate the causal effect (e.g., no unmeasured confounding)
7. **Estimate** the causal effect (i.e., choose a particular technique, such as regression, matching, weighting, etc.)

Causal Identification Steps

c.f., Goetghebeur et al., 2020:

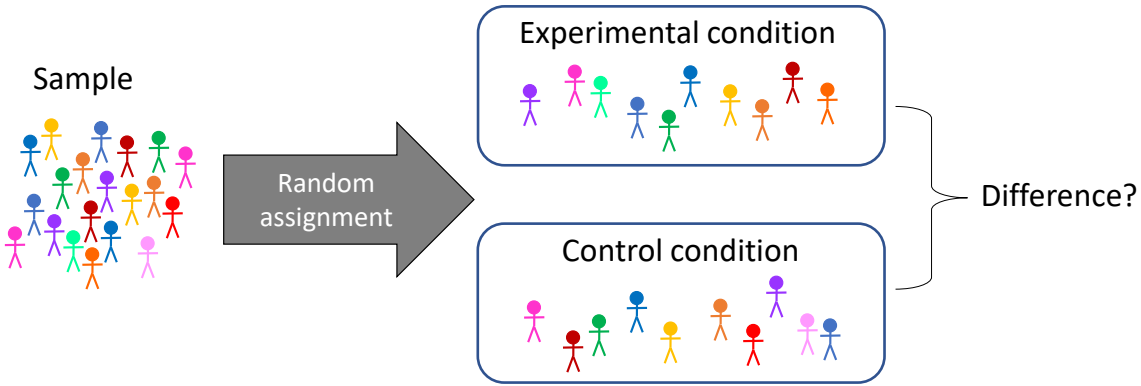
1. Define **exposure**, and two levels of interest (e.g., aspirin vs. no aspirin)
2. Define **outcome variable** (e.g., headache 1 hour later)
3. Define **population** of interest (e.g., people who have a headache)
4. Formalize the **potential outcomes**, one for each level of treatment
5. Specify the causal effect in terms of a parameter to estimate: the **estimand** (e.g., the difference in means between potential outcome distributions)
6. State the **assumptions** needed to estimate the causal effect (e.g., no unmeasured confounding)
7. **Estimate** the causal effect (i.e., choose a particular technique, such as regression, matching, weighting, etc.)
8. **Sensitivity analysis** (important, but not covered in this course)

Overview

- Potential Outcomes & Causal Effects
- **Estimating Causal Effects: Assumptions**
- Estimating Causal Effects: Controlling for Confounders

Rubin's Potential Outcome Framework: Assumptions for identification of a causal effect based on observational data

RCTs: The gold standard for estimating ACEs



Essentially, we want to mimic the situation we have in an RCT with random assignment e.g., every unit should be equally likely to be in the treatment or control group.

From Observation to Causation Assumption 1: Exchangeability

Exchangeability:

For each unit Treatment is independent of their potential outcomes: $Y_i^1, Y_i^0 \perp\!\!\!\perp X_i$

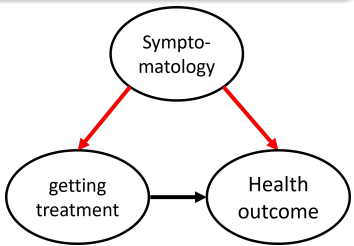
From Observation to Causation Assumption 1: Exchangeability

Exchangeability:

For each unit Treatment is independent of their potential outcomes: $Y_i^1, Y_i^0 \perp\!\!\!\perp X_i$

Violation of Exchangeability:
When there is a relation between people's assignment to treatment and their potential outcomes.

Example:



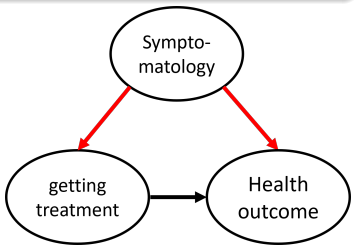
From Observation to Causation Assumption 1: Exchangeability

Exchangeability:

For each unit Treatment is independent of their potential outcomes: $Y_i^1, Y_i^0 \perp\!\!\!\perp X_i$

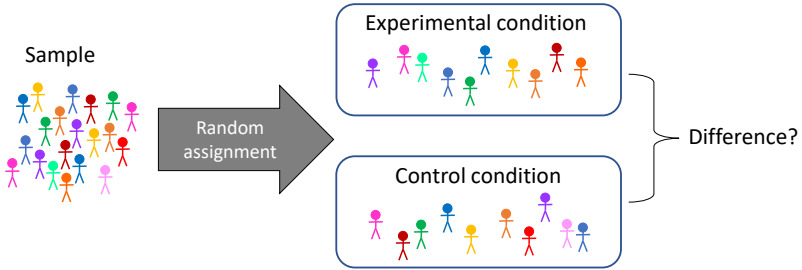
Violation of Exchangeability:
When there is a relation between people's assignment to treatment and their potential outcomes.

Example:



- The same idea of no backdoor paths between X and Y!
- SO, exchangeability = no confounding!

Exchangeability related to RCTs



Due to random assignment, in RCTs individuals receiving treatment ($X = 1$) are **exchangeable** with respect to their potential outcome with those who do not receive treatment ($X = 0$):

$E[Y^1|X = 1] = E[Y^1|X = 0]$ and
$E[Y_i^0|X = 1] = E[Y^0|X = 0]$

Without random assignment (observational data), what treatment people get may depend on third variables that also relate to the potential outcomes...Confounding!

Exchangeability and Missing Data

Only part (at best half) of the potential outcomes is observed.
This makes causal inference a **missing data problem**.

Exchangeability and Missing Data

Only part (at best half) of the potential outcomes is observed.
This makes causal inference a **missing data problem**.

Exchangeability is about: What is the **missing data mechanism** for the potential outcomes?

Exchangeability and Missing Data

Only part (at best half) of the potential outcomes is observed.
This makes causal inference a **missing data problem**.

Exchangeability is about: What is the **missing data mechanism** for the potential outcomes?

- Missing Completely At Random (MCAR)?
- Missing At Random (MAR; random, once accounted for observed covariates)
- Missing Non At Random (MNAR; missing patterns depend on unobserved covariates or the outcome)
- Check out: https://stefvanbuuren.name/fimd/sec-MCAR.html

Exchangeability implies a MCAR assumption for the missing potential outcomes, conditional on the Treatment variable.

Our Headache Example: Exchangeability VIOLATED

Z is pre-treatment headache severity

	Unobserved			Observed		Confounder
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i	Z_i
Charles	**1**	1	0	1	1	9
Susan	**0**	1	-1	1	0	7
Tracy	**1**	1	0	1	1	8
Ken	**1**	1	0	1	1	8
Pete	0	**0**	0	0	0	6
Helen	0	**1**	-1	0	1	3
Kate	0	**0**	0	0	0	4
George	0	**1**	-1	0	1	5

People who took aspirin had **more severe headaches pre-treatment** (Z_i).

Assumption: Conditional Exchangeability

Instead of assuming exchangeability, we may assume **conditional exchangeability**: Conditional on a set of observed covariates, the potential outcomes are independent of treatment assignment.

Conditional Exchangeability:

$Y_i^1, Y_i^0 \perp\!\!\!\perp X_i | Z_i$; given a specific value of Z, the assignment is random.

Assumption: Conditional Exchangeability

Instead of assuming exchangeability, we may assume **conditional exchangeability**: Conditional on a set of observed covariates, the potential outcomes are independent of treatment assignment.

Conditional Exchangeability:

$Y_i^1, Y_i^0 \perp\!\!\!\perp X_i | Z_i$; given a specific value of Z, the assignment is random.

That is, no **unobserved** confounding.

Assumption: Conditional Exchangeability

Instead of assuming exchangeability, we may assume **conditional exchangeability**: Conditional on a set of observed covariates, the potential outcomes are independent of treatment assignment.

Conditional Exchangeability:

$Y_i^1, Y_i^0 \perp\!\!\!\perp X_i | Z_i$; given a specific value of Z, the assignment is random.

That is, no **unobserved** confounding.

This implies *missing at random* (MAR) assumption!
(rather than MCAR, missing completely at random)

From Observation to Causation Assumption 2: Positivity

There must be exposed and unexposed participants at every combination of values of our observed confounders Z in the population under study.

From Observation to Causation Assumption 2: Positivity

There must be exposed and unexposed participants at every combination of values of our observed confounders Z in the population under study.

In an RCT, positivity should be present by design.

From Observation to Causation Assumption 2: Positivity

There must be exposed and unexposed participants at every combination of values of our observed confounders Z in the population under study.

In an RCT, positivity should be present by design.

Violations can be spotted by:

- making tables of each categorical covariate and treatment (should be no empty cells)
- categorize a continuous covariate and make table (but this depends on number and width of categories)
- considering all combinations of covariates (becomes impossible, but then we use propensity scores, tbd later!)

From Observation to Causation Assumption 2: Positivity

There must be exposed and unexposed participants at every combination of values of our observed confounders Z in the population under study.

In an RCT, positivity should be present by design.

Violations can be spotted by:

- making tables of each categorical covariate and treatment (should be no empty cells)
- categorize a continuous covariate and make table (but this depends on number and width of categories)
- considering all combinations of covariates (becomes impossible, but then we use propensity scores, tbd later!)

Positivity and exchangeability combined are also known as **strong ignorability**.

From Observation to Causaion Assumption 3: Consistency

Consistency links observed outcomes Y_i to potential outcomes Y_i^X, through:

- $Y_i = Y_i^1$ for individuals with $X_i = 1$
- $Y_i = Y_i^0$ for individuals with $X_i = 0$

From Observation to Causaion Assumption 3: Consistency

Consistency links observed outcomes Y_i to potential outcomes Y_i^X, through:

- $Y_i = Y_i^1$ for individuals with $X_i = 1$
- $Y_i = Y_i^0$ for individuals with $X_i = 0$

Consistency:

$$Y_i = Y_i^x \text{ for } X_i = x_i$$

In words: the observed outcome Y_i equals the potential outcome for the treatment level that was observed.

From Observation to Causaion Assumption 3: Consistency

Consistency links observed outcomes Y_i to potential outcomes Y_i^X, through:

- $Y_i = Y_i^1$ for individuals with $X_i = 1$
- $Y_i = Y_i^0$ for individuals with $X_i = 0$

Consistency:

$$Y_i = Y_i^x \text{ for } X_i = x_i$$

In words: the observed outcome Y_i equals the potential outcome for the treatment level that was observed.

This assumption requires:

- well defined treatment and no-treatment conditions
- one treatment level implies one specific version of treatment
- no measurement error

Consistency Violation: Different versions of treatment

Published in final edited form as:
Ann Epidemiol. 2016 October ; 26(10): 674–680. doi:10.1016/j.annepidem.2016.08.016.

Does water kill? A call for less casual causal inferences

Miguel A. Hernán[1,2]

If there are multiple ways to raise X from 0 to 1, this means:

- there are multiple treatments (i.e., multiple versions of $X = 1$)
- these may have different causal effects (i.e., multiple versions of Y_i^1 for one person)
- this renders the causal question **ill-defined**

Consistency Violation: Different versions of treatment

Published in final edited form as:
Ann Epidemiol. 2016 October ; 26(10): 674–680. doi:10.1016/j.annepidem.2016.08.016.

Does water kill? A call for less casual causal inferences

Miguel A. Hernán[1,2]

If there are multiple ways to raise X from 0 to 1, this means:

- there are multiple treatments (i.e., multiple versions of $X = 1$)
- these may have different causal effects (i.e., multiple versions of Y_i^1 for one person)
- this renders the causal question **ill-defined**

E.g.: Changing BMI to affect weight - lower BMI achieved via Muscle loss vs Fat loss

Consistency Violation: Different versions of treatment

Published in final edited form as:
Ann Epidemiol. 2016 October ; 26(10): 674–680. doi:10.1016/j.annepidem.2016.08.016.

Does water kill? A call for less casual causal inferences

Miguel A. Hernán[1,2]

If there are multiple ways to raise X from 0 to 1, this means:

- there are multiple treatments (i.e., multiple versions of $X = 1$)
- these may have different causal effects (i.e., multiple versions of Y_i^1 for one person)
- this renders the causal question **ill-defined**

E.g.: Changing BMI to affect weight - lower BMI achieved via Muscle loss vs Fat loss

E.g.: Different therapists providing the treatment, some therapists are better than others.

Consistency Violation: Different versions of treatment

Published in final edited form as:
Ann Epidemiol. 2016 October ; 26(10): 674–680. doi:10.1016/j.annepidem.2016.08.016.

Does water kill? A call for less casual causal inferences

Miguel A. Hernán[1,2]

If there are multiple ways to raise X from 0 to 1, this means:

- there are multiple treatments (i.e., multiple versions of $X = 1$)
- these may have different causal effects (i.e., multiple versions of Y_i^1 for one person)
- this renders the causal question **ill-defined**

E.g.: Changing BMI to affect weight - lower BMI achieved via Muscle loss vs Fat loss

E.g.: Different therapists providing the treatment, some therapists are better than others.

Hernán: Specify a **target trial** with a causal question: A detailed description of the RCT one *would* have done, had there been no ethical/practical limitations.

Podcast about target trial with Hernán: https://casualinfer.libsyn.com/casual-inference-talking-target-trials-with-miguel-hernan-episode-01

SUTVA

Others (including Rubin) use the Stable Unit Treatment Assumption (SUTVA);
it stems from going from the ICE to the ACE.

Stable unit treatment value assumption (SUTVA):

For each unit:

- The potential outcomes do not vary with the treatments assigned to other units (i.e., RCT: no interference)
- there are **no** different versions of each treatment level that lead to different potential outcomes (i.e., the consistency assumption).

Overview

- Potential Outcomes & Causal Effects
- Estimating Causal Effects: Assumptions
- **Estimating Causal Effects: Controlling for Confounders**

Controlling for Confounders

Many different techniques available.

We follow Schafer and Kang (2008) who discuss 9 different techniques, which can be divided into 4 main strategies:

- **No confounders** included: naive approach (Method 1)
- Model the relation between **confounders and Y**: account for confounding by including confounders as covariates (Methods 2 and 3)
- Model the relation between **confounders and X**: adjust the data to what would have been obtained in an RCT (Methods 4, 5, and 6)
- Model the relation between **confounders and X AND Y**: assumptions need to be correct for at least one of the two (Methods 7, 8 , and 9; not this course).

Note: In the lab, you will apply methods 1-6.

1. Naive Estimator: Prima Facie Effect

Simply the observed difference in means between treatment groups.

Prima facie effect

$$PFE = E[Y_i^1|X_i = 1] - E[Y_i^0|X_i = 0]$$

i.e., we assume exchangeability:

- $E[Y_i^1|X_i = 1] = E[Y_i^1|X_i = 0]$
- $E[Y_i^0|X_i = 0] = E[Y_i^0|X_i = 1]$

that is, there is no confounding.

This mean difference could, for example, be estimated using regression analysis with a dummy predictor.

Model the relation between confounders (Z_i) and outcome (Y_i)

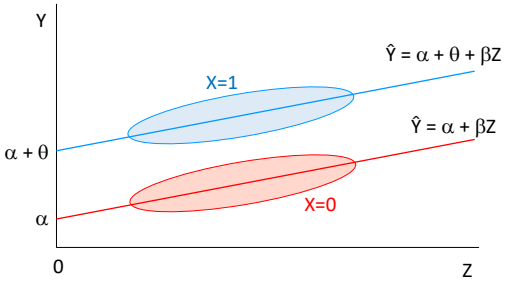
Account for imbalance by including confounders as covariates in a model that relates the outcome and treatment.

2a. ANCOVA (here, linear regression without interactions)

Confounders are added as predictors ('control variables') to a linear regression of the outcome on the treatment variable.

ANCOVA

$$Y_i = \alpha + \theta X_i + \beta_1 Z_{i1} + \cdots + \beta_p Z_{ip} + e_i$$



Here θ is the estimator of the (causal) effect of X.
It is the effect of X when Z are equal to zero.

2a. ANCOVA: Estimator Average Causal Effect

Assumptions:

- conditional exchangeability - no unobserved confounders
- consistency
- correct model specification
- (Also, are confounders and not colliders/mediators; no unobserved conditioning on colliders)

$$
\begin{aligned}
\text{ACE} &= E[Y_i^1 - Y_i^0] \\
&= E[Y_i^1] - E[Y_i^0] \\
&= E[Y_i^1 | X_i = 1, Z_i] - E[Y_i^0 | X_i = 0, Z_i] && \text{No unobserved confounding} \\
&= E[Y_i | X_i = 1, Z_i] - E[Y_i | X_i = 0, Z_i] && \text{Consistency} \\
&= \{\alpha + \theta + Z_i'\beta\} - \{\alpha + Z_i'\beta\} && \text{Linearity, no interactions} \\
&= \theta
\end{aligned}
$$

2b. Regression analysis (linear regression more general than ANCOVA)

For example interactions between covariaties, treatment and covariates, non-linear effects, etc.

2b. Regression analysis (linear regression more general than ANCOVA)

For example interactions between covariaties, treatment and covariates, non-linear effects, etc.

Regression model with treatment*covariates interactions:

$$Y_i = \alpha + \theta X_i + \beta_1 Z_{i1} + \cdots + \beta_p Z_{ip} + \gamma_1 X_i \times Z_{i1} + \cdots + \gamma_p X_i \times Z_{ip} + e_i$$

where

- β's are the slopes for outcome with covariates in the control group ($X_i = 0$)
- γ's are the interaction effects; the differences in slopes between the treatment group ($X_i = 1$) and the control group ($X_i = 0$).

Treatment*Covariate interactions

With treatment*covariates interactions the treatment effect depends on the values of the covariates.

$$\begin{aligned} E\big[Y_i^1 - Y_i^0\big] &= E\big[Y_i^1 | X_i = 1, Z_i\big] - E\big[Y_i^0 | X_i = 0, Z_i\big] = E\big[Y_i | X_i = 1, Z_i\big] - E\big[Y_i | X_i = 0, Z_i\big] \\ &= \big\{\alpha + \theta + Z_i'(\beta + \gamma)\big\} - \big\{\alpha + Z_i'(\beta)\big\} = \theta + Z_i'\gamma \end{aligned}$$

Average causal effect - for the average Z:

$$ACE = \theta + E\big[Z_i\big]'\gamma$$

Treatment*Covariate interactions

With treatment*covariates interactions the treatment effect depends on the values of the covariates.

$$E\left[Y_i^1 - Y_i^0\right] = E\left[Y_i^1 | X_i = 1, Z_i\right] - E\left[Y_i^0 | X_i = 0, Z_i\right] = E\left[Y_i | X_i = 1, Z_i\right] - E\left[Y_i | X_i = 0, Z_i\right]$$
$$= \left\{\alpha + \theta + Z_i'(\beta + \gamma)\right\} - \left\{\alpha + Z_i'(\beta)\right\} = \theta + Z_i'\gamma$$

Average causal effect - for the average Z:

$$ACE = \theta + E\left[Z_i\right]'\gamma$$

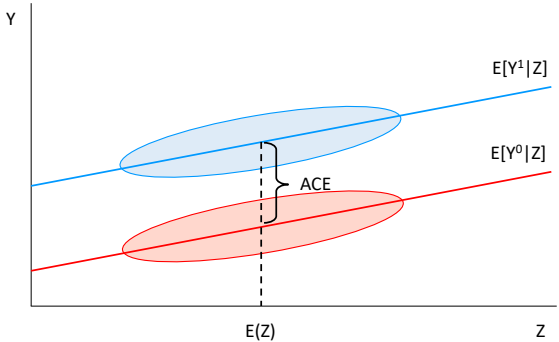
Average causal effect for the treated - for the average Z when X=1

$$ACE_1 = \theta + E\left[Z_i | X_i = 1\right]'\gamma$$

Average causal effect for the controls - for the average Z when X=0:

$$ACE_0 = \theta + E\left[Z_i | X_i = 0\right]'\gamma$$

Case 1: ACE in the ANCOVA model

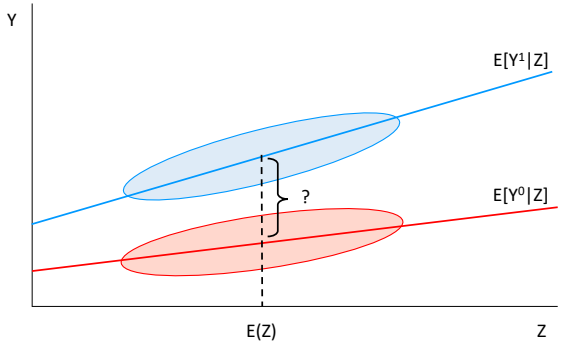


$$ACE = \theta + E\big[Z_i\big]\gamma = \theta$$

This is the classical ANCOVA scenario in which:

- there are no differences on the covariate Z between the treatment groups (akin to expectation under a RCT)
- there is no interaction between treatment X and covariate Z - regression lines are parallel.

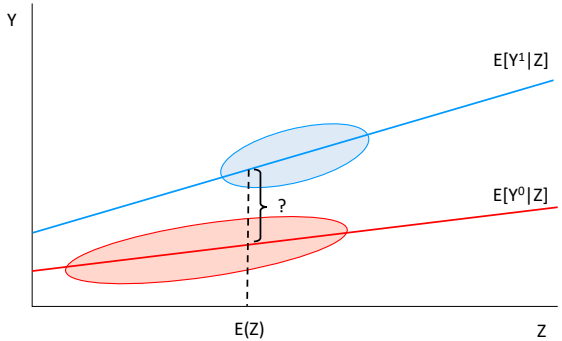
Case 2: What is it?



Explain:

- what kind of scenario this is
- what the quantity denoted by ? represents

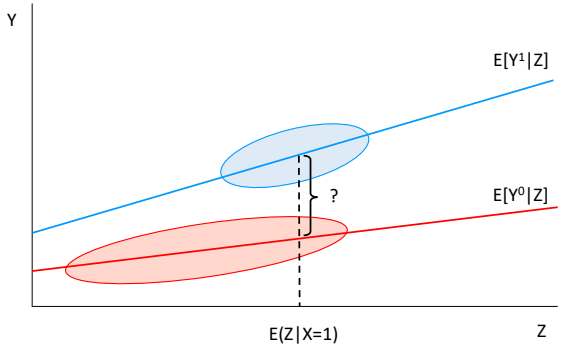
Case 3: What is it?



Explain:

- what kind of scenario this is?
- what does the quantity denoted by ? represent?

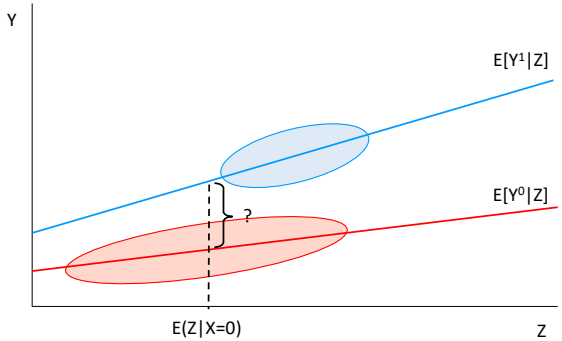
Case 4: What is it?



Explain:

- what kind of scenario this is?
- what does the quantity denoted by ? represent?

Case 5: What is it?



Explain:

- what kind of scenario this is?
- what does the quantity denoted by ? represent?

Method 3. "Regression estimation"

Idea: Estimate unobserved potential outcomes using the confounders

1. get parameter estimates for the confounders for units with $X_i = 1$: $\hat{\beta}^1$
2. Use this estimated model to predict for (the unobserved) potential outcomes when treated: $\hat{Y}_i^1 = Z_i' \hat{\beta}^1$
3. get parameter estimates for the confounders for units with $X_i = 0$: $\hat{\beta}^0$
4. Use this estimated model make prediction for (the unobserved) potential outcome when not treated: $\hat{Y}_i^0 = Z_i' \hat{\beta}^0$

Using these (estimates of) both potential outcomes for every person:

Method 3. "Regression estimation"

Idea: Estimate unobserved potential outcomes using the confounders

1. get parameter estimates for the confounders for units with $X_i = 1$: $\hat{\beta}^1$
2. Use this estimated model to predict for (the unobserved) potential outcomes when treated: $\hat{Y}_i^1 = Z_i'\hat{\beta}^1$
3. get parameter estimates for the confounders for units with $X_i = 0$: $\hat{\beta}^0$
4. Use this estimated model make prediction for (the unobserved) potential outcome when not treated: $\hat{Y}_i^0 = Z_i'\hat{\beta}^0$

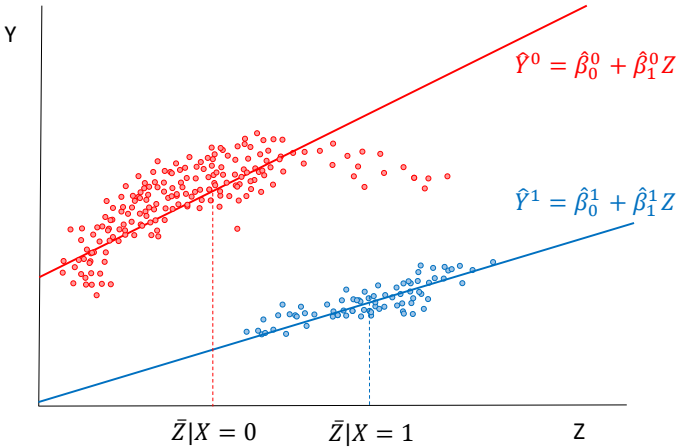
Using these (estimates of) both potential outcomes for every person:

Regression estimate of ACE:

$$\hat{ACE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i^1 - \hat{Y}_i^0)$$

Note: We can either use predicted potential outcomes for only the unobserved potential outcomes, or for all -both observed an unobserved- potential outcomes.

Method 3. Regression Estimation



Important: Positivity Assumption
For example: We need $\hat{\beta}^0$, based on observations around $\bar{Z}|X = 0$ to make predictions around $\bar{Z}|X = 1$.

If positivity is violated, this requires questionable extrapolation.

Model the relation between
confounders (Z_i) and treatment (X_i)

Propensity Scores

All remaining techniques use **propensity scores**: the probability of each unit of being treated.

Propensity scores (assuming conditional exchangeability - no unobserved confounding):

$$\pi_i = P[X_i = 1|Z_i]$$

Goal: Use propensity scores instead of confounder covariates in model in some way or form (depends on exact method).
For this to work, we want $Z_i \perp\!\!\!\perp X_i|\pi_i$.

Propensity Scores

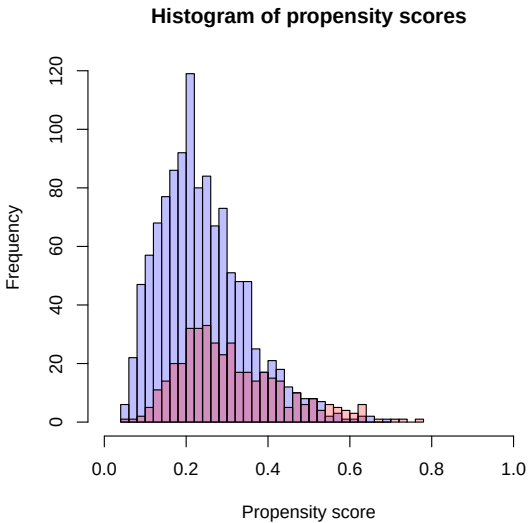
π_i are probabilities, hence often estimated with logistic regression.

Propensity scores (assuming conditional exchangeability - no unobserved confounding):

$$\pi_i = P[X_i = 1|Z_i] = \frac{exp(Z_i'\phi)}{1+exp(Z_i'\phi)}$$

- Outcome variable: Treatment group (0,1)
- Predictors: Identified confounding variables
- Fit logistic regression model and save predicted probabilities per unit.

Show overlap of propensity scores



The distributions of the propensity scores may differ for the treated and the untreated, but should overlap to serve causal inference.

- Non-overlapping areas imply a violation of positivity assumption.
- which can lead to problems for the methods using propensity scores
- e.g., issues with extrapolation.

Method 4. Matching

Matching implies you create pairs that consist of a treated and a non-treated unit, who have **identical covariate scores**.
Hard in case of many covariates: Propensity scores offer a solution.

Background: In an RCT we have: $P\big(Z|X=1\big) = P\big(Z|X=0\big)$

Method 4. Matching

Matching implies you create pairs that consist of a treated and a non-treated unit, who have **identical covariate scores**.
Hard in case of many covariates: Propensity scores offer a solution.

Background: In an RCT we have: $P\big(Z|X=1\big) = P\big(Z|X=0\big)$

Balancing property:

$$P\big(Z|\pi=c, X=1\big) = P\big(Z|\pi=c, X=0\big)$$

If the propensity model is correct, then comparing treated and untreated individuals with the same π is a way of mimicking an RCT - you balance out the covariates among the (new) two groups.

Method 4. Matching

Matching implies you create pairs that consist of a treated and a non-treated unit, who have **identical covariate scores**.
Hard in case of many covariates: Propensity scores offer a solution.

Background: In an RCT we have: $P\big(Z|X=1\big) = P\big(Z|X=0\big)$

Balancing property:

$$P\big(Z|\pi=c, X=1\big) = P\big(Z|\pi=c, X=0\big)$$

If the propensity model is correct, then comparing treated and untreated individuals with the same π is a way of mimicking an RCT - you balance out the covariates among the (new) two groups.

!Note! We match units from the largest group to the characteristics of the units of the smallest group.

Matching provides us not with the ACE, but with **ACE1 or ACE0**

Method 4. Matching

Many important matching techniques.
Important: Evaluating how successful matching was.

	Stratified by DIET		
	0	1	SMD
n	1220	1220	
DISTR.1 (mean (SD))	0.71 (0.45)	0.71 (0.45)	0.007
BLACK (mean (SD))	0.18 (0.38)	0.17 (0.38)	0.004
NBHISP (mean (SD))	0.16 (0.36)	0.15 (0.36)	0.007
GRADE (mean (SD))	9.37 (1.35)	9.37 (1.34)	0.002
SLFHLTH (mean (SD))	2.36 (0.95)	2.35 (0.91)	0.011
SLFWGHT (mean (SD))	3.82 (0.69)	3.84 (0.70)	0.033
WORKHARD (mean (SD))	2.07 (0.86)	2.05 (0.85)	0.022
GOODQUAL (mean (SD))	1.81 (0.65)	1.84 (0.71)	0.049
PHYSFIT (mean (SD))	2.53 (0.97)	2.53 (0.93)	0.007
PROUD (mean (SD))	1.85 (0.77)	1.86 (0.79)	0.011
LIKESLF (mean (SD))	2.46 (1.05)	2.52 (1.06)	0.057
ACCEPTED (mean (SD))	2.33 (1.03)	2.35 (1.06)	0.023
FEELLOVD (mean (SD))	1.92 (0.87)	1.93 (0.90)	0.010

Podcast about matching (and propensity scores, simulations, identifiability assumptions, etc.): https://seriousepi.blubrry.net/2021/03/01/1-16-finding-the-perfect-match-requires-common-support-matching-with-dr-anusha-vable/

Method 5. Inverse probability weighting

The **probability of received treatment** is:

- π_i for those who were **treated** $(X_i = 1)$
- $1 - \pi_i$ for those who were **NOT treated** $(X_i = 0)$

Method 5. Inverse probability weighting

The **probability of received treatment** is:

- π_i for those who were **treated** $(X_i = 1)$
- $1 - \pi_i$ for those who were **NOT treated** $(X_i = 0)$

Among treated individuals $(X = 1)$, those with large π_i are *overrepresented* in comparison to those with small π_i.

Among untreated individuals $(X = 1)$, those with large $1 - \pi_i$ are *overrepresented* in comparison to those with small $1 - \pi_i$.

Method 5. Inverse probability weighting

The **probability of received treatment** is:

- π_i for those who were **treated** $(X_i = 1)$
- $1 - \pi_i$ for those who were **NOT treated** $(X_i = 0)$

Among treated individuals $(X = 1)$, those with large π_i are *overrepresented* in comparison to those with small π_i.

Among untreated individuals $(X = 1)$, those with large $1 - \pi_i$ are *overrepresented* in comparison to those with small $1 - \pi_i$.

To account for this inbalance, we:

- create a pseudo-population
- by weighing each unit by the inverse probability of their received treatment:
- weight $\frac{1}{\hat{\pi}_i}$ for units with $X_i = 1$
- weight $\frac{1}{1-\hat{\pi}_i}$ for units with $X_i = 0$

Method 5. Inverse probability weighting - Pseudo Population

Let's focus on the individuals in our sample who have $\hat{\pi}_i = \frac{2}{3}$.

Received treatment	$X_i = 1$	$X_i = 0$
Observed Y_i's	• • • •	• •
Prob. of received treatment	$\hat{\pi}_i = \frac{2}{3}$	$1 - \hat{\pi}_i = \frac{1}{3}$
Inverse prob. weight	$\frac{1}{\hat{\pi}_i} = \frac{3}{2} = 1\frac{1}{2}$	$\frac{1}{1-\hat{\pi}_i} = \frac{3}{1} = 3$
Pseudo-population	• • • • • •	• • • • • •

Method 5. Inverse probability weighting - Pseudo Population

Let's focus on the individuals in our sample who have $\hat{\pi}_i = \frac{2}{3}$.

Received treatment	$X_i = 1$	$X_i = 0$
Observed Y_i's	● ● ● ●	● ●
Prob. of received treatment	$\hat{\pi}_i = \frac{2}{3}$	$1 - \hat{\pi}_i = \frac{1}{3}$
Inverse prob. weight	$\frac{1}{\hat{\pi}_i} = \frac{3}{2} = 1\frac{1}{2}$	$\frac{1}{1-\hat{\pi}_i} = \frac{3}{1} = 3$
Pseudo-population	● ● ● ● ● ●	● ● ● ● ● ●

Do this for each case (based on their own $1/\hat{\pi}_i$); the resulting pseudo-population:

- is (about) twice as large as the original sample
- makes $\pi_i = 0.5$ for every unit (as in an RCT)
- and thus should have $Z_i \perp\!\!\!\perp X_i$ (as in an RCT)
- (should be checked if succesful)

Check balancing in pseudo-population

To check whether our pseudo-population **mimics an RCT**, we could check standardized mean differences on the covariates:

	Stratified by DIET				
	0		1		SMD
n	6014.23		6368.45		
DISTR.1 (mean (SD))	0.64	(0.43)	0.63	(0.43)	0.028
BLACK (mean (SD))	0.24	(0.43)	0.25	(0.44)	0.035
NBHISP (mean (SD))	0.15	(0.35)	0.13	(0.33)	0.061
GRADE (mean (SD))	9.20	(1.38)	9.26	(1.41)	0.037
SLFHLTH (mean (SD))	2.23	(0.94)	2.25	(0.91)	0.020
SLFWGHT (mean (SD))	3.33	(0.80)	3.15	(1.00)	0.196
WORKHARD (mean (SD))	2.12	(0.90)	2.14	(0.86)	0.017
GOODQUAL (mean (SD))	1.81	(0.67)	1.79	(0.68)	0.036
PHYSFIT (mean (SD))	2.30	(0.95)	2.31	(0.90)	0.014
PROUD (mean (SD))	1.78	(0.77)	1.77	(0.76)	0.018
LIKESLF (mean (SD))	2.18	(1.02)	2.14	(1.00)	0.045
ACCEPTED (mean (SD))	2.18	(1.01)	2.16	(1.02)	0.024
FEELLOVD (mean (SD))	1.81	(0.84)	1.79	(0.83)	0.028

Note: The original sample was 6000 girls in total (i.e., 4780 with $X = 0$ and 1220 with $X = 1$).

IPW estimate of ACE

Computing the ACE using inverse probability weighting by hand:

For individuals with $X_i = 1$:

An estimate of $E\left[Y_i^1\right]$ is $\frac{\sum_i X_i Y_i/\hat{\pi}_i}{\sum_i X_i/\hat{\pi}_i}$

For individuals with $X_i = 0$:

An estimate of $E\left[Y_i^0\right]$ is $\frac{\sum_i (1-X_i) Y_i/(1-\hat{\pi}_i)}{\sum_i (1-X_i)/(1-\hat{\pi}_i)}$

The IPW estimate of the ACE is now:

$$\hat{ACE} = \frac{\sum_i X_i Y_i/\hat{\pi}_i}{\sum_i X_i/\hat{\pi}_i} - \frac{\sum_i (1-X_i) Y_i/(1-\hat{\pi}_i)}{\sum_i (1-X_i)/(1-\hat{\pi}_i)}$$

In practice with R: Compute the weights, and use the package survey (in exercises). Sensitive to outliers.

6. Stratification

Stratification is also referred to as blocking or subclassification. Propensity scores can also be used to:

- create strata (e.g., 5 strata of 20% scores each)
- estimate the ACE in each stratum separately (e.g., with mean difference or regression): $\hat{\theta}_s$
- combine the stratum-specific ACEs in an overall ACE: $\hat{ACE} = \sum_i \frac{N_s}{N} \hat{\theta}_s$

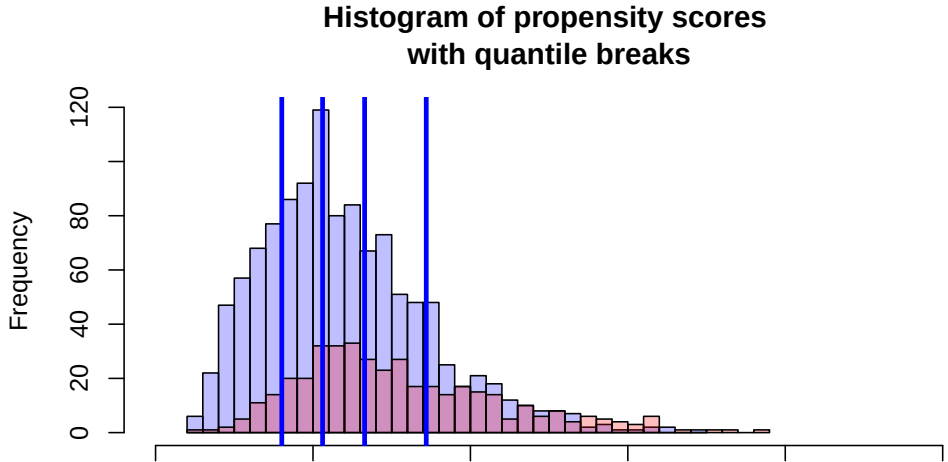
Strata should be **narrow enough**:

- within each stratum covariates do not make a difference; that is, it mimics an RCT
- Again, should be checked how succesful this was

Overlap

To allow for sensible comparisons within each stratum:

- there need to be treated and non-treated individuals in each stratum
- strata should not be too wide (otherwise make more strata)



Controlling for Confounders: Assumptions

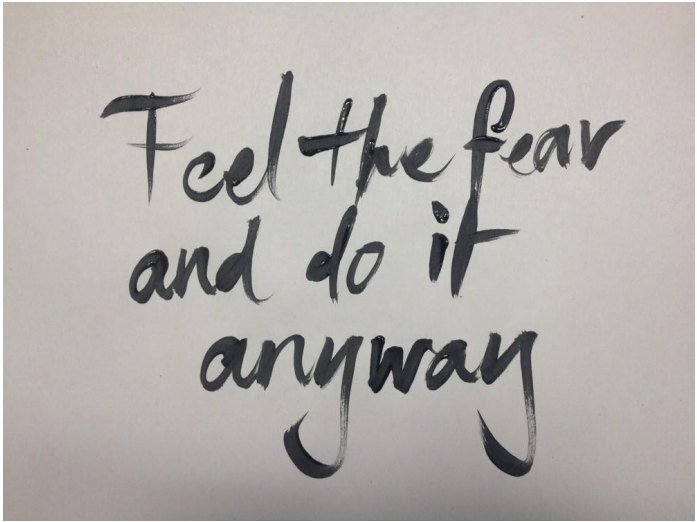
All the methods discussed require the identifiability assumptions: conditional exchangeability (no unobserved confounding), positivity, and consistency (and/or SUTVA).

Additional method-specific assumptions:

- Prima facie effect (method 1) requires the assumption of no confounding at all - exchangeability
- ANCOVA and regression estimation (methods 2 and 3) require correct specification of the outcome model
- matching, IPW, and stratification (methods 4, 5, and 6) require that (the use of) $\hat{\pi}_i$ balances the confounder distribution; it thus requires correct specification of the propensity score model
- dual-modeling strategies (methods 7, 8, and 9; not this course) require correct specification of either outcome model or propensity score model

Of course, it also requires to correctly identify confounders (rather than mediatiors, colliders).
Helpful if possible: Ensuring that the supposed confounders are measured prior to treatment.

Causal Inference: It's hard



Causal Inference Assignment Part I

- Check out the assignment on BB if you haven't yet
- Form a pair if you haven't yet (check out the file on Teams channel "working together")
- Trouble finding a partner? Let me know.

Causal Inference Assignment Part I

- Check out the assignment on BB if you haven't yet
- Form a pair if you haven't yet (check out the file on Teams channel "working together")
- Trouble finding a partner? Let me know.

Next week: Lab meeting 2

- Practicing 6 methods for controlling for confounders...

Next time: Lab

- Practicing 6 methods for controlling for confounders...