27–28. Semester Review

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

29 Nov and 1 Dec 2022

Learning goals for the semester

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

- evaluate the credibility of causal claims
- answer causal questions in your own research
- engage with new methods for causal inference

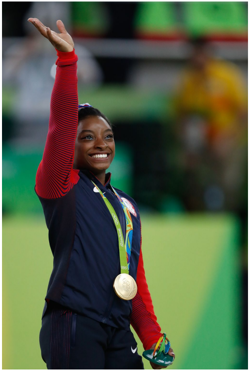
Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal claims require an argument



Left photo: By Fernando Frazão/Agência Brasil - http://agenciabrasil.ebc.com.br/sites/_agenciabrasil2013/files/fotos/1035034-_mg_0802_04.08.16.jpg,CCBY3.0br,https://commons.wikimedia.org/w/index.php?curid=50548410
Right photo: By Agencia Brasil Fotografias - EUA levam ouro na ginástica artística feminina; Brasil fica em 8 lugar, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=50584648

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	?	?
Ian	?	No (0)	?

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)	?
Ian	?	No (0)	?

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	?	No (0)	?

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	No (0)	No (0)	?

Causal claims require an argument

1. Statistical evidence
   - Simone Biles swung on the uneven bars. She won a gold medal.
   - I did not swing on the uneven bars. I did not win a gold medal.
2. Possible causal claim
   - Swinging on the uneven bars causes a person to win a gold medal.

	Do you win gold if you:		Causal effect
	Swing	Do not swing	of swinging
Simone Biles	Yes (1)	No (0)	$+1$
Ian	No (0)	No (0)	0

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Consistency

Consistency

Y_i^1 Y_i^2

Y_i^3 Y_i^4

Potential Outcomes

Consistency

Y_i^1 Y_i^2

Y_i^3 Y_i^4

Potential Outcomes

Y_i

Factual Outcomes

Consistency

Consistency Assumption

$$Y_i^{A_i} = Y_i$$

Y_i^1 Y_i^2

Y_i^3 Y_i^4

Potential Outcomes

Y_i

Factual Outcomes

Consistency

$$Y_i = Y_i^{A_i}$$

The factual outcome Y_i equals the potential outcome $Y_i^{A_i}$ under the factual treatment A_i

Requires

1. Potential outcomes are well-defined
2. Treatments are measured in observed data

Consistency Threat: Interference

Consistency Threat: Interference

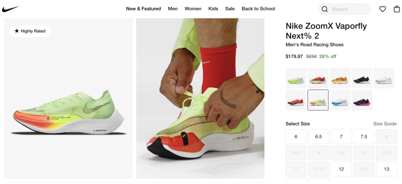


Image source: Nike

Consistency Threat: Interference

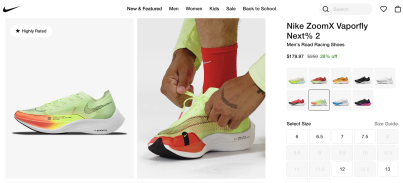


Image source: Nike

Suppose my challenger is a tiny bit faster than me

Consistency Threat: Interference

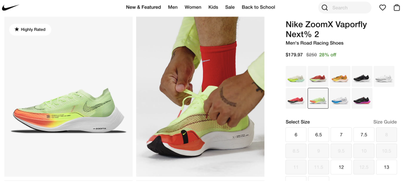


Image source: Nike

Suppose my challenger is a tiny bit faster than me

		I wear springy shoes	
		Yes	No
Challenger wears springy shoes	Yes		
	No		

Consistency Threat: Interference

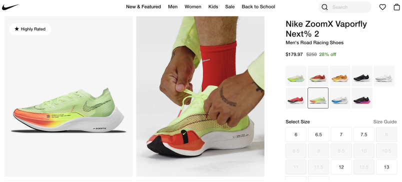


Image source: Nike

Suppose my challenger is a tiny bit faster than me

Challenger wears springy shoes	I wear springy shoes: Yes	I wear springy shoes: No
Yes	**I lose**	**I lose**
No		

Consistency Threat: Interference

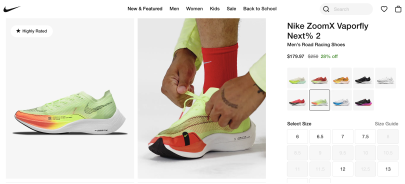


Image source: Nike

Suppose my challenger is a tiny bit faster than me

Challenger wears springy shoes	I wear springy shoes: Yes	I wear springy shoes: No
Yes	**I lose**	**I lose**
No	**I win!**	**I lose**

Consistency Threat: Interference

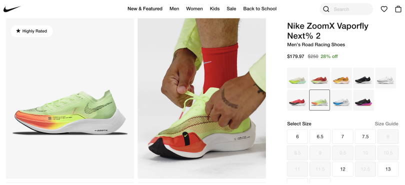


Image source: Nike

Suppose my challenger is a tiny bit faster than me

Challenger wears springy shoes	I wear springy shoes: Yes	I wear springy shoes: No
Yes	I lose	I lose
No	I win!	I lose

Interference:
The outcome of unit i depends on the treatment of unit j

Consistency Threat: Interference

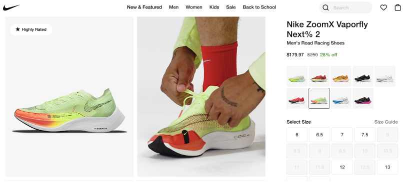


Image source: Nike

Suppose my challenger is a tiny bit faster than me

Challenger wears springy shoes	I wear springy shoes: Yes	I wear springy shoes: No
Yes	**I lose**	**I lose**
No	**I win!**	**I lose**

Interference:
The outcome of unit i depends on the treatment of unit j

$Y_i^{A_i=a}$ is undefined

Need $Y_i^{A_i=a, A_j=a'}$

Consistency Threat: Variations of Treatment

Consistency Threat: Variations of Treatment

- Finding: A Cornell degree causes higher earnings, compared to no college degree

Consistency Threat: Variations of Treatment

- Finding: A Cornell degree causes higher earnings, compared to no college degree
- For-profit college makes an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Consistency Threat: Variations of Treatment

- Finding: A Cornell degree causes higher earnings, compared to no college degree
- For-profit college makes an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Problem: The college you attend is relevant

Consistency Threat: Variations of Treatment

- Finding: A Cornell degree causes higher earnings, compared to no college degree
- For-profit college makes an ad:
  - Pay us lots of money
  - Evidence shows a degree will raise your earnings

Problem: The college you attend is relevant

Formally, you can collapse treatments only if we believe treatment variation irrelevance (VanderWeele 2009)

$$Y_i^{a,k_a} = Y_i^{a,k'_a} \quad \text{for all} \quad k_a, k'_a$$

where

- a is the treatment — college degree
- k_a and k'_a are versions — which college

Consistency. Summary

$$Y_i^{A_i} = Y_i$$

Consistency. Summary

$$Y_i^{A_i} = Y_i$$

Consistency holds when

- potential outcomes are well-defined and
- treatment is measured

Consistency. Summary

$$Y_i^{A_i} = Y_i$$

Consistency holds when

- potential outcomes are well-defined and
- treatment is measured

Consistency would not hold to the degree that

- The outcome of unit i depends on the treatment of unit j
- The treatment A_i is a coarse aggregation over relevant version

Consistency. Summary

$$Y_i^{A_i} = Y_i$$

Consistency holds when

- potential outcomes are well-defined and
- treatment is measured

Consistency would not hold to the degree that

- The outcome of unit i depends on the treatment of unit j
- The treatment A_i is a coarse aggregation over relevant version

Those settings require careful definitions of the potential outcomes

Consistency. Summary

$$Y_i^{A_i} = Y_i$$

Consistency holds when

- potential outcomes are well-defined and
- treatment is measured

Consistency would not hold to the degree that

- The outcome of unit i depends on the treatment of unit j
- The treatment A_i is a coarse aggregation over relevant version

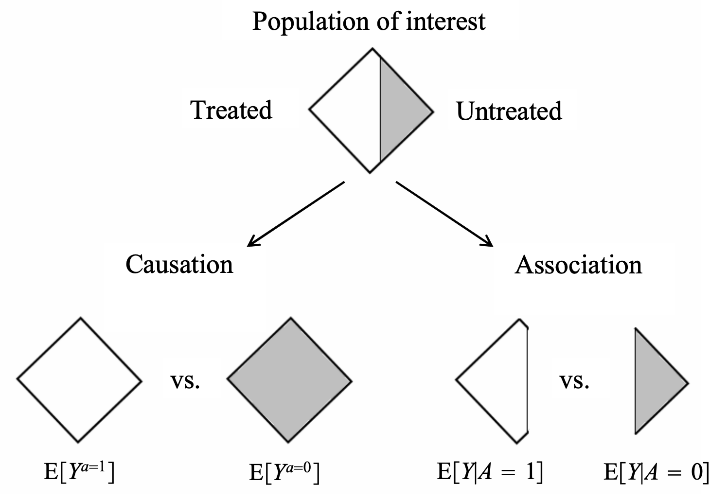
Those settings require careful definitions of the potential outcomes

Important: Consistency is a concern that **precedes** confounding

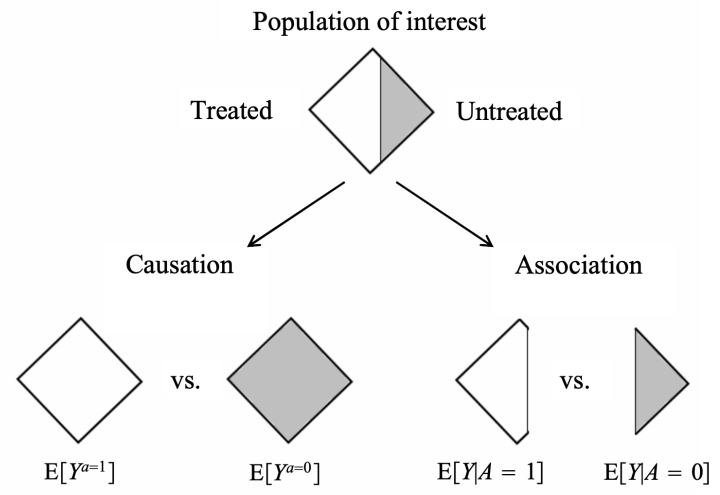
- First define the treatment and outcome
- Then move on to subsequent assumptions

Exchangeability

Exchangeability



Exchangeability



Exchangeability. Under randomization, the left equals the right:

$$E(Y^{a=1} - Y^{a=0}) = E(Y \mid A = 1) - E(Y \mid A = 0)$$

Exchangeability: In math

Exchangeability: $A \perp\!\!\!\perp \{Y^a\}$

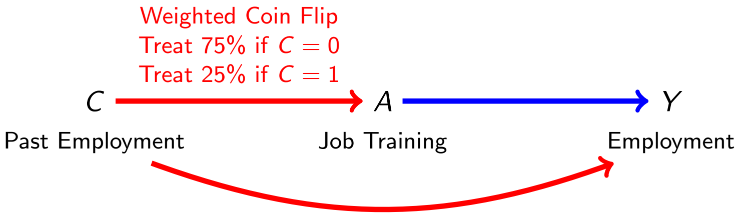
- Treatment is independent of potential outcomes
- Achieved in expectation in a simple randomized experiment

Conditional exchangeability: $A \perp\!\!\!\perp \{Y^a\} \mid \vec{L}$

- Treatment is independent of potential outcomes within strata of $\vec{L}$
- Achieved in expectation in a randomized experiment with unequal assignment probabilities that are a function of $\vec{L}$

Exchangeability: Using DAGs

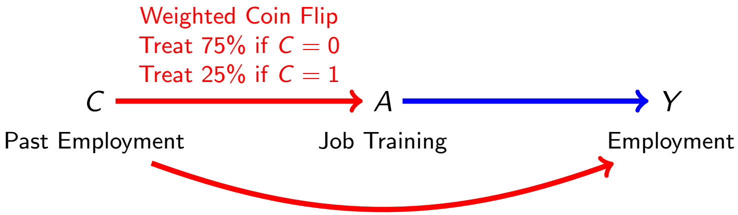
A and Y are associated through two paths



Exchangeability: Using DAGs

A and Y are associated through two paths

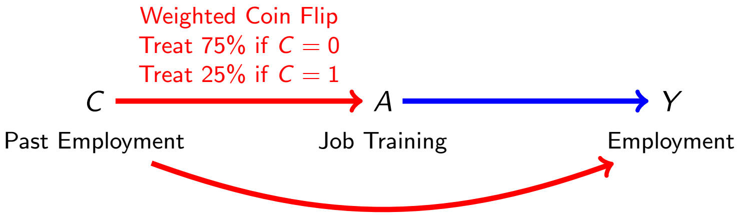
- A **causal path** $A \to Y$



Exchangeability: Using DAGs

A and Y are associated through two paths

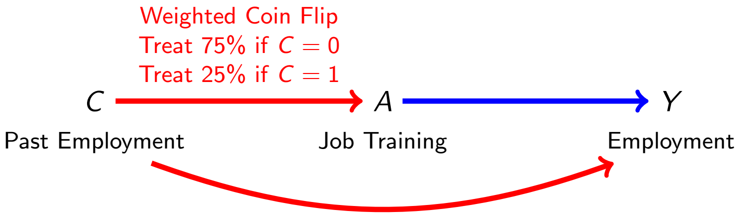
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$



Exchangeability: Using DAGs

A and Y are associated through two paths

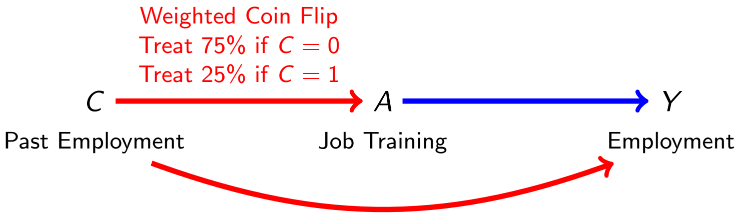
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A



Exchangeability: Using DAGs

A and Y are associated through two paths

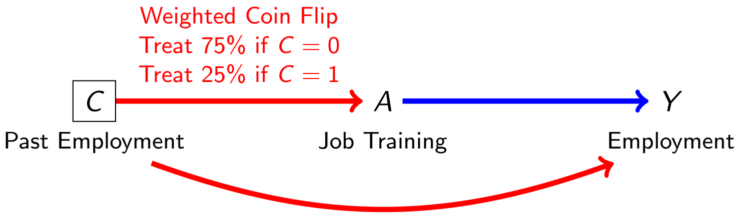
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y



Exchangeability: Using DAGs

A and Y are associated through two paths

- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y



To block the backdoor path, condition on C

- Analyze within strata of C
- Denoted by the box

Exchangeability: Using DAGs

Three sources of association in DAGs

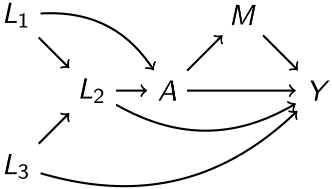
1. Causal path: $A \to Y$, $A \to M \to Y$, etc.
2. Backdoor path: $A \leftarrow L \to Y$
   - Conditioning on L blocks the path
3. Path through an adjusted collider: $A \to \boxed{C} \leftarrow Y$
   - Conditioning on C opens the path

The effect $A \to Y$ is identified if the only source of association between A and Y comes from causal paths

Exchangeability: Using DAGs

What adjustment set identifies the total effect of A on Y?

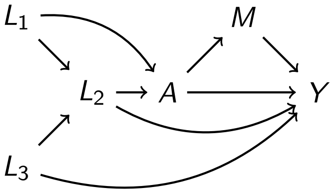
- (3 correct answers)



Exchangeability: Using DAGs

What adjustment set identifies the total effect of A on Y?

- (3 correct answers)

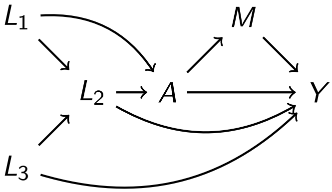


- We have to condition on L_2 to block $A \leftarrow L_2 \rightarrow Y$

Exchangeability: Using DAGs

What adjustment set identifies the total effect of A on Y?

- (3 correct answers)



- We have to condition on L_2 to block $A \leftarrow L_2 \rightarrow Y$
- That opens the path $A \leftarrow L_1 \rightarrow \boxed{L_2} \leftarrow L_3 \rightarrow Y$

Exchangeability: Using DAGs

What adjustment set identifies the total effect of A on Y?

- (3 correct answers)

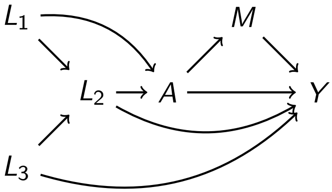
$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & & M \arrow[dr] & \\
& L_2 \arrow[r] \arrow[rrr, bend right] & A \arrow[ur] \arrow[rr] & & Y \\
L_3 \arrow[ur] \arrow[urrrr, bend right] & & & &
\end{tikzcd}
$$

- We have to condition on L_2 to block $A \leftarrow L_2 \rightarrow Y$
- That opens the path $A \leftarrow L_1 \rightarrow \boxed{L_2} \leftarrow L_3 \rightarrow Y$
- We can block that by conditioning on L_1, L_3, or both

Exchangeability: Using DAGs

What adjustment set identifies the total effect of A on Y?

- (3 correct answers)



- We have to condition on L_2 to block $A \leftarrow L_2 \rightarrow Y$
- That opens the path $A \leftarrow L_1 \rightarrow \boxed{L_2} \leftarrow L_3 \rightarrow Y$
- We can block that by conditioning on L_1, L_3, or both

Three sufficient sets: $\{L_1, L_2\}$, $\{L_1, L_3\}$, $\{L_1, L_2, L_3\}$

Positivity

For every value a of the treatment A
and every confounder value $\vec{\ell}$ for the sufficient adjustment set $\vec{L}$,

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$$

Examples where positivity does not hold:

- $A = a$ is employment as a surgeon
  $\vec{L} = \vec{\ell}$ is completion of medical school
- $A = a$ is chewing a tough steak
  $\vec{L} = \vec{\ell}$ is being an infant with no teeth

What to do:
Redefine the causal question to only involve counterfactuals for which positivity holds

Causal inference by adjustment for confounding: Three key assumptions

1. Consistency $Y_i = Y_i^{A_i}$
2. Exchangeability $A \perp\!\!\!\perp \{Y^a\} \mid \vec{L}$
3. Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$

Addressed in that order: each requires the previous

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$\mathrm{E}(Y^a)$$

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$\mathrm{E}(Y^a) = \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}) \qquad \text{rules of probability}$$

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$\begin{aligned}
\mathsf{E}(Y^a) &= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell})\mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}) && \text{rules of probability} \\
&= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell})\mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) && \text{exchangeability}
\end{aligned}$$

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$
\begin{aligned}
\mathrm{E}(Y^a) &= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}) && \text{rules of probability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) && \text{exchangeability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a) && \text{consistency}
\end{aligned}
$$

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$
\begin{aligned}
\mathrm{E}(Y^a) &= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}) && \text{rules of probability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) && \text{exchangeability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a) && \text{consistency} \\
&= \frac{1}{n} \sum_{i} \mathrm{E}(Y \mid \vec{L} = \vec{\ell}_i, A = a)
\end{aligned}
$$

Nonparametric identification: A proof

Linking counterfactual to factual outcomes

$$
\begin{aligned}
\mathrm{E}(Y^a) &= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}) && \text{rules of probability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) && \text{exchangeability} \\
&= \sum_{\vec{\ell}} \mathrm{P}(\vec{L} = \vec{\ell})\mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a) && \text{consistency} \\
&= \frac{1}{n}\sum_{i} \mathrm{E}(Y \mid \vec{L} = \vec{\ell}_i, A = a)
\end{aligned}
$$

Intuition

For each unit i, take the mean outcome
among units who look like them $\vec{L} = \vec{\ell}_i$
but who got the treatment of interest $A = a$.

Positivity ensures that there are some such units

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

Outcome modeling: The parametric g-formula

Outcome modeling: The parametric g-formula

Nonparametric Estimation

$\hat{\text{E}}()$ is the sample
mean in the subgroup

$$\downarrow$$

$$\text{E}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \text{E}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

$$\uparrow$$

Parametric Estimation

$\hat{\text{E}}()$ is a prediction
from a regression model

The parametric g-formula: Simple case

$$
\begin{tikzcd}
L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

Parametric model $\quad \mathsf{E}(Y \mid L, A) = \alpha + \gamma L + \beta A$

The parametric g-formula: Simple case

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

Parametric model $\quad \mathrm{E}(Y \mid L, A) = \alpha + \gamma L + \beta A$

Estimator for the potential outcome under treatment ($A = 1$):

$$\hat{\mathrm{E}}(Y^1) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)$$

Estimator in words:

1. Estimate the regression model
2. Change all treatment values to 1
3. Predict for everyone
4. Take the sample mean

The parametric g-formula: Any estimator

Let $\hat{f}(\vec{\ell}, a)$ be a prediction function

$$\hat{f}(\vec{\ell}, a) \approx \mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a)$$

The parametric g-formula: Any estimator

Let $\hat{f}(\vec{\ell}, a)$ be a prediction function

$$\hat{f}(\vec{\ell}, a) \approx \mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a)$$

We can convert that to an estimator for a post-intervention mean

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}(\vec{\ell}_i, a) \approx \mathrm{E}(Y^a)$$

The parametric g-formula: Any estimator

Let $\hat{f}(\vec{\ell}, a)$ be a prediction function

$$\hat{f}(\vec{\ell}, a) \approx \mathrm{E}(Y \mid \vec{L} = \vec{\ell}, A = a)$$

We can convert that to an estimator for a post-intervention mean

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}(\vec{\ell}_i, a) \approx \mathrm{E}(Y^a)$$

Any prediction function is allowable

Hill, Jennifer L. 2011.
"Bayesian nonparametric modeling for causal inference."
Journal of Computational and Graphical Statistics 20.1:217-240.

- Binary treatment (simulated)
- Continuous confounder X (simulated)

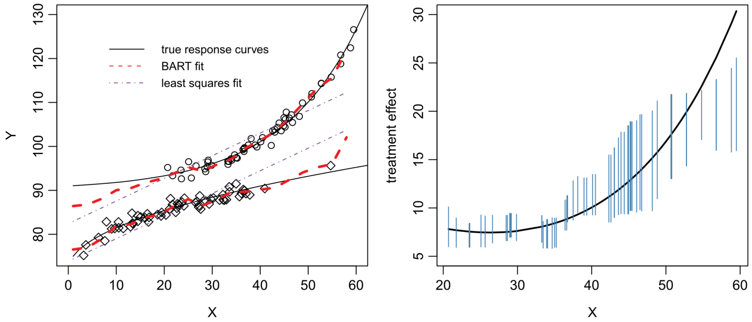


Figure 1. Left panel: simulated data with linear regression and BART fits. Right panel: BART inference for treatment effect on the treated. A color version of this figure is available in the electronic version of this article.

Statistical Science
2019, Vol. 34, No. 1, 43–68
https://doi.org/10.1214/18-STS667

Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition[1]

Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone

Abstract. Statisticians have made great progress in creating methods that reduce our reliance on parametric assumptions. However, this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to at best 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, "Is Your SATT Where It's At?", launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from 30 competitors across the two versions of the competition (black-box algorithms and do-it-yourself analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies and settings that affect performance. The most consistent conclusion was that methods that flexibly model the response surface perform better overall than methods that fail to do so. Finally new methods are proposed that combine features of several of the top-performing submitted methods.

Key words and phrases: Causal inference, competition, machine learning, automated algorithms, evaluation.

1. INTRODUCTION

In the absence of a controlled randomized or natural experiment,[2] inferring causal effects involves the difficult task of constructing fair comparisons between ob-

Vincent Dorie is Associate Research Scientist, Data Science Institute, Columbia University, 475 Riverside Drive, Room 320L, New York, New York 10115, USA (e-mail: vdorie@gmail.com). Jennifer Hill is Professor of Applied Statistics and Data Science, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: jennifer.hill@nyu.edu). Uri Shalit is Assistant Professor, Faculty of Industrial Engineering and Management, Technion, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: urishalit@technion.ac.il). Marc Scott is Professor of Applied Statistics, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: marc.scott@nyu.edu). Dan Cervone is Director of Quantitative Research, Los Angeles Dodgers, Dodger Stadium, 1000 Vin Scully Ave., Los Angeles, California 90012, USA (e-mail: dcervone@gmail.com).

[1]Discussed in 10.1214/18-STS684; 10.1214/18-STS680; 10.1214/18-STS690; 10.1214/18-STS689; 10.1214/18-STS679; 10.1214/18-STS682; 10.1214/18-STS688

[2]We use natural experiment to include (1) studies where the causal variable is randomized not for the purposes of a study (for instance, a school lottery), (2) studies where a variable is randomized but the causal variable of interest is downstream of this (e.g., plays the role of an instrumental variable), and (3) regression discontinuity designs.

The parametric g-formula by matching

$$\vec{L} \longrightarrow A \longrightarrow Y$$

For each i with $A = 1$,

- Find a unit j with $A = 0$ that minimizes

$$d(\vec{\ell}_i, \vec{\ell}_j)$$

  for some distance metric d

- Impute $\text{E}(Y \mid \vec{L} = \vec{\ell}_i, A = 0)$ with the observed Y_j
- Estimate $\hat{\delta}_i = Y_i - Y_j$

Estimate $\hat{\text{E}}(Y^1 - Y^0 \mid A = 1) = \frac{1}{n_1} \sum_{i:A_i=1} \hat{\delta}_i$

Outcome modeling and treatment modeling

Outcome modeling

$$
\begin{tikzcd}
L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

1) Model $\{L, A\} \to Y$
2) Impute unobserved Y^a

Outcome modeling and treatment modeling

Outcome modeling

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

1) Model $\{L, A\} \to Y$
2) Impute unobserved Y^a

Treatment modeling

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

1) Model $L \to A$
2) Create a pseudo-population where $A \perp\!\!\!\perp L$
3) Take means in the pseudo-population

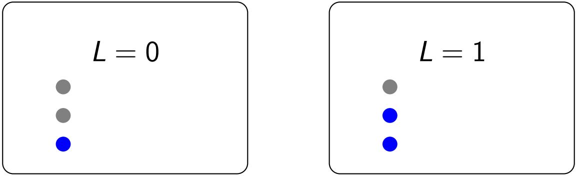
Intuition: Weight to a pseudo-population

Intuition: Weight to a pseudo-population

Factual population:

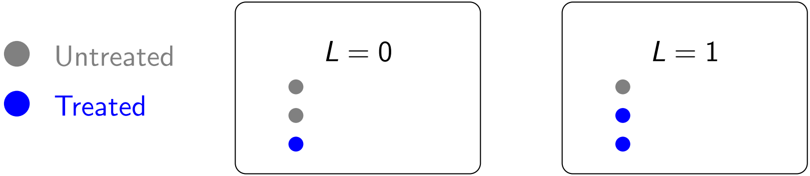
Untreated

Treated

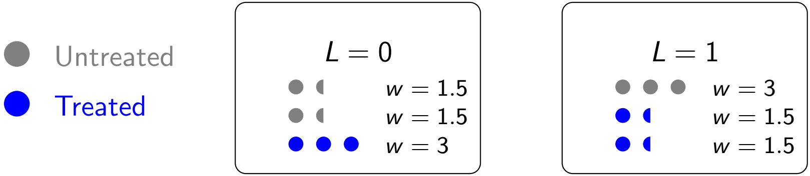


Intuition: Weight to a pseudo-population

Factual population:

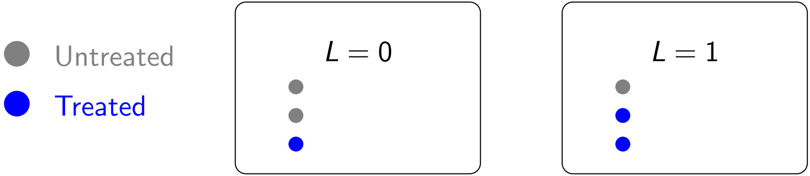


Pseudo-population:

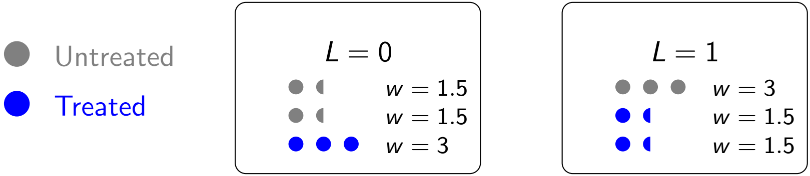


Intuition: Weight to a pseudo-population

Factual population:



Pseudo-population:



Each weight is the inverse probability of the observed treatment:

$$w_i = \frac{1}{\mathrm{P}(A = a_i \mid L = \ell_i)}$$

Intuition: Weight to a pseudo-population

In the factual population

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population

$$\begin{tikzcd} L \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

Marginal structural models

In the factual population

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population

$$\begin{tikzcd} L \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

Marginal structural models

In the factual population

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population

$$\begin{tikzcd} L \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population, estimate

$$\mathrm{E}(Y^a) = g(a)$$

for some parametric function $g()$

Marginal structural models

In the factual population

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population

$$\begin{tikzcd} L \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

In the pseudo-population, estimate

$$E(Y^a) = g(a)$$

for some parametric function $g()$

Example:

$$E(Y^a) = \alpha + \beta a$$

Marginal structural models

Useful when

- Treatment A contains values that are sparsely populated
  - Need to pool information from other treatment values
- You have theory about the shape of the causal response surface $\text{E}(Y^a)$
- You are able to model the treatment assignment rule $\vec{L} \rightarrow A$ to create the weights

Less useful when

- Outcome under $A = a$ is not informative about outcome under $A = a'$
- It is hard to model $\text{P}(A = a \mid \vec{L})$
  - Example: Continuous rather than categorical A

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

We first addressed treatments at one time point

$$\begin{tikzcd} A \arrow[rr] & & Y \\ L \arrow[u] \arrow[urr] & & \end{tikzcd}$$

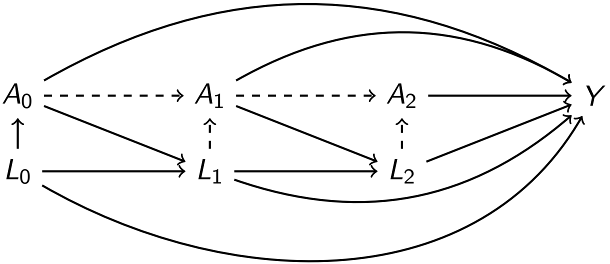
We first addressed treatments at one time point

$$
\begin{tikzcd}
A \arrow[r] & Y \\
L \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

Then, we addressed treatments at many time points

$$
\begin{tikzcd}
A_0 \arrow[r] \arrow[dr] \arrow[rrr, bend left] & A_1 \arrow[r] \arrow[dr] \arrow[rr, bend left] & A_2 \arrow[r] & Y \\
L_0 \arrow[u] \arrow[r] \arrow[urrr, bend right] & L_1 \arrow[u] \arrow[r] \arrow[urr, bend right] & L_2 \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

Treatments at many time points: Estimation strategy



1) Weight to a pseudo-population without the dashed arrows

stabilizing term

$$w_i = \prod_{t=0}^{2} \frac{\text{P}(A_t = A_{ti})}{\text{P}(A_t = A_{ti} \mid \bar{L}_{t,i} \bar{A}_{t-1})}$$

weight for i

product over periods

treatment probability at period t

2) Apply a marginal structural model

$$\text{E}(Y^{\vec{a}}) = g(\vec{a})$$

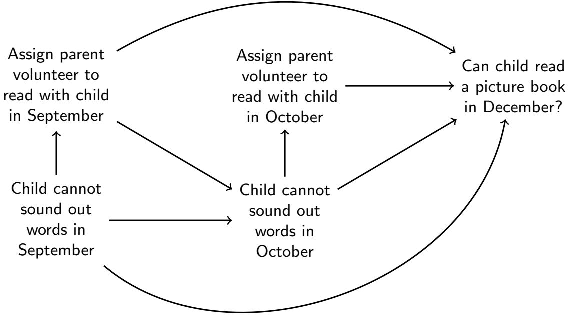
Example:

$$\alpha + \beta\left(\frac{a_0 + a_1 + a_2}{3}\right)$$

Step back: When would I study treatments over time?

- Treatment is assigned at many time points
- as a function of observables up to that point

Our example from class:



Mediation

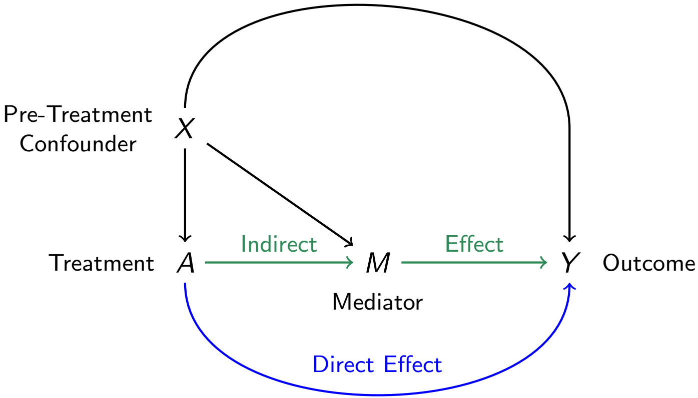
Mediation

Methods to this point have emphasized total effects

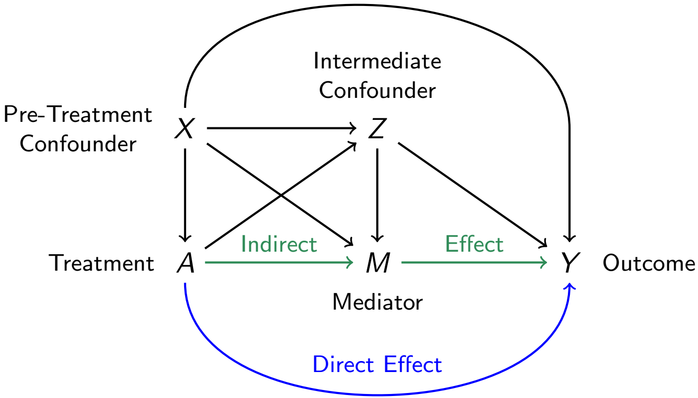
- How big is the effect of A on Y?
- In what subgroups of $\vec{L}$ is that effect large?

Mediation asks questions about how those effects come to be

- By what causal process does A affect Y?



Pre-Treatment
Confounder
X
Treatment
A
Indirect
M
Effect
Y
Outcome
Mediator
Direct Effect



Intermediate
Confounder
Pre-Treatment
Confounder
X
Z
Treatment
A
Indirect
M
Effect
Y
Outcome
Mediator
Direct Effect

Mediation: Controlled direct effects

Test for causal paths that remain after one is blocked by an intervention.

Causal claim

If the school librarian visits a class, that causes kids to read more.

Mediation: Controlled direct effects

Test for causal paths that remain after one is blocked by an intervention.

Causal claim

If the school librarian visits a class, that causes kids to read more.

“How” statement

Happens because they become more likely to visit the school library

Mediation: Controlled direct effects

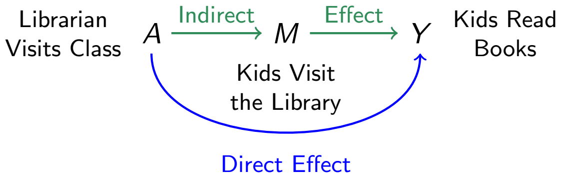
Test for causal paths that remain after one is blocked by an intervention.

Causal claim

If the school librarian visits a class, that causes kids to read more.

"How" statement

Happens because they become more likely to visit the school library



Mediation: Controlled direct effects

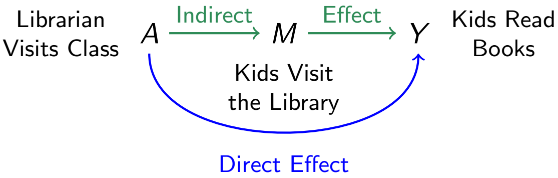
Test for causal paths that remain after one is blocked by an intervention.

Causal claim

If the school librarian visits a class, that causes kids to read more.

"How" statement

Happens because they become more likely to visit the school library



Experimental test (0) Close the library. Does A affect Y?

Experimental test (1) Mandate library visits. Does A affect Y?

Note: Both physically manipulate M

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values.

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values. But what if we could?

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values. But what if we could?

$$\mathrm{E}\left(Y^1 - Y^0\right) \qquad \text{Total effect}$$

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values. But what if we could?

$$
\begin{aligned}
&\text{E}\left(Y^1 - Y^0\right) && \text{Total effect} \\
&= \text{E}\left(Y^{1M^1} - Y^{0M^0}\right) && \text{since } Y^a = Y^{aM^a}
\end{aligned}
$$

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values. But what if we could?

$$
\begin{aligned}
&\mathrm{E}\left(Y^1 - Y^0\right) && \text{Total effect} \\
&= \mathrm{E}\left(Y^{1M^1} - Y^{0M^0}\right) && \text{since } Y^a = Y^{aM^a} \\
&= \mathrm{E}\left(Y^{1M^1} \underbrace{- Y^{1M^0} + Y^{1M^0}}_{\text{Add 0}} - Y^{0M^0}\right)
\end{aligned}
$$

Mediation: Natural direct effects

Decompose a total effect into direct and indirect effects

Each child has a mediator value realized under each treatment

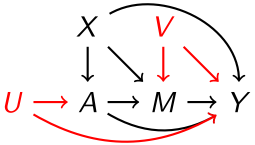
- M_i^0: Would they visit the library if $A = 0$?
- M_i^1: Would they visit the library if $A = 1$?

We cannot observe both mediator values. But what if we could?

$$
\begin{aligned}
&\text{E}\left(Y^1 - Y^0\right) && \text{Total effect} \\
&= \text{E}\left(Y^{1M^1} - Y^{0M^0}\right) && \text{since } Y^a = Y^{aM^a} \\
&= \text{E}\left(Y^{1M^1} \underbrace{-Y^{1M^0} + Y^{1M^0}}_{\text{Add 0}} - Y^{0M^0}\right) \\
&= \underbrace{\text{E}\left(Y^{1M^1} - Y^{1M^0}\right)}_{\substack{\textbf{Indirect Effect} \\ \text{(effect through } M\text{)}}} + \underbrace{\text{E}\left(Y^{1M^0} - Y^{0M^0}\right)}_{\substack{\textbf{Direct Effect} \\ \text{(effect not through } M\text{)}}}
\end{aligned}
$$

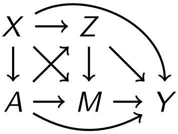
Identification: Controlled and natural direct effects

All mediation estimands require no unmeasured confounding



Unmeasured
U and V would
undermine identification

CDE allows a measured intermediate confounder Z
NDE does not allow this



Examples of mediation questions

Mediation questions examine how an effect arises:
the causal process by which a treatment affects an outcome

- Zhou (2022): The effect of enrolling in college (A) on earnings (Y) only partially operates through BA completion (M)
- Wodtke & Parbst (2017): The effect of neighborhood disadvantage (A) on children's academic achievement (Y) does not operate through school poverty (M)
- Brand, Moore, Song, & Xie (2019): The effect of parental divorce (A) on educational attainment (Y) operates partially through a drop in family income (M)

Causal Inference in Observational Settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

When treatment causes death: Principal stratification

$$\begin{tikzcd}
A \arrow[rr] \arrow[dr] & & Y \\
& S \arrow[ur] & U \arrow[l] \arrow[u]
\end{tikzcd}$$

$\leftarrow$ Y only defined
when $S = 1$

- Suppose you have a new exercise program $A = 1$ vs $A = 0$
- You study the effect on heart rate after 12 months $Y^1 - Y^0$
- But only some of the people survive $S = 1$
  - And survival was higher in your program

When treatment causes death: Principal stratification

Four principal strata

1. Would survive regardless of treatment
   - Our focal group: Effect of exercise on heart rate is well-defined!
2. Would not survive regardless of treatment
3. Would survive if and only if they receive the exercise program
4. Would survive if and only if the do not receive the program

Difficulty: We don't observe the strata.

When treatment causes death: Principal stratification

Assume **monotonicity**:
Exercise never hurts people. It only does nothing or helps.

Survivors in the control condition would have survived regardless

Survivors in the treatment condition are a mix of

- Survive regardless — Proportion π_1
- Survive if and only if treated — Proportion π_3

By estimating $A \rightarrow S$, you can estimate π_1 and π_3
Then, we can set-identify a range of values for

$$\text{E}(Y^1 - Y^0 \mid S = \text{Survive Regardless})$$

When treatment causes death: Principal stratification

The big idea

1. Sometimes your sample is inherently selected (e.g., survival)
2. That needs to go on the DAG
3. If treatment causes selection, then you unavoidably condition on a post-treatment variable
   - Which you know causes bad problems!
4. Principal stratification is the tool to set-identify an effect in a pre-treatment subgroup (e.g., survive regardless) for whom the selection variable is not post-treatment

$$\begin{tikzcd} A \arrow[rr] \arrow[dr] & & Y \\ & S \arrow[ur] & U \arrow[l] \arrow[u] \end{tikzcd}$$

$\leftarrow$ Y only defined when $S = 1$

When variables are measured with error

When variables are measured with error,
the DAG should include the true variable X
and the measured variable X^*

When variables are measured with error

Hernan & Cole (2009) Fig 2

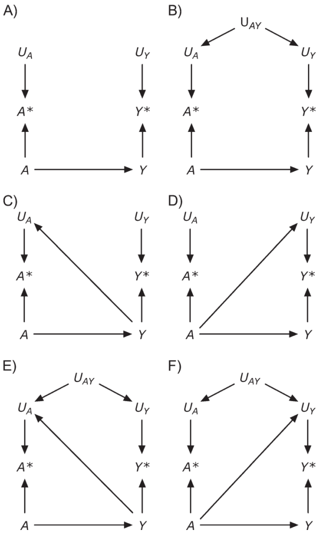


Figure 2. A structural classification of measurement error.

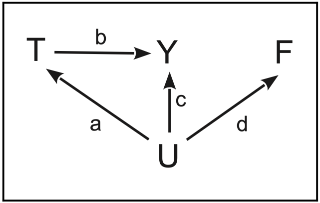
When variables are measured with error

In some cases, you can make strong assumptions
to achieve identification despite measurement error

When variables are measured with error

In some cases, you can make strong assumptions to achieve identification despite measurement error

One example: Elwert & Pfeffer 2022

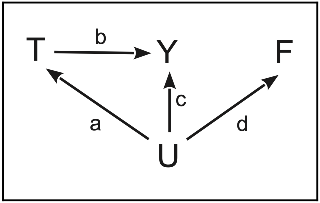


$$E(Y \mid T, F) = \alpha + \beta T + \gamma F$$

When variables are measured with error

In some cases, you can make strong assumptions
to achieve identification despite measurement error

One example: Elwert & Pfeffer 2022



$$\mathrm{E}(Y \mid T, F) = \alpha + \beta T + \gamma F$$

Difference estimator

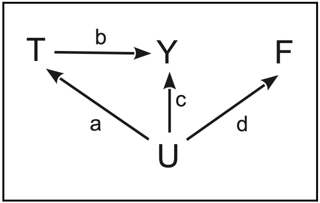
$$\beta - \gamma = b + \underbrace{ac}_{\substack{\text{Bias}\\\text{without}\\\text{control}}} \underbrace{\frac{a-d}{a-a^2 d}}_{\substack{\text{Bias}\\\text{Multiplier}\\|M|}}$$

Eliminates bias in special case

When variables are measured with error

In some cases, you can make strong assumptions to achieve identification despite measurement error

One example: Elwert & Pfeffer 2022



$$E(Y \mid T, F) = \alpha + \beta T + \gamma F$$

Difference estimator

$$\beta - \gamma = b + \underbrace{ac}_{\substack{\text{Bias}\\\text{without}\\\text{control}}} \underbrace{\frac{a-d}{a-a^2d}}_{\substack{\text{Bias}\\\text{Multiplier}\\|M|}}$$

Eliminates bias in special case

Control estimator

$$\beta = b + \underbrace{ac}_{\substack{\text{Bias}\\\text{without}\\\text{control}}} \underbrace{\frac{1-d^2}{1-a^2d^2}}_{\substack{\text{Bias}\\\text{Multiplier}\\|M|<1}}$$

Always reduces bias

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

When exchangeability does not hold: Bounds

You will meet critics who do not believe exchangeability.
You will want to show them bounds!

Setting:

- A binary treatment
- Y bounded outcome in range $y_{\text{Min}} \leq Y \leq y_{\text{Max}}$

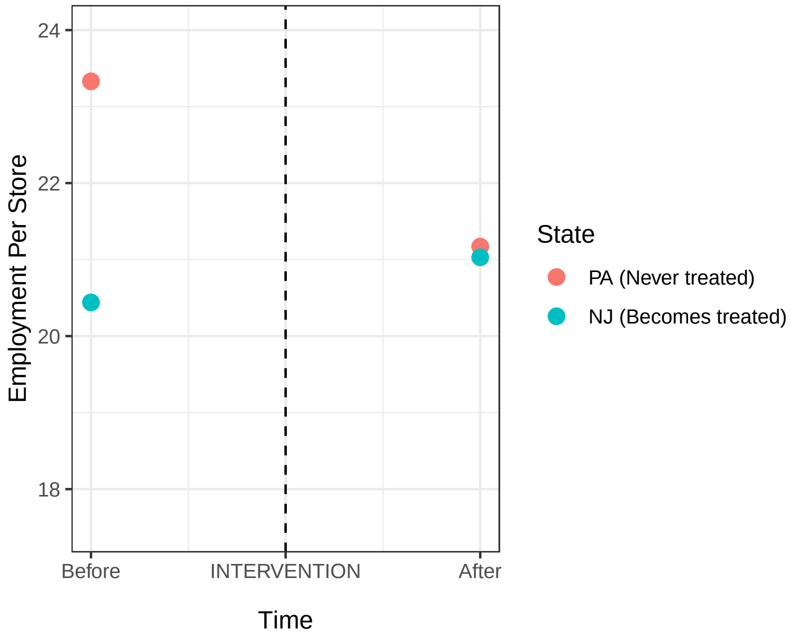
To simplify notation,

- Let $\hat{\bar{y}}_a = \hat{\text{E}}(Y \mid A = a)$ be the sample mean of Y among those with $A = a$
- Let $\hat{\pi}_a = \hat{\text{P}}(A = a)$ be the sample proportion with $A = a$

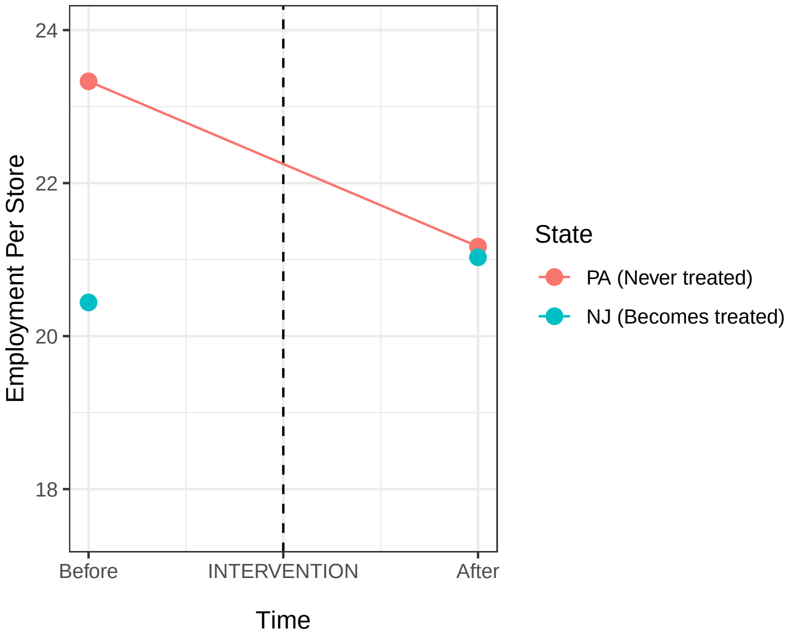
$$\hat{\text{E}}^{\text{LowerBound}}(Y^a) = \hat{\pi}_a \hat{\bar{y}}_a + (1 - \hat{\pi}_a) y_{\text{Min}}$$

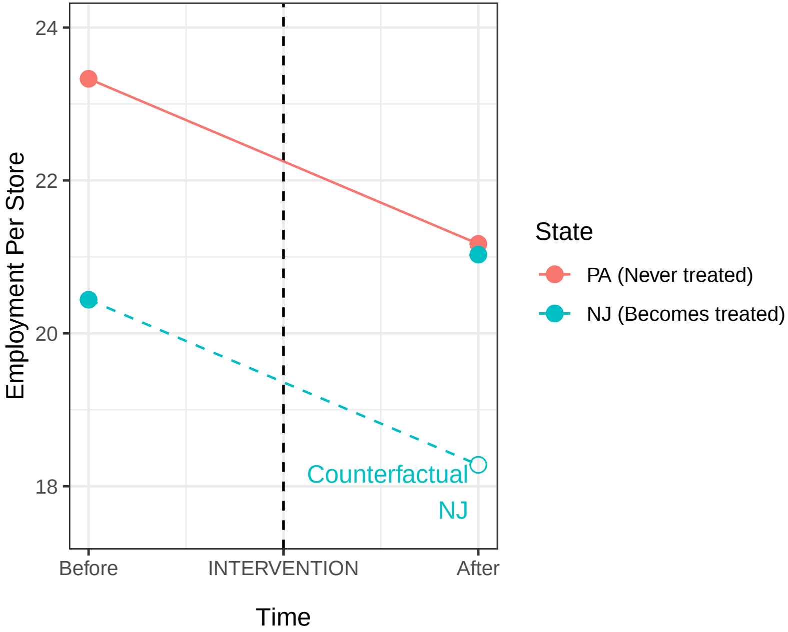
$$\hat{\text{E}}^{\text{UpperBound}}(Y^a) = \hat{\pi}_a \hat{\bar{y}}_a + (1 - \hat{\pi}_a) y_{\text{Max}}$$

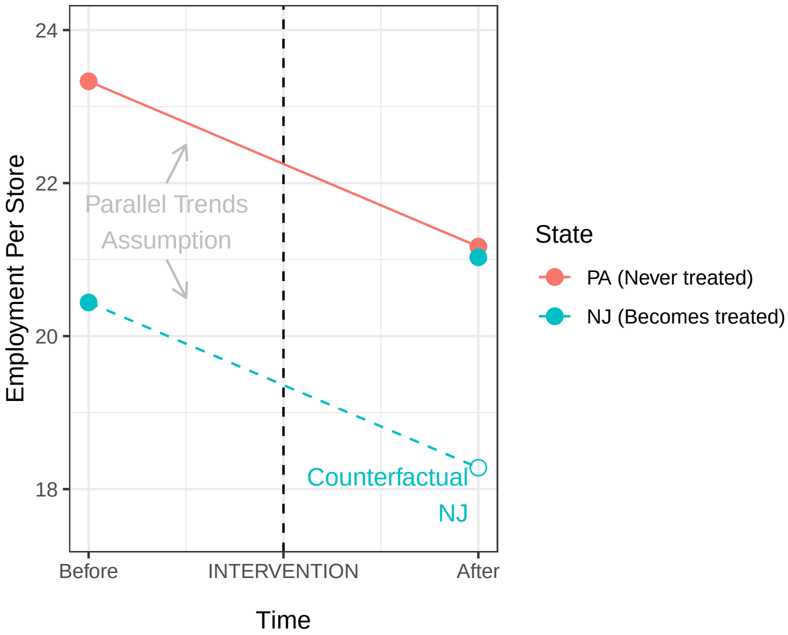
Difference in difference



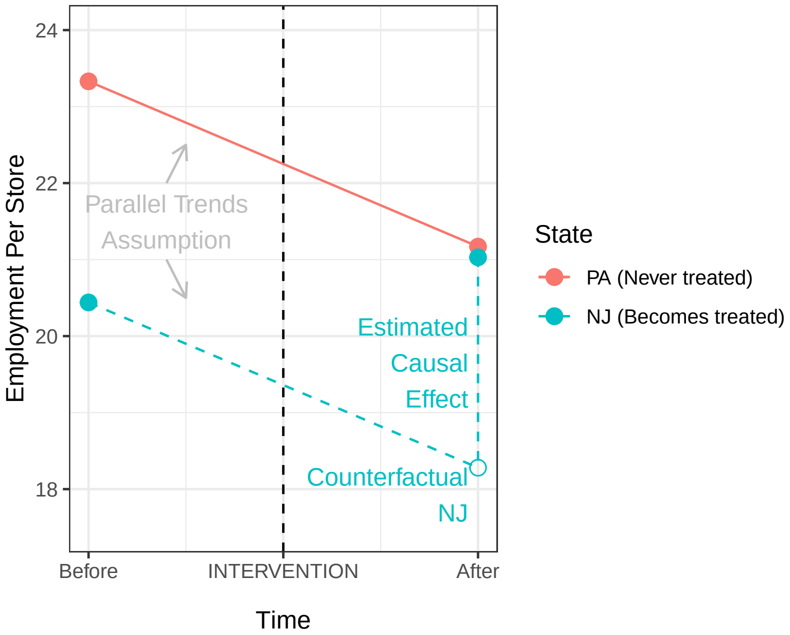
24
22
20
18
Employment Per Store
Before
INTERVENTION
After
Time
State
PA (Never treated)
NJ (Becomes treated)


Employment Per Store
24
22
20
18
Before
INTERVENTION
After
Time
State
PA (Never treated)
NJ (Becomes treated)


Employment Per Store
24
22
20
18
Counterfactual
NJ
Before
INTERVENTION
After
Time
State
PA (Never treated)
NJ (Becomes treated)



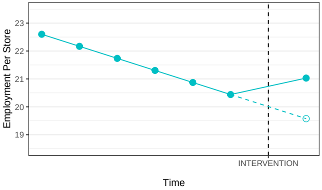
Employment Per Store
24
22
20
18
Parallel Trends
Assumption
Counterfactual
NJ
Before
INTERVENTION
After
Time
State
PA (Never treated)
NJ (Becomes treated)



24
22
20
18
Employment Per Store
Parallel Trends
Assumption
Estimated
Causal
Effect
Counterfactual
NJ
Before
INTERVENTION
After
Time
State
PA (Never treated)
NJ (Becomes treated)

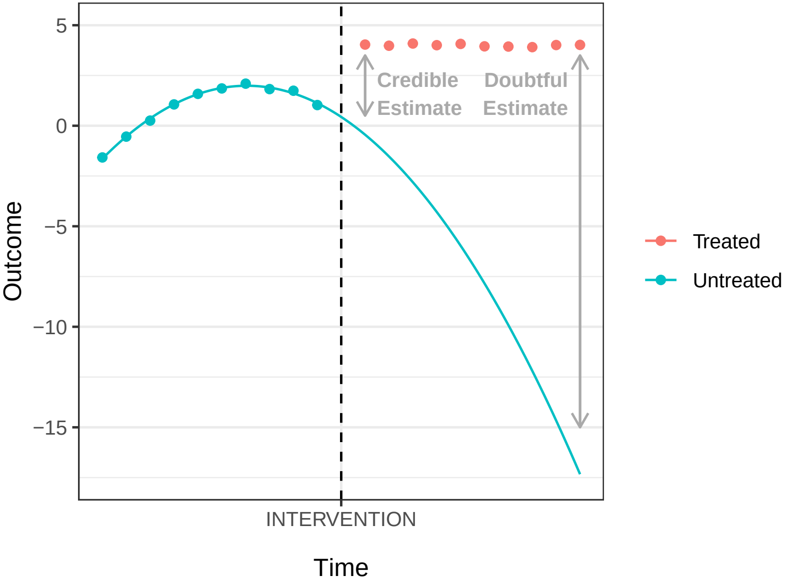
Interrupted time series[1]

You study one unit. It is untreated. Then it is treated.



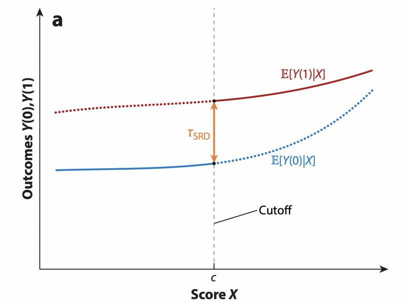
[1] Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. International Journal of Epidemiology, 46(1), 348-355.

Interrupted time series: When it becomes doubtful



Regression discontinuity[2]

Cattaneo & Titiunik 2022 Fig 1a



Theoretical Estimand

$E(Y(1) - Y(0) \mid X = c)$

Empirical Estimand

$\lim_{x \downarrow c} E(Y \mid X = x)$

$-$

$\lim_{x \uparrow c} E(Y \mid X = x)$

Identifying Assumptions

$E(Y(1) \mid X = x)$ and $E(Y(0) \mid X = x)$ are continuous at $x = c$

and $f_X(x) > 0$ for x near c

[2] Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. Annual Review of Economics, 14, 821-851.

Synthetic control[3]

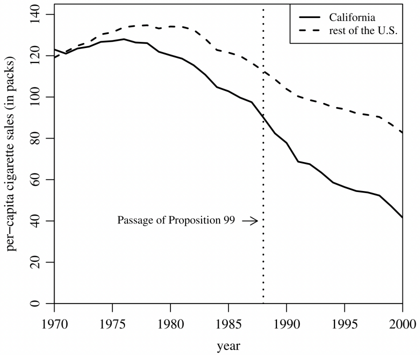


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Can't use RD

— Effect at 1988 not of interest

Can't use ITS

— Hard to extrapolate Y^0 trend

Can't use DID

— No other state like CA

Idea: Create a **synthetic CA** to estimate $Y^0_{\mathrm{CA},t}$ for $t \geq 1988$

[3] Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association, 105(490), 493-505.

Synthetic control[4]

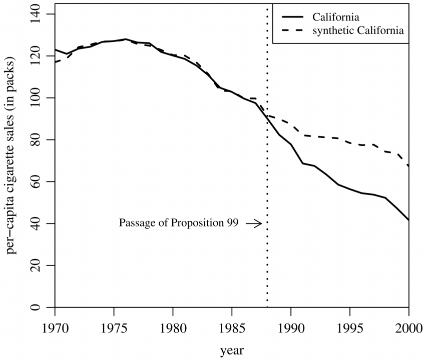
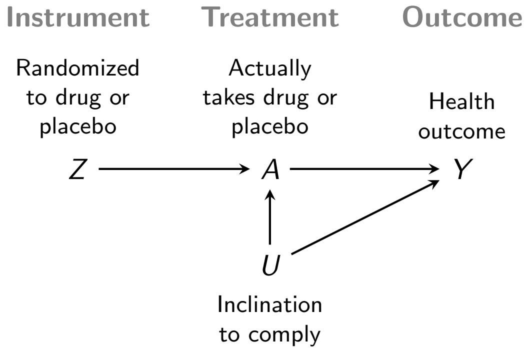


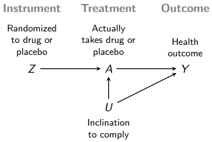
Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

[4] Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association, 105(490), 493-505.

Instrumental variables: Experiment with noncompliance



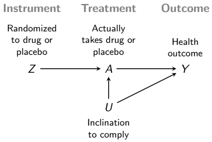
Instrumental variables



Four principal strata

Compliers	$(Z \to A) = +1$	$(Z \to Y) = (A \to Y)$
Always takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
Never takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
~~Defiers~~	$(Z \to A) = -1$	$(Z \to Y) = -(A \to Y)$

Instrumental variables

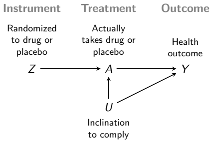


Four principal strata

$$
\begin{array}{rll}
\text{Compliers} & (Z \to A) = +1 & (Z \to Y) = (A \to Y) \\
\text{Always takers} & (Z \to A) = 0 & (Z \to Y) = 0 \\
\text{Never takers} & (Z \to A) = 0 & (Z \to Y) = 0 \\
\cancel{\text{Defiers}} & (Z \to A) = -1 & (Z \to Y) = -(A \to Y)
\end{array}
$$

- The effect $Z \to Y$ arises entirely from compliers

Instrumental variables

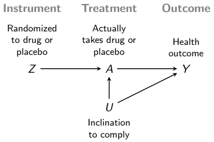


Four principal strata

Compliers	$(Z \to A) = +1$	$(Z \to Y) = (A \to Y)$
Always takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
Never takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
~~Defiers~~	$(Z \to A) = -1$	$(Z \to Y) = -(A \to Y)$

- The effect $Z \to Y$ arises entirely from compliers
- The effect $Z \to A$ tells us how many compliers there are

Instrumental variables



Four principal strata

Compliers	$(Z \to A) = +1$	$(Z \to Y) = (A \to Y)$
Always takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
Never takers	$(Z \to A) = 0$	$(Z \to Y) = 0$
~~Defiers~~	$(Z \to A) = -1$	$(Z \to Y) = -(A \to Y)$

- The effect $Z \to Y$ arises entirely from compliers
- The effect $Z \to A$ tells us how many compliers there are

$$\mathrm{E}(Y^{a=1} - Y^{a=0} \mid S = \text{Complier}) = \frac{\mathrm{E}(Y^{z=1} - Y^{z=0})}{\mathrm{E}(A^{z=1} - A^{z=0})}$$

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Causal Inference in Observational Settings

- Causal claims require an argument
- Causal inference without models (nonparametric)
  - Consistency
  - Exchangeability
  - Positivity
- Causal inference with models (parametric)
  - Outcome modeling: The parametric g-formula
  - Treatment modeling: Propensity scores
  - Marginal structural models
- Treatments that unfold over time
  - Treatments in many time periods
  - Controlled direct effects
  - Natural direct effects
- Complexities that arise in real settings
  - When treatment causes death: Principal stratification
  - When variables are measured with error
- When exchangeability does not hold
  - Bounds
  - Difference in difference
  - Regression discontinuity
  - Synthetic control
  - Instrumental variables
- Recap: Causal inference in observational settings

Randomized experiments are the gold standard in causal inference, and for many good reasons

Phase 3 Clinical Trial of Investigational Vaccine for COVID-19 Begins

Multi-Site Trial to Test Candidate Developed by Moderna and NIH

July 27, 2020 [5]

[5] Published 27 July 2020. https://www.niaid.nih.gov/news-events/phase-3-clinical-trial-investigational-vaccine-covid-19-begins

An experimental protocol clarifies the causal estimand

- Target population ✓
- Well-defined contrast between treatments ✓
- Well-defined outcome after follow-up period ✓
- Attrition is transparent ✓
- Assumptions (randomized design) support causal claims ✓

[6] Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology, 183(8), 758-764.

An experimental protocol clarifies the causal estimand

- Target population ✓
- Well-defined contrast between treatments ✓
- Well-defined outcome after follow-up period ✓
- Attrition is transparent ✓
- Assumptions (randomized design) support causal claims ✓

An observational study cannot randomize.

[6] Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology, 183(8), 758-764.

An experimental protocol clarifies the causal estimand

- Target population ✓
- Well-defined contrast between treatments ✓
- Well-defined outcome after follow-up period ✓
- Attrition is transparent ✓
- Assumptions (randomized design) support causal claims ✓

An observational study cannot randomize.
Otherwise, an observational study can do all of the above.[6]

[6] Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology, 183(8), 758-764.

A few key principles for causal inference in observational settings

When you **evaluate a causal claim**, ask yourself

1. What is the unit of analysis?
2. What is the treatment?
3. Do the potential outcomes make sense? consistency
4. Can I draw the DAG? exchangeability
5. Are all treatment values plausible? positivity

When you answer causal questions in **your own research**, consider

1. What is my causal contrast?
    - What would it mean to assign someone to my treatment?
2. What is my target population?
3. What assumptions do I need for inference?
    - If these assumptions, then this causal conclusion
    - The if-then statement should be airtight!
    - Then argue for those assumptions

When you **engage with new methods**, ask yourself

1. Counterfactuals always require assumptions.
   What assumption does this method make?
2. In what settings would this method perform well?
3. In what settings would this method perform poorly?
4. How does this method invoke key concepts I already know?

Learning goals for the semester

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

- evaluate the credibility of causal claims
- answer causal questions in your own research
- engage with new methods for causal inference

What comes next?

- Comments on two peers' proposals due 5pm Monday Dec 5
- Course evaluations
- Go out and conduct causal research!

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!