

# 1. Causal Questions: Observing and Intervening

Ian Lundberg

Cornell Info 6751: Causal Inference in Observational Settings  
Fall 2022

23 Aug 2022

Why causal inference?

What motivated you to take this course?

Why causal inference?

Why am I teaching this course?

# Why causal inference?

Why am I teaching this course?

Causal inference provides tools to

- ▶ Speak to policy interventions
- ▶ Understand social systems

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

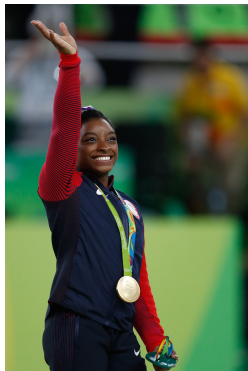
Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

# Causal claims hinge on arguments, not on data



Left photo: By Fernando Frazão/Agência Brasil - [http://agenciabrasil.ebc.com.br/sites/\\_agenciabrasil2013/files/fotos/1035034-\\_mg\\_0802\\_04.08.16.jpg](http://agenciabrasil.ebc.com.br/sites/_agenciabrasil2013/files/fotos/1035034-_mg_0802_04.08.16.jpg), CC BY 3.0, <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 hinge on arguments, not on data

## 1. Statistical evidence

- ▶ Simone Biles swung on the uneven bars. She won a gold medal.



# Causal claims hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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 hinge on arguments, not on data

## 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

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus



Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

## Notation: Potential outcomes

$Y_i$	Outcome	Whether person $i$ won a gold medal
$A_i$	Treatment	Whether person $i$ swung on the bars
$Y_i^a$	Potential Outcome	Outcome person $i$ would realize if assigned to treatment value $a$

### Examples:

$Y_{\text{Biles}} = 1$	Simone Biles won gold
$A_{\text{Biles}} = 1$	Simone Biles swung on the bars
$Y_{\text{Biles}}^1 = 1$	If she swings, Simone Biles would win gold
$Y_{\text{Biles}}^0 = 0$	If she does not swing, Simone Biles would not win gold

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

- ▶ Draw a random person from the population

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

- ▶ Draw a random person from the population
- ▶ Put them on the uneven bars



# Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

- ▶ Draw a random person from the population
- ▶ Put them on the uneven bars
- ▶ The outcome  $Y^1$  is a random variable:
  - ▶ It takes the value 1 if we have sampled Simone Biles
  - ▶ It takes the value 0 if we have sampled anyone else

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

- ▶ Draw a random person from the population
- ▶ Put them on the uneven bars
- ▶ The outcome  $Y^1$  is a random variable:
  - ▶ It takes the value 1 if we have sampled Simone Biles
  - ▶ It takes the value 0 if we have sampled anyone else

We omit the subscript  $i$  when talking about random people

## Notation: Potential outcomes

A person's potential outcomes are **deterministic mappings**

- ▶ Simone Biles would win gold if she participates  $Y_i^1 = 1$
- ▶ Simone Biles would not win gold without participating  $Y_i^0 = 0$

These are just numbers

The outcome for a random person is a **random variable**

- ▶ Draw a random person from the population
- ▶ Put them on the uneven bars
- ▶ The outcome  $Y^1$  is a random variable:
  - ▶ It takes the value 1 if we have sampled Simone Biles
  - ▶ It takes the value 0 if we have sampled anyone else

We omit the subscript  $i$  when talking about random people

**Check for understanding:**

Does it make sense to write  $V(Y_i^1)$ ? How about  $V(Y^1)$

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Notation: The consistency assumption

$Y_i^1$

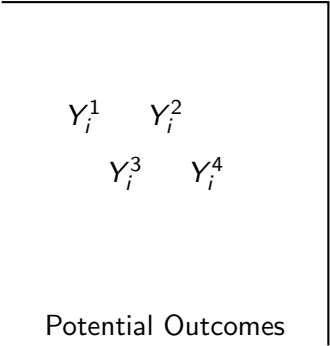
$Y_i^2$

$Y_i^3$

$Y_i^4$

Potential Outcomes

Notation: The consistency assumption



A diagram consisting of a horizontal line at the top and a vertical line on the right, forming an L-shape. Inside the corner, four labels are arranged in a 2x2 grid:  $Y_i^1$  and  $Y_i^2$  in the top row, and  $Y_i^3$  and  $Y_i^4$  in the bottom row.

$Y_i^1$     $Y_i^2$   
 $Y_i^3$     $Y_i^4$

Potential Outcomes



A diagram consisting of a horizontal line at the top and a vertical line on the left, forming an L-shape. Inside the corner, a single label  $Y_i$  is centered.

$Y_i$

Factual Outcomes

Notation: The consistency assumption

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



Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

## Notation: Expectation operator

The **expectation operator**  $E()$  denotes the population mean

$$E(Y^a) = \frac{1}{n} \sum_{i=1}^n Y_i^a$$

The quantity  $Y^a$  inside the expectation must be a random variable

## Notation: Expectation operator

The **expectation operator**  $E()$  denotes the population mean

$$E(Y^a) = \frac{1}{n} \sum_{i=1}^n Y_i^a$$

The quantity  $Y^a$  inside the expectation must be a random variable

A **conditional expectation** is denoted with a vertical bar

$$E(Y \mid A = a) = \frac{1}{n_a} \sum_{i:A_i=a} Y_i$$

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Practice: How would you say this in English?

1.  $E(\text{Earnings} \mid \text{Degree} = \text{TRUE}) > E(\text{Earnings} \mid \text{Degree} = \text{FALSE})$

2.  $E(\text{Earnings}^{\text{Degree}=\text{TRUE}}) > E(\text{Earnings}^{\text{Degree}=\text{FALSE}})$

## Practice: How would you say this in English?

1.  $E(\text{Earnings} \mid \text{Degree} = \text{TRUE}) > E(\text{Earnings} \mid \text{Degree} = \text{FALSE})$

► Average earnings are higher among those with college degrees

2.  $E(\text{Earnings}^{\text{Degree}=\text{TRUE}}) > E(\text{Earnings}^{\text{Degree}=\text{FALSE}})$



## Practice: How would you say this in English?

1.  $E(\text{Earnings} \mid \text{Degree} = \text{TRUE}) > E(\text{Earnings} \mid \text{Degree} = \text{FALSE})$

- ▶ Average earnings are higher among those with college degrees

2.  $E(\text{Earnings}^{\text{Degree}=\text{TRUE}}) > E(\text{Earnings}^{\text{Degree}=\text{FALSE}})$

- ▶ On average, a degree causes higher earnings

A **descriptive** statement

Language: among, disparity

	Degree <sub>i</sub>	Earnings <sub>i</sub> <sup>Degree</sup>	Earnings <sub>i</sub> <sup>No degree</sup>
Person 1	TRUE	\$100k	\$90k
Person 2	TRUE	\$90k	\$80k
Person 3	FALSE	\$60k	\$50k
Person 4	FALSE	\$50k	\$40k

A **causal** statement

Language: effect, leads to, produces, benefits

	Degree <sub>i</sub>	Earnings <sub>i</sub> <sup>Degree</sup>	Earnings <sub>i</sub> <sup>No degree</sup>
Person 1	TRUE	\$100k	\$90k
Person 2	TRUE	\$90k	\$80k
Person 3	FALSE	\$60k	\$50k
Person 4	FALSE	\$50k	\$40k

## Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.
2. On average, doing the homework causes more learning.

## Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.

$$E(\text{Learning} \mid \text{HW} = \text{TRUE}) > E(\text{Learning} \mid \text{HW} = \text{FALSE})$$

2. On average, doing the homework causes more learning.

## Practice: How would you write this in math?

1. On average, students who do the homework learn more than those who don't.

$$E(\text{Learning} \mid \text{HW} = \text{TRUE}) > E(\text{Learning} \mid \text{HW} = \text{FALSE})$$

2. On average, doing the homework causes more learning.

$$E(\text{Learning}^{\text{HW}=\text{TRUE}}) > E(\text{Learning}^{\text{HW}=\text{FALSE}})$$

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Fun example: Does A affect Y?



## Fun example: Does A affect Y?

Treatment	$A_i$	Receives heart transplant (1) vs. not (0)
Outcome	$Y_i$	Dead (1) or alive (0) after five days

## Fun example: Does A affect Y?

Treatment  $A_i$     Receives heart transplant (1) vs. not (0)  
Outcome  $Y_i$     Dead (1) or alive (0) after five days

Unit	$Y_i^1$	$Y_i^0$	Causal Effect
Zeus	0	1	Transplant would save his life
Hera	0	0	Would survive regardless
Poseidon	1	0	Transplant would kill him

## Fun example: Does A affect Y?

Treatment  $A_i$     Receives heart transplant (1) vs. not (0)  
Outcome  $Y_i$     Dead (1) or alive (0) after five days

Unit	$Y_i^1$	$Y_i^0$	Causal Effect
Zeus	0	1	Transplant would save his life
Hera	0	0	Would survive regardless
Poseidon	1	0	Transplant would kill him

**Question:** Does a heart transplant cause death?

## Fun example: Does A affect Y?

Treatment  $A_i$     Receives heart transplant (1) vs. not (0)  
Outcome  $Y_i$     Dead (1) or alive (0) after five days

Unit	$Y_i^1$	$Y_i^0$	Causal Effect
Zeus	0	1	Transplant would save his life
Hera	0	0	Would survive regardless
Poseidon	1	0	Transplant would kill him

**Question:** Does a heart transplant cause death?

No.

The average causal effect is  $\frac{1}{n} \sum_{i=1}^n \left( Y_i^1 - Y_i^0 \right) = 1 + 0 - 1 = 0$

## Fun example: Does A affect Y?

Treatment  $A_i$     Receives heart transplant (1) vs. not (0)  
Outcome  $Y_i$     Dead (1) or alive (0) after five days

Unit	$Y_i^1$	$Y_i^0$	Causal Effect
Zeus	0	1	Transplant would save his life
Hera	0	0	Would survive regardless
Poseidon	1	0	Transplant would kill him

**Question:** Does a heart transplant cause death?

No.

The average causal effect is  $\frac{1}{n} \sum_{i=1}^n \left( Y_i^1 - Y_i^0 \right) = 1 + 0 - 1 = 0$

Yes.

Potential outcomes are unequal.  $Y_i^1 \neq Y_i^0$  for at least some  $i$

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

Why causal inference?

Central ideas for today

Causal claims hinge on arguments, not on data

Notation

Potential outcomes

Consistency assumption

Expectation operator

Practicing notation

Fun example: Does A affect Y?

Logistics: Syllabus

# Syllabus: Learning goals

[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

Students will learn to

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



# Syllabus: Readings

[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

A combination of articles and this textbook:

Hernán, M.A., and J.M. Robins. 2020. *Causal Inference: What If?*  
Boca Raton: Chapman & Hall / CRC. PDF available at  
[hsph.harvard.edu/miguel-hernan/causal-inference-book/](https://hsph.harvard.edu/miguel-hernan/causal-inference-book/)

# Syllabus: Support

[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

- ▶ Ask questions on [Ed Discussion](#)
- ▶ Ask questions in office hours
  - ▶ 1 hour after each class
  - ▶ [calendly.com/ianlundberg/office-hours](https://calendly.com/ianlundberg/office-hours)

# Syllabus: Assignments

[tinyurl.com/Info6751](http://tinyurl.com/Info6751)

1) Problem sets	Weekly	50%
2) Ideas for the research proposal	Oct 31	10%
3) Final research proposal	Nov 21	30%
4) Feedback to two peers	Dec 5	10%

# Syllabus: Type and/or Handwrite

[tinyurl.com/Info6751](http://tinyurl.com/Info6751)

I type math in  $\text{\LaTeX}$ .

You can type, handwrite, or combine the two. All equally good.

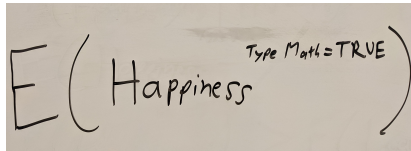
# Syllabus: Type and/or Handwrite

[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

I type math in  $\text{\LaTeX}$ .

You can type, handwrite, or combine the two. All equally good.

**Example.** A student thought about their expected happiness under an intervention to make them type math.



A handwritten expression on a light brown background. It features a large capital letter 'E' followed by an opening parenthesis '('. Inside the parentheses, the word 'Happiness' is written in a cursive script. Above 'Happiness', the text 'Type Math = TRUE' is written in a smaller, more formal script. The expression is closed with a large closing parenthesis ')'. The entire expression is  $E(\text{Happiness}^{\text{Type Math} = \text{TRUE}})$ .

Then the student decided to handwrite.

Syllabus: Suggested workflow  
[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

Lecture → Reading → Problem Set

# Syllabus: Suggested workflow

[tinyurl.com/Info6751](https://tinyurl.com/Info6751)

Lecture → Reading → Problem Set

You are now ready to

- ▶ Read Hernán and Robins Ch 1
- ▶ Complete Problem Set 1 Part 1

After Thursday, you will be ready to

- ▶ Read Hernán 2016
- ▶ Complete Problem Set 1 Part 2

Problem set is due on Canvas Monday at 5pm

Let me know what you are thinking

[tinyurl.com/CausalQuestions](https://tinyurl.com/CausalQuestions)

Office hours TTh 11am-12pm and at  
[calendly.com/ianlundberg/office-hours](https://calendly.com/ianlundberg/office-hours)  
Come say hi!