

# Correlation to Causation

---

QTM 110

KEVIN MCALISTER

# Administrative Stuff

---

Don't forget that the welcome quiz is due tonight!

I will use this to make group assignments. A/D/S ends a little later than usual this semester, so I'll probably post these groups next Thursday after the end. Your first PS will go live the same day.

# Administrative Stuff

---

The first content quiz of the semester will be posted on Canvas after class.

- One week to complete
- No time limit, but don't spend too much time on it
- Shouldn't take more than 30 minutes for a well-prepared student
- Correlation and Discussions of Causation and Counterfactual Dependence

# Previously on...

---

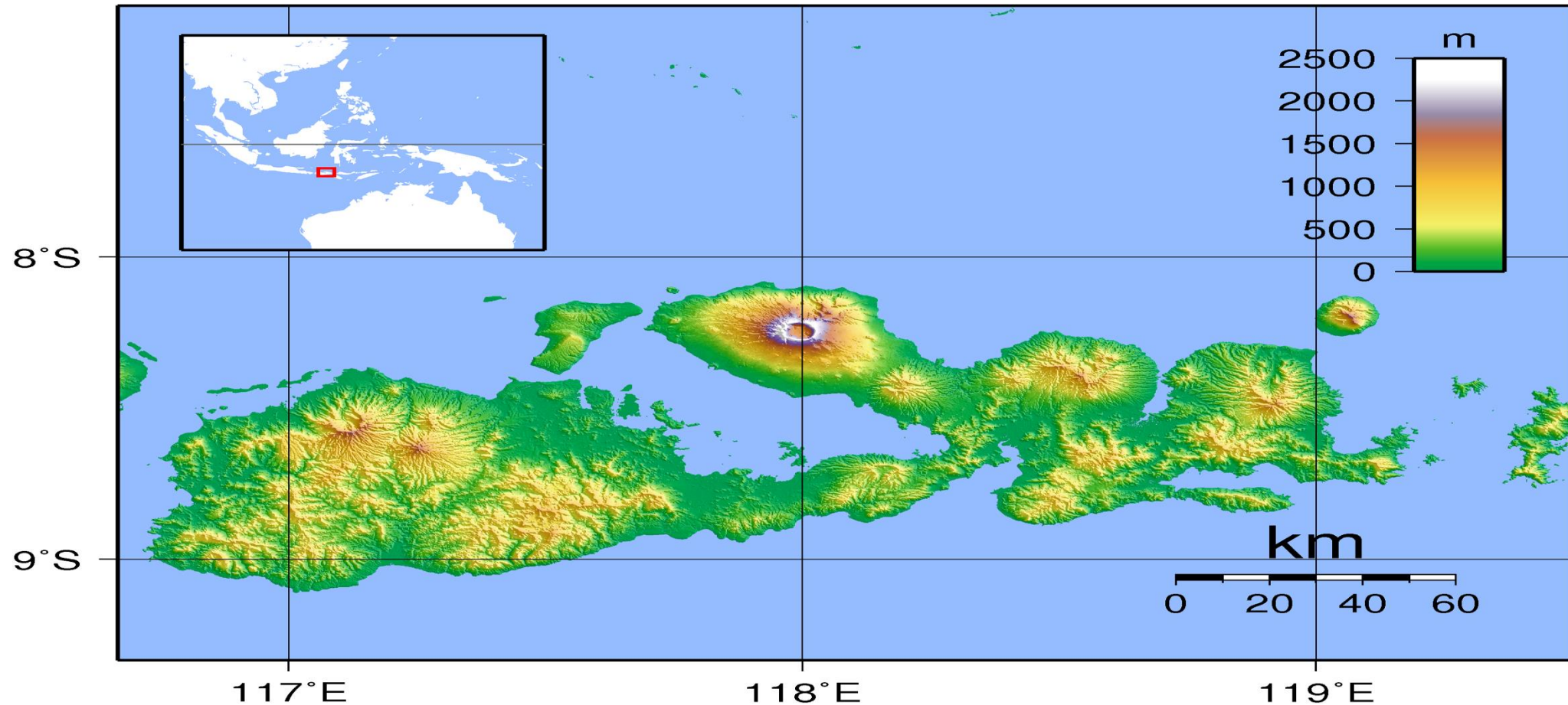
- Two variables are correlated when an increase in one corresponds to a predictable change in the other
- Three ways to measure correlation: covariance, correlation coefficient, and regression coefficient (pros and cons for each)
- Correlation lets us explore relationships
  - Descriptive
  - Predictive
  - Causal
- Causal relationships are hard!

# Today's Takeaways

---

- Correlation and causation are different things
- Pinning down a definition of causation that always makes sense is difficult
- We can define causation a number of different ways, but there is one that makes a lot of sense for identifying causal effects
- Counterfactual dependence requires correlation **and** identification of outcomes in both the observed **and** counterfactual worlds
- Potential outcomes are the basis for our estimates of causal effects
- No silver bullet – counterfactual dependence comes with its own problems.

# Sumbawa, Indonesia

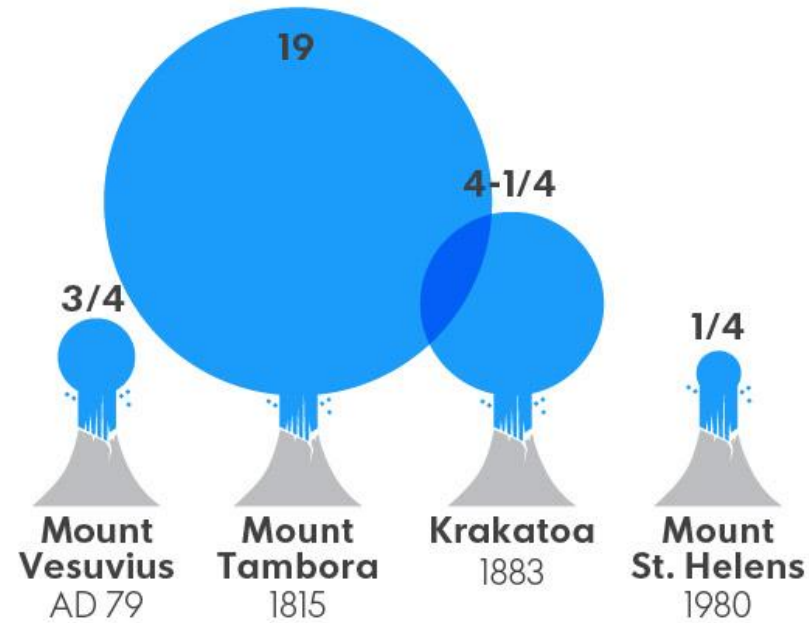


---

## VIOLENT VOLCANOES

The eruption of Mount Tambora in 1815 was the biggest in recorded human history.

Cubic miles of ejecta:



SOURCE: USA TODAY Research  
Janet Loehrke, USA TODAY



# April 10<sup>th</sup> , 1815

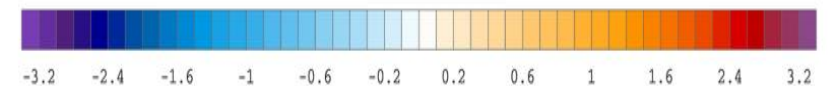
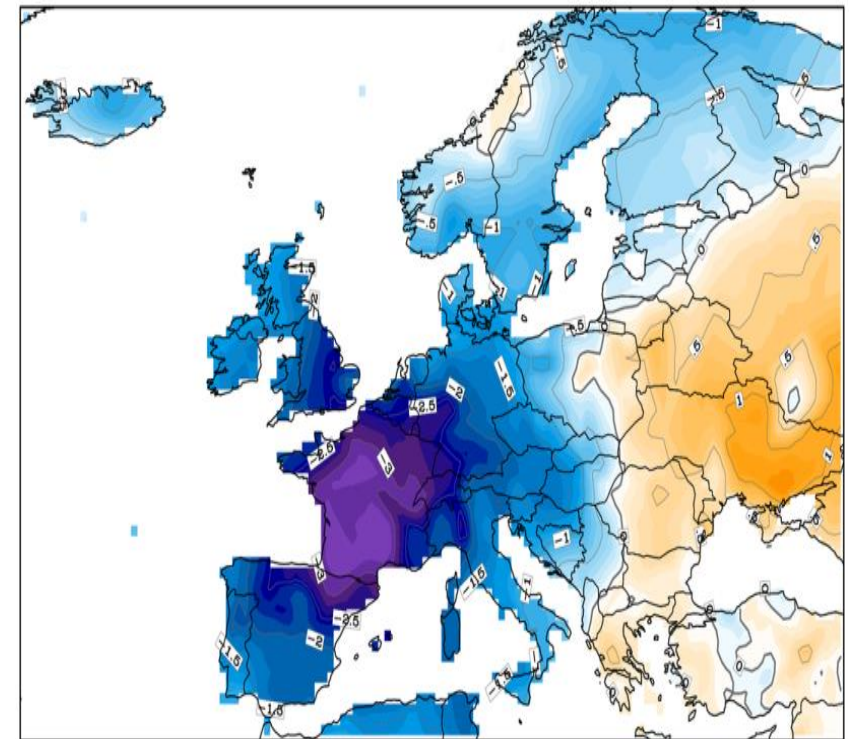
---



# 1816: Year Without a Summer



1816 Summer temperature anomaly



# Lake Geneva, Switzerland

---

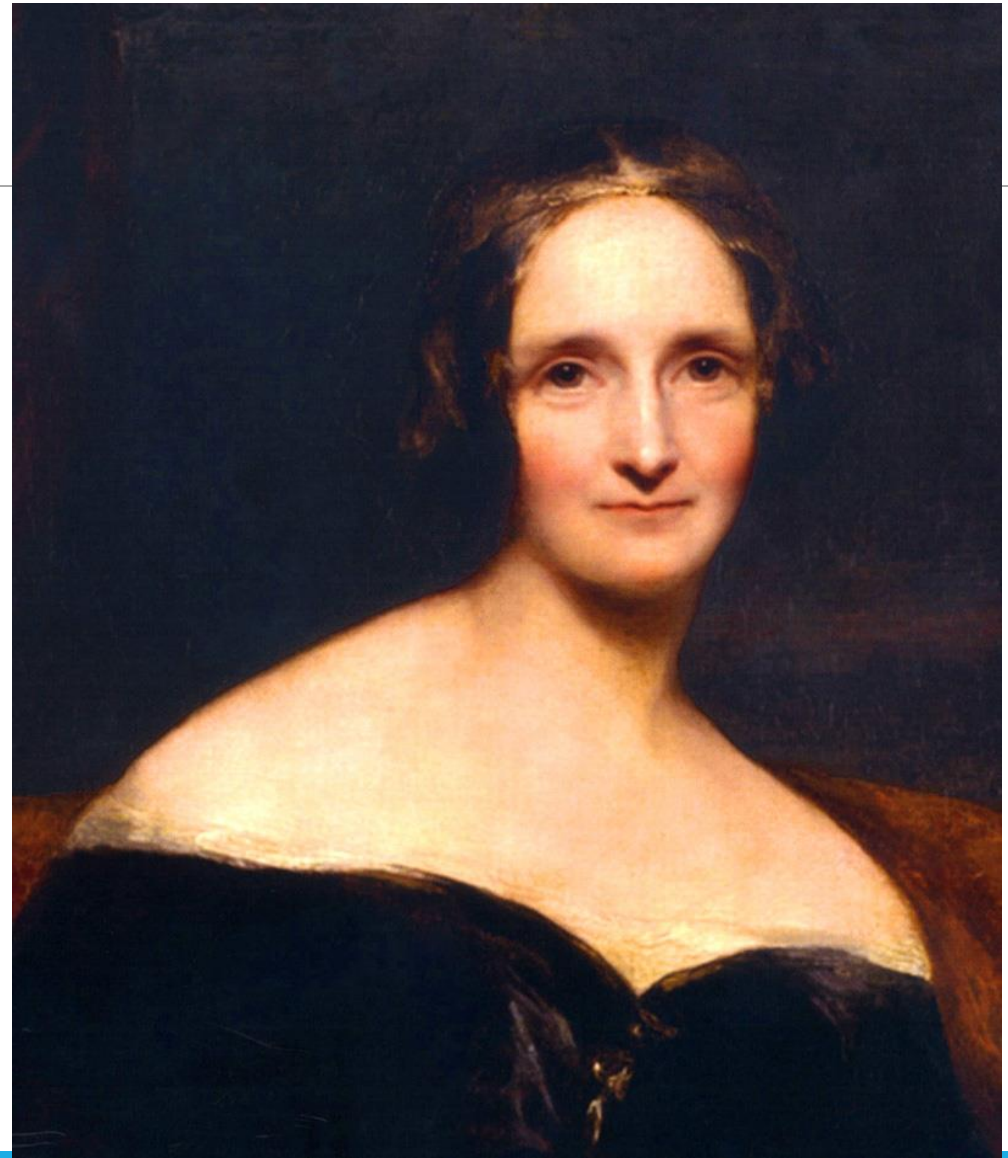


# Lake Geneva, Switzerland, 1816

---



Mary Shelley



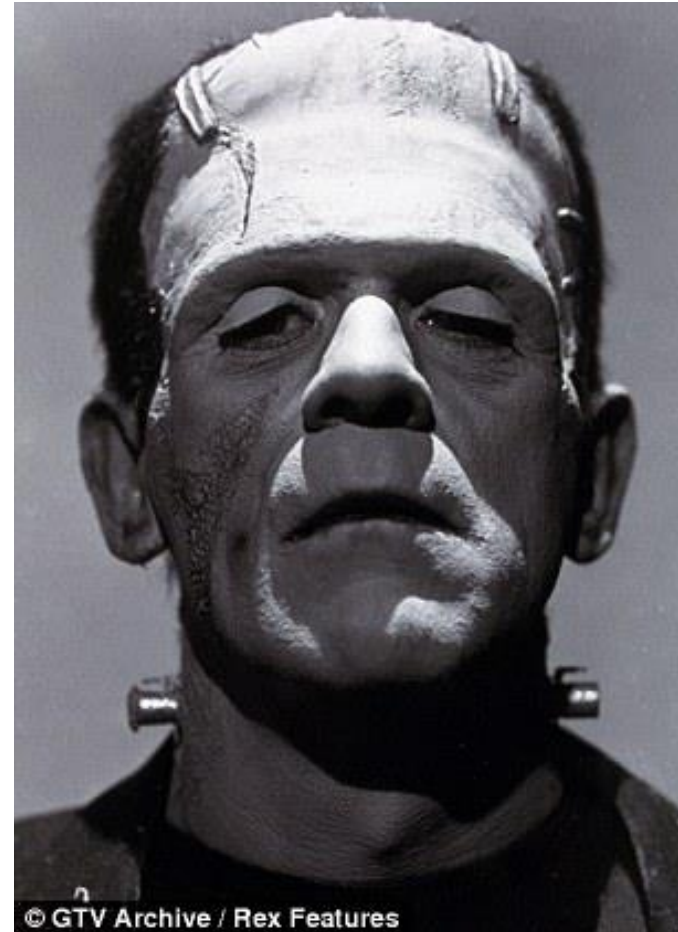
# Scary Stories

---



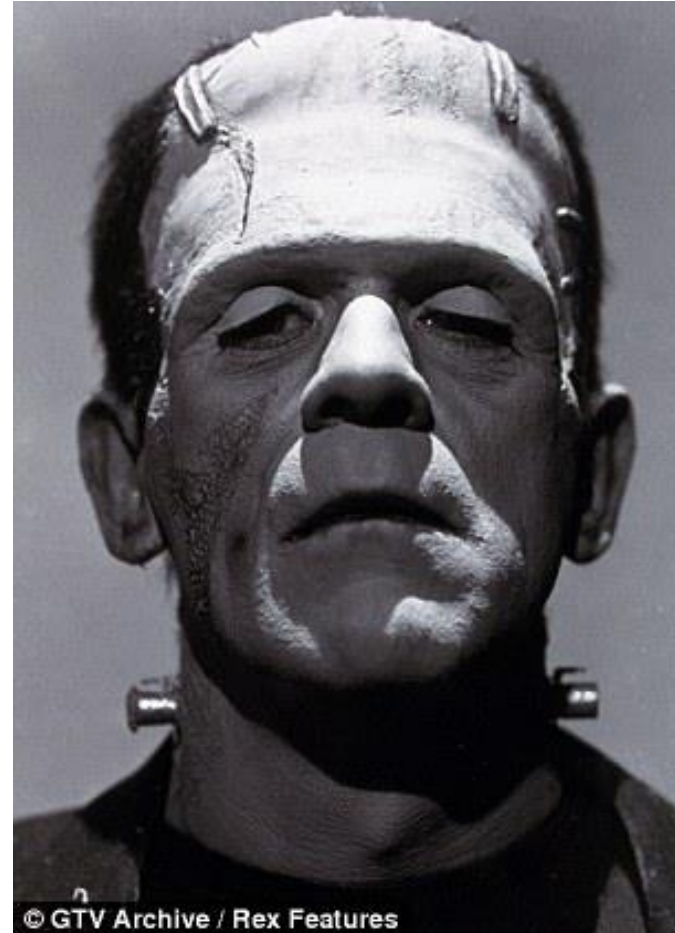
# Frankenstein

---



# Just Kidding: Frankenstein's Monster

---



# Did the Eruption of Tambora cause Frankenstein?

---

# Did the Eruption of Tambora cause Frankenstein?

---

What's our definition of cause?

Surprisingly hard to pin down exactly.

# Our Goal

---

Arrive at a coherent definition of causation

# What we want

---

We want a formal definition that captures the following:

**A causal effect is a change in one feature of the world that results from a change in another feature of the world**

Why does it matter?

- Can we say that smoking causes cancer?
- Does Ozempic cause maintainable weight loss for the obese?
- Do stricter voter ID laws suppress turnout?
- Job training vs. lump sum payments – which decreases extreme poverty most? If payments, how much is the correct amount?

# Causation 1: Correlation

---

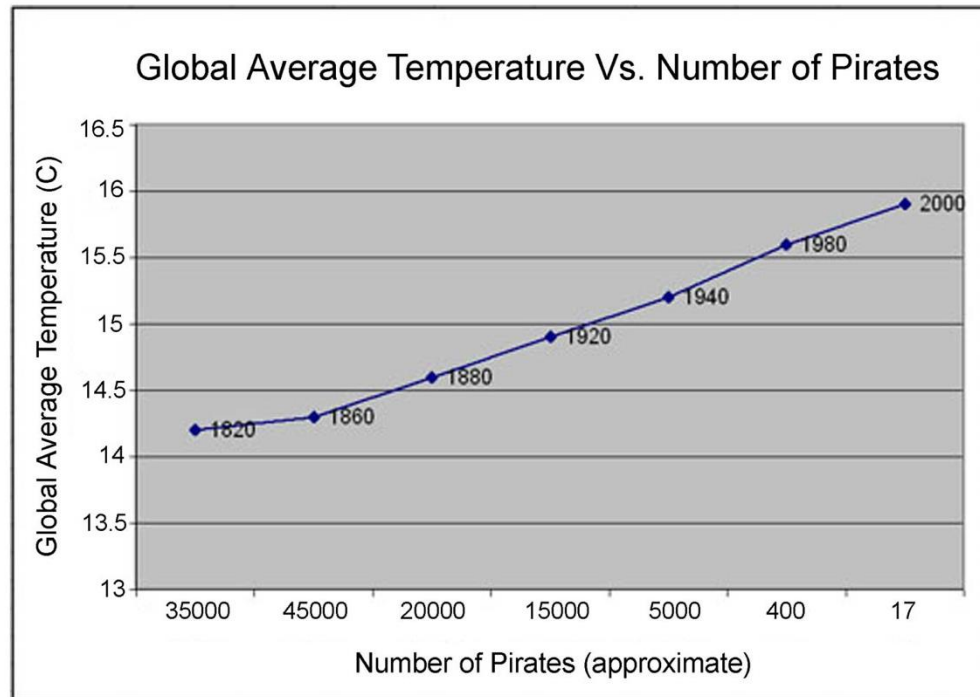
One attempt at a definition of causation might be:

“Can we predict one quantity by knowing another, using a known correlation between the two?”

Reasoning: if we understand the cause, we should be able to forecast the outcome.

# Causal Inferences

**STOP GLOBAL WARMING: BECOME A PIRATE**



WWW.VENGANZA.ORG



# Causal Inferences

---

## Pirates and Global Warming

- There is a correlation between pirates and global warming.
- Are they divine beings?



# Causal Inferences

---

## Pirates and Global Warming

- Even though there's a correlation, do you think that there is a causal relationship?
- Put another way, did the decrease in pirates **cause** an increase in global temperatures?

# Causation 1: Correlation

---

## Problem 1: Correlation without Causation

- Common Cause
  - Omitted Variable, Confounders, Selection
    - Dressing for Success, Job Training Programs, Elite Colleges, Handwashing and Peanut Allergies, Any parenting choice

## Problem 2: Direction

- Correlation of X with Y is the same as Y with X
  - $COV(X,Y) = COV(Y,X)$
  - Reverse Causation!
    - Spinning windmills and wind
    - Mother's Day cards and Mother's Day

# Causation 2: Regularity

---

Every time X happens, Y happens, but not the other way around

Addresses issue of direction!

Problems?

# Causation 2: Regularity

---

Every time X happens, Y happens, but not the other way around

Addresses issue of direction!

## **Problem 1: Deterministic**

- Gunshots cause death, but not every time

## **Problem 2: Trivial relationships**

- Every time I ring a bell, I am human
- Every time I put gas in my car, I drive

# Causation 3: Temporal Order

---

Correlation wasn't sufficient, but what if we added another condition:

## The Arrow of Time

- Event A can cause event B only if event A precedes event B in time.

Problems?

# Causation 3: Temporal Order

---

Correlation wasn't sufficient, but what if we added another condition:

## The Arrow of Time

- Event A can cause event B only if event A precedes event B in time.

## Problems:

- Christmas cards cause Christmas
- Attendance cause shows
- Barometers and Hurricanes
- In general: prediction  $\neq$  causation

# Causation 4: Physical Connection

---

Moving billiard ball touches resting ball and sends it off.

- We can experience with our sense
- Maybe we should require that causes physically produce effects through some observable mechanism.
- **Problem:**
- Hard to verify in many cases
- Requires more convoluted stories

## Ex: Police Deter Crime

- Police arrested other criminals in the past
- Media Reports on Crime
- Criminal Reads report
- Deterred from committing crime

# Causation 5: Counterfactual Dependence

---

## Definition

X causes Y if and only if

1. Y occurs when X occurs, and
2. Y would not have occurred in the **counterfactual world** where X did not occur

# Causation 5: Counterfactual Dependence

---

## What is a counterfactual world?

The counterfactual world is one identical to the observable one **except** for a single change in X.

A bizarro world, if you will...



# Counterfactual Dependence

---

Let's think about some counterfactual worlds.

Counterfactuals are easiest to think about when the causal variable can only take one of two values:

- Receive the pill or don't receive the pill
- Terry McLaurin signs a new contract before the start of the season or he doesn't
- A student goes to college or doesn't

We're mostly going to be thinking about these situations in this class.

# Counterfactual Dependence

---

Causal variables can also have more than two categories, be continuous, etc.

Causal Variable: 0 mg of drug, 1 mg of drug, 2 mg of drug, etc.

A little more difficult to think about, but possible

# Causation 5: Counterfactual Dependence

---

Why is counterfactual dependence a good definition of causation?

- X and Y should demonstrate some level of **correlation**
- In a world where X did not occur **and everything else remained the same**, we should expect that Y does not occur.

Under this definition, the only possible cause of the change in Y is the change in X!

# Causation 5: Counterfactual Dependence

---

The only requirement in this definition is that we should be able to envision a counterfactual world:

- Identical to the observed world in every way except for X

# Causation 5: Counterfactual Dependence

---

Regularity is not required!

Think about graduating from college and making more than 100k at 30

- Some people will graduate from college but not make a lot of money at 30

Reworded question:

Does graduating from college **cause** an **increased chance** of making more than 100k at 30?

- Increased chance is the same as cause even though it won't happen with regularity!

# Causation 5: Counterfactual Dependence

---

Briefly, think about the following causal claims. Does the claim plausibly meet the conditions of counterfactual dependence? Why or why not?

- Decrease in pirates caused an increase in average global temperature
- Windmills cause wind vs. Wind causes windmills
- Vaccines cause increased chance of Autism
- 5G (at the level that most humans receive) causes an increase in likelihood of cancer
- Smoking cigarettes causes an increased chance of lung cancer

# Causal vs Non-Causal Stories

---

Causal stories should have a clear connection to counterfactual dependence!

- Cigarettes and Lung Cancer
- A person smokes cigarettes. This leads to lesions on the lungs that are more likely to grow cancerous cells. These cells lead to lung cancer.
- In a world where someone **doesn't** smoke, they don't develop lesions. No lesions means no lung cancer.
- Lesions are a **mechanism** in this story – a variable through which the causal variable leads to the outcome.

# Causal vs Non-Causal Stories

---

Causal stories should have a clear connection to counterfactual dependence!

- Cigarettes and Lung Cancer (the R.A. Fisher story – more to come soon!)
- Some people are born with “bad” genetics. These genetics lead people to be prone to addiction and be more likely to smoke cigarettes.
- These same “bad” genetics lead people to be more likely to develop lung cancer.
- In a world where someone doesn’t smoke, the “bad” genetics still lead to lung cancer – whether or not someone smokes is irrelevant!
- Genetics is an example of a **confounder** – a third variable at play that is correlated with both the causal variable and the outcome. Non-causal!

# Causal vs Non-Causal Stories

---

How can we know which story is true?

- This is what we'll be exploring for the next few weeks in this class!
- Spoiler – both stories are probably true (to an extent), but smoking cigarettes has a much larger effect than genetics!

# Counterfactual Dependence

---

The causal variable is then associated with a **potential outcome**

**For each possible value of the causal variable, there is an associated measure of the outcome. These values are the potential outcomes.**

Example: If I receive the pill, then I will live for an additional 20 years.  
If I don't receive the pill, then I will live for an additional 2 years.

# Counterfactual Dependence

---

Counterfactual dependence is the **potential outcomes framework!**

- Because we require that there are clear counterfactual worlds, there is an observed outcome and an outcome that **could have happened** had the alternative condition been true
- The key aspect that gives us a framework for assessing causal effects

# Counterfactual Dependence

---

Key point: Everything prior to the time at which the causal variable is changed is **identical**. Things after can change.

Because everything was identical prior to the causal variable, then all changes after must be due to the change.

Change in one variable results in change of another. **Causation!**

It's a rigid definition, but one that will serve us well in our goal of making causal inferences.

# Counterfactual Dependence

---

Let  $X$  be binary event

- College, drug treatment, batting position

**$X = 1$  “Treated”**

- Went to college, Received treatment, Batted left

**$X=0$  “Untreated”**

- Didn't go to college, Didn't get treatment, Batted right

Let  $Y$  be some outcome of interest

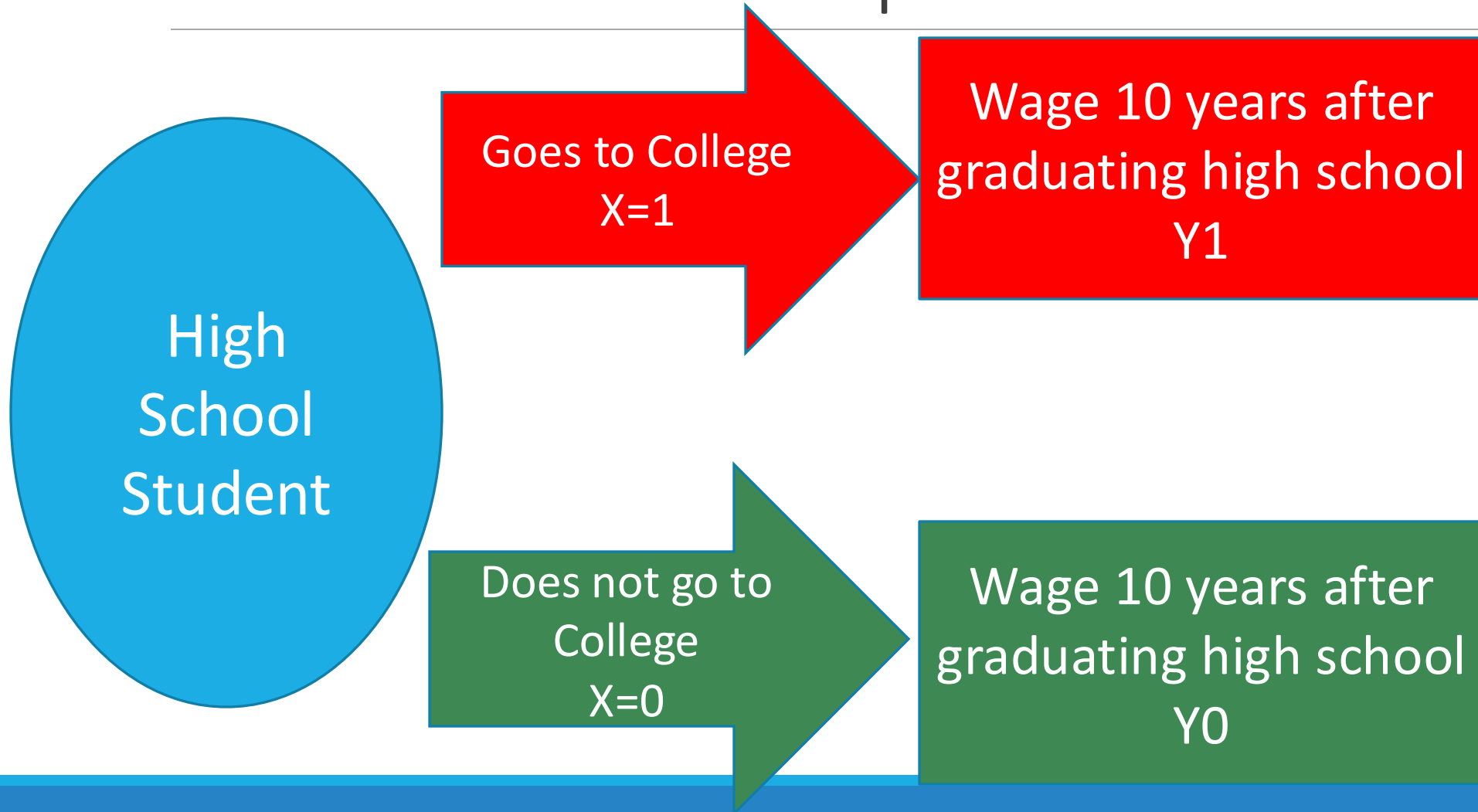
- Wage, Survival in years, Distance of hit ball

**$Y1 \equiv Y$  in the counterfactual world where  $X = 1$**

**$Y0 \equiv Y$  in the counterfactual world where  $X = 0$**

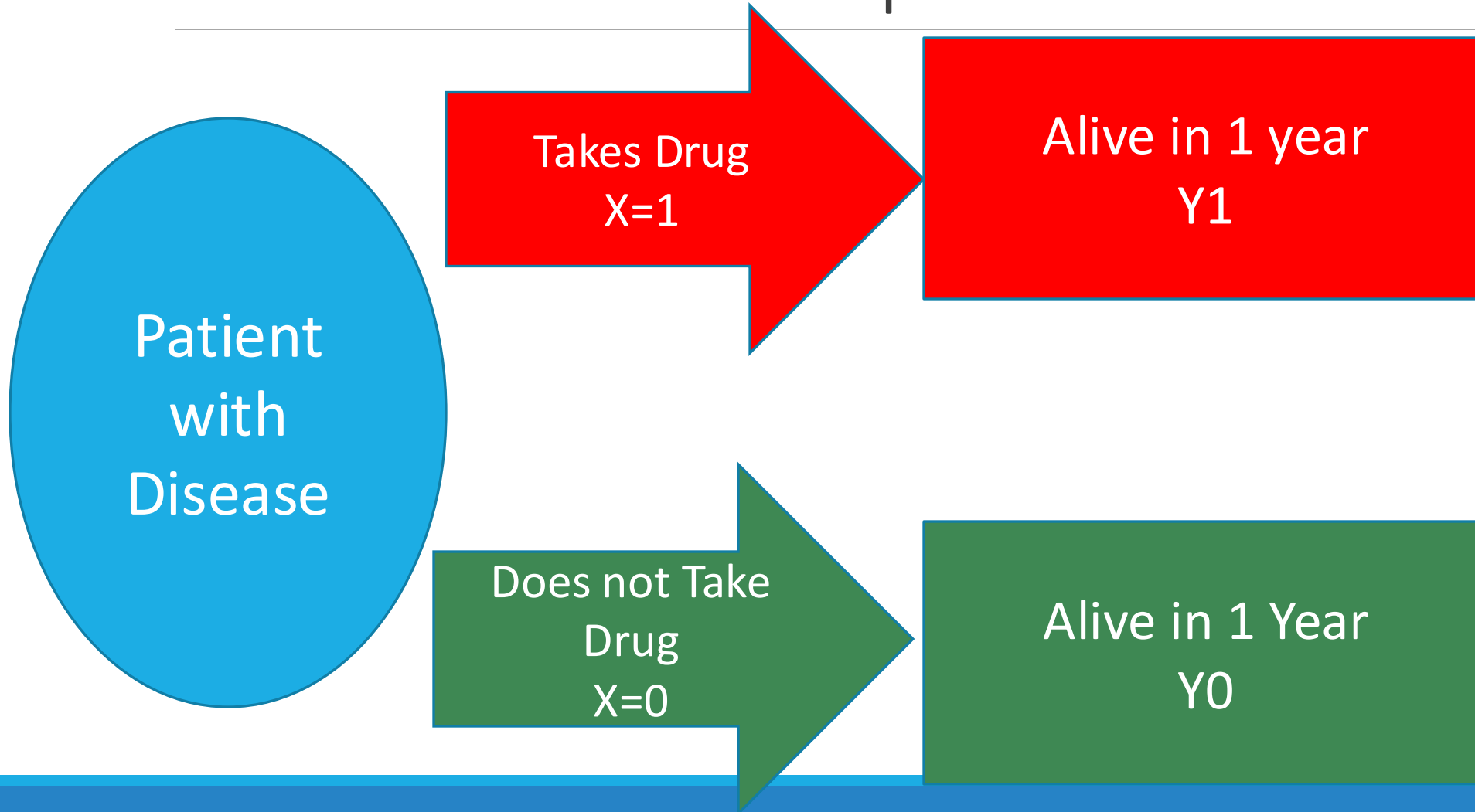
# Counterfactual Dependence

---



# Counterfactual Dependence

---



# Counterfactual Dependence

---

Suppose that we observe the outcome when  $X = 0$  ( $Y_0$ ) and  $X = 1$  ( $Y_1$ ).

What is a reasonable way to assess the effect of  $X$  on  $Y$ ?

# Counterfactual Dependence

---

Suppose that we observe the outcome when  $X = 0$  ( $Y_0$ ) and  $X = 1$  ( $Y_1$ ).

What is a reasonable way to assess the effect of  $X$  on  $Y$ ?

$$Y_1 - Y_0$$

The difference in the potential outcomes is equal to the causal impact of  $X$  on  $Y$ .

# After School Programs and Test Scores

---

Claim: Attending after school programs sponsored by the school or other community organizations causes children to have higher test scores.

Mechanism: Adding structure to a child's afternoon, even if their parents work until 5 PM, establishes a routine and children are more likely to complete homework assignments.

Let's systematically look at this question and use the model of counterfactual dependence to determine if the claim is plausible!

# After School Programs and Test Scores

---

Step 1: Is the first condition met? On average, will we see that a child who attends after school programs has higher test scores?

# After School Programs and Test Scores

---

Step 1: Is the first condition met? On average, will we see that a child who attends after school programs has higher test scores?

Yes! There is a positive correlation between these two variables.

# After School Programs and Test Scores

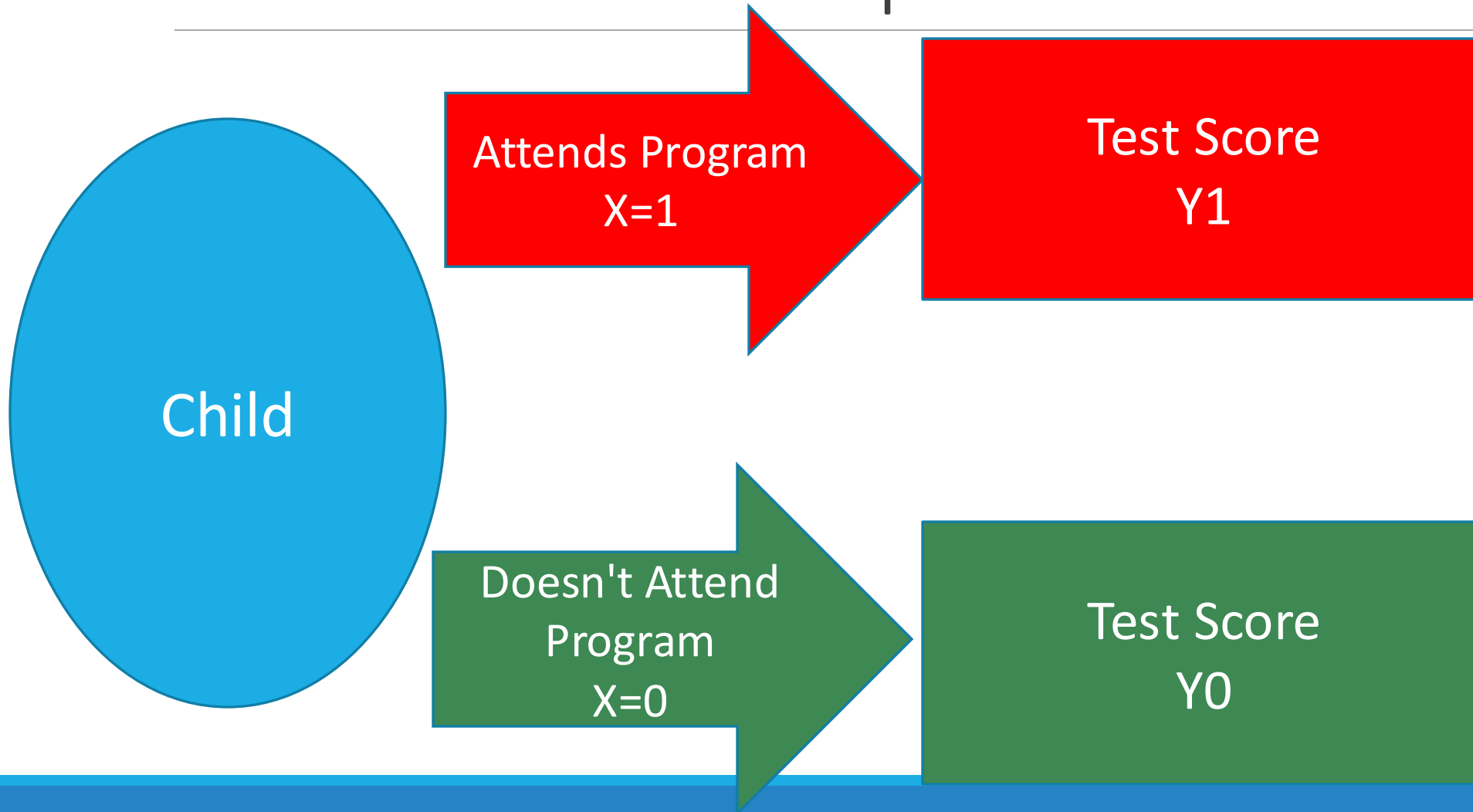
---

Let's say that the causal variable is whether a child attends an after school program. The outcome of interest is their test score on a standardized end-of-year test.

What are the potential outcomes in this scenario?

# Counterfactual Dependence

---



# After School Programs and Test Scores

---

We have a clear definition of the causal variable and a clear statement of the potential outcomes.

Measure the impact of attending an after school program as:

$$Y1 - Y0$$

The model of counterfactual dependence makes this easy, right? Any problems?

- Counterfactual: everything is the same **except for the value of the causal variable**

# After School Programs and Test Scores

---

Remember that a counterfactual world is one where the only difference between the compared units is the causal variable.

Child in a world where they attend the program vs. Child in a world where they don't

The tricky part...

# Alcoholism and Family Dinners

---



How many did a DARE program in elementary or middle school?

A nationwide public-service campaign urged parents to eat dinner with their children.

“There is no more important thing a parent can do.”

“The key to ridding our nation of the scourge of substance abuse.”

Eating dinner as a family **causes** a decrease in the chance that a child will use drugs/develop substance dependencies (the claim said this was reduced by 50%!)

# Alcoholism and Family Dinners

---

In groups:

- Clearly outline the causal claim
- What is the causal variable and its possible values?
- What are the potential outcomes?
- Can we realistically observe both the true and counterfactual outcomes? Why or why not? If not, what are some realistic ways we could **approximate** an observable counterfactual world for a child who does not frequently eat dinners with family?
- Use intuition to point out reasons that this claim may be false. In particular, think about the problems you found in the previous part.

# Summary

---

- We need a coherent and rigid definition of causality to move forward with causal inferences
- Correlation alone is not sufficient (confounding, reverse causation, etc.)
- Counterfactual dependence works.
- Correlation and a meaningful difference in potential outcomes
- Requires clearly outlining counterfactual and potential outcome expectations
- Not a solve-all: limitations on what we can actually observe.