Why do we need causal inference?

We spent a lot of time thinking about how to measure relationships.

However, merely showing a relationship is often not sufficient.

Often, "associations" or "relationships" are not indicative of a causal effect.

$\hookrightarrow$ "Correlation does not imply causation."

Why do we need causal inference?

A typical issue we have to worry about: **confounding.**

Assume we show a relationship between variables X and Y.

But there is another variable Z that **causes** both X and Y.

$\hookrightarrow$ Example: X is a surgical procedure and Y is a score that measures how healthy a person is.

$\hookrightarrow$ What is the likely relationship between X and Y?

$\hookrightarrow$ What could Z be in this example?

Why do we need causal inference?

Typically, we want to know about the **causal effect** of X on Y.

However, results are often discussed in a causal manner when we have reasons to doubt the causal interpretation.

Some examples from a field that is notorious for this: observational epidemiology, particularly studies of diet and health.

Coffee Drinkers May Live Longer

Caffeine is probably not responsible for the healthful effects of coffee.

Eat Natto, Live Longer?

Men and women who ate fermented soy products like natto, miso and tempeh had lower rates of cardiovascular disease and early death.

Story at a glance:

- **Hot dogs could shorten one's life by 36 minutes.**

What is Causality?

We spent a lot of time thinking about how to measure relationships.

For example, at this point you know that simply measuring a regression slope (or a correlation coefficient) isn't sufficient.

We want to know: is an estimate a "good" reflection of the DGP quantity we care about? That is: can we make an **inference** about an estimand from an estimate?

Two fundamental issues of inference:

1. Is our estimate noisy?
2. Is our estimate biased?

What is Causality?

Given the social science discourse of the past 20 years, it'd be understandable to think bias is the more important issue.

That is because people want to know about **causal relationships**.

Over the last 1-2 decades: **credibility revolution** or the **causal inference movement**.

↪ In 2021, there was even a Nobel Prize in Economics for causal inference!

Much of research before ∼2000 (or even before 2010) uses different (read: less credible) standards of inference.

You absolutely must learn to think about social science research from a causal inference perspective.

What is Causality?

Unfortunately: causality is a complex topic!

Our basic definition comes from BdM and Fowler (2021, p. 38):

"a causal effect is a change in $\overbrace{\text{some feature of the world}}^{\textbf{outcome variable } (Y)}$ that $\underbrace{\text{would result}}_{\textbf{counterfactual reasoning}}$ from a change to $\underbrace{\text{some other feature of the world}}_{\textbf{treatment variable } (T)}$."

To think about a causal effect, we need the **potential outcomes**:

— Y_{1i} is the outcome for unit i if $T = 1$

— Y_{0i} is the outcome for unit i if $T = 0$

(Note: this is a binary treatment variable!)

Potential Outcomes

Does assigning a case i to a Republican appointed judge *cause* that case to have a pro-plaintiff outcome?

$\hookrightarrow$ In this situation T is whether the case is assigned to a Republican and Y is whether a case has a pro-plaintiff outcome.

When we ask this question, we want to know $\tau_i = Y_{1i} - Y_{0i}$.

$\hookrightarrow$ This is called the **unit level treatment (or causal) effect**.

If $\tau_i = 0$, there is no treatment effect, but if $\tau_i = 1$ or $\tau_i = -1$, there is a positive or negative treatment effect, respectively.

$\hookrightarrow$ A positive or negative treatment effect is a "causal effect."

Potential Outcomes

Remember: we're committed to stochastic DGPs!

So, even if we see a specific treatment effect for one case, random chance could generate a different treatment effect in others.

We'll typically want to know about *average* treatment effects.

One estimand of interest is the **average treatment effect (ATE)**:

$$\overline{\tau} = E(Y_{1i} - Y_{0i}) = E(Y_{1i}) - E(Y_{0i})$$

There are others: average treatment effect on the treated (ATT), average treatment effect on the untreated (ATU), etc.

$\hookrightarrow$ BdM and Fowler (2021) explains these in more detail.

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	1	1	0
2	0	0	0
3	0	0	0
4	1	1	0
5	1	1	0
6	0	1	-1
7	1	1	0
8	0	0	0
9	0	1	-1
10	0	0	0

Suppose we have a dataset of 10 court cases, whose potential outcomes look as they do in this table.

$\hookrightarrow$ The average treatment effect (ATE) is $\overline{\tau} = -0.2$.

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	?	?	1	1
2	?	0	?	0	0
3	0	?	?	1	0
4	?	1	?	0	1
5	1	?	?	1	1
6	0	?	?	1	0
7	?	1	?	0	1
8	0	?	?	1	0
9	?	1	?	0	1
10	?	0	?	0	0

The **fundamental problem of causal inference**: we never observe all the potential outcomes for a unit.

$\hookrightarrow$ Potential outcome(s) we don't observe: **counterfactual(s)**.

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	?	?	1	1
2	?	0	?	0	0
3	0	?	?	1	0
4	?	1	?	0	1
5	1	?	?	1	1
6	0	?	?	1	0
7	?	1	?	0	1
8	0	?	?	1	0
9	?	1	?	0	1
10	?	0	?	0	0

Can we just ignore the missing potential values and calculate the difference in means of the two columns? That is:

$$E(Y_{1i}|T = 1) - E(Y_{0i}|T = 0)$$

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	?	?	1	1
2	?	0	?	0	0
3	0	?	?	1	0
4	?	1	?	0	1
5	1	?	?	1	1
6	0	?	?	1	0
7	?	1	?	0	1
8	0	?	?	1	0
9	?	1	?	0	1
10	?	0	?	0	0

If we do this calculation, then we get an estimate of -0.2, which is the same as before.

BUT THAT IS JUST A COINCIDENCE.

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	?	?	1	1
2	?	0	?	0	0
3	0	?	?	1	0
4	1	?	?	1	1
5	1	?	?	1	1
6	?	1	?	0	1
7	?	1	?	0	1
8	0	?	?	1	0
9	?	1	?	0	1
10	?	0	?	0	0

Two small changes and we get a biased estimate of zero!

What happened? I didn't tell you this, but the potential outcomes in the first dataframe were *missing at random*; these aren't.

Potential Outcomes

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	?	?	1	1
2	?	0	?	0	0
3	0	?	?	1	0
4	1	?	?	1	1
5	1	?	?	1	1
6	?	1	?	0	1
7	?	1	?	0	1
8	0	?	?	1	0
9	?	1	?	0	1
10	?	0	?	0	0

General lesson, applied to potential outcomes: when data is missing at random, then estimates will be unbiased.

$\hookrightarrow$ Try it in R to see for yourself!

Sources of Bias

As we've seen, if potential outcomes are not missing at random, then you will likely have a biased estimate.

But what does it mean for potential outcomes to be missing at random?

↪ It means that whether a unit ended up in the treatment or control group was (as-if) random!

When would potential outcomes be missing non-randomly?

1. If there is a **confounding variable** that is correlated with the treatment and the outcome variable.

↪ Also known as **omitted variables bias** or **selection bias**.

Sources of Bias

As we've seen, if potential outcomes are not missing at random, then you will likely have a biased estimate.

But what does it mean for potential outcomes to be missing at random?

$\hookrightarrow$ It means that whether a unit ended up in the treatment or control group was (as-if) random!

When would potential outcomes be missing non-randomly?

2. If there is **reverse causality** in the sense that the outcome variable has a causal effect on the treatment variable.

Another example

Let's say we assess the effect of eating 3 pounds of Kale a day on health outcomes. Let the outcome be weather a person is healthy or not. The treatment (T_i) is eating 3 pounds of Kale a day (or more).

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

What is average unit-level treatment effect?

What is the average treatment effect?

Another example

Let's say we assess the effect of eating 3 pounds of Kale a day on health outcomes. Let the outcome be weather a person is healthy or not. The treatment (T_i) is eating 3 pounds of Kale a day (or more).

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

What is average unit-level treatment effect? **0**

What is the average treatment effect? **2/3**

Another example

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

What is average unit-level treatment effect? **0**

What is the average treatment effect? **2/3**

From the unit-level effects, we know that eating Kale does not make you healthier. However, the average treatment effect is positive. Why?

Another example

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

There is a correlation between the potential outcomes and the treatment!

People who are "always" healthy (i.e. $Y_{1i} = Y_{0i} = 1$) are more likely to eat Kale.

Another example

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

People who are "always" healthy (i.e. $Y_{1i} = Y_{0i} = 1$) are more likely to eat Kale.

Potentially, people who are generally in good health want to stay in good health and therefore eat Kale.

$\hookrightarrow$ **Reverse causation**!

Another example

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$	T_i	Y_i
1	1	1	0	1	1
2	1	0	1	0	0
3	0	1	-1	0	1
4	1	1	0	1	1
5	0	0	0	0	0

People who are "always" healthy (i.e. $Y_{1i} = Y_{0i} = 1$) are more likely to eat Kale.

People who have higher incomes can afford to buy more Kale or live closer to a farmer's market; therefore, they are more likely to eat Kale.

However, people with higher incomes are also more likely to be healthy.

$\hookrightarrow$ **Confounding**!

Summary so far

We usually assume that there is an unobserved ("counterfactual") potential outcome for each unit.

$\hookrightarrow$ If the unit is treated: this is what happens if the unit does not receive the treatment.

We call those potential outcomes Y_{1i} and Y_{0i}.

Ideally, we would like to know the following:

$$\tau_i = Y_{1i} - Y_{0i}$$

Which we can average to calculate the average unit-level treatment effect:

$$\overline{\tau} = E(Y_{1i} - Y_{0i})$$

Summary so far

Ideally, we would like to know the following:

$$\tau_i = Y_{1i} - Y_{0i}$$

Which we can average to calculate the average unit-level treatment effect:

$$\overline{\tau} = E(Y_{1i} - Y_{0i})$$

Problem: we never observe both Y_{1i} and Y_{0i} for the same unit.

$\hookrightarrow$ This is called the **fundamental problem of causal inference**.

Summary so far

If the treatment itself is uncorrelated with the potential outcomes, then we can estimate the average treatment effect by simply taking the difference in means of the observed outcomes for the treatment and control groups.

$$\overline{\tau} = E(Y_{1i}|T = 1) - E(Y_{0i}|T = 0)$$

However, if the treatment is correlated with the potential outcomes, then this estimate will be biased.

Bias means: the estimate will not be equal to the average unit-level treatment effect.

Summary so far

If the treatment itself is uncorrelated with the potential outcomes, then we can estimate the average treatment effect by simply taking the difference in means of the observed outcomes for the treatment and control groups.

However, if the treatment is correlated with the potential outcomes, then this estimate will be biased.

We can also say: the potential outcomes are not missing at random.

Summary so far

Two common reasons why treatment assignment is not random:

1. **Confounding**: there is a variable that is correlated with the treatment and the potential outcomes.
2. **Reverse causality**: the outcome variable has a causal effect on the treatment variable.

Ensuring Unbiased Effects

In real research, how do you ensure that potential outcomes are missing at random?

Ideally, you randomly assign the treatment to each unit!

↪ Then the researcher is randomly deciding which potential outcome is visible for each unit.

There are many kinds of randomized experiments out there, but they all involve a researcher *controlling* who gets treatment.

Cheryl's experiments course (POL 290G) is all about how to run randomized experiments. Take it!

Ensuring Unbiased Effects

But it is not always possible to randomize treatment.

The core thing you need to do to try to prevent bias is to prevent the units from "selecting" into the treatment or control groups.

You have to exploit some **exogenous source of variation** that takes control away from the units. (Obv. randomization does this.)

You can search for a naturally occurring source of exogenous variation. This is often called a **natural experiment**.

But this usually comes at a cost, narrowing the scope of analysis.

Ensuring Unbiased Effects

Lauren P.'s causal inference (POL 285) is all about natural experiments.

Definitely take this!

This is a de facto requirement, and may become a de jure requirement at some point in the not-to-distant future.

To take the methods exam, you should take both classes.

Ensuring Unbiased Effects

So far, you haven't seen me talk about "controlling for" confounding variables.

It used to be common-place for scholars to say that one way to solve for omitted variable bias was to control for confounders.

We didn't talk much about this, but this mostly entails adding additional variables to a regression. (Details in POL 213.)

For example if X is a confounder for the relationship between T and Y, then you might run this regression:

$$Y = \alpha + \beta_1 T + \beta_2 X + \varepsilon$$

Ensuring Unbiased Effects

Recall that the sample linear regression slope on an independent variable T is an unbiased estimator if

$$T \perp\!\!\!\perp \varepsilon$$

When you control for confounders, you are (often implicitly) invoking a **conditional independence assumption** (or a **selection on observables** assumption):

$$T|X \perp\!\!\!\perp \varepsilon$$

Because people tended to include a ton of controls to satisfy this, the practice is often (pejoratively) called **kitchen sink regression**.

Ensuring Unbiased Effects

This is no longer considered a good way to deal with bias.

There are three main issues:

1. Controlling for confounders creates an illusion of satisfying the exogeneity assumption. How do you know you've controlled for every confounder? (You can't know this!)
2. Controlling for arbitrary sets of confounders can cause your main estimate to fluctuate wildly. So, your estimate will depend on which specific controls you include.
3. Estimates are much less transparent and more difficult to interpret and understand.

Instead, you have to design your research to satisfy the exogeneity assumption, e.g. via randomization.

Averaging at a Cost

Earlier, I said we average because we think DGPs are stochastic.

That's not the only reason: averaging also helps us mitigate the FPCI by freeing us from having to identify *unit* effects.

As you've seen, this only works when potential outcomes are missing at random; otherwise our estimate will be biased.

But even when they are missing at random, focusing on average effects comes at cost.

Average effects are not the same thing as individual effects.

$\hookrightarrow$ Averaging "destroys" some information that could be useful.

Averaging at a Cost

These yield the same ATEs, but are clearly very different:

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	1	1	0
2	0	0	0
3	0	0	0
4	1	1	0
5	1	1	0
6	0	1	−1
7	1	1	0
8	0	0	0
9	0	1	−1
10	0	0	0

Case	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	1	0	1
2	0	0	0
3	0	0	0
4	1	0	1
5	0	1	−1
6	0	1	−1
7	1	0	1
8	0	1	−1
9	0	1	−1
10	0	1	−1

Estimating **heterogeneous treatment effects** is the process of trying to see these kinds of differences.

Averaging at a Cost

Because you always average, you have to be very careful and precise in describing any effects that you estimate.

You simply *cannot* say that an average treatment effect is uniform across all units.

But, people do try to see if the treatment effect varies within certain subgroups of the sample.

On the other hand, you can't undermine an average causal effect by simply finding a counterexample.

↪ Example: "The Pfizer vaccine doesn't work because my friend was vaccinated and was hospitalized with COVID."

Changing How You Think

Throughout this course, we've talked a lot about faulty inferences.

To remind you: the two major issues are noise and bias.

You know that it's bad to have biased estimates.

The potential outcomes framework (and causal inference more generally) focuses our attention on this issue.

You can't establish a causal relationships with biased estimates!

$\hookrightarrow$ You can have an unbiased estimate that is not causal.

Changing How You Think

It requires a really important change in perspective.

You should *start* a research project by figuring out how you will convince yourself that you will have unbiased estimates.

You do not need any data to do this!

↪ The way you convince yourself is your **causal identification strategy**. We often call research that has a good identification strategy "causally identified" or "well identified."

↪ The "easy" thing to do: run randomized experiments!

This is known as the **design-based approach**.

↪ Main idea behind **pre-registration** and **results-blind review**.

Changing How You Think

But even if you cannot feasibly run a randomized experiment, thinking in these terms will help you ask answerable questions.

Always ask yourself: what is the hypothetical randomized experiment I would run?

↪ If you cannot imagine a hypothetical experiment, then you might have a **fundamentally unanswerable question** (FUQ).

I should be clear here: there's a value judgment built into this phrase, but it's a value judgment that is widely held.

Changing How You Think

One very common reaction to this stuff: "I don't think potential outcomes or causal inference is useful."

Potential outcomes framework reveals many questions we ask about causality are conceptually hazy or just plain weird.

But in reality: people like this because it disciplines them and pushes them to clarify what they mean.

Its narrowness is a feature, not a bug.

↪ Remember: no bullshit!

All is not lost. There are other ways to make progress on FUQs.

Changing How You Think

This all may seem a little abstract, but consider the following IRL example.

What is the effect of a judge's race on their decision-making?

The "treatment variable" is the race of a judge.

The "outcome variable" is how cases get decided by the judge.

What is the hypothetical experiment you'd run in which you randomly assign the treatment?

Hmm....

Changing How You Think

Another common reaction: "I am not trying to establish a causal effect, only an association/correlation."

This often verges on gaslighting.

In context, it is often clear the researcher wants a reader to infer a causal relationship. Otherwise, why would we care?

↪ Think about any relationship you care about. Would you feel satisfied if someone told you the effect was **spurious**?

Note: the word "effect" implies a causal relationship.

There's still lots of non-identified work out there; be careful evaluating it.

Causal Inference and Generalizability

A somewhat more interesting criticism of the causal inference approach to social science goes something like this:

"Sure, causal inference allows us to estimate unbiased effects, but the questions you can ask are too narrow to be interesting."

In other words, causal effects have **internal validity** (unbiasedness) but not **external validity** (generalizability).

So, you see that judge partisanship affects decision making in a sample of civil rights cases.

$\hookrightarrow$ But we really want to know if politics *generally* affects judging.

Common retort: why should we want to generalize bullshit?

Concluding Remarks and Advice

Quantitative social science is hard.

In this class, I hope you've gotten a flavor for what is ahead of you.

I also hope you have built a solid foundation in probability, statistics and R that you can keep building your toolkit.

Your job going forward is to learn how to both evaluate and produce research. To think clearly with data.

I highly recommend reading BdM and Fowler (2021) cover to cover, if you haven't already.

Always happy to help you plan your future coursework. And let's talk if you want to do the methods field exam.

Concluding Remarks and Advice

What we do is important, so...

no bullshit!