Lecture 5 – Causality

The Problem

A student starts going to the tutoring center and her grades go up. Did tutoring cause it?

To answer that we would need to know what her grades would have been if she hadn't gone, and we can't observe that world because she went. Maybe she was already turning things around, or maybe whatever pushed her into the tutoring center would have pushed her grades up on its own.

Does Robert Frost understand this?

> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

How does he know?

Potential Outcomes

This thinking is formalized in the **potential outcomes framework**, which underlies most modern empirical work in economics. Here is the notation:

> D_i is whether student i was treated. $D_i = 1$ if they were tutored, $D_i = 0$ if not.
>
> Y_{i1} is student i's GPA *if they are tutored*.
>
> Y_{i0} is student i's GPA *if they are not tutored*.
>
> Y_i is the GPA we actually observe.

Both Y_{i1} and Y_{i0} exist for every student, but we only ever see the one that happened.

$$Y_i = D_i \cdot Y_{i1} + (1 - D_i) \cdot Y_{i0}$$

The causal effect of tutoring for student i is the difference between her two worlds, $Y_{i1} - Y_{i0}$. Notice what that requires – for every student, one of the two numbers is missing.

Student	Y_{i1} (tutored)	Y_{i0} (not tutored)	Effect
Maria	3.4	?	?
Devon	?	2.8	?
Carlos	3.1	?	?
James	?	3.3	?

The Effect column is empty and it will stay empty. We will never recover the causal effect for an individual student, which is the **fundamental problem of causal in-**

ference. This is not a data problem and more data will not fix it. What we can do instead is learn about averages across groups.

Think about this in your own life. How would your career be different if you had gone to a different university? Can you know the answer to that?

Selection Bias

So we give up on individual students and compare groups instead. If we take the average GPA among tutored students and subtract the average GPA among untutored students, does that give us the effect of tutoring?

Only if the two groups would have looked the same without tutoring. Would they?

Who shows up at the tutoring center? If it is students who are struggling, then the tutored group would have had lower grades anyway and the comparison makes tutoring look worse than it is. If it is the conscientious students who go to office hours and do the reading, then the tutored group would have had higher grades anyway and tutoring looks better than it is.

Both stories are plausible and they point in opposite directions. We call this difference in the outcomes based on who chooses to be treated **selection bias**. A simple comparison of means gives us the causal effect plus selection bias, and there is no way to separate the two.

A Puzzle

A puzzle: across students, more grant aid is associated with *lower* graduation rates. More grants, less likely to finish.

Should a policymaker conclude that grants hurt students? Why or why not? Who gets grants?

Why might two things move together?

When you see a correlation there are at least four things it could be.

> X causes Y.
>
> Y causes X, or reverse causality. Do small classes raise achievement, or do districts put students with high performance in small classes?
>
> Something else causes both, which we call omitted variables. Family income affects both the grant and the graduation rate.
>
> They are determined jointly in a market. Teacher salaries and teacher quality are set together in the same labor market.

When you read a study, ask which of these you are looking at.

Randomized Controlled Trials

RCTs fix this problem of selection bias, because people don't get to decide whether they are treated.

It solves the *identification problem*. This means, identifying a causal effect. You may also hear discussion of "identifying assumptions," this just means what do you have to assume to get a causal effect.

If we flip a coin to decide who gets tutoring, then who gets tutored has nothing to do with who they are, and the struggling students and the conscientious students end up split evenly across both groups.

Why does this work?

It relies on the Law of Large Numbers, which states that if we sample from a population our sample average will converge to the population average as the sample grows. If we randomize people into two groups, both groups converge to the same thing.

This includes things that are easy to observe like prior GPA or free lunch status, and we could have controlled for those anyway. The real magic of RCTs is that the Law of Large Numbers also balances the things we *cannot* observe, like motivation and family support. Nothing else in this unit does that.

Checking it worked

One of the important things to check in an RCT is that predetermined covariates are "balanced," meaning that treatment and control look similar. You do this by comparing the two groups on characteristics measured *before* treatment, such as prior test scores, demographics, and attendance.

If they don't look similar, something went wrong with the randomization, and you should wonder what else went wrong with it.

Estimation

Estimating the effect just means comparing two averages. People usually run a regression to do it, which we will cover next time, but the answer is the same.

Researchers often add controls for pre-treatment characteristics. They don't need to, since randomization already handled that, but it generally increases statistical precision.

What an Experiment Does and Doesn't Tell You

Internal vs. external validity

Internal validity is about how well the study establishes the causal relationship in the setting studied. Were the randomization rules followed? Was the data collected in a reliable way? A well run experiment should always be internally valid.

External validity is how generalizable the result is. Suppose UT Austin runs an experiment offering extra aid and finds no effect on graduation. Does that tell us what would happen at Austin Community College? At UT San Antonio? At a for-profit?

Charter school lotteries are a classic case; the lottery is internally valid because the randomization procedures are followed, but only certain families apply to charter schools. What the lottery tells us is the effect of a charter school among students who tried to enroll, which may not be the effect on anyone else.

You can only know the effect of the thing that was randomized

I often summarize this by saying, "You can only know the effect of the thing that was randomized."

Treatment is usually a bundle. A program gives students extra aid *and* an advisor *and* a summer orientation, so when you randomize the bundle you learn about the bundle. You cannot separate out the advisor unless the advisor was randomized separately.

The same issue shows up downstream. Suppose a job training program raises both earnings and community college credits. Can you say the earnings went up because of the credits? (No, credits were not randomized) You can say the program raised both.

Spillovers

We usually assume that treating one person does not affect anyone else. If you tutor half a classroom, the tutored kids may help the other kids learn.

If that happens, would comparing the tutored students to their classmates uncover the causal effect of tutoring? Would it be an underestimate or an overestimate?

This should be considered when you design the experiment. Should treatment be at the student level, the classroom level, or the school level?

Multiple testing

If you look at twenty outcomes, you should expect one of them to come back significant by chance, because we typically construct 95% confidence intervals. Splitting the sample twenty ways causes the same problem.

A few defenses:

> Use theory to tell you where you expect to find effects, rather than running many specifications and then deciding on your theory.

Pre-specify your analysis plan. The AEA maintains a registry where researchers describe their outcomes, populations, and how they will handle multiple testing https://www.socialscienceregistry.org/

Use a multiple testing correction.

Use common sense and don't torture the data to make your preferred story appear.

Ethics

Some experiments are unethical, and historically researchers have run unethical experiments including the Tuskegee syphilis experiment and others. As a result, Institutional Review Boards now approve research involving human subjects before it begins. Economics also employs a no deception policy, which means researchers cannot lie to subjects and makes researchers credible in the long run.

There is also a reason to randomize that has nothing to do with research. When access to a program is limited by funding, as with charter seats or a scholarship with forty slots and two hundred applicants, randomization has two benefits – easy evaluation and fairness.

Wrapping Up

Every student has two **potential outcomes** and we only ever see one, which is the fundamental problem of causal inference.

Comparing people who chose treatment to people who did not gives you the causal effect **plus selection bias**.

Randomization removes selection bias, including on the things you cannot measure.

An experiment tells you the effect of **what was randomized, on whoever was in the study**. Internal validity is about the first part and external validity is about the second.

Most education policy questions cannot be randomized. Nobody is going to randomly assign school funding, teacher certification rules, or a state accountability system. The rest of this unit is about what you do then.