**Black Introduction to Main Workshop Themes**

Mon 7 August 2023

**I want to offer a preview of some of the takeaway lessons for the week**

**Message 1: Keep learning.** The methods for Research Design for Causal Inference continue to evolve. For the graduate students in the room, it’s amazing how many people never learn much in the way of empirical methods beyond what they learned in graduate school. Their training gets stale. If you keep learning, you’ll have a big comparative advantage, that will grow over time.

There is a lot of stuff that I would do today that I didn’t know *how to do* 5 years ago, or three years ago, or didn’t know I *should do*. There are things that I would *not do* today, or would consider insufficient, that I thought were fine 5 years ago, or 3 years ago.

There is stuff I’m experimenting with today, that may make it into published research or not, depending on how these efforts turns out.

**Related Message 2: Let your research problem guide your analysis. Don’t assume a canned method will work.** The *goal* is credible causal inference. The methods often reflect the goal, plus the strengths and weaknesses of the data you’ve got. Good research design is about having a research project in mind, and then understanding what method, or combination of methods, will move you toward credible causal inference. I’ll offer a personal example or two later this week.

Although, sometimes you start with “here is a **shock** to the world, that creates a nice natural or quasi-experiment. Is there an interesting research question it can help to answer? Some research starts with the research question, and the researcher needs to find a basis for credible causal inference. But some research starts with a basis for causal inference, and the researcher needs to decide whether that can help to answer an interesting research question.

And some research questions don’t have causal answers. As Don stressed this morning, to have credible causal inference, you need to have several things:

1. A control group

2. A shock to the treatment group, but not to the control group

Ideally an external shock, close to randomly assigned

If you have 1 and 2, then you have a **chance** of constructing a credible estimate for the missing potential outcomes for the treatment group; and sometimes for the control group also. How good a chance – that’s what the rest of this week is about.

**Message 3: “Everyone does it” is *not* a justification for using a poor method.** Even if it is in fact true that everyone else is using said method, which happens with some frequency. We’ll see examples during the week where everyone who uses a standard method is wrong. Eventually someone writes a paper saying so, and blows up lots of highly placed papers.

**Message 4: Assumption free estimation doesn’t exist.** We’ll devote a lot of effort to making assumptions we make as *weak* and *realistic* as we can, but one can’t avoid them. The best you can hope for is to know which assumptions you are making, not make them any stronger than you need to, and think carefully about whether they are plausible, and how you can convince a skeptic that they are plausible.

The first and most important skeptic is yourself. Credible causal inference depends on *believable* assumptions. Lots of times, researchers don’t realize what assumptions they are making, let alone defend them. Other times, they don’t believe the assumptions themselves, but hope that the referees and journal editors will let them get away with them. Sometimes they are right about that.

**Message 5: LATE is as good as it gets**

Code words: LATE = **local** average treatment effect. In the Angrist-Imbens-Rubin paper on causal IV analysis, this concept is called CACE for Complier Average Causal Effect; but LATE is the term that has stuck.

“treatment heterogeneity” is what one worries about.

Drug good for men may not work for women. Statins are a good example. Drug good for whites may not work for blacks; Drug good for young adults may have strong side effects for the elderly, or maybe the elderly need a lower dosage to limit side effects. Or a larger dose, as is the case for flu vaccine, and probably COVID vaccine too. And so on.

All credible causal inference is local. Sadly.

**Related Message 6: Distrust OLS** and most other regression methods. They **ignore** treatment heterogeneity. Sometimes, they provide unbiased estimates of an **average** treatment effect anyway; sometimes not. Sometimes regression gives you an estimate of the average treatment effect for the *whole population*, sometimes it gives an estimate of the average treatment effect for the *treated*; but often it provides neither. Often, regression methods, especially the fancier ones, embed strong assumptions you didn’t know you were making, that likely aren’t true.

You can’t **assume** that regression is giving you a sensible estimate, or that a regression coefficient will answer the question you think you are asking.

**Related Message 7: Credible internal validity is possible. External validity rarely is.** And never from a single study.

***Internal validity*** = You have a credible estimate of a local causal effect of a treatment on an outcome, for your sample. Or sometimes a subset of your sample.

***External validity* =** You have a solid basis for believing that you can extrapolate from your sample to a larger universe.

**Message 8: We don’t live in Asymptopia**

Asymptotically unbiased does not mean unbiased, especially for small samples. Methods for infinite samples are of little use unless they converge soon enough.

Some do: The sample mean converges to the normal distribution (that’s the central limit theorem) and does so reasonably rapidly in most real-world datasets.

Some don’t: Famous example: Paper by Josh Angrist uses quarter of birth as instrument for whether kids graduate from high school. Sample size ~300,000. They authors surely thought the asymptotics for instrumental variables had kicked in. They hadn’t – which is now known as the weak instruments problem. Eric French will discuss this paper, and the weak instruments problem on Friday.

Another example: Alberto Abadie and Guido Imbens have an important paper which shows that bootstrap methods for estimating standard errors, which are used for many empirical methods when you can’t directly compute the standard errors, **don’t work** for most matching methods – they don’t converge quickly enough to be useful.

**Message 9: Randomization is your friend, when you can find it or create it.**

Core goal of causal inference: someone does something to someone, and we want to obtain a credible estimate of the average effect on the “treated”, or perhaps on the whole sample.

This is very hard without a proper “control” group, often impossible. Randomized experiments are great, when you can find or create them, because they provide a good control group. Although you have to make sure the randomized experiment is really random; sometimes it isn’t.

Conversely, unobservables are your (hidden) enemy. The risk is indirect causation through unobservables. Randomization ensures that unobservables are balanced *in expectation*. Even then, they may not be balanced *in fact*, especially in a small sample.

If you have a randomized experiment and you have chance imbalance on observables, you can re-randomize to improve balance. How to do that – an advanced topic, that we won’t cover. The core insight is that if you start with a randomized experiment, then even if you re-randomize to improve balance on observables, you still get balance **in expectation** on unobservables. Maybe even better balance than you would otherwise get, if the unobservables correlate with the observables.

In contrast, if you have an observational study you can – and usually **should** -- balance the sample on observed covariates, and then **hope** that you then also have balance on the unobservables, but hope is all you can ever have, never proof.

Although “bounds” methods for estimating how likely it is that unobservables explain your observed results can be valuable. Another important topic we won’t reach, but you ought to know exists.

**Message 10: Rules, and changes in rules, are often your friends too**

If you can’t run a randomized experiment, rules can sometimes offer a good substitute. They often have hard edges, which often create good treatment and control groups. We call these “quasi-experiments” or natural experiments.

Many of the best observational studies rely on legal shocks – changes to legal rules; or on discontinuities in rules. But: you always need to check for what is known as “**covariate balance**”: Are the treated and controls really similar?

With a warning that covariate balance is the term that is used, but it is a poor term. If you use panel data, you will also usually want pre-treatment balance on the **outcomes**.

You need to check for “**common support**” – do the treated and controls live in the same covariate space? If you don’t have covariate balance, including common support, trim your sample until you have reasonable overlap. Your estimate will then be **local** to the region of common support. Sometimes, there won’t be a respectable sample left. That’s too bad, but if that happens, you didn’t have a respectable sample for the study to begin with – you only thought you did.

Conversely, continuous “instruments,” that aren’t based on shocks of some kind, are rarely credible. It’s too easy for them to correlate with an unobservable and thus indirectly predict your dependent variable.

**Message 11: Graph your results.** Some of my favorite papers are ones that one where one can summarize the main idea in a single graph, or maybe two. Sometimes, you can see things in graphs, and scatterplots, that you’ll never see any other way.

Including problems with the data, that need to be addressed.

**Message 12: Care with data:** Easy to say, but hard to do. **Really understanding** your data will pay off, surprisingly often. Graph it. Check for missing data and outliers. Decide what to do about the missing data and the outliers, if anything. We won’t formally discuss either topic in this workshop, but Don Rubin has also developed central methods for handling missing data; see his book with Rod Little. Methods for assessing and reducing sensitivity to outliers exist, and are often worth using.

I’ll give a lunch talk later this week, on bloopers due to bad research design. I might have given a second talk on bloopers due to bad data, that researchers didn’t realize was bad.

I’ll stop there, a dozen lessons is enough.

Well, actually, I won’t quite stop there.

A wall with many images on it

Description automatically generated

This is a photo I took, a few months ago, of a wall in Jerusalem, with photos of the hostages that Hamas took. The Hebrew reads:

Men, women, children, elderly.

Still hostages of Hamas

Then the English: Bring them home now

Slide prepared for talk not long ago at NBER, in Boston. Appropriate at Northwestern too.

I wanted to not be silent.

I am a dual US-Israeli citizen.

I’m not religious. I wore a kippa today to celebrate Israel’s existence.

We (Jews) aren’t going anywhere. Not in the U.S., and not in Israel.

And yes, anti-Zionism is antisemitism.

Not least because the Jews in Israel have nowhere else to go.

Israel, and the vast majority of Israelis, want peace.

Hamas and Hezbollah want war.

So war we have. Israel’s job is to beat Hamas, so that we don’t have another war in 10 years, and in 20 years. A cease-fire now, on the terms that Hamas wants, is a victory for Hamas, and a guarantee of another worse war down the road. No one should want that, especially not anyone who cares about the lives of both Palestinians and Israelis.

I’ll be happy to say more, offline, for anyone who is interested.