**Monday morning introduction**

**Zoom logistics**. This is our third time now, offering this workshop as a hybrid: in-person and Zoom. We’re doing our best to make the Zoom part go smoothly, including learning from what didn’t go smoothly last year.

During each session, please send questions through the chat. Scott Cunningham will monitor the chat, answer himself the questions he can answer, and will ask the speaker the others -- as time permits. Scott will use his judgment on whether and when to ask a question – please recognize that he can’t convey every question to the speaker. This is true for the in-person attendees as well; the speakers will exercise judgment on when to take questions, and when to move on.

For any periods where Scott is not available, the Zoom will be monitored by post-doc who is working for me, Sadia Farzana, **[\*introduce her]**. That’s for Monday-Wednesday. I have another post-doc, Andy Yuan, who will handle Thursday and Friday.

**[\*Introduce co-organizer, Scott Cunningham.]**

**Logistics**

***Friday presentations:*** We will run parallel sessions, on Friday from 1:30-5. Details to come. At least one will be open to remote attendees. We have invited you to sign-up by email. Please email me or Sebastian Bujak directly if you want to present and have not already signed up. **Deadline**: **tomorrow afternoon**. Sooner if possible. With a title, and ideally an abstract.

Session leaders:

**Friday session leaders**

Me

Scott Cunningham

Eric French

Josh Lerner

Dhruv Aggarwal [if needed]

I encourage those of you who can to attend one of the sessions, even if you are not presenting. From experience, we expect that you will find the sessions valuable.

**Reception today:**  after the afternoon session, starting roughly 4:30: In the central law school internal courtyard, unless it’s raining – **[\*Bernie to point to roughly where]**.

Reception also on Thursday, same logistics.

And again for the advanced workshop, next Monday, for those of you who are attending both.

**Advanced workshop:** For thoseofyou are staying for the advanced workshop. There will be some overlap in topics, especially for difference-in-differences and instrumental variables, where the advanced workshop will cover topics that didn’t fit in the one day we allotted during the main workshop. Integration across different speakers is tricky, but we’ll do what we can. The speakers at the advanced workshop will know what was and wasn’t covered this week.

***Breakfast, coffee, tea, and snacks:***  In Rubloff RB 155 right outside our room **[\*Bernie to point to where]**.

***Lunch****.* There will be a vegetarian option every day. More than that I can’t tell you; but Sebastian can.

***Rest rooms***. **[\*Bernie to point to where]**.

**Internet and zoom:** If you have any issues, please mail me, Sebastian, or [lawAV@law.northwestern.edu](mailto:lawAV@law.northwestern.edu). We will have a law school AV person here throughout the workshop, to address issues as they arise; who will also monitor this email address.

***Zoom links***: Zoom link should have been provided to all attendees separately by email. Anyone who didn’t get one isn’t listening to me now, but hopefully that is a null set.

All sessions will be recorded, but the video won’t track the speaker if the speaker walks around the room.

***Recordings links*:** All sessions will be recorded and posted.Separate link for each day; will be provided as soon as we can, usually by the next morning.

***Keeping on Schedule.*** We **won’t**. We may run long. Please expect this. The morning sessions nominally end at noon, but in practice, we will often go to 12:15 or 12:30.

The plan is roughly three academic hours in the morning and the same in the afternoon. With one or two breaks in the middle, depending on speaker preference. Plus separate Stata and R sessions at the end of the day Tuesday-Wednesday-Thursday; and at lunch on Friday.

Often with a lunch talk – a brief overview today from me; tomorrow by Don Rubin, Wednesday from Scott, and Thursday from me.

***Lunch:* Will be outside this room.** We will have a vegetarian option for those who want this. For those of you who will eat fish or chicken but not meat, there might be something on a particular day; if not, we have vegetarian.

***Readings, slides and such:*** Readings and slides either are or will be posted to the workshop site.

[**Readings**](https://nuwildcat.sharepoint.com/sites/LAW-Shared-Folders/Shared%20Documents/Forms/AllItems.aspx?ga=1&id=%2Fsites%2FLAW%2DShared%2DFolders%2FShared%20Documents%2FBBlack%5FShared%2FCausal%20Inference&viewid=d7ce2306%2D55d5%2D4127%2D8675%2D3f2bf41a3754)**:** Go to the workshop website at <https://www.law.northwestern.edu/research-faculty/events/conferences/causalinference/>

**[show navigation on screen]**

We’ll do our best to keep the presentation slides and other materials updated on the website. The slides for today are posted. We can make paper copies for those who want them. Please ask by emailing Sebastian Bujak.

We **know** you like to have access to the slides in advance. Sometimes that happens, sometimes the slides arrive “just in time.” When we get them, we post them promptly.

We’ve posted the suggested readings to the workshop website, except for full books, like the Imbens and Rubin book. Along with much else. We realize that the readings are more than you can plausibly do during the workshop. The books you will have to buy, but you should, and the prices are sensible.

***Internet access.*** Most of you should have access through Eduroam. For those who don’t, there is a guest network called, of all things, Northwestern Guest, that you can access.

***Rest rooms:* [\*Bernie to point to locations for large and small ones]**

***Logistics during the workshop:*** Email me, Sebastian Bujak, Scott Cunningham, or some combination **[\*emails on board]:**

[bblack@northwestern.edu](mailto:bblack@northwestern.edu) [bblack@law.northwestern.edu and bblack@kellogg.northwestern.edu all go to the same mailbox]

[sebastian.bujak@law.northwestern.edu](mailto:sebastian.bujak@law.northwestern.edu)

[scunning@gmail.com](mailto:scunning@gmail.com)

[causalinference@law.northwestern.edu](mailto:causalinference@law.northwestern.edu)  should go to both me and Sebastian

If something is not right during the workshop, and you tell us early enough, we might be able to fix it. No guarantees, but we’ll do what we can.

Remembrances:

Mat McCubbins: my co-organizer for 2010-2019. He died in 2021.

Daniel Black: My son, who died suddenly in September2021, at age 15. No more to be said. But saying this much and showing a photo is part of how I remember him.



**Introduce faculty:**

**Scott Cunningham:** Co-organizer. Scott has his own Mixtape book, if I can call it that, which is free online. The paper copy looks like this. Scott will say more about this project on Wednesday. **[\*show book]**

We try to choose our speakers well, and when we find a good one, we often invite them back. The speakers are world class at what they do, and I predict that you will like them a lot.

**Don Rubin:** With that,I’m delighted to be able to introduce Don Rubin as our first day speaker. Don is a grandmaster of causal inference. He developed a number of central methods that are in common use today starting in the 1970s, and has been active ever since, in expanding the boundaries of what counts as good research design. Much of what we will do builds on what is often called the Rubin Causal Model, which treats causal inference as a missing data problem, in which, to estimate treatment effects, we must impute a missing **potential outcome**. He will say much more about that concept.

In 2021, Don came as close as it is possible to come to winning the Nobel Prize in Economics without winning it. Solely because, as best I can tell, his professional degree is in statistics rather than economics. But his work is central to our understanding of how to conduct causal inference with observational data.

His work includes the core idea, from his 1973 paper, that you **can** **do** serious causal inference with observational data. Jerzey Neyman developed the potential outcomes notation in the 1920s, but didn’t apply it to observational data.

It includes Don’s work in the 1980s with Paul Rosenbaum developing the concept of the propensity score. And his work in the 1990s with Josh Angrist and Guido Imbens on causal instrumental variable analysis. And much more. The Nobel prize committee decided to award the Nobel prize to Angrist and Imbens, and to cite two earlier Angrist and Imbens papers, instead of the Angrist-Imbens-Rubin paper, but the 1996 Angrist Imbens Rubin paper is the one you should read.

Let me say a few words about what Don and I hope you take away from his lectures – and those of our other speakers. What I find most valuable is learning how Don *thinks* about causal inference. What are the core problems to be solved? How can one think about them? That understanding doesn’t come all at once, and you need to listen for it, rather than for specific methods.

You will get a flavor for Don’s thinking today. His approach to thinking about causal inference and research design is also reflected in his book with Guido Imbens. We’ve assigned some chapters as suggested reading for today, and more tomorrow, but the full book will repay careful reading.

My major regret is that, in the interests of finishing the book, Guido and Don did not cover difference-in-differences (DiD, I will call it) or regression discontinuity designs (RD or RDD, although the name is a bad one, because RD designs are not really about regression at all).

But not to worry about what’s not in the Imbens and Rubin book: Yiqing Xu, our speaker Wednesday, is one of the leaders in a flurry of recent DiD work on DiD designs which has clarified when the classic approaches work, and when they need to be revised. So is Andrew Goodman-Bacon, who will be our advanced DiD speaker next week.

**Overall approach: emphasize intuition.** Throughout the week, we’ve asked our speakers to emphasize intuition, and go light on algebra and formal proofs. How far to tilt in that direction is a judgment call, that each speaker makes for himself, with some guidance from me and Scott. Part of what I want to learn myself from these workshops, and hope you can learn, from Don Rubin and our other speakers, is how people who are really good at this stuff *think* about causal inference, and what they think the important moves are, much more than particular proofs.

Don Rubin will tilt especially toward intuition, both because he goes first and because that is his style. In our experience, that will be just right for many of you, but some will want more algebra. Some of you will be thinking, this is too slow and simple for me. My advice: Don’t worry. There will be plenty of more technical “meat” to follow. Listen to how Don Rubin *thinks* about causal inference. There is plenty of math in the Imbens and Rubin book.

**Some cautions and comments**. We’ll cover a lot of ground, we’ll push you pretty hard, and you won’t get everything in one go. You’ll need to read more, both now and later.

We make judgments, and the speakers make judgments, about what to cover, and what not to reach. We cover as much as we can in one week, but there will be a lot we don’t reach. If you want to know more about Topic X, please ask me or Scott. We might know something.

***Asking questions.*** A request as we go along. We’re pushing our speakers hard too – 5 or 6 lecture hours is a lot. I want to **insist** that you respect the breaks between the morning sessions and between the afternoon sessions, and not use them as question time. Our speakers need the breaks. Better times to ask questions will be at the end of the morning, during lunch, or at the end of the afternoon sessions.

During the sessions, **clarifying** questions can be valuable. But before asking a question, please ask yourself, is this question mostly about my particular research project, or is it of broad interest to the group. For narrower questions, please hold them for lunch or after the sessions; email is a good option, too.

If you can’t ask the speaker, email me and Scott, and we’ll try to provide answers, and revert to the speakers when we can’t.

**Stata, R, and Python materials:** A recurring request from many attendees is: Give us specific code to implement the methods you are discussing. What we do this year is a combination of things:

First, many of the speakers will build code into their slides, and/or provide some on screen examples.

For RD, two wonderful primer books by Matias Cattaneo and Rocio Titiunik are posted to the readings, and have lots of code in both languages; they have also written Stata and R packages.

For advanced DiD, there are Stata and R packages from Yiqing Xu and, at this point, at least a half-dozen other sets of authors that provide somewhat different approaches; Andrew Goodman-Bacon will discuss these and other approaches in the advanced workshop. You can ask me or Scott as well. Scott knows this area much better than I do, so you might ask him first. But I know something, including about which approaches **don’t work** if you have a nonstandard research design, which I have for a current project.

After each day, we will give you a set of related coding questions, and ask you to look at them overnight as your energy level permits. Answers will be supplied the next day, with brief afternoon discussion by me for Stata, and by my colleague and coauthor Josh Lerner for R. We will also experiment this year by providing Python based solutions as well.

Beyond that, for any of the methods discussed in class, if you have a specific question, please email me and Scott and we will do our best to provide code, a paper, or both, that uses this particular method. But that won’t be in real time.