SISCER Module 13
Causal Inference with Observational Data: Common Designs and Statistical Methods

Ting Ye & Qingyuan Zhao

University of Washington & University of Cambridge

July 23, 2025

About Ting

Ting Ye, PhD (she/her); I am an Assistant Professor in Biostatistics at UW

- Research: causal inference (observational studies and clinical trials)
  - I develop new methods and apply causal inference methods to answer scientific questions
- Background
  - PhD in Statistics, University of Wisconsin-Madison
  - Postdoc in Statistics, UPenn

About Qingyuan

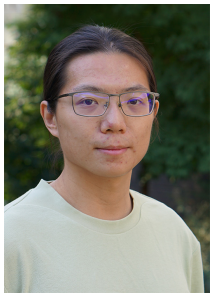
Qingyuan Zhao, PhD (he/his); I am a Professor of Statistics in the Department of Pure Mathematics and Mathematical Statisics at the University of Cambridge.

Research interest

Causal inference, sensitivity analysis, selective inference. I am increasingly interested in how "design"—a fundamental yet elusive principle—shapes the practice of statistical applications in biomedical and social sciences.

Our fabulous TA

Yuhan Qian; PhD student in Biostatistics at UW.



Who are you?

We have a wonderful diverse set of background here

- Bioinformatics, Pediatrics, Epidemiology, Oncology, Environmental Health Science, Medicine, *definitely others...*
- Academic institutions, health providers, biopharmaceutical companies.

Learning goal and learning process

The objective is to study methods for **designing and analyzing** observational studies to make them **provide more reliable evidence**.

To achieve this goal, we will

- Learn common designs of observational studies and how to analyze them
- Gain experience from practicing with real data

You are highly encouraged to apply the methods to your own data!

What are we going to learn?

6 lectures over 3 days

- Lecture 1: Randomized experiments and randomization inference.
- Lecture 2: Matching and balancing weights for observational studies.
- Lecture 3: Sensitivity analysis; Intro to DAG models.
- Lecture 4: Estimation methods under no unmeasured confounding.
- Lecture 5: Instrumental variables & Mendelian randomization.
- Lecture 6: Difference-in-differences; time-varying exposures.

Achieving our goals

Each day will include two sessions (30min break in between), each structured as follows:

- 50 minutes lecture
- 30 minutes practical data analysis with R
  - We will use breakout rooms so you can work in small groups
- 10 minutes discussion of exercises

We will assume basic background in statistical inference and some knowledge of R.

Please note: the 3 day course moves quickly. So, **please ask questions!** – as "chat" or "raise hand" in Zoom calls, or on our Slack channel.