BSTA 7900: Causal Inference in Biomedical Research
Fall 2025

Instructors:
Nandita Mitra, PhD
Professor of Biostatistics
Co-Director Center for Causal Inference
622 Blockley Hall
nanditam@pennmedicine.upenn.edu

Wei (Peter) Yang, PhD
Professor of Biostatistics
632 Blockley Hall
weiyang@pennmedicine.upenn.edu

Teaching Assistant:
Jimmy Kelliher, MS
PhD Candidate, Biostatistics
Jimmy.Kelliher@Pennmedicine.upenn.edu

Class time: T/Th 10:15 – 11:45am
Location: John Morgan, M100

Office Hours:
Instructors: By appointment
Teaching Assistant: 2 hours/week (time and location will be posted on Canvas)

Overview:
This course considers approaches to defining and estimating causal effects in various settings. The potential-outcomes approach provides the framework for the concepts of causality developed here. Topics will include: the definition of effects of scalar or point treatments; nonparametric bounds on effects; identifying assumptions and estimation in simple randomized trials and observational studies; alternative methods of inference and controlling confounding; propensity scores; sensitivity analysis for unmeasured confounding; graphical models; instrumental variables estimation; joint effects of multiple treatments; direct and indirect effects; intermediate variables and effect modification; randomized trials with simple noncompliance; principal stratification; effects of time-varying treatments; time-varying confounding in observational studies and randomized trials; nonparametric inference for joint effects of treatments; marginal structural models; and structural nested models.

Recommended books:
Most of the course readings will be from journal articles, but the following books provide useful background information.

Hernán MA, Robins JM (2022). Causal Inference. Boca Raton: Chapman & Hall/CRC.

Pearl, J. Causality (2009). Models, Reasoning and Inference, Second Edition. Cambridge University Press.

Ding P. (2024). A First Course in Causal Inference. CRC Press.

Grading:
70% homework (4 assignments)
30% final written project and oral presentation

Students are encouraged to discuss homework problems collaboratively; however, all submitted work, including written responses and computer code, must be completed independently and reflect each student's own understanding. The use of generative AI tools (such as ChatGPT, Copilot, etc.) to produce or edit any part of the assignment is not permitted.

All course materials will be placed in the Course Canvas site.

Lecture Schedule

DATE		TOPICS	Instructor
Aug	26	Overview: potential outcomes, causal effects	Mitra
	28	Observational studies: confounding and causal assumptions	Mitra
Sep	2	DAGs	Mitra
	4	DAGs: do-calculus	Mitra
	9	DAGs: backdoor path criterion	Mitra
	11	Matching	Mitra
	16	Matching	Mitra
	18	Matching	Mitra
	23	Inverse probability of treatment weighting	Mitra
	25	Doubly robust estimators	Kelliher
	30	Trial emulation and RCTs with noncompliance	Mitra
Oct	2	Principal stratification	Mitra
	7	Instrumental variables	Mitra
	9	Instrumental variables	Mitra
	14	Diff-in-Diff and synthetic controls	Hettinger
	16	Interference	Kelliher
	21	Direct and Indirect Effects	Yang
	23	Direct and Indirect Effects	Yang
	28	Time-dependent confounding overview / g-methods	Yang
	30	G-formula	Yang
Nov	4	IPTW and marginal structural models	Yang
	6	IPTW and marginal structural models	Yang
	11	MSM doubly robust estimation	Yang
	13	Structural nested models	Yang
	18	Dynamic treatment regimes	Yang
	20	Failure time analysis	Yang
	25	Bayesian causal inference	Oganisian (by zoom)
	27	Thanksgiving	No Class
Dec	2	Student presentations	
	4	Student presentations	
	9	University reading day	No Class
	11	Student presentations (if needed)	

Final Project

Written:

Due Monday Dec 15

- Choose a causal inference topic / method that we haven't directly covered in class.
- The topic must be approved by course instructors via email.
- Read 3 journal articles on the topic (perhaps 1 main article that you focus on, but a couple of others to give you broader knowledge of the topic).
- Write a short paper (generally 5-8 pages) summarizing the work. You should clearly describe the following:
    - For what types of studies would this method be applicable?
    - What are the potential outcomes?
    - What causal parameters would they like to estimate? What is the interpretation of these parameters?
    - What is the biggest challenge for estimating these types of causal effects from the types of studies for which the proposed methods would be applied?
    - What causal assumptions do they make to identify the causal parameters? Do they seem plausible? In what situations might they be violated?
    - What models for observed data do they use? What statistical modelling assumptions, if any, do they make?
    - Briefly describe the inference algorithms. Are there challenges with implementation? Can you think of scenarios in which the algorithm wouldn't converge or would be too computationally demanding to be feasible?
    - Describe your overall opinion of the method. Strengths, weaknesses, limitations, etc

Presentation:

- 15 minute in-class presentation + 5 minutes of Q&A
- Dec 2 and 4 (and Dec 11 if needed)
- Just highlight key points: what is the causal question? What types of studies is this applicable to? What assumptions do they make? How do they carry out point and interval estimation? Brief description of algorithm. Does it seem to work well? Alternatives? Pros/cons. The audience should at least understand what the research topic is, what makes the problem challenging, and what was the gist of the proposed solution.