BSTA 790: Causal Inference in Biomedical Research
Fall 2020

Instructors:

Nandita Mitra, PhD
Professor of Biostatistics and Chair of GGEB
622 Blockley Hall
nanditam@upenn.edu

Peter Yang, PhD
Associate Professor of Biostatistics
632 Blockley Hall
weiyang@mail.med.upenn.edu

Class time: T/Th 9:00 – 10:20am, Virtual via BlueJeans

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 (2017). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming.
https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

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

P.R. Rosenbaum, Observational Studies (2nd ed.), Springer-Verlag, New York, 2002.

Software:
R (http://www.r-project.org/)

Coursera:
https://www.coursera.org/learn/crash-course-in-causality/
This is a less technical version of the first part of this course, and so could make for a helpful supplement to the main lectures.

Grading:
70% homework
30% final project

While students are encouraged to discuss homework problems together, the actual document that is turned in (including computer code) must be each student's own work.

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

Tentative Lecture Schedule

DATE		TOPICS
Sept	1	Overview: potential outcomes, causal effects
	3	Observational studies: confounding and causal assumptions
	8	DAGs
	10	DAGs: do-calculus
	15	DAGs: backdoor path criterion; SWIGs
	17	Matching
	22	Matching
	24	Matching
	29	IPTW, regression
Oct	1	Doubly robust estimators
	6	Randomized trials with noncompliance
	8	Principal stratification
	13	Instrumental variables
	15	Instrumental variables
	20	Mediation
	22	Sensitivity analysis
	27	Interference (Youjin)
	29	Bayesian Causal concepts (Arman)
Nov	3	Time-dependent confounding overview / g-methods
	5	G-formula
	10	IPTW and marginal structural models
	12	IPTW and marginal structural models
	17	Doubly robust estimation
	19	Structural nested mean models
	24	Survival outcomes
	26	Thanksgiving Break (no class)
Dec	1	Dynamic Treatment Regimes
	3	Semiparametric inference (influence functions, TMLE)
	8	Student presentations
	10	Student presentations

Final Project

Written:

- Due Friday Dec 18
- Choose a causal inference topic / method that we haven't directly covered in class
- 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 (a few 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 8 and 10
- 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.