

# PHP 2610 Causal Inference and Missing Data

Fall Semester 2023

## General Information

**Canvas:** <https://canvas.brown.edu/courses/1093082>

**Class time and location:** 121 South Main Street 241, Tuesdays and Thursdays 9:00 – 10:20am

**Instructor:** Youjin Lee (youjin\_lee@brown.edu)

**Teaching Assistant:** TJ Tan (chichun\_tan@brown.edu)

## Office Hours:

Instructor: Fridays 9-10am at SPH 704 or via Zoom (<https://brown.zoom.us/my/ylee209>)

TA: Wednesdays 2-3pm at SPH 700 or via Zoom (<https://brown.zoom.us/j/3364863742>)

## Course Summary

This course provides conceptual and methodological guidelines on how to define and estimate causal effects in various settings. The course begins by examining the differences between randomized designs and observations studies, as well as some challenges in causal effect estimation with observational studies from a missing data perspective. It then covers useful causal inference methods for evaluating policy effects under different types of designs. Finally, the course covers advanced topics, including mediation analysis and time-varying treatments.

## Why Should You Take This Course?

Most of everyday decisions and public policies are based on questions about the potential effect of a certain cause. Did the pandemic reduce air pollution? Do gun control laws cause more or fewer crimes? Answering these questions requires causal reasoning, which is often not simply achieved by naïve comparison in observed phenomenon, such as comparing air quality before and after the pandemic without considering other factors that could affect the outcome of interest in ways other than the pandemic.

Many studies in public health and social science often overlook missing observations and confuse associations with causation without careful examinations. This can lead to the design of ineffective – and sometimes harmful – interventions and public policies. This course provides guidance on when we can connect what we observe in the real world possibly with missing data to the causal effects of interest and how we can read and analyze causal effects from empirical evidence. To that end, the course helps develop causal reasoning and provides methodological background of missing data and causal inference so that students will be able to perform valid practices of causal inference as a public health researcher. After taking this course, students will be able to connect causal inference to missing data problems, separate causal questions from statistical inferences, describe the assumptions required to identify the causal effects, and estimate the causal effects from observational data in diverse settings.

## Goals and Expectations

The course mainly consists of three parts: lectures, readings, and assignments, all of which are closely connected with each other. First, students are expected to come to class on time, ready to study each week's topics. To do so, students are highly encouraged to read the textbooks or articles assigned at the previous lecture. Students from the Department of Biostatistics are highly recommended to read advanced readings. Assignments are evaluating three parts: (a) comprehending causal problems in articles, (b) representing and deriving causal effects through mathematical formulations, and (c) applying the methods using statistical programming R. For part (a), students are highly encouraged to hone the ability to understand the causal problems from real data examples provided in the reading assignment. For part (b), the lecture will provide some guidance on representing causal concepts using mathematical notation, assuming that students are already equipped with a basic background in probability and statistics. Lastly, the course will cover practical implementations of the methods and provide relevant R code and datasets for part (c).

### *Credit hours and time expectations*

Over 13 weeks, students are expected to spend 3 hours in class, 4 hours on assigned reading and discussion, and 3 hours on Problem Sets per week (130 hours total). Students are expected to spend up to 45 hours reviewing the course materials and taking part in office hours for the exam preparation. The midterm and final exam will take about 2 hours and 3 hours, respectively (180 hours total).

### **Pre-requisites**

Students should have completed a one-year biostatistics sequence at the level of PHP 2510-2511, or equivalent courses. In addition, students must be comfortable with programming in R. Some useful websites are here:

[https://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)

<https://www.r-bloggers.com/2015/12/how-to-learn-r-2/>

Some useful sample data and code are available at <https://github.com/youjin1207/CausalInference2610>.

### **Assessment:**

20% Participation

25% Problem sets

25% Midterm Exam (in-class)

30% Final Exam (in-class)

### *Participation*

We expect students to take an active role in learning in lecture and online/offline discussion. Engagement both with the teaching assistant and the instructor by asking and answering questions will contribute to this grade. Students can also earn participation credit by interacting with their classmates on the canvas discussion board. After the lecture, the instructor will post three questions from the reading materials (a total of 14 articles; not from the advanced readings) that are assigned to each lecture. Students are encouraged to answer all of the three questions for a minimum of three articles throughout the course. Each discussion thread will expire in one week.

### *Problem sets*

Students will complete a total of five problem sets throughout the semester. Problem sets will primarily cover the topics from the lectures for the most recent two weeks (as of the due date). Each problem set will equally affect the course grade by 5% for each. Problem sets will be assigned on Thursdays and will be due on the following Friday by 11:59pm. While students are encouraged to discuss homework problems together, the actual document that is turned in must be each student's own work.

### *Midterm Exam (in-class only, October 24 Tuesday, 60 minutes)*

The format of the Midterm Exam will be similar to that of Problem Sets, with a mixture of multiple-choice questions, mathematical derivations, and essay questions. It will encompass the subjects up until instrumental variable methods (Lecture 11). The exam will be conducted under closed-book conditions.

### *Final Exam (in-class only, December 19 Tuesday, 90 minutes)*

The format of the Final Exam will be similar to that of the Midterm exam. It will cover the subjects up until Lecture 25. A double-sided one-page A4 size cheat sheet will be allowed.

### *Late or missed assignments*

Problem sets and the Final Exam must be turned in online at or before the posted due date. Every one-day (24 hours) of delay will result in a ten-point (out of 100) downgrade.

Letter grades are based on the following cut-offs:

A = 90% of higher  
 B = 80-89%  
 C = 70-79%  
 D = 60-69%  
 F = less than 60%

### Course materials

All course lecture notes will be placed in the Course Canvas site.

Here is the book available online.

Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. CRC Press.  
 Available at: <https://theeffectbook.net/>

Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.  
 Available at: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

The reading materials are available online in the Brown library.

### Course outline

Detailed lecture outlines with reading assignments are available at Canvas.

Dates	Lecture
9/7	Introduction and course overview
9/12, 9/14	Randomized experiments, observational studies, and confounding
9/19, 9/21	Propensity score weighting and matching methods
9/26, 9/28	Missing data, multiple imputations, and inverse probability weighting
10/3, 10/5	Sensitivity analysis (10/3), Noncompliance and encouragement design (10/5)
10/10, 10/12	Instrumental variables
10/17, 10/19	Instrumental variables and Regression discontinuity
10/24	In-class midterm
10/26	Interrupted time series
10/31, 11/2	Difference-in-differences and synthetic control approaches
11/7, 11/9	Causal mediation analysis
11/14	Implementation with R
11/16, 11/21	Direct acyclic graph and time-varying treatments
11/28, 11/30	G-formula and marginal structural models
12/5	Dynamic treatment regimes
12/7	Review session

### Books, Supplies, and Materials

If your Brown undergraduate financial aid package includes the Book/Course Material Support Pilot Program (BCMS), concerns or questions about the cost of books and course materials for this or any other Brown course (including RISD courses via cross-registration) can be addressed to [bcms@brown.edu](mailto:bcms@brown.edu). For all other concerns related to non-tuition course-related expenses, whether or not your Brown undergraduate financial aid package includes BCMS, please visit the Academic Emergency Fund in E-GAP (within the umbrella of “E-Gap Funds” in UFunds) to determine options for financing these costs, while ensuring your privacy.

### Accessibility and Accommodations

Brown University is committed to full inclusion of all students. Please inform me early in the term if you may require accommodations or modification of any of course procedures. You may speak with me after class, during office hours, or by appointment. If you need accommodations around online learning or in classroom accommodations, please be sure to reach out to Student Accessibility Services (SAS) for their assistance

(seas@brown.edu, 401-863-9588). Students in need of short-term academic advice or support can contact one of the academic deans in the College.

### **COVID-19 Plans and Protocols**

We will follow University requirements and guidance.

### **Class Recording and Distribution of Materials**

Lectures delivered in-person will not be recorded. Some lectures that cannot be delivered in-person will be recorded and posted on Canvas. The Canvas site can only be accessed by the TA and by students enrolled in the class. Lectures and other course materials are copyrighted. Students are prohibited from reproducing, making copies, publicly displaying, selling, or otherwise distributing the recordings or transcripts of the materials. The only exception is that students with disabilities may have the right to record for their private use if that method is determined to be a reasonable accommodation by Student Accessibility Services. Disregard of the University's copyright policy and federal copyright law is a Student Code of Conduct violation.

### **Use of Technology to Support Your Learning in This Course**

This course will use Canvas and Zoom. The instructor and the TA are committed to ensuring access to online course resources by students. If you have any concerns or questions about access or the privacy of any of these platforms, please reach out to the instructor. The IT Service Center (<https://it.brown.edu/get-help>) provides many IT Services including remote assistance, phones, tickets, and chat. Please also see the Online and Hybrid Learning Student Guide.

### **Students with Special Needs**

Brown University is committed to full inclusion of all students. Students who, by nature of a documented disability, require academic accommodations should contact the professor during office hours. Students may also speak with Student and Employee Accessibility Services at 401-863-9588 to discuss the process for requesting accommodations.

### **Diversity Statement**

This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is our intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, previous computing skills, gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

### **English Language Learners**

Brown University welcomes students from around the world, and the unique perspectives international students bring enrich the campus community. To empower students whose first language is not English, an array of ELL support is available on campus including language and culture workshops and individual appointments. For more information about English Language Learning at Brown, contact the ELL Specialists at [ellwriting@brown.edu](mailto:ellwriting@brown.edu).

### **Guidance on the use of AI tools for assignments**

The instructor acknowledges the utility of AI tools (e.g., ChatGPT) for educational purposes. However, it should be used with cautions, particularly when students may not have the ability to determine whether the AI-generated responses are correct or not. The primary purpose of assignments (Problem Sets and Reading Questions) is to support students in learning and receiving timely feedback from both the instructor and the TA, rather than evaluation. These assignments are also important for the instructor to learn how or whether students understand the materials so that she can adjust the level of exams. In that sense, it is highly encouraged for students to initially tackle the assignments unaided by AI so that they could identify parts that require additional time to understand.