CLASS SESSIONS

Tue 10:00-11:20 on zoom
Link: https://columbiacuimc.zoom.us/j/96152268342
Passcode: P8122-001

Thu 10:00-11:20
Classroom: Russ Berrie 2

INSTRUCTOR
Linda Valeri, Ph.D.
lv2424@cumc.columbia.edu
Office hours by appointment.

TEACHING ASSISTANT(S)
Amy Pitts
ajp2257@cumc.columbia.edu
Office hour: Mon, 9-10 am ET
(zoom link: https://columbiacuimc.zoom.us/j/95105464068
Meeting ID: 951 0546 4068)

Melanie Mayer
mm4963@cumc.columbia.edu
Office hour: Wen, 9-10 am ET
(zoom link: https://columbiacuimc.zoom.us/j/96017037463?pwd=NnE3YlhnL2Y5RUdBRmpFcENneFljdz09,
passcode: 653538)

COURSE DESCRIPTION
Substantive questions in empirical scientific and policy research are often causal. This class will introduce students to different theoretical frameworks for causal inference and statistical methods important to the practice of causal inference. As theoretical frameworks, we will discuss potential outcomes, causal graphs, randomization and model-based inference, and causal mediation. We will cover various methodological tools including randomized experiments, matching, and inverse probability weighting, dynamic causal models, and statistical methods for mediation. We will analyze the strengths and weaknesses of these methods. The course will draw upon examples from social sciences, public health, and other disciplines. The instructor will illustrate application of the approaches using R software. Students will be evaluated and will deepen the understanding of the statistical principles underlying the approaches as well as their application in homework assignments, group presentations, a take home midterm, and a take home exam.

COURSE STRUCTURE

We will first introduce the potential outcomes framework and focus on statistical approaches to estimate average causal effects. We will then discuss concepts and statistical methods to investigate causal effects of time-varying exposure and causal mechanisms of mediation and interaction. The class meets twice a week. We will alternate live instruction sessions (which will be recorded), recorded lectures followed by Q&A sessions (asynchronous learning), discussion of articles based on group presentations and practice labs. Attendance to all sessions (Tuesday on zoom and Thursday in class) is mandatory. The instructor and teaching assistants are available for office hours.

QUESTIONS, ANNOUNCEMENTS AND SUBMISSIONS
In addition to office hours, please use the Courseworks Discussion Board when asking questions about lectures, homework, and other course materials. This allows all students to benefit from the discussion and to help each other understand the materials. Students are encouraged to participate in discussions and answer any questions that are posted. All class announcements as well as electronic submissions of assignments will be made through Courseworks.

PREREQUISITES
Applied Regression I (P8100), Applied Regression II (P8110). Or Biostatistical Methods I (P8130) and Biostatistical Methods II (P8131)

COURSE LEARNING OBJECTIVES
Students who successfully complete this course will be able to:

> Formalize scientific questions about etiology, intervention, and mechanism using the potential outcomes framework.
> Elicit and evaluate assumptions needed to identify average causal effects in randomized and observational studies
> Articulate assumptions, relative strengths and weaknesses of matching and weighting estimation approaches
> Evaluate when appropriate and implement matching and weighting approaches
> Interpret results of causal analyses for the investigation of fixed time and time-varying exposure effects
> Elicit and evaluate assumptions necessary to identify mediating mechanisms
> Implement and interpret results of statistical methods for mediation

Required Textbooks
Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming (available online at the authors' website).

Suggested Textbooks
Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
VanderWeele. T.J. (2015) *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press.

ASSESSMENT AND GRADING POLICY
Student grades will be based on:
Homework 25%
Group presentation and in class discussion 25%
Midterm Take Home Exam..25%

Final Project 25%

Incomplete grades will not be given except for a condition or circumstance documented by the Office of Students Affairs in advance of the final examination.

A+   Reserved for highly exceptional achievement
A   Excellent. Outstanding achievement
A-   Excellent work, close to outstanding
B+   Very good. Solid achievement expected of most graduate students
B   Good. Acceptable achievement
B-   Acceptable achievement, but below what is generally expected of graduate students
C+   Fair achievement, above minimally acceptable level
C   Fair achievement, but only minimally acceptable
C-   Very low performance
F   Failure. Course usually may not be repeated unless it is a required course.

COURSE REQUIREMENTS
• Class will be recorded and posted after 2-4 hours.
• Tuesday class will be on zoom and Thursday class will be in person. Class will be recorded and posted after 2-4 hours.
• We will notify when it is mandatory to listen to a recording prior to the class.
• Portions of R output that are relevant to homework assignments should be electronically cut and pasted into the appropriate part of the body of the homework and NOT attached as an appendix or an additional attachment.
• Relevant code for each homework should be included in the appendix.
• Un-annotated output will not be accepted or graded.
• Homework are submitted electronically (on Courseworks) as a PDF to avoid unwanted format changes. Do NOT print extensive outputs (e.g. dataframes, warnings, etc). A single output should never exceed a page.
• Homework are due by 5 pm ET on days indicated in the Assignments section on Courseworks and in the syllabus. Late homework will not be accepted under ANY circumstances.
• The best six out of seven Homework will count towards grading.
• Group presentation. Please, see guidelines at the end of the syllabus. Guidelines and expectations are also posted on Courseworks in the Group Presentation folder.
• The midterm take-home exam is similar to a homework in its format. However, you should not consult with anyone including teaching assistants. The midterm take-home exam will be posted on 10/20 and will be due on 10/26 at 11:59pm ET. The detailed instruction will be given later in the semester.
• The final take-home exam will involve a data analysis. Again, you should not consult with anyone including teaching assistants. The final take home exam will be posted on 12/15 and will be due on 12/20 at 11:59pm ET. The detailed instruction will be given later in the semester.
• Late midterm and final will not be accepted except for a condition or circumstance documented by the Office of Students Affairs
• Class does not meet on November 8th and November 24th.
• Last class meets on December 15th

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath.
http://mailman.columbia.edu/about-us/school-mission/

Academic Integrity
Students are required to adhere to the Mailman School Honor Code, available online at
http://mailman.columbia.edu/honorcode.

Disability Access
In order to receive disability-related academic accommodations, students must first be registered
with the Office of Disability Services (ODS). Students who have or think they may have a disability
are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or
by email at disability@columbia.edu. If you have already registered with ODS, please speak to your
instructor to ensure that s/he has been notified of your recommended accommodations by Lillian
Morales (lm31@columbia.edu), the School's liaison to the Office of Disability Services.

COURSE SCHEDULE
Please see the lecture section of Courseworks to download the readings, exams, and lecture slides.

* indicates required reading.

Weeks 1 to 2 (9/5 ~ 9/16) – Introduction: Overview of the course and potential outcomes

Learning Objectives:
Distinction between scientific questions of description, prediction, and explanation.
Counterfactual responses and the fundamental identification problem
  - Formalize estimands that arise in causal inference
  - Articulate assignment mechanisms.
 	
Reading:
  * Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming, Chapters 1 p. 3-12, 6 p. 69-81.
  - Holland, P. W. 1986. Statistics and Causal Inference. Journal of the American Statistical Association, Vol. 81, No. 396: 945-960.
 	
Asynchronous learning:
  - DUE Wed week 1 (9/7): listen to recording on potential outcomes and causal inference principles and *post a question or comment by 5pm ET*
  - DUE Mon week 2 (9/12): listen to recording on identification assumptions and assignment mechanism and *post a question or comment by 5pm ET*
 	
Live instruction session:
  - Tue week 1 (9/6): Introduction to the course. Discussion on scientific questions of description, prediction, and explanation
 	
Lab activity:
  - Thu week 1 (9/8): Q&A potential outcomes and causal inference principles
  - Tue week 2 (9/13): Q&A identifiability assumptions and assignment mechanism
  - Thu week 2 (9/15): HW1- potential outcomes assignment review

Assignment:
- DUE Wed week 1 (9/7) at 5pm ET: write 1 sentence on the discussion board about why you are interested in causal inference; write 1 sentence introducing yourself to the class.
- DUE Tue week 2 (9/13) at 5pm ET: HW1- potential outcomes and assignment mechanism

Weeks 3 and 4 (9/19 ~ 9/30) – Randomized Experiments

Learning Objectives:
Randomized Experiments
- Articulate and analyze randomized experiments from Fisher's perspective (Permutation test, randomization inference)
- Articulate and analyze randomized experiments from Neyman's perspective (Inference for the average treatment effect)

Reading:
* Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. Chapter 2.1-2.3 p. 13-20.
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press. Chapters 4-6.

Asynchronous learning:
- DUE Wed week 3 (9/21): Listen to recording on Randomization inference and *post a question or comment by 5pm.*

Live instruction session:
- Tue week 3 (9/20): Randomized designs and review hypothesis testing
- Tue week 4 (9/27): Inference for the average treatment effect, Neyman vs Fisher

Lab activity:
- Thu week 3 (9/22): Q&A Randomization inference and practice exercise in R
- Thu week 4 (9/29): Q&A on group presentations

Assignment:
- DUE Fri week 3 (9/23) at 5pm ET by Group #1 & #2: post recording of a 10 minutes presentation
- DUE Tue week 4 (9/27) by 5pm ET (everyone): Read article assigned to Group #1 & #2, listen to the recording of the presentation, and post question/comment on discussion board.
- DUE Fri week 4 (09/30) at 5pm ET: HW2 - Randomization inference

Weeks 5 and 6 (10/3 ~ 10/14) – Observational Studies

Learning Objectives:

Observational studies and the problem of confounding bias
- Interpret assumptions encoded in directed acyclic graphs
- Articulate identification under selection on observables and the back-door criterion
- Motivate and implement regression adjustment for confounders

Reading:
- * Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. Chapter 7 p. 83-96.
- Shrier I and Platt R. (2008) Reducing bias through directed acyclic graphs. BMC Medical Research Methodology. 8:70-85.

Asynchronous learning:
- DUE Mon week 6 (10/10): listen to recording on Directed Acyclic Graphs (DAGs) and *post a question or comment by 5pm ET*

Live instruction session:
- Tue week 5 (10/4): Observational studies and confounding

Lab activity:
- Thu week 5 (10/6): practice exercise - Observational studies and g-formula
- Tue week 6 (10/11): Practice exercise – DAGs
- Thu week 6 (10/13): Q&A on group presentations on confounding

Assignment:
- DUE Fri week 5 (10/7) at 5pm ET by Group #3 & #4: post recording of a 10 minutes presentation
- DUE Tue week 6 (10/11) by 5pm ET (everyone): Read article assigned to Group #3 & #4, listen to the recording of the presentation, and post question/comment on discussion board.

Weeks 7 and 8 (10/17 ~ 10/28) – Subclassification and Matching

Learning Objectives:

Subclassification and Matching Methods for confounding adjustment
- Implement Covariate Matching
- Implement Balance Checks
- Articulate Properties of Matching Estimators

Reading:
- * Stuart. 2010. Matching methods for causal inference: A review and a look forward. Statistical Science. Vol. 25, No. 1: 1–21.
- Rubin, D. B. (2006). Matched sampling for causal effects. Cambridge University Press. Chapters 3-5.

Live instruction session:

- Tue week 7 (10/18): Subclassification
- Thu week 8 (10/27): Matching

Lab activity:
- Thu week 7 (10/20): Midterm review
- Tue week 8 (10/25): Q&A and practice exercise - Propensity score subclassification

Assignment:
- DUE Tue week 7 (10/18) at 5pm ET: HW3 - confounding and the g-formula
- Midterm take home exam handed out on Thurs, 10/20 after class and due on Wed, 10/26 at 11:59pm ET

Week 9 - 11 (10/31 ~ 11/18) – Weighting Method and practice of propensity score approaches

Learning Objectives:
- Understand and apply weighting based on the Propensity Score

- Use and interpret marginal structural models

- Understand Dynamic and static treatment regimens

- Use and interpret marginal structural models in longitudinal settings

Reading:
* Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. Chapter 12-13 p. 11-30,15 p. 43-51.
- Robins, J.M. Hernán, M.A. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11:550-560.

Live instruction session:
- Tue week 9 (11/1): Propensity score matching lecture cont'd
- Tue week 10 (11/8, Election Day): no class
- Thu week 10 (11/10): Propensity score weighting

Note: Nov 8 is election day - holiday

Lab activity:
- Thu week 9 (11/3): Practice of propensity score matching
- Tue week 11 (11/15): Practice of propensity score weighting
- Thu week 11 (11/17): Q&A on group presentations on propensity score approaches for causal inference

Assignment:
- due Fri week 10 (11/11) by Group #5 & #6: post recording of presentation

---

- due Tue week 11 (11/15) (everyone): Read article assigned to Group #5 & #6 and post question/comment on discussion board.
- due Fri week 11 (11/18) at 5pm ET: HW4 – propensity score subclassification assignment

Week 12-13 (11/21 ~ 12/02) – Mediation Concepts

Learning Objectives:
  - Articulate assumptions under which traditional approaches to mediation analysis are valid and implement such approaches

  - Articulate main concepts and identifiability conditions for causal mediation analysis under the potential outcome framework

  - Articulate assumption of regression approaches for mediation analysis for continuous and categorical outcomes.

Reading:
  * Valeri, L. and VanderWeele, T.J. (2013). Mediation analysis allowing for exposure -mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods, 18:137-150.
  - VanderWeele. T.J. (2015) Explanation in Causal Inference: Methods for Mediation and Interaction, Oxford University Press (chapter 2).

Asynchronous learning:

  - Mon week 12 (11/21): listen to recording on traditional approaches for mediation. *Post a question or comment by 5pm*

  - Mon week 13 (11/28): listen to recording on concepts of causal mediation under the potential outcome framework. *Post a question or comment by 5pm*

Lab activity:
  - Tue week 12 (11/22): practice exercise - Traditional approaches for mediation
  - Tue week 13 (11/29): Q&A on Introduction on Mediation analysis
  - Thu week 13 (12/01): practice exercise – Concepts of Mediation

Thu week 12 (11/24): Thanksgiving - holiday

Assignment:
  - due Fri week 13 (12/02) at 5pm ET: HW5 – propensity score weighting assignment

Week 14 - 15 (12/05 ~ 12/16) – Regression approaches for Mediation

Learning Objectives:

- Regression approaches for mediation analysis
- Conduct mediation analysis using CMAverse R package

Reading:
* Shi, B., Choirat, C., Coull, B. A., VanderWeele, T. J., & Valeri, L. (2021). Cmaverse: a suite of functions for reproducible causal mediation analyses. Epidemiology, 32(5), e20-e22.
* https://bs1125.github.io/CMAverse/

Asynchronous learning:
- Mon week 14 (12/05): Regression methods for causal mediation analysis. *Post a question or comment by 5pm*

Live instruction session:
- Thu week 15 (12/15): course re-cap

Lab activity:
- Tue week 14 (12/06): Q&A - regression approaches for mediation
- Thu week 14 (12/08): practice exercise - regression approaches for mediation
- Tue week 15 (12/13): Q&A on group presentations

Assignment:
- due Fri week 14 (12/9) by Group #7 & #8: post recording of presentation
- due Mon week 15 (12/12) (everyone): Read article assigned to Group #7 & #8, listen to the recording and post question/comment on discussion board.
- due Thu week 15 (12/15) at 5pm ET: HW6 – mediation analysis

- Final Take Home Exam handed out Thu, 12/15 and due Tue, 12/20 at 11:59pm ET

GROUP PRESENTATION GUIDELINES AND EXPECTATIONS

Group members: Please see course works on the People page → group reading assignment tab.

Activity Aims: to apply and critically analyze scientific literature that uses causal inference methods.

General instruction

We have randomly assigned each member of the class to groups with 5-6 participants. We have assigned a reading to each group about topics relevant to classes modules. On weeks with presentations, two groups will present.

Group #	Topic	Weeks	Paper
1		3-4	**Group #1.** Fujiwara, T., & Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? Experimental evidence from Benin. American Economic Journal: Applied Economics, 5(4), 241-55.
2			**Group #2.** Alsan, M., Garrick, O., & Graziani, G. (2019). Does diversity matter for health? Experimental evidence from Oakland. American Economic Review, 109(12), 4071-4111.
3	Confounding	5-6	**Group #3.** Combined Postmenopausal Hormone Therapy and Cardiovascular Disease: Toward Resolving the Discrepancy between Observational Studies and the Women's Health Initiative Clinical Trial. *Am J Epidemiol.* 2005; 162(5), 404-414.
4			**Group #4.** Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol.* 2006;35(2):337-344.
5	Matching & Weighting	9-11	**Group #5.** Chan, G.J., Stuart, E.A., Zaman, M. *et al.* The effect of intrapartum antibiotics on early-onset neonatal sepsis in Dhaka, Bangladesh: a propensity scores matched analysis. *BMC Pediatr* 14, 104 (2014).
6			**Group #6.** Geleris J, Sun Y, Platt J, et al. Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. *N Engl J Med.* 2020;382(25):2411-2418.
7	Mediation	14-15	**Group #7.** Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? *Am J Epidemiol.* 2006;164(11):1115-1120.
8			**Group #8.** VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, et al. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol.* 2012;175(10):1013-1020.

Lab structure section.

- Presentations. Each group will upload a recording of the presentation on the due date indicated in the modules. The presentation will last 10 minutes maximum for each group.
- All students will read the articles, listen to the presentations, and post a question/comment on the discussion board.
- Class discussion based on students' questions on Courseworks: This discussion should be guided by the paper presenters. Each group will give a very brief summary of the paper and share the summary of the questions posted on Courseworks with the whole class.
- Conclusions: Instructor will give some final remarks.

Group instructions

The group will critically analyze the paper and present the paper. Consider structuring the presentation with the following section:
> 1) Introduction.
> 2) Methods.
> 3) Results.
> 4) Discussion.

- All team members should be responsible of one section and at least two team members should present. Team members coordinate to share responsibilities in leading the in class discussion. (at least two team members)
- The presentation should last 10 minutes, and students should use slides, infographics, or other communication means.
- Group will upload the presentation by the due date indicated
- Students will post questions and comments by the due date indicated
- Group will answer questions posted by fellow students on Courseworks prior to the in-class discussion
- You could ask your TAs clarifications or explanation about the paper during office hours.

Evaluation criteria.
This group assignment will be graded according next criteria:
1. Critical analysis of the paper
2. Clear and concise explanation of methods relevant to the course
3. Proper use of presentation time
4. Quality of answer to students' questions