Epi265: Research Methods in Chronic Disease Epidemiology - Spring

2 Units

Course Outline

Instructors:
Maria Glymour
Associate Professor

Contact:
mglymour@epi.ucsf.edu
office hours: http://mglymour.youcanbook.me/

Hours:
Mondays 10:30-12:20

Audience:
Graduate students, including professional students pursuing research activities, interested in quantitative epidemiologic studies of the determinants and consequences of chronic diseases, including dementia, stroke, cardiometabolic, and musculoskeletal disease. .

This course assumes a basic foundation in epidemiology and quantitative research methods. Concurrent enrollment or previous completion of a regression course is strongly recommended.

Summary:
This course will focus on clearly articulating and testing research hypotheses related to the determinants and consequences of chronic conditions. Most examples will be drawn from literature on social and lifecourse determinants of dementia, stroke, and cardiometabolic disease.

Learning objectives:
At the conclusion of the class, students should be able to:

1. Articulate specific, testable hypotheses regarding determinants of chronic disease and how chronic diseases influence functional outcomes
2. Propose alternative study designs (e.g., case-control, cohort, quasi-experimental, or randomized trial) to test such hypotheses
3. Articulate advantages and disadvantages of alternative designs, considering the research question, exposure, and outcome under consideration.
4. Select appropriate statistical approaches for data analysis
5. Describe and estimate the magnitude of potential sources of bias in observational, quasi-experimental, or randomized studies, including confounding, selection bias, missing data, and incorrect measurement.
6. Review applied quantitative article in chronic disease epidemiology and summarize research question, and pros and cons of: the data set, study design, measurement approach, and analytic approach for the specific research question.

Student commitment:
The course will involve two hours of class time per week, with accompanying reading, peer-review comments, take-home quizzes, and homeworks. In addition, there will be a term-long project to apply skills developed in class.

Grades for the class will be based on the following:

- Five reading responses. Responses are due each week the evening (8pm) before class, but you may skip 1 week. Please let me know in advance the week you are skipping. (25% of the grade).
- Three take-home quizzes (10% of the grade each)
- Final research project write-up (30%)
- Feedback to other students (15% of the grade)

Reading response topics are listed below and should be posted on the class website for everyone to review. **You are welcome and encouraged to discuss these with classmates (or anyone who will listen). Part of the class grade depends on commenting on one another's reading responses. Obviously, be nice.**

You should always be prepared to discuss your response in class.

A note about the quizzes: They are intended to be completed in 45 minutes. These will be take-home rather than in class, and you may use any materials you wish (it is open-book). I ask you to commit to completing the quiz in 45 minutes. Please use a timer and stop 45 minutes after you have started.

Final research project write-up:
Read recent issues of a journal in your field (e.g., Am J Epi; Epidemiology; Neurology; Pediatrics). Choose an article that addresses a research topic of interest to you using an observational study design. Summarize the study design, sampling, and analytic plan used in the paper. Draw a DAG corresponding with the author's implicit or explicit causal hypotheses and specify which part of the DAG they are testing in this article. Review their handling of statistical conclusion validity, internal validity and external validity Describe what you consider the greatest weaknesses of the paper. (This should be ~2000 words).

Write the methods section for an alternative study that would address the same research topic with either a stronger design or a design that addressed the most important limitations of the initial study. This should be ~1500-2000 words.

Session/Lecture Title	Session/Lecture Content	Session/Lecture Leader	Reading Response Topic
Class 1. Causal inference in the context of observational data	• Shadish Cook and Campbell Validity Approaches • Linking with DAGs, Doll and Hill • When do we do well?	M Glymour	N/A
Class 2 Longitudinal study designs: data sources	• Alternative approaches • Tradeoffs in sample size, representativeness, and measures • Goal is to see multiple approaches, readings focus on applied examples	M Glymour	Choose at least 3 distinct data sources (e.g., ARIC, HRS, death certificate data, NHS, etc), and give an example of a research question (e.g., a hypothesis about the effect of a specific exposure on a specific outcome) you consider the study exceptionally strong to address. For each, provide an example of a research question you consider the design very weak to address. Explain why the data source is strong or weak for each question. Do not just discuss the questions addressed in the readings, think of new questions, preferably things you might be interested in. This is not supposed to be a commentary related to the substantive questions in the readings: the goal is to focus on the pros and cons of various data sources. For hypotheses each study would not be well equipped to address, if possible describe another study that could address the hypothesis.
Class 3 Clustered Data	• Neighborhood Effects • Repeated Measures • Growth curves **Quiz 1**	M Glymour	Find any article using clustered data and describe: the unit of clustering; the hypothesized effects and the level at which the exposure is measured (is it a characteristic of the cluster or the observation within the cluster); and the statistical model used to estimate the effect. Describe whether there are any other statistical models that might be appropriate and whether they would be preferable (e.g., GEE vs mixed).
Class 4 Interactions	•	Tyler VanderWeele	
Class 5: Lifecourse in	• Matching the analysis to the	M Glymour	For a specific exposure-outcome combination of interest to

longitudinal data analysis	theoretically relevant causal question • Specify the research question and the exposure of interest (cumulative, point exposure, direct effect, indirect effect, subgroup effects) • Is time to event the right framework? • Age, period, cohort effects		you, specify which lifecourse model is likely most appropriate and why you think this is the case. Describe the regression models you could use to test your hypothesis. Are there any possible data sets in which this test could be conducted, and if so, what concerns would you have about interpreting your proposed test of the lifecourse model?
Class 6. Nontraditional methods	• Basics of IVs • Links with RCTs • Links with DID, equivalence of fuzzy regression discontinuity • Assumptions for interpretation • Non-adherence in RCTs • ITT vs IV estimates **Quiz 2**	M Glymour	When is a quasi- or natural-experiment more appropriate than a randomized experiment? When is a quasi- or natural-experiment more informative than a conventional observational study? Give an example of a substantive question and a stakeholder (e.g., policymaker, patient, clinician) who would be more interested in an ITT effect estimate vs an IV effect estimate. Discuss how each (ITT and IV) correspond to effect estimates from conventional studies.
Class 7 Targeted Maximum Likelihood Estimation	•	Kara Rudolph	
Class 8 Bias Analysis	•	Sander Greenland	
Class 9 Evaluating Uncertainty	• CIs vs p-values • Bootstrapping • Power calculations • Publication bias (if time) • Comparing subgroup effects **Quiz 3**	M Glymour	Specify a hypothesis regarding a particular exposure and outcome and a binary effect modifier including specific measures of association (specify the magnitudes of that association you anticipate: I suggest making everything cross-sectional). Using the software of your choice, generate a population with 1000 people under a causal structure consistent with this hypothesis. Draw a simple random sample 100 individuals from this population and estimate the population average exposure-outcome association and the association stratified by your modifier of interest within this

			subset. Repeat this 10 times and write the parameter estimates and CI each time.
Class 10 Sampling	• SRS • Complex: clustering and stratification • Systematic • Capture Recapture • (RDS and time location sampling)	M Glymour	For any chronic condition (stroke, heart disease, breast cancer, hypertension, etc), choose a *national* data set and a geographically localized data set and describe the sampling approach used in each. Explain the motivations for the particular sampling approach used in the study design. Explain how this sampling affects the analysis and interpretation of results.

Three optional textbooks:

1. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008
2. Cook T, Campbell D, Shadish W. *Experimental and quasi-experimental designs for generalized causal inference*: Houghton Mifflin; 2002. Angrist and Pischke Mostly Harmless Econometrics; 2009 Princeton University Press.

Class 1:

•

1. Cook T, Shadish W, Wong V. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. Journal of Policy Analysis and Management. 2008;27(4):724-750.
2. Winship C, Morgan SL. The estimation of causal effects from observational data. Annual Review of Sociology. 1999;25:659-706.

Class 2: Alternative Data Sources

1. Wu T. The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives. *American Journal of Epidemiology.* 1989;129(4):687.
2. Pagidipati and Gaziano. Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement. *Circulation*. 2013; 127:749.
3. Pickett KE, Luo Y, Lauderdale DS. Widening social inequalities in risk for sudden infant death syndrome. *American Journal of Public Health*2005;**95**(11): 1976.
4. Stroud N, Mazwi TML, Case LD, et al. Prestroke physical activity and early functional status after stroke. *Journal of Neurology, Neurosurgery & Psychiatry* 2009;**80**(9): 1019.
5. Lee S, Colditz GA, Berkman LF, Kawachi I. Caregiving and risk of coronary heart disease in US women* 1:: A prospective study. *American Journal of Preventive Medicine* 2003;**24**(2): 113-9.
6. Hauser R, Willis R. Survey design and methodology in the Health and Retirement Study and the Wisconsin Longitudinal Study. *Population and Development Review.* 2004;30:209-235. (this is a long article – skim it for key features of study design)
7. Banks J, et al. Disease and Disadvantage in the United States and England. JAMA 2006; 295 (2037)

Class 3: Clustered Data

1. Hanley et al., Statistical analysis of correlated data using GEE: an orientation. Am J Epi 2003 v 157, pg 364.
2. Singer, J. D. (1998). "Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models." Journal of Educational and Behavioral Statistics **24**(4): 323-355.
3. Wilson RS, Hebert LE, Scherr PA, Barnes LL, Mendes de Leon CF, Evans DA. Educational attainment and cognitive decline in old age. *Neurology.*2009;72(5):460.
4. Arcaya M (2013) "Effects of Proximate Foreclosed Properties on Individuals' Weight Gain in Massachusetts, 1987–2008" Am J Public Health

Class 4: Tyler

(there will be readings, but I do not know them yet. I encourage you to participate in the upcoming journal club on this topic).

Class 5: Statistical models corresponding with substantive hypotheses: lifecourse

1. Wills AK, Lawlor DA, Matthews FE, Aihie Sayer A, Bakra E, Ben-Shlomo Y, Benzeval M, Brunner E, Cooper R, Kivimaki M, Kuh D, Muniz-Terrera G, Hardy R. Life Course Trajectories of Systolic Blood Pressure Using Longitudinal Data from Eight UK Cohorts. PLoS Med 2011;8(6):e1000440.
2. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International journal of Epidemiology* 2002;**31**(2):285-293.
3. Mishra G, Nitsch D, Black S, De Stavola B, Kuh D, Hardy R. A structured approach to modelling the effects of binary exposure variables over the life course. *International journal of epidemiology* 2009;**38**(2):528-537.
4. Naumova, E., A. Must, et al. (2001). "Tutorial in Biostatistics: Evaluating the impact of critical periods' in longitudinal studies of growth using piecewise mixed effects models." International Journal of Epidemiology **30**(6): 1332.
5. Fitzpatrick A, Kuller L, Lopez O, et al. Midlife and late-life obesity and the risk of dementia: cardiovascular health study. *Archives of Neurology.*2009;66(3):336
6. Wilson RS, Hebert LE, Scherr PA, Barnes LL, Mendes de Leon CF, Evans DA. Educational attainment and cognitive decline in old age.*Neurology.* 2009;72(5):460.

- **Class 6: Non-traditional designs**

1. Shadish W, Cook T. The renaissance of field experimentation in evaluating interventions. *Annual review of psychology.* 2009;60:607-629.ONLY NEED TO READ THROUGH PAGE 615 (ON EXPERIMENTAL DESIGNS)
2. Brookhart et al., Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiology and drug safety 2010 (19): 537-554
3. Banks and Mazzonna. The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. The economic journal, 2012, vol 122 pg 418-488
4. Ludwig et al., Neighborhoods, Obesity, and Diabetes -- a randomized social experiment. NEJM 2011; 365:1509-19
5. Glymour, Tchetgen Tchetgen, and Robins. Credible Mendelian Randomization. Am J Epi.

Optional Reading:

6. ***Two optional articles you may find useful someday: e,w:*** Murray, D., S. Varnell, et al. (2004). "Design and analysis of group-randomized trials: a review of recent methodological developments." American Journal of Public Health **94**(3): 423.
7. Hayes R, Bennett S. Simple sample size calculation for cluster-randomized trials. *Intl J Epidemiol.* 1999;28(2):319.

Class 7: Kara Rudolph
Class 8: Sander Greenland

Class 9: Uncertainty

1. Sterne JAC, Smith GD. Sifting the evidence - what's wrong with significance tests? *British Medical Journal.* Jan 27 2001;322(7280):226-+.
2. Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2006;2(8):e124: 0696-0701
3. Grunkemeier, G. and Y. Wu (2004). "Bootstrap resampling methods: something for nothing?" The Annals of thoracic surgery **77**(4): 1142.
4. Ertel et al., Frailty modifies effectiveness of psychosocial intervention in recovery from stroke. Clin Rehabil. 2007. 21:511.

(we will add another applied article)

Class 10: Sampling.
Trying to find readings not in a textbook.