HEALTH SERVICES RESEARCH DESIGN (HPM 225B)
UCLA – School of Public Health

Winter 2022
(syllabus v04 2021_0107)

Class Sessions: Monday and Wednesday, 1:00 – 2:50 pm
Location: 61-269

If zoom:

Faculty: Jack Needleman
Department of Health Policy and Management
Phone: 310.267.2706
Email: needlema@ucla.edu
Office: Room 31-269C
Assistant: Maria Porras, mporras@ph.ucla.edu

Lab Sessions: Dhruv Khurana MS MA PhD
Senior Economist-Statistician
UCLA Integrated Substance Abuse Programs
10911 Weyburn Ave, Suite 200, Los Angeles, CA 90024
Ofc: (310)-267-5251

DKhurana@mednet.ucla.edu

Class:
In classroom: 61-269
Zoom: https://ucla.zoom.us/j/99237721167?pwd=VXhHSmg1UmROKzQxZzhjSnZTZm9iZz09

Needleman Office hours:
Zoom:
Monday 4-5:
https://ucla.zoom.us/j/93820113959?pwd=YWVRUVU1OWdsZjhrYU9zT3JhZ0FqZz09
Thursday 2-3:
https://ucla.zoom.us/j/92582625833?pwd=aU1xMGgrSGJCYVNZckg1VFdDTXA3UT09

And by appointment, in person or via zoom.

Khurana office hours:
calendly.com/dhruvkhurana

Course Overview:

SYLLABUS OUTLINE

    I.   **COURSE OVERVIEW, GRADING AND OTHER GENERAL MATERIAL**
    II.   **ASSIGNMENTS/GRADING**
    III.   **RESEARCH PHILOSOPHY AND APPROACH TO THE CLASS**

IV. **CLASS SESSIONS, READINGS AND NOTES ON PREPARATION FOR SPECIFIC SESSIONS**

V. **INSTRUCTIONS AND TEMPLATE FOR HS225B TERM PAPER**

I. **COURSE OVERVIEW, GRADING AND OTHER GENERAL MATERIAL**

Course Overview

This course focuses on the challenges of implementing research designs in practice, including the application of conceptual models in empirical analyses, the operationalization of study design through analysis, and preparation of research articles. It is the second of a two-quarter sequence on the design of health services research. Focusing on the design of the research and analyses required to address mediation, moderation and test complex conceptual models, and presenting an approach to exploratory analysis, it complements rather than directly supports the material on specific statistical analytic methods presented in Biostatistics 201AB and HPM 237C.

Class sessions

The class meets three times a week, Mondays and Wednesdays 1-3pm and a lab Thursday, 8-10am. There will be 20 regular class sessions and 8 labs. Because there are two Monday holidays, some regular class sessions will be held during the lab period. The lab session may be rescheduled with the concurrence of the class.

How Classes will be Conducted:

Course format will be in-class unless University requires switch to remote (Zoom-based) learning. In order to accommodate students who may not be able to attend in-person classes because they are quarantined, in-person classes will be recorded and zoomed synchronously BUT the instructor will not attempt to conduct a hybrid class with active participation from those not in the class-room.

Some sessions with guest presenters may be conducted via zoom.

Last year, based on mid-course feedback, some lectures were shifted from in-class to pre-recorded asynchronous lectures to allow class time to focus on student questions and clarification. This approach was well received, and will be discussed at the end of week 1.

Learning Objectives for the Course

Upon completion of HS 225B, student should be able to:
- Comprehend, formulate, and employ theory, models, and variables in the conceptualization of a health services research problem
- Know how to identify and use secondary data sources relevant to health services research
- Translate health services research designs into regression models
- Design and conduct exploratory analyses
- Implement, test and interpret moderated and mediated relationships and models
- Write a research manuscript describing the methods and results of an empirical data analysis corresponding to an appropriate conceptual model and study design

Texts and readings

The required text for the course is
- Remler DK, Van Ryzin G. Research Methods in Practice: Strategies for Description and Causation. Sage Publications, 2015.

Other readings available online or on reserve and distributed through the Biomedical Library or on the course CCLE website.

If you have a different edition of the textbooks, please check with the instructor to make sure that you are reading the correct sections.

You will need to set up a UCLA VPN account with multifactor authorization in order to access some of the online publications for free from your home. Readings not available for free online will be placed on reserve in the Biomedical Library or made available through the class CCLE website. **All reading should be done in advance of class.** Lectures will generally not cover all that is in the readings. You will be asked for your questions about the readings and concepts, and these will form the basis for modifying lectures and class discussion.

Communications and email

Coming to office hours is encouraged, and if the posted time is not convenient, please request a separate appointment.

Email is encouraged. Regarding email, I get too many emails a day and you should assume yours will be lost unless you start the subject "hpm225b."

Academic Honesty and Plagiarism

In all course assignments, it is expected that you will give proper credit to others whose language or ideas you use; failure to do so may result in a lower grade or in more severe consequences. Please use quotation marks for sentences or phrases that you are quoting verbatim, and indicate the source. You do not need to use quotation marks if you paraphrase, but you must still cite the source, regardless of whether it is from a journal article, the Internet, a report, etc. Citations should be provided whenever you assert something as fact, unless it is widely known and accepted. Turnitin will help identify any issues.

I would prefer you use a reference style which includes (name, date) in the text, as this is easier to track moving back and forth in a document and reference list. You should use the same style consistently throughout papers and provide complete and appropriately formatted references. Use Endnote, Refworks, or Zotero to automate and standardize in-text citations and for formatting reference lists.

I don't anticipate plagiarism will be a problem, but one not uncommon problem is a paraphrase so close that it is virtually a quote. Whether cited or not, this can cross the line to appropriation of another person's words. Some common phrases or sentences, like "Health care spending in the US is X, Y% of the GDP" are realistically how that sentence should be written and there is no need to worry about it. Other paraphrases, or uncited ideas, can be more troublesome. The UCLA library has some resources for students, including this somewhat condescending site:

http://unitproj.library.ucla.edu/col/bruinsuccess/03/01.cfm

The slide pack provides links to other resources, including this useful link:

http://unitproj.library.ucla.edu/col/bruinsuccess/06/01.cfm

An internet search will identify more.

Technology use in class

HPM225B
Instructor: Needleman

Please be sure to access and set up Zoom (via CCLE) and enable virtual backgrounds if you want additional privacy. Cell phones and other electronic devices should be silenced or turned off prior to the start of class sessions.

Disability and Inclusivity

UCLA ADA Policy	Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE) at (310) 825-1501 or in person at Murphy Hall A255. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit www.cae.ucla.edu.
ADA Contact	Nickey Woods Center for Accessible Education A255 Murphy Hall Phone: (310) 825-1501 TTY / TTD: (310) 206-6083 Fax: (310) 825-9656
Inclusivity	UCLA's Office for Equity, Diversity, and Inclusion provides resources, events, and information about current initiatives at UCLA to support equality for all members of the UCLA community. I hope that you will communicate with me if you experience anything in this course that does not support an inclusive environment, and you can also report any incidents you may witness or experience on campus to the Office of Equity, Diversity, and Inclusion on their website (https://equity.ucla.edu/).

II.   **ASSIGNMENTS/GRADING**

Activity/Assignment	Due Date	% of Grade
Class participation	N/A	See discussion of grading
1. Article critique	1/25	15%
2. Exploratory analysis presentation	2/14-2/16	15%
3. Analytical Approach for Exploring a Conceptual Model	2/25	20%
4. Draft version of research manuscript (see syllabus on what to include)	3/1	0% (goal is to provide feedback on paper analysis prior to final)
5. Final version of research manuscript (see detailed instructions on what to include)	3/18	50%

Overview

The goals for this quarter include developing the skills to construct measures, do preliminary and exploratory analysis, conduct a regression-based analysis that adequately considers moderated and mediated pathways, and

interpret it in light of your conceptual model, preliminary analysis and results. Assignments are key to helping achieve these goals.

Please submit all assignments via the link provided on CCLE. When relevant, include *clean* versions of all Stata programs and output (do and log) files. (Use of statistical software other than Stata must be pre-approved by Dr. Khurana.) It is expected that all assignments will be turned in on time; late assignments may be subject to a grading penalty. That said, things happen. If you will be late with any assignment, please contact me **before the due date** to set a new deadline.

Active class participation is expected. While there is no specific grade for class participation, grades may be raised or lowered from the calculated grade by a half grade for high or low class participation.

Research Paper (Assignments 4 and 5)

The core assignment will be completion of a research paper with statistical analysis and its interpretation. Consistent with the learning objectives above, the paper should reflect adequate exploratory analysis, and at least one analysis presented should involve an interaction (moderation), mediation, or direct comparison of estimates from alternative models. You will be expected to bring relevant data and analyses from your class project in as examples throughout the quarter.

The paper you will prepare by the end of the quarter will be a research paper, but slightly different in scope from those you are likely to submit to a journal. Specifically, it will have more detail on the conceptual model, measures and preliminary analysis than most journal articles. It can be less detailed in its introduction or rationale for the research. See accompanying instructions and template at the end of the syllabus.

The conceptual model may be original or an extension of a standard model. Thus, while you can begin with a standard model like the Andersen model, it needs to be customized to your research. You must have at least one clear, testable hypothesis. Although your regression model will contain a number of predictors, it is recommended that the paper focus on testing a narrow range. Focusing on one regressor plus its mediation, moderation, and collinearity with one variable would be a reasonable scope. **Ignoring the possibility of mediation, moderation or collinearity on the right-hand side of your model is a major weakness in the paper for this course.**

Papers are limited to 6000 words plus tables, references, appendices, and as a broad guideline you might want to think about 3500-4500 words or 15-20 pages of double-spaced text (not including tables and references) as a target. Much of it can be taken from previous class assignments although if you do so it needs to be woven together carefully.

Please read carefully and adhere to the guidelines for the HPM 225B research manuscript.

Based on feedback from students in prior years, I have modified the paper assignment in two ways. First, students will be permitted to work in small groups of 2 to 3 students on the paper. No one is obligated to work with others, but students in prior classes have found this helpful. Second, a common data set is provided and a general common research topic using this data set is posed. This can facilitate learning about the data and measures together and establish a conversation around a common research challenge. **Many students have come into class with fa specific paper topic or data set they wish to work on.** If you are not planning to use the datasets provided, consult with the Dr. Khurana or me about the data set you would like to use. You will want to choose it early on in the quarter. **Given how short a quarter is, you should only pursue a paper with your own data if you have the data in hand when the quarter begins.** Paper topics using your own data sets should be approved by the end of the second week of classes.

The common research topic I offer is motivated by the following: Hospitals that serve large proportions of African-American and Hispanic patients or Medicaid and uninsured patients are often scored lower on

Medicare quality measures such as readmissions than other hospitals. Yet risk adjustment methods that include patient race/ethnicity and dual-eligibility status (the only measure of socioeconomic status available in the Medicare claims data) or more recent analyses that use neighborhood SES measures (typically at the five or nine digit zip code level, coded from the American Community Survey census tract or block level data) find that only a small proportion of the variance in outcomes is attributable to patient SES. Some critics of these analyses argue that the differences across hospital performance relate to differences in resources available within hospitals that serve large numbers of economically disadvantaged patients, an issue masked by CMS's use of standardized DRG prices to measure resource use at the hospital level, or by variations in the neighborhood resources available to potential or discharged patients that influence the quality of care pre- and post-hospitalization. Others, more critical of these hospitals, argue that discrimination against minority patients contributes to the differences observed, and/or these hospitals are less quality oriented or effective in delivering high quality care.

If you take the class project as your topic, you are being asked to design a research project examining some dimension of the issues discussed above, and explore the association among (1) variations in hospital outcomes for minority and disadvantaged populations, (2) variations in hospital quality, (3) variations across patients in the severity and complexity of their disease, (4) discrimination in care between minority and white patients, and (5) neighborhood effects influencing patient presentation of disease and ability to access relevant services in their community. You don't have to address all these associations or try to resolve all the issues raised in the literature, so there is considerable variation in the research you might design and conduct.

The following datasets are being made available for carrying out this research project:
- A modified version of the California OSHPD 2010 public use discharge data file for California, with patients with 3 digit or no zip codes, unknown race or ethnicity, children, and from zip codes with fewer than 100 cases excluded. (The reason for making the 2010 dataset and not a later year available, is that the 2010 public use dataset was the last to provide detailed information at the patient/admission level on variables for five digit zip code and detailed ethnicity and age data. The more constrained data sets make it difficult to explore issues of geographic or demographic associations of care.) Variables available include: hospital, patient demographics, insurance status, zip code of residence, primary diagnoses and secondary diagnoses, including present on admission comorbidities and hospital acquired complications (with a present on admission indicator for each diagnosis), primary procedure and secondary procedures, length of stay, discharge status and location, source of admission (home, nursing facility, etc., although these are often imprecisely coded), and total charges among other variables.
- The OSHPD hospital financial report data set with detailed hospital data, including data on size, ownership, location and payer mix. There is a great deal of detailed information here, organized by page, column and row for individual data elements. Using and understanding the hard (i.e., pdf) copy of the reporting forms is strongly recommended and consultation with the instructor on how to interpret the financial data if you haven't had a course in financial management or previously used hospital financial and statistical data is strongly encouraged.
- American Community Survey aggregated at the zip code level for merger of some community/neighborhood variables.

Documentation files are available.

Additional Assignments

In addition to the paper, as noted in the table and the syllabus, there are 3 additional graded assignments, 2 individual and 1 group:

Assignment #1: Article Critique

HPM225B
Instructor: Needleman

Please provide a critique for the following article:

> Wu VY, Fingar KR, Jiang HJ, Washington R, Mulcahy AW, Cutler E, Pickens G. Early Impact of the Affordable Care Act Coverage Expansion on Safety-Net Hospital Inpatient Payer Mix and Market Shares. Health Serv Res. 2018 Oct;53(5):3617-3639. doi: 10.1111/1475-6773.12812. Epub 2018 Jan 21. PMID: 29355927; PMCID: PMC6153150.

In your critique, briefly critique the study in the following categories (keep the summary brief, focus on the critique):
   (a) question addressed (how clear)
   (b) theoretical/ conceptual framework used (completeness, adequacy for constructing and interpreting model)
   (c) basic research strategy to answer the question (e.g., for a study of the impact of Medicaid expansion on health care use: pre-post with control, or Diff in Diff, analysis of physician services for those eligible for Medicaid under expansion in expansion states to those not in expansion states; purpose is to assess whether the approach, if well executed, could provide answer to the question)
   (d) data sources
   (e) measures employed (and how well and appropriately they measure what they are asserted to measure)
   (f) appropriateness of the analysis in answering the research question and testing the conceptual model
   (g) adequacy and accuracy of interpretation and discussion (did the researchers correctly and accurately interpret their results).

With respect to item (f), we are interested principally in whether the analytic strategy can resolve competing/conflicting explanations and hypotheses, not whether the astatistical methods chosen were optimal.

On balance, how confident are you that the conclusions are correct?

Note: What you submit should be a critique/assessment of the research and not a just a summary. This article, on the reading list for week 8, may be helpful in this assignment:
▪ Greenhalgh T, "How to Read a Paper: Assessing the Methodological Quality of Published Papers," British Medical Journal 1997; 315:305-8. Available for free online at http://www.bmj.com/cgi/content/full/315/7103/305

Please limit your critique to 3 pages, approximately 1200-1400 words, with 1-inch margins, single spaced.

Assignment #2: Exploratory Analysis Presentation

In week 3, you will present your project description. For Assignment 2, please work with your project group to prepare a 15 minute presentation in which you:
   (a) **Briefly** remind the class of the research question, design and data sample, conceptual model;
   (b) Present the results of at least two exploratory analyses, which might include univariate analysis (means, percentages, distributions, etc.); bivariate analysis (cross-tabs, chi-square, correlations, simple regressions, stratified analyses, etc.), two-way or three-way graphs, nested cross tabs, and/or factor analysis. As you present the analyses, explain why you did the analysis, what you hoped to learn from the analysis, and what you found. Among the topics on which you might present are:
      a. What you have learned about key variables and measures, such as degree of missingness, extent of variation and important features of the measure (large numbers of zeros, small rates, etc.)

      b. Bivariate relationships of the outcome with individual measures and stratified or trivariate relationships such as nested cross tabs.

      c. The relationships of variables on the RHS of your model (extent of collinearity, factor or scale structure, associations).

(c) Describe how the results from these analyses have (or will) influence your subsequent data analysis decisions, for example, helping you decide which variables to keep or drop, whether and how to create categories from continuous data, or combine categories in data with ordered or unordered categories, how to treat your outcome, the type of relationship between your outcome and primary regressor, how to construct a scale, identify potential moderators or mediators, etc.

Please prepare either PowerPoint slides or a handout that shows the relevant Stata code and output.
Focus on describing what analyses you did (and why), what your learned from the analyses, and (most importantly!) the implications of what you've learned for next steps in your analysis. The presentation should be well thought through, but is an opportunity to get feedback from the class on your approach/decisions.

Sign up will be on the CCLE site. The powerpoint or handout should also be posted to the CCLE website.

The setting for the presentations will depend on University COVID protocols in place. Based on past years, I want to encourage you to focus primarily on content so you can get feedback rather than trying for elaborate powerpoints. Raw stata output can be sufficient if in landscape mode, increased in size to 12 or 14 point, and the font bolded.

Assignment #3: Analytical Approach for Exploring a Conceptual Model

Please develop and submit Stata do and log files that details your approach for conducting exploratory analysis to examine **four inter-related boxes (pathways)** in the following conceptual model.
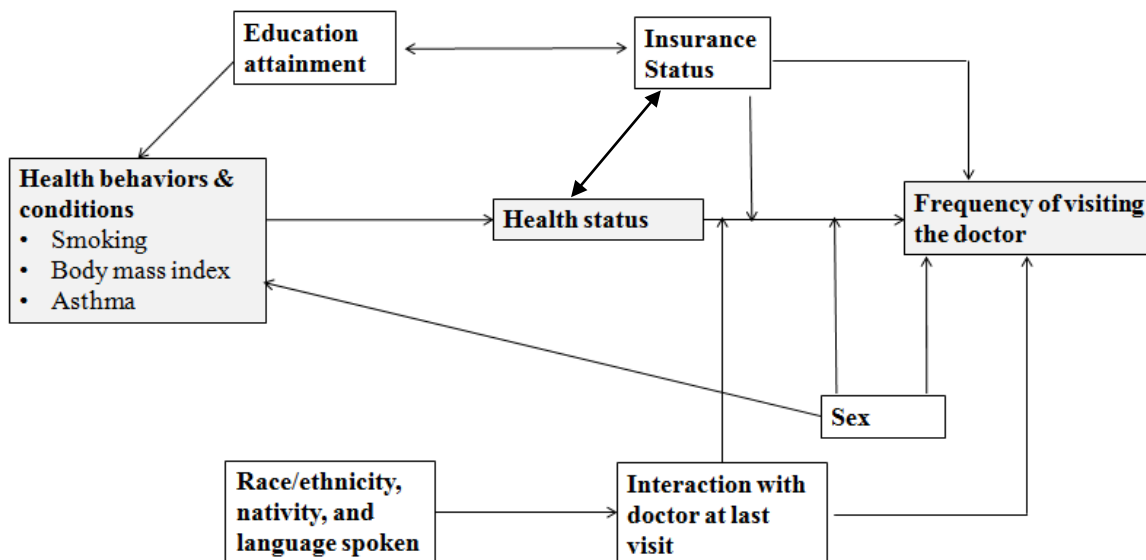
Please use the CHIS dataset (located on CCLE). Provide the code that you would use to begin exploring the data and testing relationships. In your code, please annotate:
- Why you are conducting each analysis: what you are hoping to learn/examine

- What you learned from the results, including major take aways & implications/next steps for analysis – for example, whether/how you would use the variable, limitations you would note, how you might create a scale, whether this variable would be a contender for a mediator/moderator, etc.

Please submit both your do-file and log files.

III.   RESEARCH PHILOSOPHY AND APPROACH TO THE CLASS

There is a **philosophy/attitude for this course about how research should be done**. **In its positive expression**, it emphasizes the idea of intellectual play and curiosity in guiding the choice of research questions, and efforts to rigorously draw conclusions from data through thoughtful analysis. It views research design as an active process, applying conceptual models to the specific issues and challenges of a specific project or question. Our conceptual models are often rich, complex and subtle, and the analysis to explore them should have similar complexity and richness. **In its negative expression, the philosophy of the course abhors research done by rote**. By that I mean research in which a left-hand side outcome is selected, a right-hand side variable of interest is identified believed to be associated with the outcome, a series of control variables selected because they are expected, a regression or similar analysis run, and the outcome assessed in terms of the sign of the regressor of interest and whether it is statistically significant. In rote research, the magnitude of the effect and interactions, collinearities or relationships or conceptual and structural relationships among the RHS variables are unexplored or underexplored, and often there is an undue focus on p-values and statistical significance rather than substantive findings. This type of research is boring and, when done on observational data as most HSR is, can lead to incorrect inferences and recommendations on policy and practice.

So, I invite you as you develop your approach to research and as researchers to keep the following in mind. I hope these themes and their implication for carrying out research will be explored in the class.
--**Remain intellectually playful and curious**
-Ask questions worth answering
-Construct answers worth paying attention to.
The research literature is a conversation, with each study trying to move the conversation forward. Research doesn't need to be perfect, but it needs to be sound enough that it should be part of the conversation.

--**Don't conduct research by rote**.

--**In a research environment in which you are using observational data, context matters**. And all the context that matters cannot be incorporated into the analysis, because of its complexity and data limitations.
Therefore:
-**Use your conceptual model** to define your understanding of how you think the world works with respect to what you are studying.
-Think about how you will **test that model in your research**, and how it might be changed by what you find.
-**Think about what you can't test or measure** and how it might affect what you observe in your analysis.
-**Your conceptual model and empirical analysis should be in a dialog** with one another.
-**Measures matter**. If readers don't believe your measures, they won't believe your results.
-**Samples matter**. Selection effects are rampant in observational data. The statistical methods for adjusting for selection, in which you will be well trained, are not perfect. One of the powerful tools you have for reducing the impact of selection on your results is finding appropriate samples and populations to study.

--**Understand your data and what you are finding**.
    -**Exploratory analysis is critical. Don't rush to the regression!**
    -**Think about the interactions, mediation, collinearity and confounding of your right hand side** regressors, including among your control variables. (Long aside: There is a tendency to assume/interpret results as if all relationships between a covariate and the regressor of interest and outcome are confounding relationships. As a confounded relationship (i.e., the covariate is upstream from and causally associated with both the regressor and outcome), that the association found when the covariate included is a more accurate measure of the "true" association. But this will not be the case if the relationship is a mediating relationship with the covariate downstream from the regressor of interest, or if the relationship is a moderated relationship). In light of this reflection, what relationships among them and your regressor of interest need to be examined?
    -**You should be looking for robust relationships**, not relationships dependent on the precise sample you have, the specific definition of the measures chosen, and the statistical method you have chosen. The goal is not to find (or create) a statistically significant relationship and publish it, it is to understand what is going on in the world. Let's not torture the data to make it tell the story we want to hear.
    -**Conduct sensitivity analyses and specification tests**.
    -**Magnitudes matter**. Analyze the impact of your regressors, not just the sign of the relationship.
        -ORs, RRs, elasticities can be misleading. Translate your results into natural units.
        -Graph results for insight.

--**Identify the audience for your research and write a clear story of your research for them**. Your journey as the researcher to the findings is not necessarily of great interest or relevance to your audience. They want to have confidence your results are right and will examine your methods to assure this, but their primary concern is the implications of the findings for use. This will help you not only in your writing but also interpreting your results and shaping the analysis.

Research design is an active process, applying conceptual models to the specific issues and challenges of a specific project or question. The class will reflect this. I will try to keep lecturing to a minimum. Lecture slides will be posted to the CCLE website approximately 1 week prior to the class. Some lectures will be synchronous in class; others posted ahead and asynchronous. The lecture notes complement the readings and both should be read/reviewed prior to class. You are expected to be familiar with the material when you come into the class. Please post any questions you have in feedback in advance of class and I will be asking for additional questions you identified on the readings and lectures or lecture notes at the start of class. The model only works if you come to the class prepared.

The syllabus may be changed during the quarter. An updated syllabus will be posted to the CCLE site and an announcement sent out.

    IV.   **CLASS SCHEDULE**

Session *Date*   *Topic*

Week 1:

1   01/03   **Introduction to HS225B**

Prior to class:
Read syllabus
Prep questions in introductory email:
    1. When you read a paper, what do you look at and look for in a table of regression results?

2. When you start working with a new data set (like NHIS in Macinko's 225A), what are the first things you do to look at the data?

3. The Randomized Controlled Trial is usually presented as the standard for good research. It's basic structure is a) select an outcome, b) select a treatment that is randomly assigned, c) trust that the other variables influencing the outcome are balanced by randomization, d) assess the magnitude of the difference between the randomly assigned groups and whether these differences can be ruled out as random due to statistical variation (i.e., are statistically significant). In 225A, in this course, and elsewhere, you have been asked to develop and apply a conceptual model for the phenomena you are trying to study. How do you think about the role and analysis of the conceptual model in light of the ideal of the randomized controlled trial?

Read:
- Remler & Van Ryzin, Chapter 2 ("Theory, Models and Research Questions")
- Gelman, Andrew, "The failure of null hypothesis significance testing when studying incremental changes, and what to do about it," July 3, 2017, available on CCLE for Week 1
- Chang and Meyerhoefer, "Do Elections Make You Sick," NBER Working Paper 26697, 2020. Available on CCLE. You only need to look at the abstract, page 5 and page 32.
- http://andrewgelman.com/2017/12/24/walk-crooked-miie/ [note the 3 key leaps of statistics]

In class:
- Overview of class
- Overview of assignments
- Discussion of class project
- Discussion of role of conceptual models
- Discussion of Chang and Meyerhoefer

Review/recommended:
The following blog entries
- http://andrewgelman.com/2017/12/24/walk-crooked-miie/
- http://errorstatistics.com/2015/07/17/statistical-significance-according-to-the-u-s-dept-of-health-and-human-services-i/
- http://andrewgelman.com/2015/09/04/p-values-and-statistical-practice-2/
- http://errorstatistics.com/2015/08/31/the-paradox-of-replication-and-the-vindication-of-the-p-value-but-she-can-go-deeper-i/
- http://theincidentaleconomist.com/wordpress/youre-probably-thinking-of-p-values-all-wrong-p-0-05/

Also, review the following discussions of p values and their interpretation:

- http://jnci.oxfordjournals.org/content/108/8/djw194.full
- http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877414/

2   01/05   Aligning research questions, study design, conceptual models, data sources and analysis strategies

Prior to class
Read the following:
- Remler & Van Ryzin, Chapter 11 ("Observational Studies") and Chapter 15 ("Natural and Quasi Experiments")
- Bauman AE, Sallis JF, Dzewaltowski DA, and Owen N. "Toward a Better Understanding of the Influences on Physical Activity: The Role of Determinants, Correlates, Causal Variables, Mediators, Moderators, and Confounders" American Journal of Preventive Medicine. 2002 23(2S)
- Ginsburg LR et al., "The Relationship between Organizational Leadership for Safety and Learning from Patient Safety Events," Health Services Research. 2010 Jun; 45(3):607-32.

Question for discussion: What is the conceptual model that guides this research? Is it adequate? Is it adequately reflected in the regression design?

In class:
Lecture/discussion of role of conceptual models
Discussion of Ginsburg conceptual model and its implementation in the models

Lab 1   01/06   Review of foundational Stata concepts (Dr. Khurana)

Week 2:

3   01/10   Testing coefficients and models
**   Interactions, moderation (and stratification), and confounding**

Prior to class:
Read:
Testing coefficients and models

- Xiao Chen, Philip B. Ender, Michael Mitchell and Christine Wells, Regression with Stata, Chapter Two, Regression Diagnostics. Available on ucla website at: https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-statachapter-2-regression-diagnostics/

Note: The UCLA idre website (formerly ats) offers many resources on statistical analysis, methods, code and instructional material.

- Wu, H-S. "How to Model Mediating and Moderating Effects." CFDR Workshop Series. 2011. Available on CCLE website
- Ginsburg LR et al., "The Relationship between Organizational Leadership for Safety and Learning from Patient Safety Events," Health Services Research. 2010 Jun; 45(3):607-32.
  Question for discussion: Consider Table 2 Model 5. What would Ginsburg's results be if large hospitals were the reference group? What would the coefficient on Formal Leadership, Hospital Size (now small==1), and Formal LeadershipXHospital size? How would it change the finding in the abstract: "Formal organizational leadership for patient safety is an important predictor of learning from minor, moderate, and major near-miss events, and major event dissemination. This relationship is significantly stronger for small hospitals (>100 beds)."

In class:
Lecture on testing coefficients and models
Discussion of interactions
Discussion of Ginsburg (including Stata analysis of recoded variables (size: small==1, large==0)

Reference/recommended:
- Fairchild AJ and MacKinnon DP. "A General Model for Testing Mediation and Moderation Effects." Prevention Science. 2009 Jun;10(2):87-99. doi: 10.1007/s11121-008-0109-6. Available online at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2908713/pdf/nihms213778.pdf

4   01/12   Missing data (Scott Comulada)
Prior to class:
Read:
- Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol, 17(1), 162. doi:10.1186/s12874-017-0442-1

- Li, P., Stuart, E. A., & Allison, D. B. (2015). Multiple Imputation: A Flexible Tool for Handling Missing Data. JAMA, 314(18), 1966-1967. doi:10.1001/jama.2015.15281
- Ayilara, O. F., Zhang, L., Sajobi, T. T., Sawatzky, R., Bohm, E., & Lix, L. M. (2019). Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry. Health Qual Life Outcomes, 17(1), 106. doi:10.1186/s12955-019-1181-2

Skim:

- https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/

In class:
Lecture, TBD, Missing Data

Reference/recommended
- White, I. R., Royston, P. and Wood, A. M. (2011), Multiple imputation using chained equations: Issues and guidance for practice. Statist. Med., 30: 377–399. doi:10.1002/sim.4067 http://onlinelibrary.wiley.com/doi/10.1002/sim.4067/pdf
- Pigott T.D. (2001), "A Review of Methods for Missing Data," Educational Research and Evaluation, 7 (4), 353-383.
- Rubin, D.B. (1976), "Inference and Missing Data,"Biometrika, 63, 581–592. http://www.jstor.org/stable/pdfplus/2335739.pdf?acceptTC=true
- Schenker et al.: Multiple Imputation of Missing Income Data. Journal of the American Statistical Association, 2006, vol. 101, pages 924-933. http://www.jstor.org/stable/pdfplus/27590772.pdf

Lab 2   01/13   Data management and missing data (Dr. Khurana)

Week 3:

==**Reminder: For discussion next week: Be prepared to present your draft research question/hypothesis; conceptual model; main outcome and primary regressor of interest and how they will be interpreted and measured. See session 6 below.**==

NA   01/17   NO CLASS: Martin Luther King Day

5   01/19   Exploratory data analysis 1: Basic approaches

Prior to class:
Read
- Remler & Van Ryzin, Chapter 8 ("Making Sense of the Numbers"), lingering over their discussion of practical significance and use of graphs to explore the data
- Aneshensel CS. 2013. "Theory-Based Data Analysis for the Social Sciences," 2nd Edition. Chapter 1 ("Introduction to Theory Based Data Analysis") and Chapter 3 ("Relationships as Associations")
- National Institute of Standards and Technology, NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ Chapter 1, Exploratory Data Analysis. Look at section 1 introduction and look around section 3 EDA Techniques. The chapter, which is very long, is posted to the CCLE web site. The handbook can also be accessed on the web, which makes exploring a bit easier: http://www.itl.nist.gov/div898/handbook/index.htm

  Reflect on the question: how would you decide if you should stratify an analysis based on exploratory analysis? What would you look at? What would you look for? Please post your thoughts to the Discussion post: Exploratory Analysis on the CCLE site.

In class:
Lecture

Discussion of question on deciding on stratification.

Lab 3 01/20   **Implementing exploratory analysis** (Dr. Khurana)

<mark>*Reminder: Article critique due 1/25*</mark>

<u>Week 4</u>

6   01/24   Project discussions.

Prior to class:
Post powerpoints to Assignment 1.5 Powerpoints of proposed research projects.

In class:
Each group/individual will briefly describe their research project for the quarter. Your presentation should include:
• Research question (1-2 sentences)
• Basic research design/strategy to address question (overview, with description of key analyses that will allow you to answer your question)
• Data to be used
• Conceptual model, highlighting pathways of interest, possible mediated, moderated and confounded relationships
• To the extent time permits, more detail on design: outcome, key individual variables, measures.

7   01/26   Exploratory analysis 2: Application to the class project.

Prior to class:
Read:
▪ Weiner, Handbook of Psychology, Volume 2, Chapter 1, Exploratory Data Analysis. Available on CCLE website.
▪ Exploratory data analysis. Available on CCLE website.

In class:
Presentation and discussion of exploratory analysis relevant to an illustrative project related to the class project, and invite real time exploration of the data.

8   01/27   Measurement 1 (in lieu of lab, during lab time)
Note: Measurement was discussed in 225A, and the goal here is not to repeat what has been covered, but anything that is relevant to measurement that you want addressed in these sessions should be noted in the CCLE Feedback by 1/25.
Current topics to be discussed in two sessions:
    Validity and Reliability
    Assessing the quality of scales
    Latent variables and factor analysis

Prior to class:
Read:
▪ Remler & Van Ryzin Chapter 4 ("Measurement") or skim if read in 225A
▪ Clark, L. A., & Watson, D. (1995). "Constructing validity: Basic issues in objective scale development." Psychological Assessment, 7, 309-319.
▪ Stata manual kappa-interrater reliability, Remarks and examples

In class:
Lecture

Reference/recommended
- Larson, J.S., The Measurement of Health (New York: Greenwood Press, 1991), Chapter 3 "Measurement of Morbidity and Disability"
- McDowell, Ian, Measuring Health: A Guide to Rating Scales and Questionnaires, third edition (New York: Oxford University Press, 2006).
- Fryback, Dennis G., et al., "US Norms for Six Generic Health-Related Quality-of-Life Indexes From the National Health Measurement Study," Medical Care. 2007 Dec:45(12):1162-70
- Kaplan, RM & Ries, AL (2007) Quality of Life: Concept and Definition. Journal of Chronic Obstructive Pulmonary Disease, 4, 1-9.
- Moon, M & F.T. Juster, "Economic Status Measures in the Health and Retirement Study," Journal of Human Resources, 30(Supplement 1995) S138-S157. http://www.jstor.org/stable/pdfplus 6281.pdf?acceptTC=true

Week 5:

9   01/31   Measurement 2

Prior to class:
Read:
- Bollen, K., & Lennox, R. (1991) "Conventional wisdom on measurement: A structural equation perspective." Psychological Bulletin, 110, 305-315.
- Floyd, F. J., & Widaman, K. F. (1995). "Factor analysis in the development and refinement of clinical assessment instruments." Psychological Assessment, 7, 286-299.
- Clark, L. A., & Watson, D. (1995). "Constructing validity: Basic issues in objective scale development." Psychological Assessment, 7, 309-319.

In class:
- Lecture

Reference/recommended
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. Psychological Assessment, 12, 287-297.
- Cappelleri JC, Lundy JJ, Hays RD. Overview of Classical Test Theory and Item Response Theory for Quantitative Assessment of Items in Developing Patient-Reported Outcome Measures. Clinical Therapy 2014 May ; 36(5): 648–662.
- DeVellis RF. Scale Development: Theory and Applications, 3rd Edition. 2012. Sage Publications Inc.

10   02/02   Mediation

Prior to class:
Skim from 225A and session 3
- Barron RM and Kenny DA. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." J Pers Soc Psychol. 1986 Dec;51(6):1173-82. available online through UCLA library.
- Wu, H-S. "How to Model Mediating and Moderating Effects." CFDR Workshop Series. 2011. Available on CCLE website
Read:

- Fairchild AJ and MacKinnon DP. "A General Model for Testing Mediation and Moderation Effects." Prevention Science. 2009 Jun;10(2):87-99. doi: 10.1007/s11121-008-0109-6. Available online at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2908713/pdf/nihms213778.pdf
- Wu AD and Zumbo BD. "Understanding and Using Mediators and Moderators." Social Indicators Research. 2008 July: 87:367–392. doi:10.1007/s11205-007-9143-1

Skim:
On computational approaches
Methods for assessing mediation in non-linear procedures:

Stata: medeff
Imai, K., L. Keele, D. Tingley. A General Approach to Causal Mediation Analysis. *Psychological Methods. 2010;15(4):309-34.*
Stata: khb
Breen, R., K. B. Karlson, A. Holm. Total, Direct, and Indirect Effects in Logit and Probit Models. Sociological Methods & Research. 2013;42(2):164-91.

g-estimation (Implemented in Stata as gformula)
Daniel, R. M., B. L. De Stavola, S. N. Cousens. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata Journal. 2011;11(4):479-517.
Goetgeluk, S., S. Vansteelandt, E. Goetghebeur. Estimation of controlled direct effects. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2008;70:1049-66.

Watch: video lecture

In class:
Examples of formal analysis of mediated relationships

Reference/recommended:
- James LR and Brett Jeanne M. "Mediators, Moderators and Tests for Mediation." Journal of Applied Psychology. 1984; 69(2):307-321. Available on CCLE website.
- James LR, Mulaik, SA and Brett JM. "A Tale of Two Methods." Organizational Research Methods 2006 9: 233. Available at: http://orm.sagepub.com/content/9/2/233.full.pdf+html
- Iocabucci D. "Mediation analysis and categorical variables: The final frontier." Journal of Consumer Psychology. 2012; 22(4):582-94.
- MacKinnon DP and Cox MC. Commentary on "Mediation Analysis and categorical variables: The final frontier," by Dawn Iacobucci." Journal of Consumer Psychology. 2012; 22(4):600-602.

Lab 4   02/03   Confounders & interactions (moderation) & mediation (Dr. Khurana) & factor analysis

Week 6:
Reminder: Next week you will be presenting your exploratory analyses for your research projects

11   02/07   Generating and interpreting predicted levels and probabilities

Prior to class:
Read:
- Stata manual, margins command
- Williams, Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects, Presentation to Stata Users Conference, July, 2011. On CCLE.
- Exploring Regression Results Using Margins, University of Wisconsin SSCC, 2016. On CCLE.

- UCLA ats Stata FAQ How can I use the margins command to understand multiple interactions in regression and anova?
- UCLA idre Stata FAQ How can I use the margins command to understand multiple interactions in regression and anova? [different from ats document; both on ccle]

In class:
Lecture and examples of margin command
Discussion of margins

12   02/09   Difference in difference models and regression discontinuity models

Prior to class:
Lecture:TBD
Read:
- Wing C, Simon K, Bello-Gomez RA. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. Annu Rev Public Health. 2018 Apr 1;39:453-469. doi: 10.1146/annurev-publhealth-040617-013507. Epub 2018 Jan 12. PMID: 29328877.
- Gelman blog posts:
  o https://statmodeling.stat.columbia.edu/2019/06/25/another-regression-discontinuity-disaster-and-what-can-we-learn-from-it/
  o https://statmodeling.stat.columbia.edu/2018/08/02/38160/
  o http://www.stat.columbia.edu/~gelman/research/published/2018_gelman_jbes.pdf

  Examples to be discussed in class:
- Stevenson, A. J., Flores-Vazquez, I. M., Allgeyer, R. L., Schenkkan, P., & Potter, J. E. (2016). Effect of Removal of Planned Parenthood from the Texas Women's Health Program. N Engl J Med, 374(9), 853-860. doi:10.1056/NEJMsa1511902
- Perraillon, M. C., Konetzka, R. T., He, D., & Werner, R. M. (2019). Consumer Response to Composite Ratings of Nursing Home Quality. Am J Health Econ, 5(2), 165-190. doi:10.1162/ajhe_a_00115
Skim:
- Chang and Meyerhoefer, "Do Elections Make You Sick," NBER Working Paper 26697, 2020. Available on CCLE (From week1)

In class:
Be prepared to discuss Stevenson and Parraillon as examples of Diff in Diff and Regression Discontinuity.

Additional readings/reference material:

- Another example of DinD design: Slusky, David, Measuring the ACA's Parental Mandate Using Difference-in-Differences (August 22, 2014). Available at SSRN: https://ssrn.com/abstract=2177436 or http://dx.doi.org/10.2139/ssrn.2177436
- Another example of regression discontinuity: Wherry, Laura R., & Meyer, Bruce D. (2016). Saving Teens Using a Policy Discontinuity to Estimate the Effects of Medicaid Eligibility. Journal of Human Resources, 51(3), 556-588. doi:10.3368/jhr.51.3.0913-5918R1

Lab 5   02/10   Post-regression commands, predicted probabilities & presenting results

Week 7:

HPM225B
Instructor: Needleman

This week will focus on your presentation of the exploratory analyses from you term research projects. The goal for this week is to give you feedback on the exploratory analysis and reflections on what might be learnt that will influence the shape of your research. The instructions for the presentations are in the assignment discussion above in Section II. I will repeat here the final element of those instructions: Describe how the results from these analyses have (or will) influence your subsequent data analysis decisions, for example, helping you decide which variables to use, how to treat your outcome, the type of relationship between your outcome and primary regressor, how to construct a scale, identify potential moderators or mediators, etc.

All students are expected to be active participants in the discussion of the exploratory analyses of their classmates.

Screen sharing will be available so you can make your presentation. If all presentations are not completed, we will finish next week.

13	02/14	**Exploratory analysis presentations (1)**
14	02/16	**Exploratory analysis presentations (2)**
15	02/22	**Writing and presenting research 1 (In lieu of lab session this week)**

The next two sessions focus on writing and presenting findings. This is a large subject and could be an entire class. Goal is to provide an orientation to writing and presenting, with some attention paid to visualization and data presentation.

Clear writing is essential to the presentation of research findings. Clarity in writing begins with clarity in understanding what you have found, what your audience wants to learn about what you have found, and how to structure your written and in-person presentation of findings to communicate these.

Visualization and data presentation are increasingly being recognized as critical skills. You need to learn how to deliver engaging and captivating presentations to your audience – whether in an academic setting or professionally as policymakers, public health scholars, or as researchers. You must also find ways to visualize your data to provide an optimal experience to the reader – most information for the least effort. The Schwabish videos and required and recommended readings provide good orientation and advice in these task.

Prior to class:
Watch: Needleman video lecture (short, without examples)
Watch either:
  Schwabish, NBER 2017 lecture on data communication:
https://www.nber.org/lecture/summer-institute-2017-methods-lecture-better-data-communication
   Or Schwabish 2021 seminar at Pomona College:
   https://www.youtube.com/watch?v=P3k-QkeEmPc

Read:
▪ Greenhalgh T, "How to Read a Paper: Assessing the Methodological Quality of Published Papers," British Medical Journal 1997; 315:305-8. Available for free online at http://www.bmj.com/cgi/content/full/315/7103/305
▪ Schwabish JA. An economist's guide to visualizing data. *Journal of Economic Perspectives* 2014;28(1):209-234. https://www.proquest.com/scholarly-journals/economists-guide-visualizing-data/docview/1504127378/se-2. doi: http://dx.doi.org/10.1257/jep.28.1.209.
Read introduction and skim:

- Special Issue on Reviewing in: Academic Medicine September 2001 - Volume 76 - Issue 9, pp. 863-975. Introduction, Chapters 1-4, and Appendices (all articles). Available for free online at http://journals.lww.com/academicmedicine/toc/2001/09000

Reflect on how you organize yourself to write a research paper. To what extent do you outline? Does your outline include full ideas or just topics? Do you outline from beginning to end sequentially, or work out of order? How do you think about the scope of the analysis to be presented, introduction and discussion?

Independently, Andrew Gelman and I, arrived at similar recommendations on how to write an article. Mine are in the presentation. His are here:

   https://statmodeling.stat.columbia.edu/2009/07/30/advice_on_writi/

In class:
Examples of issues that emerged in review of papers in work-in-progress seminar and as reviewer, tied to key points in lecture
Discussion of challenges you have faced in writing.

Recommended/reference:
- Good PI and Hardin JW. Common Errors in Statistics (and How to Avoid Them). Hoboken, NJ: John Wiley & Sons, 2003. Chapter 8.
- Lang TA and Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers. Philadelphia, PA: American College of Physicians, 1997. Will be on reserve in biomedical library.
- Other Greenhalgh papers in the 1997 British Medical Journal series on How to Read A Paper.

Worksheets for presentation preparation, available on CCLE:
- Better presentation worksheet
- Better presentation supplies checklist

The following books are recommended and e-copies are available for download through the UC library
- Schwabish, J. A. (2017). *Better presentations : a guide for scholars, researchers, and wonks*. New York City: Columbia University Press.
- Schwabish, J. A. (2021). *Better data visualizations : a guide for scholars, researchers, and wonks*(pp. 1 online resource).
- Schwabish, J. A. (2020). *Elevate the debate : a multi-layered approach to communicating your research*(First edition. ed., pp. 1 online resource).

Reference list of standards for writing:
- STROBE Strengthening the Reporting of Observational Studies in Epidemiology
  - http://www.equator-network.org/index.aspx?o=1032
- Quasi experimental / non-randomized evaluations
  - TREND - Transparent Reporting of Evaluations with Non-randomized Designs
  - http://www.equator-network.org/index.aspx?o=1032
- Randomised (and quasi-randomised) controlled trial
  - CONSORT – Consolidated Standards of Reporting Trials
  - http://www.equator-network.org/index.aspx?o=1032
- Study of Diagnostic accuracy / assessment scale
  - STARD Standards for the Reporting of Diagnostic Accuracy studies
  - http://www.equator-network.org/index.aspx?o=1032
- Systematic Review of Controlled Trials
  - PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses
  - http://www.equator-network.org/index.aspx?o=1032
- Systematic Review of Observational Studies
  - MOOSE Meta-analysis of Observational Studies in Epidemiology

- o http://www.equator-network.org/index.aspx?o=1032
- Qualitative studies
  - o COREQ: Consolidated criteria for reporting qualitative research Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care 19 (6), 349-357. (http://dx.doi.org/10.1093/intqhc/mzm042)

Week 8:
==Reminder Due 2/25: CHIS Exploratory Assignment==

==Reminder: Submit topics for discussion/presentation in session 19 (3/8) no later than 3/1. Can be submitted by email or on CCLE site in dedicated feedback activity.==

NA   02/21   No class: President's Day Holiday

16   02/23   Writing and presenting research 2

Prior to class:
Read:
- Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, Moyer V, Guyatt G, for the Evidence-Based Medicine Teaching Tips Working Group. "Tips for Learners of Evidence-Based Medicine: 1. Relative Risk Reduction, Absolute Risk Reduction and Number Needed to Treat," Canadian Medical Association Journal 2004 August 17;171 (4) 353-358.
  Available for free online at http://www.cmaj.ca/cgi/reprint/171/4/353
- United Nations Economic Commission for Europe. Making Data Meaningful: Part 2: A Guide to Presenting Statistics. 2009. Available at:
  http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf

Reflect on challenges of presenting quantitative findings clearly so they are understood and their intended impact on the reader is realized.
Look at the good example of a chart with high cognitive load on page 26 on the UNECE document. How would you improve it.

In class:
Continued discussion of presentation of findings, focusing on presentations at conferences and presenting quantitative data.
Discussion of responding to reviewers
Continued discussion of challenges you have faced in writing.

Recommended/reference:
- 11 tips on how to present research findings. Available at:
  https://www.socialsciencespace.com/2010/09/11-tips-on-how-to-present-research-findings/
- Dijkers, MPJM, Brown, M, Gordon, WA. "Getting Published and Having an Impact: Turning Rehabilitation Research Results Into Gold." Available at:
  https://ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/focus19/

Lab 6   02/24   Topic TBD (Kha)

Week 9:
==Reminder: Due 03/01: Draft of paper except results, but including descriptive statistics and exploratory analysis. Ungraded but opportunity for early feedback==

17   02/28   Discussion of how papers going
Or Discussion of CHIS conceptual model exercise

18   03/02   Survey weights (TBD)

Prior to class:
Review:
- http://www.stata.com/meeting/dcconf09/dc09_pitblado_svy.pdf
- http://www.stata.com/meeting/canada09/ca09_pitblado_handout.pdf

In class:
Lecture

Lab 7   03/03   Survey weights (Khurana)

Week 10:

19   03/07   Student-requested topics
No assignment.

20   03/09   Course review/Evaluation

Prior to class:
Watch: lecture
Reflect on course. Be prepared to discuss key lessons for you in how to design and conduct research.
Complete course evaluation

In class
Summary of course.
Discussion

Lab 8   03/10   Khurana office hours on research papers.

Reminder: Final paper due 03/18

V.   **INSTRUCTIONS AND TEMPLATE FOR HS225B TERM PAPER**
Revised 2017_1227

1.INSTRUCTIONS

An electronic copy of the research paper and supporting materials (see below) should be submitted to the instructors by the due date.

The paper should not exceed 6000 words + tables, graphs, and references, and shorter manuscripts are encouraged. An appendix with exploratory analysis, crosstabs or other preliminary analysis should also be provided and can be up to 2000 words + tables and graphs. A detailed template for the paper is provided below.

The emphasis for this assignment is a technical exercise to demonstrate that you understand how to formulate research questions and testable hypotheses, develop a conceptual model, design a research study that can provide a reasonable answer to your questions, interpret results, and describe the nature of any biases. The point is <u>not</u> to write a publishable paper, so you should not provide an extensive discussion of the background or why the topic is compelling.

Common problems seen in years past:

1) Inconsistencies between the research question, hypotheses, conceptual model and regression (e.g., describing constructs in one section differently from in another section, or including variables in your regression model that you never described in the conceptual and measurement models, or wording the research question as though you're testing moderation but then only including main effects in your regression equation).
2) Providing an extensive literature review and discussion of why your question is really important, but having minimal description of your study design, variables, and methods.
3) Developing a measurement model but calling it a conceptual model (e.g., including a box called "demographics" without thinking about what those variables are actually measuring – is it truly skin color or sexual organs or number of years lived that are driving different healthcare utilization patterns? Or do race and gender and age actually proxy for other conceptual factors, such as preferences, discrimination, health status, income, etc.?)
4) Justifying inclusion of variables based on significant effects found in prior literature without thinking about what the variable means conceptually in your particular regression (e.g., maybe race was significant in the last dozen studies because it proxied for SES and nobody controlled adequately for financial resources, but you're using data with detailed income and wealth data, so therefore race may either be unnecessary to include, or at least interpreted differently)
5) Failing to think about differences in sampling units that could affect results, or conducting analysis in ways that addresses these problems. For example, comparing quality outcomes in public and private hospitals for all patients, ignoring information in the descriptive statistics that the payer mix and age distribution of the patients are very different. (Large public hospitals, such as LA County, often have very low proportions of Medicare and privately insured patients, and high numbers of Medicaid and uninsured, compared to private nonprofit or for-profit hospitals. This affects the age distribution of the patients, and sometimes the disease distribution. Stratifying analysis by age or by payer would provide additional control for these differences.)

6) Failing to ensure that the sample size is the same for all of your analysis variables before you run your programs (needs to be addressed either through use of complete-case analysis or another missing data method). Descriptive statistics should be based on the analysis sample. One approach is to conduct a regression with all the variables to be included, identify the sample based using the e(sample) function, e.g.: gen byte samp=e(sample) generates a variable with 1 if in the estimation sample, and 0 if not in the estimation sample.

7) Describing every possible omitted variable without explaining whether it's actually likely to cause a bias or in which direction (if the direction of the bias is indeterminate, explain why).

8) Incomplete presentation of results. For the class, as opposed to a journal, papers should include the full regression specification. See item 3) below on how regression results should be presented. As noted in that discussion, the stata estimates command makes preparing regression tables relatively straightforward and I will accept a table as constructed by estimates without additional formatting changes **if** your variable names are largely self-explanatory.

You <u>must</u> include the following tables in your paper:

1) A table describing all your measures, with columns for: **concept** (from the conceptual model); **measure proxy** or proxies (descriptions of the variables to be constructed to capture the concept, e.g. if the concept is socio-economic status, there might be measures for income, education, occupation, and race/ethnicity); description of the **construction or coding** of the measure (e.g., a specific variable in the dataset, in which case indicate how it is scaled, a constructed variable, in which case describe the variables in the data set from which the measure is derived, such as "which will be a series of indicator variables derived from variable education, with separate indicates for less than high school, high school graduate, some college, college graduate, post-baccalaureate" or "number of visits will be the sum of number of primary care visits and number of visits to specialists"; **rationale for using this proxy**, i.e., why it is a good proxy for the underlying concept (some description here, although a citation to the literature on the measure can also be included and serve as some of the rationale for the measure); and **limitations**, a discussion of the potential weaknesses of the measure, based on its incomplete coverage of the concept, its proxying for other concepts as well as that you are using it for, weaknesses in the reliability of the data used, etc. Measurement tables can be long; they do not count against the word count. They can also be important bridges between your conceptual model and empirical analyses. While published papers do not contain tables such as this, it is a critical component of the paper and you should invest adequate time in its preparation. It is also a paper component that can be constructed while you are conducting your design and analysis, and need not wait until the end of the quarter.

2) A table summarizing your regression specification, with columns representing each of the following:

- A list of all of the variables that theoretically influence your outcome measure, including any important interactions

- The hypothesized direction of the effect of each variable.

- A brief rationale for this hypothesis; alternatively, this information can be moved to the text, if there is insufficient room in the table. Do NOT rely solely on prior empirical literature as

the rationale for a hypothesized regressor effect, but rather, briefly explain the conceptual reason in your own words.

- The measured variable(s) (proxy or proxies) for the theoretical variable. If there is no empirical variable corresponding to the theoretical measure, write N/A (and make sure you discuss the likely direction of the omitted-variable bias in the limitations section). This is the map of a relatively complete ideal regression model based on your conceptual model and an indication of what variables you will be able to include in the modeling.

Do not use any form of stepwise regression analysis unless you have multiple empirical proxies for the same theoretical variable and your sample size is too small (or the collinearity too high) to include all of them in the empirical specification.

3)   A table with descriptive statistics for the analysis sample, showing each dependent and independent variable used in the analysis, the sample size for the variable, and the mean and standard deviation (or frequency) for the variable. Present unweighted data, so that the true distributions of the variables can be determined. If useful for presenting results, the table can report these statistics for subsamples defined by values of the dependent variable or the primary regressor of interest, using separate columns for the subsamples. Constructing tables with columns for subsamples can serve as the basis for some tests of the equivalence of covariates or bivariate analysis.

4)   One or more tables with estimates based on the regression models. These tables **must** include the following information:

The dependent variable

- A row for each variable included in the regression, whether it is the primary regressor or not. If you are examining the impact of including or excluding specific variables through multiple models, present the models side by side in a table. **Include the full specification!** Some journals encourage truncated regression tables with just the primary regressors of interest and for journal submission, that's fine. For this course, the full regression should be provided, including interactions.

- You are also encouraged to do multiple specifications, including and excluding variables to test the substantive impact on other variables and to model full stratification. Specifications can and should be presented in the same table to the extent possible so that coefficient estimates can be compared. The Stata estimates command makes the creation of tables with multiple regression models relatively easy.

Which sample or subsample was used and sample size(s)

- An explanation of the estimates being shown, either in the title or footnote (e.g., "Relative risks and 95% confidence intervals")

- R-square or pseudo-R-square and the F or chi-square statistic for the whole model in a row at the bottom of the table if multiple models in the table, or alternatively in a footnote to the table.

- Any other tests conducted (e.g., joint tests of b2 b3 b4), as a row at the bottom of the table. You only need to report p-values for the joint tests, not the test statistics.

- Footnote with any relevant information not obvious from the title or table itself, such as the data sources, what type of regression model was used, what other variables were controlled in the regression, and what the omitted (reference) categories are. This is also a useful approach for articles for publication. They free the reader and referee from the need to find the relevant text, and when tables are excerpted or pulled from the paper, they can stand alone because of this detail. Notes to tables generally don't count against the word count, and won't in this case. You are strongly encouraged to liberally use table notes.

- Estimates demonstrating the direction, magnitude, variance and statistical significance of the effects. At a minimum the table should present coefficient estimates or exponentiated coefficients if appropriate (e.g., odds ratios or relative risk ratios), standard errors or t or z statistics). Confidence intervals or p-values would also be appreciated. For tables with multiple model specifications, it would be best if the coefficient or exponentiated coefficient, the SE or t or z statistic, and stars for p-values, all feasible using the estimates command, can be presented.

5)   As appropriate, graphs or tables with margins analysis or other data that provide a more informative presentation of results than the regression table.

6)   While exploratory analysis, crosstabs or other preliminary analyses are often not presented in a published paper, these should be included either in the body of the paper or an appendix, with a brief summary of the conclusions drawn from these analyses and how it influenced subsequent analyses or design decisions. These can be 2000 words long.

Note: Use of sample weights is not required in HPM225B.

2.TEMPLATE FOR HS22B RESEARCH PAPER

Name of document

> hpm225b_5finalpaper_yournamehere.doc [use the last name of the first alphabetized author in team]
> final should be replace by draft when you submit the draft

Identifying Information

- Title
- Authors' names and student ID numbers
- Date

Introduction

- State research question (can be worded generally)

- State specific testable hypotheses (and, if applicable, counter-hypotheses)

- Explain rationale for hypotheses

Please note: Merely citing previous literature does not constitute an explanation of your hypothesis.

- Explain why the topic is interesting/important, but keep this very brief. A few sentences are fine.

- <u>Succinctly</u> summarize the previous literature, focusing on its limitations and how/why your study makes a contribution to what is already known

Note: A comprehensive literature review is not required for the HS225B paper, since conducting literature review is not a focus in this quarter.

Conceptual Model

- Conceptual model graph or chart

- Verbal description of model

- Table summarizing conceptual variables, theoretical proxies, hypothesized direction of effect (may be ambiguous if competing hypotheses exist) and brief rationale for each (see description above)

Methods

> Study design.

Describe your identification strategy or how your design will allow you to assess your hypotheses and rule out competing hypotheses or explanations.
Sources of data

Study sample.
Indicate if this is a subsample of the full data and describe how it was constructed.

Measures.
Include table discussed above. Also discuss and present the analyses done to collapse categories, create categorical variables from continuous variables, or include quadratic or other nonlinear transformations in the data.

Dependent variable(s)

Primary regressor(s) of interest

Other control variables

Regression model

Statistical analysis (including methods used to deal with any missing data)

Results

Descriptive statistics

Multivariable statistics
Detailed description of results for main regressor(s)
Discussion of analysis of mediation, moderation and confounding
Brief summary of results for other covariates
Any tests performed, sensitivity analyses and specification checks

Bivariate relationships are not necessary to report unless useful to the interpretation of the multivariate analysis. Intermediate and exploratory analyses prior to a regression that add insight to the choice of measures or variables, decision to include interactions or other design choices are a key element of this course and should be included either in the main text or an appendix. Post-estimation analysis including joint tests, estimates of marginal effects, etc., are encouraged to the extent they provide insight into the questions initially posed.

Discussion

Brief summary of findings

Limitations – *note that this section is extremely important!*

- Internal validity, including potential for reverse causality and omitted variables and likely nature/direction of any biases

- External validity – if possible, describe likely nature of any biases or provide information regarding generalizability

Note: To get full credit for mentioning omitted variables and/or reverse causality, you need to explain likely direction of bias or describe competing biases.

Brief summary of clinical and/or policy implications and future research directions

Tables (see above)

References
No particular format required, but you should use an author date in-text citation format so I can more easily relate the reference list to the text.You should all be using Endnote or an equivalent, so you can quickly reformat citations, and the Health Services Research style would be a good choice for this paper.

STATA programs and output

You must separately include a <u>clean and final</u> copy of all STATA programs and log files used for your paper. These programs will be reviewed and the accuracy of the code will be considered in assigning a final grade. Format using a fixed width font, with a small enough point size (but no smaller than 8 point) so that most of your code is on a single line. I encourage heavy annotation of your code so that the purpose of any section is easily understood. This helps editing and modifying code either for this analysis or for reuse in subsequent analysis. An example of documentation at the beginning of a program:

```
/*
aana_anal_02.do
version 1.01.01
dates:
orig:   2005_0302
rev:    2005_0307 and version updated to 02
rev:    2005_0329 and version updated to 03
rev:    2005_0607 and version updated to 05
rev:     2007_1226 and version updated to 06
rev:     2007_1228 and version updated to 07
rev:    2008_0104 and version updated to 08
rev:    2008_0125
orig:  2008_0416 version 08a based on 08, specific analysis to respond to reviewer 2 re
selection
rev:    2010_0216 version 09 based on 08c, analysis for commentary


This program uses:
        Completed merged discharge, survey and aha file aana_pdd_surveyed.dta
        uses original dx file, ob_99_01_diag to correct & update mat_dia (maternal diabetes)
and mat_obe (maternal overweight
        saves analysis data set as:

ISSUES IDENTIFIED IN RESPONDING TO COMMENTARY
1. Diabetes
        Original code used icd9 648.0x codes to identify diabetes
        In literature, 648.8x codes for abnormal glucose tolerance are used to identify
gestational diabetes
        create mat_dia_rev measure that incorporates both codes and merge into data set, then
use in reanalysis

*/
```

Or for a specific process:

```
*********************
* PDD VARIABLES
*********************

* RECODE AGE, PAYER< ADMIT, RACE vars to numeric

destring new_age source_pay type_admit new_race died, replace
compress new_age source_pay type_admit new_race died

* BIRTHS FROM PDD COUNTS
* divide volume by 1000 so can interpret coefficient as per 1000 births

egen volume=count(master_seq_num),by(state hosp_id year)
replace volume=volume/1000

* AGE CATEGORY DUMMIES
* coded from categorical variable
* see codebook for original categories
* restricting coding of dummy variables to non-missing cases

gen byte agelt20 = new_age<=3 if new_age~=.
gen byte age2034 = new_age==4 if new_age~=.
gen byte age35_  = new_age==5 if new_age~=.
label var agelt20 "Age of Mother LT 20"
label var age2034 "Age of Mother 20-34"
label var age35_  "Age of Mother 35 or older"
```