Political Science 522: Research Design and Techniques for Political Science (Spring 2022)

Jake Bowers*

Version of 04/21/2022

1 Details:

Class meeting: We meet Wednesdays, 3:30 – 5:50pm US Central Time in 404 David Kinley Hall.

Office hours: Please make an appointment on http://calendly.com/jakebowers if you want to come to office hours to ensure that we can meet and talk. I am happy to find other times if the times listed on Calendly do not work for you. I'm also happy if you'd like to attend office hours in groups – just use the group-meeting Calendly option.

2 Overview

We will be revising the syllabus over the term. Please check in to make sure you are using the latest version (see the date stamps in the title and footer).

Political Science 522 begins with the premise that good question, good theory, good research design, and good writing go hand-in-hand. No design will overcome a muddled question, nor will a poorly planned or executed design suffice to answer a clearly stated question. In the absence of carefully-articulated theory, the researcher stands a good chance of looking for answers in the wrong places. In the presence of poor writing, not much else matters. Political Science 522 thus focuses on the question-theory-design-writing connections. Driving all of this is the need to make convincing arguments that colleagues and potential readers of your work find compelling.

The content and organization of this course have changed markedly over the years, largely because political scientists, especially during the past 15–20 years, have been seriously rethinking and debating the very foundations of empirical research. Most recently we are witnessing both a rise in the integration of in-depth studies of particular contexts and places or archival research (versions of which are often called "qualitative research") with traditional experiments, surveys, and other data measured at the level of individual humans or countries (versions of such research are often called "quantitative research") and a rise in the use of very large datasets measuring words or connections between entities for the purposes of prediction and measurement (versions of which are often called "machine learning" or "artificial intelligence"). The academic community defines what are and are not adequate explanations of political phenomena. In fact, it defines whether true explanations are even needed. Graduate students who gain a solid understanding of the lay-of-the-land, including the changes that have occurred and will likely occur, will be best positioned to publish influential work. Even more fundamentally, this course is designed to build a foundation that will serve students even as new and unanticipated changes in norms and expectations occur; and occur they will.

*jwbowers@illinois.edu. This syllabus was initially created and refined by Jim Kuklinski, and has been modified by Matt Winters, Cara Wong, Brian Gaines in addition to Jake Bowers.

2.1 The Nature and Challenges of Research Design

Research design is not statistics, even though the two are intertwined. Research design emphasizes the formulation of studies that produce convincing results. When one writes a paper and sends it out for review, the referees assigned to evaluate it will look for weaknesses: Has the author adequately considered alternative explanations? Has the author made the right comparisons to answer his or her research question? Does the author understand the process by which his or her data were generated? How does self-selection affect the interpretation of the statistical findings? Although a researcher might, and should be able to, use statistics to assist in addressing such questions when appropriate, these methodological questions themselves focus on research design, which precedes statistical analysis. Indeed, no amount of sophisticated statistical analysis can mend a poorly designed study.

There is no set recipe for formulating a good research design, which requires, first, knowing current disciplinary emphases, and, second, bringing *the right mind state* to bear. At a minimum, the researcher must learn how to anticipate and address criticisms that others will make. This course will identify some of the criticisms that are likely to arise; each student must then figure out how to make his or her research as persuasive as possible. This takes practice, lots and lots of practice and patience with oneself. It also requires, as suggested above, a broad understanding of disciplinary expectations.

Unfortunately, it is not simply a matter of the researcher identifying all potential criticisms and then addressing them all. Any research design entails making trade-offs, such as using many cases versus using a few cases and knowing them well. Students who truly enjoy research and intellectual engagement will find satisfaction in balancing the various trade-offs, frustrations notwithstanding.

2.2 Learning Goals

The **general goals** of PS 522 are to emphasize the importance of the following:

- Good writing
- Knowledge of current and past disciplinary standards and expectations
- Proper question formulation
- Continuing dialogue between question formulation and research design
- Cultivation of a “research-design state of mind”

Formulating strong research designs requires knowing fundamental principles. The following list of principles is neither complete nor “right,” nor are the listed principles mutually exclusive. View them as a general roadmap.

- Keep in mind that there is no single panacea that will completely fulfill your needs.
- Carefully explicate the question you are asking and seek to answer; and be sure that you answer the question you think you are answering.
- Convince colleagues of the importance and timeliness of your research. To this end, successful scholars know how to function as story tellers, lawyers, and detectives simultaneously.
- Carefully define and measure the core concepts in your study, and be sure always to keep the distinction between concepts and empirical measures of them (indicators and variables) close at hand. Theoretical/substantive discussions should consist of concepts only.
- Use past literature and logic to help you understand the (usually unobserved) process that generated your data and thus to guide your empirical research.
- Make the right empirical comparisons and eliminate alternative explanations, keeping in mind, especially, unmeasured factors that could be confounding your results.
- Explicitly consider whether the process by which the causally focal variable (like a new policy intervention or an experimental intervention) takes on values can be justifiably ignored. Know how the data you are

using were generated. (Scientists inevitably make implicit assumptions, and successful scholars know what they are.)

- Keep in mind that transparency is essential if others are to understand and believe your work.
- Gain intimate familiarity with your subject matter and bring to bear as many different methods and types of evidence, each with its unique bias, as possible.
- Carefully integrate multiple types of evidence (say, in-depth interviews plus survey responses, archival evidence plus administrative data) and reasoning whenever possible, and avoid simplistic, mechanical thinking.

The **specific goals** of the course are that students:

- Explain in their own words key concepts in science like "theory", "explanation", "cause", "measurement" and describe how such concepts fit together to guide applied research. Apply these concepts to how they evaluate their own work and the work of others.
- Understand the differences between concepts and measurements and be able to articulate the major criteria for good measurement. Evaluate measurement strategies in their own work and the work of others.
- Understand the purposes of randomization and articulate when randomization helps or does not help answer a research question. Evaluate the use of randomization as a tactic in their own work and the work of others.
- Understand the purposes of sampling and articulate when sampling (random or not) helps or does not help answer a research question. Evaluate the use of sampling as a tactic in their own work and the work of others.
- Practice scientific writing.
- Gain exposure to certain strategies for making causal inferences with and without randomization: If we make a comparison to learn about the implications of a scientific explanation, how can we claim that this observed comparison reflects on that focal explanation and not some alternative? Practice using such strategies in their own work as a way to remember what they have learned and create something new.
- Reflect on the role of different modes of social scientific research in the overall scientific enterprise as well as in efforts to use social science explanations and methods to directly intervene in public policy.

3 Course Requirements

3.1 Reading materials

This is a reading course. And this syllabus has many recommended readings for you to use in your future career. The required readings will usually be excerpts and chapters and articles and should be available on online. We will do our best to make them easy for you to access.

3.2 Weekly Exercises

Each week you will write a short response to the readings that allows you to reflect about the topic of the week and, ideally, connect it to what you are reading in other classes. These papers should be no more than 2 pages single spaced. Except under unusual circumstances, you should upload you answers to the weekly exercises to the class Moodle by 12 p.m. Central Time the day before class.

The point of this exercise is (1) to give you incentive to do the reading carefully, (2) a chance to practice writing, (3) a chance to connect what we are doing in this class to your other substantive classes, and (4) enhance in-class discussion using any questions, confusions or thoughts arising from this work.

3.3 Final Project: A Research Design

Most students in this class are writing a first year research paper. The final assignment for this course is the research design portion of that project or some other research design. The instructors will provide a template for a research design that will ask students to apply the concepts in research design from this course to their own work.

This research design could involve statistical power analysis, but most of the grade for this project will depend on clarity in describing the research question, scientific motivation, theory, hypotheses, concepts and measurement, and plan to learn about how observed data can shed light on the hypotheses and theory (for example, on the plans for making a case against alternative explanations).

3.4 Grading

We will assign grades as feedback to help you focus your energy and monitor your own progress. Humans need feedback to close the intention-to-action gap. They also need feedback to learn about their progress and for motivation. When you are trying to run faster, you need to measure your speed and to learn to focus your work (and also how to ensure you do not overwork your body and create injuries). Graduate school is no different — only what you are trying to achieve has more dimensions than simple running speed.

In this class I will use grades as feedback. All grades map roughly onto A=good, B=fair,C=unsatisfactory, and F=fail (i.e. you didn't try).

Because moments of evaluation are also moments of learning in this class, I do not curve. If you all perform at 100%, then I will give you all As.

You can redo any assignment in order to increase your grade on that assignment. If you want to resubmit something already graded, you need to let me know in advance so that I can make time to grade it again. I may delay a long time in re-grading an assignment.

I will calculate the grade based on the following factors:

20% Attendance ("A" if you show up, "F" if not. You can miss two classes without grade penalty.)

50% Weekly exercises ("A" excellent, creative, well-written; "B" passing; "C" not a good effort; "F" not turned in.)

30% Final project ("A" excellent, creative, well-written; "B" passing; "C" not a good effort; "F" not turned in.)

3.5 Rules and Guidelines

All assignments will be turned in on time, and unless there are unusual extenuating circumstances, there will be no incompletes.

The demands of graduate school can be overwhelming, and lead to mental health concerns. Students are encouraged to reach out to the instructor and/or other faculty members if they find themselves in such situations. There also are units on campus devoted solely to addressing such problems.

All final written work will be turned in as pdf files unless we have another specific arrangement. If you turn in work written in some other format without prior arrangement, I will not read it.[1]

All papers written in this class will assume familiarity with the principles of good writing in Becker [7].

3.6 Academic Integrity

According to the Student Code, 'It is the responsibility of each student to refrain from infractions of academic integrity, from conduct that may lead to suspicion of such infractions, and from conduct that aids others in

[1] For example, if you have some reason why pdf files make your life especially difficult, then of course I will work with you find another format.

such infractions.' Please know that it is my responsibility as an instructor to uphold the academic integrity policy of the University, which can be found here: http://studentcode.illinois.edu/article1_part4_1-401.html.

Please read and understand the UIUC student code (link opens PDF).

3.7 Disability Accommodations

To ensure that disability-related concerns are properly addressed from the beginning, students with disabilities who require assistance to participate in this class should see me as soon as possible. To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TTY), or e-mail a message to disability@illinois.edu

3.8 Emergencies

(Copied from the PS 530 Syllabus) The university requires that syllabi reference the campus policy on emergency responses, which can be found here: https://police.illinois.edu/emergency-preparedness/run-hide-fight/. A brief summary of the university recommendation is: follow your own instincts, and, as safe, run, hide and, as a last resort, fight (where applicable—there's not much to be gained in fighting a tornado). Certainly, if the tornado siren goes off while we're meeting, class will be immediately cancelled, and it would be prudent to head for a basement or interior room without windows, if possible.

4 Course Outline

Week 1: Introduction to the Course and a Discussion of Writing (19-Jan)

Introductions. What one social science paper or book have you found most inspiring or annoying in the last year?

Writing is central to science. So we will spend half of this first session focusing on scientific writing.

Required readings

- Howard S. Becker. *Writing for Social Scientists : How to Start and Finish Your Thesis, Book, or Article (Chicago Guides to Writing, Editin.* University Of Chicago Press, 1986. ISBN: 0226041085 (The whole book)
- George D Gopen and Judith A Swan. "The science of scientific writing". In: *American Scientist* 78.6 (1990), pp. 550–558

Exercise

1. Edit the introductary paragraph to an article you are reading for another class (not a methods article, but a substantive article that you are studying as you think about your own questions and research agenda) following Becker's approaches. You can turn in both the original paragraph and your edited version in your pdf document.

Be prepared to explain your decisions in class.

Week 2: What is science? What are we doing here? (26-Jan)

Philosophy of science scholars have offered a variety of views on the nature of science, both natural and social. These scholars have addressed important topics that are relevant to all practicing scientists.

Why begin a course devoted to research design with a section on philosophy of science? The readings included below address numerous "big" topics and questions that social scientists normally do not ask. Among them are: observables vs. un-observables (unobserved), explanation vs. prediction, induction and deduction, methodological holism vs. methodological individualism, indeterminism, level and unit of analysis, uniqueness of science (can it be justified?), and accumulation of knowledge (or not).

Among the specific questions that these readings raise: Is there a general pattern by which scientists conduct their work? Is the world too complex for social scientists truly to understand? Does scientific knowledge accrue in a way that leads scientists closer and closer to "the truth"? How would we know? Many contrast science and religion; are they really as different as many suppose, or is faith at work in both arenas?

Required reading:

Recommended that you read these in order. The more common idea that science produces cumulative knowledge via a process of impersonal cumulation is articulated in the recommended readings and probably implicit in discussions in your other classes.

- Stephen Thornton. "Karl Popper". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University, 2019
- Thomas S Kuhn. "Logic of Discovery or Psychology". In: *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965.* Vol. 4. Cambridge University Press. 1970, p. 1
- Paul Feyerabend. "Consolations for the Specialist". In: *Criticism and the Growth of Knowledge* 4 (1970), pp. 197–230

Exercise

Imagine: Three experts on Popper, Kuhn, and Feyerabend have been invited to advise the NSF. The experts are: Sally (the star student of Karl Popper), Huizhong (the star student of Thomas Kuhn), and Jose (the star student of Paul Feyerabend). The director of the Social, Behavioral and Economic Sciences (SBE) says, "Now that the president has made a cabinet post for science and we are in the middle of a pandemic, we need to rethink how to best organize science. In particular we want to know what one key feature we should require of a funding application and also what one or two key activities graduate students in the social sciences should do or classes they should take or skills they should master."

1. Please write one paragraph from Sally, Huizhong, and Jose explaining what Popper, Kuhn, and Feyerabend would say about how the NSF should make funding decisions on grants. The current NSF proposals require a one paragraph statement about "Intellectual Merit" and another about "Broader Impacts". Are those enough? Or should they be changed or deleted entirely? On what grounds should the NSF allocate scarce resources to social scientists?
2. Please write one paragraph from Sally, Huizhong, and Jose explaining what Popper, Kuhn, and Feyerabend would say about the most important activities (projects, classes, experiences) graduate students in the social sciences should do in order to contribute the most to society as scientists? Explain why.

Week 3: What does it mean to "explain"? (2-Feb)

Some say that the role of science in society is to explain things. What does this mean? Why might we care? How is this role different from other roles? Do we need explanation if we can predict?

Required Readings

- James Woodward. "Scientific Explanation". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University, 2019 (Scientific Explanation)
- Julie Zahle. "Methodological Holism in the Social Sciences". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Summer 2016. Metaphysics Research Lab, Stanford University, 2016 (Methodological Holism in the Social Sciences)

- L. Breiman. "Statistical modeling: the two cultures (with comments and a rejoinder by the author)". In: *Statistical Science* 16.3 (2001), pp. 199–231

Exercise

Choose one reading from one of your substantive classes from this term or last term. What kind of explanation is going on in that article? Compared to other articles in that class, is this form of explanation the same or different? How so? If the author(s) of that article could perfectly predict the outcome understudy, would that be an adequate substitute for explanation?

Week 4: Theory and Observation (9-Feb)

All sciences seek to explain, that is, to offer convincing stories of why and how phenomena occur. Most philosophy of science scholars agree that both good theory (deduction) and astute observation (induction) are essential to explaining why phenomena occur. Empirically-oriented social scientists seem to agree, at least in principle although formal deductive approaches to theory generation are not the norm in all areas of the social sciences. How important is theory in efforts by the social sciences to actively improve the world around us now (as opposed to offering and refining explanations of the world)?

Required Readings

- James Bogen. "Theory and Observation in Science". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2020. Metaphysics Research Lab, Stanford University, 2020
- Vincenzo Crupi. "Confirmation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2020. Metaphysics Research Lab, Stanford University, 2020
- Duncan J Watts. "Should social science be more solution-oriented?" In: *Nature Human Behaviour* 1.1 (2017), p. 0015
- Mita Giacomini. "Theory-based medicine and the role of evidence: why the emperor needs new clothes, again". In: *Perspectives in Biology and Medicine* 52.2 (2009), pp. 234–251
- K. Healy, "Fuck Nuance," *Sociological Theory*, 2017, pp. 118-127

Exercise

How would you characterize "good" theory?" Do you believe that "good" theory is a necessary component of good social scientific research? Why or why not?

Your fellow colleague walks around spreading the message, "it is time to drop discussions of theory, especially since scholars do not even agree on its meaning." Would you endorse or oppose his or her position? What would you say if asked to endorse? If asked to oppose?

Week 5: Concept Formation and Measurement (16-Feb)

Social scientists communicate via concepts. They do not directly observe concepts; rather they observe them through the measures and indicators they use. The chances of committing errors between defining concepts and measuring them are huge. For example, exactly what is a war? How do (should) we define a democracy?

Required readings

- Zachary Elkins. "Gradations of democracy? Empirical tests of alternative conceptualizations". In: *American Journal of Political Science* (2000), pp. 293–300
- Simon Jackman. "Measurement". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. 2008

- Gary Goertz. "Concepts, Theories and Numbers: A Checklist for Constructing, Evaluating and Using Concepts or Quantitative Measures". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. 2008
- Robert Adcock and David Collier. "Measurement Validity: A shared Standard for Qualitative and Measurement Validity: A shared Standard for Qualitative and Quantitative Research." In: *American Political Science Review* 95.3 (2001), pp. 529–546
- Tara Slough. *10 Things to Know About Measurement in Experiments*. 2020. URL: https://egap.org/resource/10-things-to-know-about-measurement-in-experiments/ a EGAP Methods Guide and see also Measurement Chapter in the EGAP Learning Days Book

Exercise

Choose another article from one of your substantive classes. (It should be an article that differs from the one you used for your last 2 exercises, and please select an article that is not a methods article.) Describe (1) what concept(s) is the focus of the theory in that article, (2) how it is conceptualized, and (3) if it is measured, how was the concept operationalized. Do you think the conceptualization is clear? Do the measures capture the concept, such that the goal of explanation has succeeded?

Week 6: Differing Notions of Causality and Why They Matter (23-Feb)

Philosophers and social scientists have emphasized different conceptions at different periods in time. Which conception prevails at any moment in time strongly influences how social scientists approach the empirical study of cause and effect.

Required readings

- Henry E Brady. "Causation and explanation in social science". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. 2008
- John Gerring. "Causal mechanisms: Yes, but..." In: *Comparative political studies* 43.11 (2010), pp. 1499–1526
- P. W. Holland. "Statistics and Causal Inference (with discussion)". In: *Journal of the American Statistical Association* 81 (1986), pp. 945–970
- Paul R. Rosenbaum. *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, 2017, p. 374. ISBN: 9780674975576, Chapter 3 (and as much of Chapters 1 and 2 as you need to understand Chapter 3)

Exercise

Choose another article that seeks to provide evidence for a causal claim from one of your substantive classes. (It should be an article that differs from the one you used for your prevoius exercises, and please select an article that is not a methods article.) Describe what kind of concept of causality is used in that article. Now choose one other approach to causality: how would the arguments in that paper differ using this different approach to causality?

Week 7: Experimental Logic (2-Mar)

Required Reading:

- R.A. Fisher. *The design of experiments. 1935.* Edinburgh: Oliver and Boyd, 1935, Chapters 1 and 2.
- Paul R. Rosenbaum. *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, 2017, p. 374. ISBN: 9780674975576, Chapters 1 to 4.

- Alan S Gerber, Donald P Green, and Edward H Kaplan. "The illusion of learning from observational research". In: *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences* (2014). Ed. by Dawn Langan Teele, pp. 9–32
- Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation.* WW Norton, 2012, Chapter 2.

Exercise

Choose yet another article from one of your substantive classes or your substantive field (it could be one you used before or a new one) that uses randomization as a central part of its research design. Describe how randomization works in that article to help those authors argue against alternative explanations for their observed comparisons — how it works to help them to argue that they are estimating a causal effect or testing a hypothesis about a causal effect.

Week 8: Experimental Implementation and Analysis (9-Mar)

Required Reading:

- Brian J Gaines, James H Kuklinski, and Paul J Quirk. "The logic of the survey experiment reexamined". In: *Political Analysis* 15.1 (2007), pp. 1–20
- Allan Dafoe, Baobao Zhang, and Devin Caughey. "Information equivalence in survey experiments". In: *Political Analysis* 26.4 (2018), pp. 399–416
- Gustavo Diaz, Christopher Grady, and James H Kuklinski. "Survey Experiments and the Quest for Valid Interpretation". In: *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications, 2020
- David E Broockman, Joshua L Kalla, and Jasjeet S Sekhon. "The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs". In: *Political Analysis* 25.4 (2017), pp. 435–464

Exercise

Choose an article from your substantive classes or your substantive field that uses a survey experiment if possible (if not, any randomized experiment will do). Assess the experiment in terms of the criteria provided in this week's readings. Did the authors design the experiment in a way that any of this week's authors would find problematic? Did they take advantage of particular strategies that allow for additional confidence in the inferences that they draw or that allow for mor efficient data collection?

Week 9: Break. No Class. (16-Mar)

Week 10: Learning from Experiments (23-Mar)

Required Readings

- Jason Barabas and Jennifer Jerit. "Are survey experiments externally valid?" In: *American Political Science Review* (2010), pp. 226–242
- Matthew S Winters and Rebecca Weitz-Shapiro. "Lacking information or condoning corruption: When do voters support corrupt politicians?" In: *Comparative Politics* 45.4 (2013), pp. 418–436
- Taylor C Boas, F Daniel Hidalgo, and Marcus André Melo. "Norms versus action: Why voters fail to sanction malfeasance in Brazil". In: *American Journal of Political Science* 63.2 (2019), pp. 385–400
- Trevor Incerti. "Corruption information and vote share: A meta-analysis and lessons for experimental design". In: *American Political Science Review* 114.3 (2020), pp. 761–774

- Christopher Grady. *10 Things to Know About Survey Experiments*. URL: https://egap.org/resource/10-things-to-know-about-survey-experiments/ (visited on 2019) 10 Things to Know about Survey Experiments

Exercise

Choose an article – experimental or not – from your substantive classes or your substantive field that uses survey data. What do the authors say that we are learning from the research? Think through other possible interpretations of the results. Provide a new hypothesis about what the reported results could mean, and then describe an experimental research design that would help you adjudicate between the authors' original interpretation and your alternative interpretation.

Week 11: Natural Experiments and Instruments (30-Mar)

Required Readings

- Thad Dunning. *Natural Experiments in the Social Sciences: A Design-Based Approach*. First. Cambridge University Press, 2012, Chapters 1–2, Chapter 5.1, Chapter 8.1, 8.3, 8.4
- Jasjeet S Sekhon and Rocio Titiunik. "When natural experiments are neither natural nor experiments". In: *American Political Science Review* 106 (2012), pp. 35–57
- José R Zubizarreta, Dylan S Small, and Paul R Rosenbaum. "Isolation in the construction of natural experiments". In: *The Annals of Applied Statistics* (2014), pp. 2096–2121

Exercise

Find an article of substantive interest to you that uses a natural experiment (defined by the authors in some way or another). Explain how they make the case for an "as-if random" research design. Do you find their case compelling? Why or why not? What else might they have done to improve their observational research design (recall that an observational design is one where we can choose what to observe but not what to manipulate or intervene).

Week 12: Quasi-Experiments (6-Apr)

Required Readings

- Donald T Campbell. "Reforms as experiments." In: *American psychologist* 24.4 (1969), p. 409
- P.R. Rosenbaum. "Choice as an Alternative to Control in Observational Studies (with discussion)". In: *Statistical Science* 14.3 (1999), pp. 259–304
- Paul R Rosenbaum. "How to see more in observational studies: Some new quasi-experimental devices". In: *Annual Review of Statistics and Its Application* 2 (Apr. 2015), pp. 21–48

Exercise

You are discussant at a conference panel on the subject of stimulus spending. Paper #1 compares annual mean unemployment rates before and after stimulus-spending packages in states that introduced them. Paper #2 compares differences in unemployment rates in the years immediately before and after stimulus packages, and then compares the mean of those differences to the mean difference before and after in other states that did **not** introduce stimulus-spending packages. Paper #3 argues that stimulus-spending packages are never introduced at random, and so matches states that did introduce them to similar states that did not. What are the strengths and weaknesses of each approach? What approach would you propose which would provide the best answer to a governmental official trying to figure out whether to use stimulus-spending to reduce unemployment?

Week 13: Causal Inference without instruments or discontinuities (13-Apr)

How to strengthen arguments against alternative explanations when you may only have one focal “treated” unit, or when you only observe a cross-section.

Required Readings

- Paul R Rosenbaum. “Modern algorithms for matching in observational studies”. In: *Annual Review of Statistics and Its Application* 7 (2020), pp. 143–176
- Paul R. Rosenbaum. *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, 2017, p. 374. ISBN: 9780674975576, Chapter 11 on Matching Techniques.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program”. In: *Journal of the American Statistical Association* 105.490 (2010)
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. “Comparative politics and the synthetic control method”. In: *American Journal of Political Science* (2014)

Exercise:

Choose one:

Choose a paper from one of your substantive classes or that you are reading for a paper that you are writing. This paper should use either stratification via modern matching or synthetic control weighting to make the case in favor of appropriate adjustment. Did the authors of this paper do a good job in making the arguments required by the techniques? If so, what in particular did they do well? Did they do a poor job in explaining and making the arguments required? If so, what in particular could they have improved upon?

OR

Choose a paper from one of your substantive classes or that you are reading for a paper that you are writing. This paper should use an observational study to make its case and use cross-sectional comparison **or** seemingly has only one focal or “treated” unit and should use neither matching nor a synthetic control approach to adjustment. Describe how you would either implement a matched design or a synthetic control study in that case (assuming that the paper you chose doesn’t use either). What would you have to show the reader (and yourself) in order to feel confident enough to proceed with estimation and testing?

Extra: For a fully model based approach to statistical adjustment see P Richard Hahn, Jared S Murray, Carlos M Carvalho, et al. “Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion)”. In: *Bayesian Analysis* 15.3 (2020), pp. 965–1056 (including the discussion from other folks working in that area).

Week 14: Sensitivity Analysis (20-Apr)

How can we convey to ourselves (and others) the role that alternative explanations might play in driving a given observed comparison as compared to our focal explanation when we cannot observe evidence about all alternative explanations? This is where sensitivity analysis comes in: reasoning about the possible influences of the unobserved on observed results helps us update our beliefs about explanation and theory and policy decision making.

Required Readings

- J. Cornfield et al. “Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions”. In: *Journal of the National Cancer Institute* (1959)
- Paul R. Rosenbaum. *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, 2017, p. 374. ISBN: 9780674975576, Chapter 9

- Carlos Cinelli and Chad Hazlett. "Making sense of sensitivity: Extending omitted variable bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1 (2020), pp. 39–67
- Stephen Chaudoin, Jude Hays, and Raymond Hicks. "Do We Really Know the WTO Cures Cancer?" In: *British Journal of Political Science* 48.4 (2018), pp. 903–928

Exercise:

Choose a paper from one of your substantive classes or that you are reading for a paper that you are writing. This paper should use an observational study or quasi-experiment or natural experiment design (this includes regression discontinuity designs, difference-in-differences designs, cross-sectional comparisons, etc.).

If the paper uses a formal sensitivity analysis, report on what they found. What kinds of unobserved variables were they most worried about? Evaluate their sensitivity analysis: would this be how you would have done it? If not, how would you have done it? If so, why?

If the paper does not involve a formal sensitivity analysis, propose one. How would you mount a sensitivity analysis for this paper? What kinds of unobserved variables would you be most worried about?

Week 15: Design Presentations Day 1 (27-Apr)

Students present drafts of the first sections of their research designs.

Week 16: Design Presentations Day 2 (4-May)

Students present drafts of the first sections of their research designs.

Week 17: DUE May 13: Final Project

5 Extra Topics

Topic 1: The Place of Qualitative Research in a Quantitatively-Oriented Discipline

Required Readings

Exercise

What kind of qualitative approach would help you in your own work the most? Why?

Topic 2: Analysis of Observational Data: Selection on Observables

As noted earlier, the term "selection on observables" refers to an assumption, often implicit, that all of the variables affecting both the outcome and selection into the treatment have been measured and fully incorporated into the analysis. This is one of the most crucial assumptions that users of observational data have routinely made, and it is reflected in the commonly stated words, with respect to regression, that "I have controlled for all confounding factors." If the assumption is wrong, the consequence can be biased estimates of the treatment effect. And, to paraphrase Rosenbaum, *proving* that the assumption is "right" fringes on the impossible, given the likelihood of the confounding effects of unobserved variables. Thus the user of observational data must convince others that his or her results are "right".

Recommended: Naïve Regression Review

G. King, R. Keohane, and S. Verba, *Designing Social Inquiry*, pp. 91-99, 150-188

Recommended: Matching on Observables

Leaders of the so-called causal inference movement have proposed careful (a key word) matching on observed variables via propensity scores as a means to increase confidence in studies that rely completely on observed variables. The first two readings set forth the rationale for propensity scores, as well as offer advice on how to maximize the value of matching. Lyall's intriguing study of the effects of violence on insurgent attacks presents a case for matching (are you convinced?). Angrist and Pischke summarize an interesting study of the effects of attending a prestigious college or university on subsequent income. The study challenged most existing work on the subject. Pay particular attention to how the authors of the study go about matching. Does it differ from Lyall's matching strategy?

P. Rosenbaum and D. Rubin, "Reducing Bias in Observational Studies Using Sub-Classification on the Propensity Score," *Journal of the American Statistical Association*, 1984, pp. 516-524

P. Rosenbaum, *Observation and Experiments*, ch. 11

J. Lyall, "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya," *Journal of Conflict Resolution*, 2009, pp. 331-362

J. Angrist and J-S. Pischke, *Mastering 'Metrics*, ch. 2

Recommended: Emergent and Continuing Concerns

Users of naïve regression emphasize the importance of controlling for all potentially confounding variables. Achen's "rule of three" challenges that emphasis. Chaudoin and Hicks demonstrate the ease with which researchers can find statistically significant results when using naïve regression. One of the arguments buttressing the rise of quantitative associational studies in the late 1950s and early 1960s was that the use of random samples of populations facilitates generalization to the populations. This rallying cry seemingly was untouchable until Aronow and Samii published their article.

C. Achen, "Toward a New Political Methodology: Microfoundations and ART," *Annual Review of Political Science*, 2002, pp. 423–450

S. Chaudoin, J. Hays, and R. Hicks, "Do We Really Know the WTO Cures Cancer?" *British Journal of Political Science*, 2018, pp. 903-928

P. Aronow and C. Samii, "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science*, 2016, pp. 250-267

Recommended: Anomaly or Lesson

An argument presumably can be made that the Wand et al. article falls into the naïve

regression category, its sophistication notwithstanding. The findings also seem

untouchable. What makes this study so convincing?

J. Wand, K. Shotts, J. Sekhon, W. Mebane, M. Herron, and H. Brady, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," *American Political Science Review*, 2001, pp. 793-810

Recommended: Other

P. Rosenbaum, "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment," *Journal of the American Statistical Association*, 1984 pp. 41-48

P. Rosenbaum and D. Rubin, "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 1985, pp. 33-38

D. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 1974, pp. 688-701

D. Rubin, "For Objective Causal Inference, Design Trumps Analysis," *Annals of Applied Statistics*, 2008, pp. 808-840

J. Zubizarretai, D. Small, and P. Rosenbaum, "Isolation in the Construction of Natural Experiments," 2015, unpublished paper

P. Rosenbaum, "Differential Effects and Generic Biases in Observational Studies," *Biometrika*, 2006, pp. 573-586 (An easier-to-read version is P. Rosenbaum, "Using Differential Comparisons in Observational Studies," *Chance*, 2013, pp. 1-14)

J. Sekhon, "The Neyman-Rubin Model of Causal Inference and Estimation Via Matching Methods," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*

K. Arceneaux, A. Gerber, and D. Green, "A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Empirical Benchmark," *Sociological Methods and Research*, 2010, pp. 256-282

Recommended: Applied Examples

S. Dale and A. Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics*, 2002, pp. 1491-1527

M. Gilligan and E. Sergenti, "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference," *Quarterly Journal of Political Science*, 2008, pp. 89-122

R. Nielsen, M. Findley, Z. Davis, T. Candland, and D. Nielson, "Foreign Aid Shocks as a Cause of Violent Armed Behavior," *American Journal of Political Science*, 2011, pp. 219-232

M. Humphreys and J. Weinstein, "Demobilization and Reintegration," *Journal of Conflict Resolution*, 2007, pp. 531-567

C. Kam and C. Palmer, "Reconsidering the Effects of Education on Political Participation," *Journal of Politics*, 2008, pp. 612-631

J. Henderson and S. Chatfield, "Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation," *Journal of Politics*, 2011, pp. 646-658

C. Kam and C. Palmer, "Rejoinder: Reinvestigating the Causal Relationship between Higher Education and Political Participation," *Journal of Politics*, 2011, pp. 659-663

R. Abramitzky, L. Boustan, and K. Eriksson, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 2012, pp. 1832-1856

Topic 3: Emerging Considerations

Four increasingly important and overlapping areas of research are treatment effect heterogeneity, spillover or interference, assessment of the problems of potential bias in observational studies, and, last and certainly not least, so-called machine learning. The focus on treatment effect heterogeneity represents a departure from the past emphasis on average causal effects. The rise of potential outcomes brought spillover, a violation of SUTVA, to the fore. Sensitivity tests of bias entail changing the assumptions of an observational study to see whether the results change, and, if so, by how much. Finally, and as statistician Leo Breiman writes in his must-read 2001 article, the rise of machine learning represents more than another methodological approach; it represents a change in culture, from stochastic data modeling to algorithmic modeling. The former makes assumptions about what goes on inside the black box of nature, the second takes the black box as complex and unknown and "let's the machine uncover the complexity." At this moment in time, machine learning emphasizes prediction, whereas traditional modeling is viewed as a tool to construct explanations of why and how phenomena occur. Within the next several years, these areas of research will become integrated in ways that currently have not yet been identified.

Recommended: Treatment Effect Heterogeneity

Determining whether different types of unit react differently to treatments, or independent variables, has a long tradition in social scientific research. Typically, scholars have identified this heterogeneity by estimating how specified variables moderate the relationship between an independent and dependent variable. This research remains important, in fact, has increased in importance. More recent research focuses on response to treatment, quite independently of demographics. The readings below report some new approaches to treatment effect heterogeneity. Note that machine learning represents an emerging approach to the topic of heterogeneity, although we will not cover this in-its-infancy approach here.

Perhaps the most interesting story surrounding the emergence of treatment effect heterogeneity is that it runs counter to what most social scientists have done for decades: identify the average treatment effect. One should not automatically assume that "more recent is better," and the subject of whether to estimate average or heterogeneous treatment effects remains open to debate. Of course, the smart answers are "both" and/or "it depends."

The Angrist piece is a classic. Gerber and Green explore heterogeneity in the experimental context. In the last two pieces (we will read only one), the authors propose an interesting and potentially valuable way to use matching of likelihood of receiving the treatment to identify heterogeneity in responses to treatment.

J. Angrist, "Treatment Effect Heterogeneity in Theory and Practice," *The Economic Journal*, 2004, pp. C52-C84

A. Gerber and D. Green, *Field Experiments,* ch. 9, pp. 289-318

J. Brand and J. Simon Thomas, "Causal Effect Heterogeneity," in S. Morgan, ed., *Handbook of Causal Analysis for Social Research*, ch. 11

U. Xie, J. Brand, and B. Jann, "Estimating Homogeneous Treatment Effects with Observational Data," *Sociological Methodology*, 2012, pp. 314-347

D. Manning, "Instrumental Variables for Binary Treatments with Heterogeneous Treatment Effects: A Simple Exposition," *Contributions to Economic Analysis*, 2004, pp. 1-14

F. Elwert and C. Winship, "Effect Heterogeneity and Bias in Main-Effects-Only Regression Models," in R. Dechter, H. Geffner, and J. Halpern, eds., *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, ch. 19

W. Rhodes, "Heterogeneous Treatment Effects: What Does a Regression Estimate?" *Evaluation Review,* 2010, pp. 334-361

C. Kam and M. Trussler, "At the Nexus of Observational and Experimental Research: Theory, Specification, and Analysis of Experiments with Heterogeneous Treatment Effects" *Political Behavior,* 2017, pp. 789–815.

J. Grimmer, S. Messing, and S. Westwood. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods," *Political Analysis*, 2017, pp. 413-434

Recommended: Spillover and Interference

Whether referred to as spillover, contagion, contamination, or interference, the violation of the assumption that one unit's treatment assignment does not affect another's outcome (SUTVA) seemingly is violated every day, in all walks of life. Social scientists have tried to overcome almost certain spillover effects by aggregating from the individual to a higher level, such as, for example, a neighborhood. The task is easier said than done.

The first three articles approach spillover and interference from an experimental perspective. Sampson uses widely-cited empirical data to document the challenges facing efforts to identify spillover empirically. Nickerson's study illustrates what can be accomplished with a little creativity.

A. Gerber and D. Green, *Field Experiments*, ch. 8, pp. 253-288

B. Sinclair, "Design and Analysis of Experiments in Multi-Level Populations," in Druckman et al., eds., *Handbook of Experimental Political Science*

J. Bowers, M. Fredrickson, and C. Panagopoulos, "Reasoning about Inference between Units: A General Framework," *Political Analysis*, 2013, pp. 97-124

R. Sampson, "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure," *American Journal of Sociology*, 2008, pp. 189-231

D. Nickerson, "Is Voting Contagious? Evidence from Two Field Experiments," *American Political Science Review*, 2008, pp. 49-55

Recommended: Machine Learning

No first-year graduate student will make it through the first year without hearing a reference to machine learning. It seemingly is "the" hot methodological topic these days. We limit our reading to a single article, one that effectively distinguishes the basic differences between machine learning and more commonly-used statistical modeling. The differences are considerable, as, Breiman documents. It is crucially important that students know and understand the fundamental differences. These differences are steeped in philosophy of science discussions.

L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science*, 2001, pp. 199-215

Recommended: Other

P. Rosenbaum, "Discussing Hidden Bias in Observational Studies," *Annals of Internal Medicine*, 1991, pp. 901-905

Topic 4: Scientific Progress in Political Science

Defining progress in any discipline borders on the impossible. If, however, scientists cannot provide compelling evidence of making progress, why bother (so would ask the ultimate skeptic)? If defining progress is next to impossible, however, scientists at least should be able to see when they are not making progress. The two subsections below approach progress from distinct perspectives.

Recommended: Advancement in Science

One of the most highly respected scholars in the philosophy has written extensively on the topic. The assigned reading is no more than a "taste."

P. Kitcher, *The Advancement of Science*, ch. 1

Recommended: Replication and Progress

Various scientific disciplines have been challenged in recent years with a "crisis of replication," where scholars have been unable to reproduce results achieved by previous scholars. What might be behind this crisis, and what are some ways to address it? The Ioannidis and Gerber and Malhotra articles provide tools for assessing publication bias. Humphreys, Sanchez de la Sierra, and van der Windt propose pre-registration as a mechanism for helping to reduce the publication of false positive findings, while Benjamin et al. propose a more restrictive critical value for statistical significance tests.

J. Ioannidis, "Why Most Published Research Findings Are False," *PLoS Medicine*, 2005, e124.

A. Gerber and N. Malhotra, "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals," *Quarterly Journal of Political Science*, 2008, pp. 313–326.

M. Humphreys, R. Sanchez de la Sierra, and P. van der Windt, "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration," *Political Analysis*, 2013, pp. 1-20

D. Benjamin et al. "Redefine Statistical Significance," *Nature Human Behavior*, 2018, pp. 6-10.

Recommended: Transparency, Replication, Publication Bias

T. Dunning, "Transparency, Replication, and Cumulative Learning: What Experiments Alone Cannot Achieve," *Annual Review of Political Science*, 2016, pp. 541-563.

A. Franco, N. Malhotra, and G. Simonovits, "Publication Bias in the Social Sciences: Unlocking the File Drawer," *Science,* 2014, pp. 1502–5.

J. Monogan, "Research Preregistration in Political Science: The Case, Counterarguments, and a Response to Critiques," *PS: Political Science & Politics*, 2015, pp. 425-429.

J. Freese and D. Peterson, "Replication in Social Science," *Annual Review of Sociology*, 2017 pp. 147–65.

Exercise

What can we do to improve political science? Think back to the readings on philosophy of science and causality and the set of methodological readings that we have done. What are your recommendations for us to be a "better" discipline? Be sure to provide your definition of what it means to be a "better" discipline.

A Recommended Readings

For anyone who wants to learn more on any of the topics we cover above, we include here some readings that we recommend.

Recommended Books

Joshua D Angrist and Jörn-Steffen Pischke. *Mastering'metrics: The path from cause to effect.* Princeton University Press, 2014

J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*

G. Imbens and Donald Rubin, *Causal Inference for Statistics, Social and Biomedical Sciences*

N. Cartwright and J. Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better*

P. Rosenbaum, *Observation and Experiment*

W. Shadish, T. Cook and D. Campbell, *Experimental and Quasi-Experimental Designs*

A. Gerber and D. Green, *Field Experiments: Design, Analysis, and Interpretation*

G. King, R. Keohane, and S. Verba, *Designing Social Inquiry*

J. Seawright, *Multi-Method Social Science*

S. Morgan and C. Winship, *Counterfactuals and Causal Inference*, 2^nd edition

H. Brady and D. Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*

Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*

Recommended: Philosophical Debates about the Nature and Meaning of Scientific Inquiry

For political scientists, at least, Popper, Kuhn, and Lakatos are some of the classic theorists of philosophy of science. Be sure to read them in the order they are listed below. Feyerabend is especially provocative, and the title of the work listed below captures his assessment of scientific inquiry. What leads Feyerabend to lead the charge against science as practiced? Johnson is a *New York Times* science writer who argues that science and religion might not differ to the extent that many suppose. Is reading the ideas and thoughts of a nonscientist worthy of our attention?

K. Popper, *The Logic of Scientific Discovery*, chs. 1-4

T. Kuhn, *The Structure of Scientific Revolutions*, 2^nd ed., pp. 92-110

I. Lakatos, "The Role of Crucial Experiments in Science," *Studies in the History of Philosophy of Science*, 1974, pp. 309-325

P. Feyerabend, *Against Method: Outline of an Anarchistic Theory of Knowledge,* analytical index, introduction, chs. 1-3

P.K. Feyerabend. "How to be a good empiricist: A plea for tolerance in matters epistemological". In: *Philosophy of Science: The Delaware Seminar* 2 (1963), pp. 3–39

G. Johnson, *Fire in the Mind*, 1995, Introduction, ch. 1

K. Popper, *Conjectures and Refutations: The Growth of Knowledge*, 4^th ed.

I. Lakatos, "Falsification and the Methodology of Scientific Research Programs," in I. Lakatos and A, Musgrave, *Criticism and the Growth of Knowledge,* 1970, pp. 91-193

C. Elman and M.F. Elman, "How Not to Be Lakatos Intolerant: Appraising Progress in IR Research," *International Studies Quarterly*, 2002, pp. 231-262.

"The Pleasure of Finding Things Out." A documentary about Richard Feynman

Recommended: Inherent Challenges to Scientific Inquiry

Scientists, social scientists included, seek to explain and understand the phenomena they study, confirm or reject theories on the basis of localized observations, and, more fundamentally, make progress toward an accumulation of knowledge. These are lofty goals that are more easily stated than achieved. Some philosophers of science devote their careers to these very questions. Approach these readings as insights "from on-high, above the fray."

A. Garfinkel, *Forms of Explanation*, pp. 49-74

B. Skow, "Scientific Explanation," in P. Humphreys, ed., *Oxford Handbooks Online*, 2016

A. Bird, "Scientific Progress," in P. Humphreys, ed., *Oxford Handbooks Online*, 2016

J. Sprenger, "Confirmation and Induction," in P. Humphreys, ed., *Oxford Handbooks Online*, 2016

J. Foder, "Observation Reconsidered," in P. Humphreys, ed., *Oxford Handbooks Online*, 2016

P. Godfrey and C. Hill, "The Problem of Unobservables in Strategic Management Research," in *Strategic Management Journal*, 1995, pp. 519-533 (the authors are not philosophy of science scholars, although their piece effectively reviews the positions of positivists versus realists)

Recommended: Social Science Perspectives

"Theory" holds a vaulted position in most scientific disciplines, and political science is no exception, at least with respect to rhetoric. Sutton and Staw argue that what is routinely offered as theory often is not. Lave and March, even though old and simplistic by today's standards, offers an interesting perspective on the relationship between theory and real-world observation. Schelling shows how good theory can help the researcher to identify the data-generating process and thus more properly to interpret the empirical findings. He uses racial segregation as his example. Garfinkel, whom we read earlier, criticizes the influence of individualism in social scientific research, arguing that the frequent result is a neglect of the larger environment within which individual behavior occurs, and, more importantly, which often is the primary causal factor of a phenomenon. Rosenbaum's discussion of "elaborate theories" is must reading, period. Healy's article has a provocative title – does it provide useful guidance on how to do theory? He clearly believes that research driven by empirics alone is destined not to achieve its purpose. Huber laments what he sees as an abandonment of theory resulting from the rise of the so-called causal inference movement. Finally, Stark's discussion of the rise of Christianity is an example of the value of using both deduction and induction when data are limited.

R. Sutton and B. Staw, "What Theory is Not," *Administrative Science Quarterly*, 1995, pp. 371-384

C. Lave and J. March, *An Introduction to Models in the Social Sciences*, pp. 1-84

T. Schelling, *Micromotives and Macrobehavior*, ch. 4

A. Garfinkel, *Forms of Explanation*, pp. 49-62 (repeated)

P. Rosenbaum, *Observation and Experiment*, ch. 7

K. Healy, "Fuck Nuance," *Sociological Theory*, 2017, pp. 118-127

J. Huber, "Is Theory Getting Lost in the 'Identification Revolution?" *Monkey Cage*, 2013

R. Stark, *The Rise of Christianity*, chs. 1, 4, 5, 6

Recommended: Other Readings on Theory and Explanation

P. Hedstrom and R. Swedberg, "Social Mechanisms: An Introductory Essay, in Hedstrom and Swedberg, eds., *Social Mechanisms*

A. Smith, "Testing Theories of Strategic Choice: The Example of Crisis Escalation," *American Journal of Political Science,* 1999, pp. 1254-1283

K. Clarke and D. Primo, "Modernizing Political Science: A Model-Based Approach," *Perspectives on Politics*, 2007, pp. 741-753

D. Lake, "Theory is Dead, Long Live Theory: The End of the Great Debates and the Rise of Eclecticism in International Relations," *European Journal of International Relations*, 2013, pp. 567-587

Recommended: Concept Formation

The following are some useful works on concept formation. Kaplan and Rosch are classic works. Adcock and Collier offer an approach that will be especially helpful to students of comparative politics.

A. Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science*, chs. 1-2

E. Rosch, "Principles of Categorization," in E. Rosch & B. B. Lloyd, eds., *Cognition and Categorization*

R. Adcock and D. Collier, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review*, 2001, pp. 529-546

Recommended: Concept Measurement

The Jackman piece is an especially useful discussion of measurement. Jackman discusses the basics of measurement and identifies the consequences of poor measurement (yes, there are consequences). Carmines and Zaller is a good introduction to the basics of concept measurement. Mondak's important piece on the measurement of respondents' knowledge in survey settings convincingly shows that poor measurement can lead to grossly wrong conclusions.

S. Jackman, "Measurement," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*, pp. 1-11

E. Carmines and R. Zaller, *Reliability and Validity Assessment*

J. Mondak, "Developing Valid Knowledge Scales." *American Journal of Political Science,* 2001, pp. 222-238

Recommended: Other Readings on Measurement

Z. Elkins, "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations," *American Journal of Political Science*, 2000, pp. 293-300

J. Cheibub, J. Gandhi, and J. Vreeland, "Democracy and Dictatorship Revisited," *Public Choice* 2010, pp. 67-101

H. Cheon and E. Machery, "Scientific Concepts," in P. Humphreys, ed., *Oxford Handbooks*

E. Rosch, "Human Categorization," in N. Warren, ed., *Advances in cross-cultural psychology*, vol. 1

D. Collier and J. Mahoney, "Conceptual 'Stretching' Revisited: Alternative Views of Categories in Comparative Analysis," *American Political Science Review*, 1993, pp. 845-855

G. Goertz, *Social Science Concepts: A User's Guide*, all

G. Goertz, "Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*

J. LaPorte and J. Seawright, "Typologies: Forming Concepts and Creating Categorical Variables," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *Oxford Handbook of Political Methodology*

C. Ragin, *Fuzzy-Set Social Science*, ch. 6

G. Sartori, "Concept Misinformation in Comparative Politics," *American Political Science Review*, 1970, pp. 1033-1053

J. Gerring, "What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences," *Polity*, 1999, pp. 357-393

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, pp. 64-82, 341-373

D. Campbell and D. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 1959, pp. 81-105

K. Imai and T. Yamamoto, "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis," *American Journal of Political Science*, 2010, pp. 543-560

Recommended: Applied Examples of Measurement

D. Collier and S. Levitsky, "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics*, 1997, pp. 430-451

G. Goertz and P. Diehl, "Enduring Rivalries: Theoretical Constructs and Empirical Patterns," *International Studies Quarterly*, 1993, pp. 147-171

Recommended: History of Causality in Academic Scholarship

The following two readings trace the ebb and flow of causality and causal analysis over time. Sociologists coauthored the first reading, an economist authored the second.

S. Barringer, S. Eliason, and E. Leahey, "A History of Causal Analysis in the Social Sciences," in S. Morgan, ed., *Handbook of Causal Analysis for Social Research*, pp. 9-26

K. Hoover, "Causality in Economics and Econometrics," *The New Palgrave Dictionary of Economics*, 2ˆnd ed.

Recommended: Conceptions of Cause and Causality

Scholars have been studying and debating causality for centuries. Brady presents an especially comprehensive overview of the various conceptions of causality. Freese and Kevern provide a "big picture" view of how scholars have characterized causes. Holland's article, in which he distinguishes "causes of an effect" from "effects of a cause", has strongly influenced thinking about causality. Mackie's INUS conditions, discussed in multiple readings, is one of the most prominent conceptions of causality among philosophers of science, although not among social scientists. Scholars generally accept the path dependence thesis, although they also seem perplexed about how to use it.

H. Brady, "Causation and Explanation in Social Science," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*

J. Freese and J. Kevern, "Types of Causes," in S. Morgan, ed., *Handbook of Causal Analysis for Social Research*, 2013, pp. 27-41,

P. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986, pp. 945-960

J. Mackie, "Causes and Conditions," *American Philosophical Quarterly*, 1965, pp. 245-264

P. Pierson, "Path Dependence, Increasing Returns, and the Study of Politics," 2000, *American Political Science Review*, pp. 251–67

Recommended: Alternative Procedures to Learn about Causal Relationships — Emphases and Assumptions

The readings immediately below offer useful overviews of, in the first case, naïve regression, and, in the second, potential outcomes/counterfactuals. They provide useful introductions to the basic rationales and workings of the two approaches. As you will read, one approach tends to be more explicit about its underlying assumptions than the other. Note that naïve regression reflects adoption of the associational view of causality, whereas causal inference reflects adoption of the potential outcomes/counterfactual view.

We will revisit both causal inference and naïve regression at various junctures throughout the semester.

J. Angrist and J. Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Economics," *Journal of Economic Perspectives*, 2010, pp. 3-30

G. King, R. Keohane, and S. Verba, *Designing Social Inquiry*, pp. 91-99, 150-188

G. Imbens and D. Rubin, *Causal Inference for Statistics, Social, and Medical Sciences*, ch. 1

S. Morgan and C. Winship," *Counterfactuals and Causal Inference*, chs. 1, 2

P. Rosenbaum, "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment," *Journal of the American Statistical Association,* 1984, pp. 41-55

J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge U.P., 2000

Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer.* John Wiley & Sons, 2016

Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect.* Basic Books, 2018

Recommended: Other Readings from Social Science about Causality

B. Braumoeller and G. Goertz, "The Methodology of Necessary Conditions," *American Journal of Political Science*, 2000, pp. 844-858 (not included in the required list of readings, although everyone should read it at some juncture)

S. Page, "Path Dependence," *Quarterly Journal of Political Science*, 2006, pp. 87-115

D. Waldner, "Process Tracing and Causal Mechanisms," in H. Kincaid, ed., *The Oxford Handbook of the Philosophy of Science*

D. Waldner, "Transforming Inferences into Explanations: Lessons Learned from the Study of Extinctions," in R. Lebow and M. Lichbach, eds., *Theory and Evidence in Comparative Politics and International Relations.*

J. Gerring, "Causal Mechanisms: Yes, But. . . ." *Comparative Political Studies*, 2010, pp. 1499-1526

Recommended: The Rise of Experimental Research in Political Science

Given the prominence of experimental research in political science today, one can easily forget (or never know) that experiments comprised a negligible part of the discipline's methodological toolbox two decades or so ago. One obvious question: Why the remarkable change? Gerber et al. offer a seething, largely dismissive, critique of the "naïve" regression analyses that have long been the dominant methodology in empirical social science research. This piece, along with a growing discontent with observational studies, appeared to motivate the "move to experiments" charge.

A. Gerber, D. Green, and E. Kaplan, "The Illusion of Learning from Observational Research," in I. Shapiro, R. Smith, and T. Massoud, eds., *Problems and Methods in the Study of Politics*

J. Druckman, D. Green, J. Kuklinski, and A. Lupia, "Experimentation in Political Science," in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*

Recommended: The Logic and Power of the Random Assignment Experiment

Why, precisely, have scholars viewed the random assignment experiment as the "gold standard" for showing cause and effect? Much of the answer presumably lies with what would seem to be an impeccable logic. Serious users of experimental designs will find Fisher's *Design of Experiments* indispensable reading. Rosenbaum discusses the logic of the random assignment experiment to set the context and, in his view, the standard for the evaluation of observational studies. Fisher's sharp null hypothesis looms large in that discussion. It also looms large in Gerber and Green's *Field Experiments.* Be sure you understand the relevance of Fisher's null hypothesis, and how it differs from the usual "average treatment effect." Angrist and Pichke illustrate both the logic and power of the random assignment experiment by walking the reader through several recent health care studies.

R. Fisher, *The Design of Experiments*

P. Rosenbaum, *Observation and Experiment*, chs. 1-3, pp. 3-52

A. Gerber and D. Green, *Field Experiments*: *Design, Analysis, and Interpretation,* chs. 1-2, pp. 1-50

J. Angrist and J-S Pischke, *Mastering 'Metrics*, Introduction, ch. 1

J. Druckman, D. Green, J. Kuklinski, and A. Lupia, "The Growth and Development of Experimental Research in Political Science," *American Political Science Review*, 2006, pp. 627-636

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch. 8

P. Rosenbaum, *Design of Observational Studies*, ch. 2

R. Sondheimer, “Analyzing the Downstream Effects of Randomized Experiments,” in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*

D. Green and A. Gerber, “The Downstream Benefits of Experimentation,” *Political Analysis*, 2002, pp. 394-402

Recommended: Applied Examples of Randomized Experiments

R. Sondheimer and D. Green, “Using Experiments to Estimate the Effects of Education on Voter Turnout,” *American Journal of Political Science*, 2010, pp. 174-189

A. Gerber, D. Green, and R. Shachar, "Voting May be Habit-Forming: Evidence from a Randomized Field Experiment, *American Journal of Political Science*, 2003, pp. 540-550

A. Coppock and D. Green, “Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities,” *American Journal of Political Science*, 2016, pp. 1044-1062

Recommended: The “Gold Standard” Revisited: Randomization under the Microscope

Cartwright and her various fellow collaborators have been critical of the widely shared assumption that the random experiment sets the standard by which alternative methods and designs should be evaluated. (To be clear, they are not arguing against the use of experiments, only the widespread view that experiments trump other methodological approaches.) Their argument is multi-faceted, focusing primarily on context-treatment interactions in their book, and a variety of other challenges in their articles. They no longer are lonely voices screaming in the wilderness. Bryan, for example, takes a seemingly balanced position, and yet finds a variety of reasons to be skeptical of randomization as the cure-all. In fact, he, along with a growing number of others, claims that there often are more effective alternatives to randomization when it comes to measuring causality. Kasy presents a much more formal argument.

N. Cartwright and J. Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better*, all

A. Deaton and N. Cartwright, “Understanding and Misunderstanding Randomized Control Experiments,” *Social Science and Medicine*, 2017, pp. 1-21

A. Deaton, “Randomization in the Tropics Revisited: A Theme and Eleven Variations,” in F. Bedecarrats, I. Guerin, and F. Roubaud, eds., *Randomized Controlled Trials in the Field of Development*, 2019, pp. 1-24

K Bryan, “A Fine Theorem—What Randomization Can and Cannot Do: The 2019 Nobel Prize,” online

M. Kasy, “Why Experimenters Should not Randomize, and What They Should Do Instead,” working paper, 2013

J.W. Tukey. “Tightening the clinical trial”. In: *Controlled Clinical Trials* 14.4 (1993), pp. 266–285. ISSN: 0197-2456

Recommended: The Push-Back against Randomization

Criticisms of the random assignment experiment have not gone unanswered. Imbens responds directly to Cartwright and Deaton, the latter himself an earlier recipient of the Nobel Prize in Economics. Banerjee and colleagues respond indirectly, in their discussions of what creative random assignment experiments can accomplish. Pay particular attention to the enlightening discussion under “Example: The Logic of Bayesian Experimentation” in Banerjee, Chassang, and Snowberg. In that discussion, they argue that an individual scholar, with a particular set of prior beliefs, could be satisfied using one or two subjects. As they go on to note, a problem arises when other individuals hold different priors; then a randomized experiment is the only feasible alternative. Does this argument make sense to you?

G. Imbens, Comments on: “Understanding and Misunderstanding Randomized Controlled Trials by Cartwright and Deaton”

A. Banerjee and E. Duflo, "An Introduction to the ‘Handbook of Field Experiments,” ’ in A. Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*, vol. 1, 2017, pp. 1-24

A. Banerjee, S. Chassang, and E. Snowberg, "Decision Theoretic Approaches to Experimental Design and External Validity," in A. Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*, vol. 1, 2017, pp. 141-174 (selected segments)

Recommended: More Issues in Randomized Experiments

H. Smith, "Specification Problems in Experimental and Nonexperimental Social Research," *Sociological Methodology*, 1990, pp. 59-91

S. Athey and G.W. Imbens, "The Econometrics of Randomized Experiments," in A. Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*, vol. 1, 2017, pp. 73-140

Recommended: The Empirical Logic(s) of Quasi-Experimentation

Two distinct strategies fall under the broad rubric of quasi-experimentation. One emphasizes eliminating alternative explanations and the other, making the right comparisons. Donald Campbell is generally viewed as the creator of the first strategy, and his "Reforms as Experiments" shows the power of such designs. In the same tradition, Shadish et al. review a large number of quasi-experimental designs, and in the process show the importance of including both a pretreatment and a control group. Rubin and Rosenbaum are most strongly associated with the second, "make the right comparisons" ("compare apples to apples") strategy. Rosenbaum's "Choice as an Alternative to Control in Observational Studies" is must-reading for all social scientists.

D. Campbell, "Reforms as Experiments," *American Psychologist*, 1969, pp. 409-429

W. Shadish, T. Cook, and D. Campbell, *Experimental and Semi-Experimental Designs*, chs. 4-6, skim

P. Rosenbaum, *Observation and Experiment*, ch. 8

D. Campbell and H. Ross, "The Connecticut Crackdown on Speeding," in E. Tufte, ed., *The Quantitative Analysis of Social Problems*

D. Pelowski, "On the Use of a Quasi-Experimental Design in the Study of International Organization and War," *Journal of Peace Research*, 1971, pp. 279-285

Recommended: Regression Discontinuity

In the words of Angrist and Pischke (*Mostly Harmless Econometrics*, p. 251), "Regression discontinuity (RD) designs exploit precise knowledge of the rules determining treatment. RD identification is based on the idea that in a highly rule-based world, some rules are arbitrary and therefore provide good experiments." In other words, RD designs attempt to use selection to advantage, rather than eliminate its effects.

The first three readings set forth the basic logic(s) of regression discontinuity. The fourth compares the accuracy of a regression discontinuity design to the random assignment experiment. Asking whether MPs can be "bought," Eggers and Hainmueller answer using both matching and regression discontinuity. Do you find their conclusions convincing?

J. Angrist and J-S. Pischke, *Mastering 'Metrics*, ch. 4, pp. 147-177

J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 6, pp. 251-267.

C. Skovron and R. Titiunik, "A Practical Guide to Regression Discontinuity: Designs in Political Science," 2015, unpublished

D. Green, T. Leong, H. Kern, A. Gerber, and C. Larimer, "Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks," *Political Analysis*, 2009, pp. 400-417

D. Eggers and J. Hainmueller, "MPs for Sale? Returns to Office in Postwar British Politics," *American Political Science Review*, 2009, pp. 1-21

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch.7

G. Imbens and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2007, pp. 615-635

T. Cook, "Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics," *Journal of Econometrics*, 2008, pp. 636-654

M. Cattaneo, L. Keele, R. Titiunik, and G. Vazquez-Bare, "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *Journal of Politics*, 2016, pp. 1229-1248

M. Cattaneo, N. Idrobo, and R. Titiunik, *A Practical Introduction to Regression Discontinuity Designs: Volume I*, forthcoming book, selections TBD

B. de la Cuesta and K. Imai, "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections," *Annual Review of Political Science*, 2016, pp. 375-396

Recommended: Applied Examples of RDD

D. Lee, E. Moretti, and M. Butler, "Do Voters Affect or Elect Policies: Evidence from the U.S. House," *Quarterly Journal of Economics*, 2004, pp. 807-859

D. Lee, "Randomized Experiments from Non-Random Selection in U.S. House Elections," *American Political Science Review*, 2008, pp. 675-697

E. Gerber and D. Hopkins, "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy," *American Journal of Political Science*, 2011, pp. 326-339

D. Gerber, D. Kessler, and M. Meredith, "The Persuasive Effects of Direct Mail: A Regression Discontinuity Design," *Journal of Politics*, 2011, pp. 140-155

A. Eggers, A. Fowler, J. Hainmueller, A. Hall, and J. Snyder, "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races," *American Journal of Political Science*, 2015, pp. 259-274

J. Angrist and M. Rokkanen, "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of American Statistical Association*, pp. 1331-1344

Recommended: Instrumental Variables

Instrumental variables and one common estimator, two-stage least squares, the most common way to implement instruments, have been around for a long time. Unlike economists, political scientists have only recently begun to understand their value and their limitations. Angrist has been an intellectual leader in IV, Rosenbaum only slightly less so. Sovey and Green introduce instrumental variables logic to political scientists. In the last reading, Miguel et al. use IV to study substantive questions of interest to political scientists. The study has strongly influenced the study of civil conflict. However, recent work (Schulz and Mankin) shows just how vulnerable the use of instrumental variables can be.

J. Angrist and J-S. Pischke, *Mastering 'Metrics,* ch. 3

J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 4

P. Rosenbaum, *Observation and Experiment*, ch. 13, pp. 258-278

A. Sovey and D. Green, "Instrumental Variables Estimation in Political Science: A Readers' Guide," *American Journal of Political Science,* 2009, pp. 188-200

E. Miguel, S. Satyanath, and E. Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, pp. 725-750

K. Schultz and J. Mankin, "Is Temperature Exogenous? The Impact of Civil Conflict on the Instrumental Climate Record in Sub-Saharan Africa," *American Journal of Political Science*, 2019, pp. 1-17

Recommended: Difference-in-differences and Instrumental Variables

S. Donald and K. Lang, "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics*, 2007, pp. 221-233

B. Lu and R. Rosenbaum, "Optimal Pair Matching with Two Control Groups," *Journal of Computational and Graphic Statistics*, 2004, pp. 422-434

J. Angrist, G. Imbens, and D. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, pp. 444-455

J. Angrist and A. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 2001, pp. 69-85

J. Angrist, "Instrumental Variables Methods in Experimental Criminology Research: What, Why, and How," *Journal of Experimental Criminology*, 2006, 23-44

L. Bartels, "Instrumental and 'Quasi-Instrumental' Variables," *American Journal of Political Science*, 1991, pp. 777-800

Recommended: Applied Examples of IV

J. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," *American Economic Review,* 1990, pp. 313-336

B. Savun and D. Tirone, "Foreign Aid, Democratization, and Civil Conflict: How Does Democracy Aid Affect Civil Conflict?" *American Journal of Political Science*, 2011, pp. 233-246

D. Acemoglu, S. Johnson, and J. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 2001, pp. 1369-1401

Recommended: Sensitivity Analysis

M. Blackwell, "A Selection Bias Approach to Sensitivity Analysis for Causal Effects," *Political Analysis*, 2013, pp. 1-14

Carrie A Hosman, Ben B. Hansen, and Paul W. Holland. "The sensitivity of linear regression coefficients's confidence limits to the omission of a confounder". In: *The Annals of Applied Statistics* 4.2 (2010), pp. 849–870

Carlos Cinelli, Jeremy Ferwerda, and Chad Hazlett. "Sensemakr: Sensitivity analysis tools for ols in r and stata". In: *Submitted to the Journal of Statistical Software* (2020)

AlexanderM Franks, Alexander D'Amour, and Avi Feller. "Flexible sensitivity analysis for observational studies without observable implications". In: *Journal of the American Statistical Association* (2019)

Recommended: Question Asked and Type of Experiment Chosen

As they advance in their research, students should never lose sight of the important connection between question asked and type of experiment chosen. Keep in mind the variety of experimental types and what sorts of question each type answers. The types are many: lab, field, survey, natural, and lab-in-the-field. All scholars will find the exchange between Winters and Weitz-Shapiro and Boas et al. both interesting and informative. Hainmueller and colleagues have been instrumental in bringing two old methods—vignettes and conjoint survey experiments –into survey experimental research, a topic on which Sniderman has written a valuable overview. Other articles compare the results of different experimental types. Gneezy and Imas discuss an increasingly important experimental type, lab-in-the-field. Rosenbaum sets forth the logic of so-called natural experiments. Other readings ask whether different types of experiment generate different conclusions. There is also the important distinction between economics- and psychology-based experiments. Dickson effectively sets forth the intriguing differences between them.

M. Winters and R. Weitz-Shapiro, "Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians?" *Comparative Politics*, 2013, pp. 418-436

R. Weitz-Shapiro and M. Winters, “Can Citizens Discern? Information, Credibility, Political Sophistication, and the Punishment of Corruption in Brazil,” *Journal of Politics*, 2017, pp. 60-74

T. Boas, F.D. Hidalgo, and M. A. Melo, “Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil,” *American Journal of Political Science*, 2018, pp. 385-400

J. Hainmueller, D. Hangartner, and T. Yamamoto, "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior," *Proceedings of the National Academy of Sciences*, 2015, pp. 2395-2400.

P. Sniderman, “Some Advances in the Design of Survey Experiments,” *Annual Review of Political Science*, 2018, pp.259-275

U. Gneezy and A. Imas, “Lab in the Field: Measuring Preferences in the Wild,” in A. Banerjee and E. Duflo, eds., *Handbook of Field Experiments*, Vol. 1, pp. 439-464

P. Rosenbaum, *Observation and Experiment*, ch. 6

J. Barabas and J. Jerit, “Are Survey Experiments Externally Valid,” *American Political Science Review*, 2010, pp. 226-242

A. Coppock and D. Green, “Assessing the Correspondence between Experimental Results Obtained in the Lab and Field: A Review of Recent Social Science Research,” *Political Science Research and Methods*, 2015, pp. 113–131

D. Broockman, J. Kalla, and J. Sekhon, “The Design of Field Experiments With Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs,” 2017, *Political Analysis*, pp. 435-464

E. Dickson, “Economics versus Psychology Experiments: Stylization, Incentives, and Deception,” in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*

Recommended: Selected Challenges to Experimental Implementation

Field experiments, especially, pose considerable challenges at the implementation and analysis stages. One challenge is that “noisy data” can prevent identifying causality. A common approach to this potential problem is to block prior to the experiment and then randomly assign within blocks. A second challenge is noncompliance; units assigned to treatment, for example, might not receive the treatment, without the researcher knowing. This has led experimenters to distinguish average treatment effects, intent-to-treat effects, and complier average effects. Be sure to understand the distinctions. A third challenge, especially in field experiments, is attrition. Green and Gerber effectively summarize the essences of all three challenges. Bowers addresses an as-yet unsettled matter regarding random assignment experiments: the use of covariates when analyzing the data. Finally, the exchange between Mutz and Pemantle, on the one hand, and Gerber et al., on the other, addresses some fundamental questions about what to report and what not to report. Pay particular attention to the exchange regarding the reporting of balance tables. Why would this issue come up at all? What is it about randomization that raises the matter of balanced covariates?

A. Gerber and D. Green, *Field Experiments: Design, Analysis, and Interpretation*, chs. 5 and 7, pp. 131-171, 211-252

J. Bowers, “Making Effects Manifest in Randomized Experiments,” in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*

D. Mutz and R. Pemantle, “Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods,” *Journal of Experimental Political Science,* 2015, pp. ???

A. Gerber et al., “Reporting Balance Tables, Response Rates, and Manipulation Checks in Experimental Research: A Reply from the Committee that Prepared the Reporting Guidelines,” *Journal of Experimental Political Science*, 2015, pp. 216-229

Recommended: The Complex Relationship between Experiment and Outside World

Scholars purportedly use experiments to understand the world. However, successfully completing this task raises some thorny issues, as the readings below address. In particular, the relationship between experiment and the outside world to which the experimenter intends to infer is complex. Diaz et al. summarize some of the relevant literature. Gaines and Kuklinski argue that using randomized experiments to infer to the external world will be problematic when self-selection occurs in that world. Dafoe, Zhang, and Caughey identify some especially important and not-easily-rectified challenges to drawing inferences from survey experiments. Broockman, Kalla, and Sehhon discuss the challenges of field experiments in which the treatment effect is measured in surveys and the treatment itself is delivered via a real-world mechanism, such as phone calls and door- to-door canvassing. In an old and often-cited article. Mook proposes that laboratory experiments, at least, frequently should not be used to infer to a world outside of the experiment.

G. Diaz, C. Grady, and J. Kuklinski, "Survey Experiments and the Quest for Valid Interpretation," in xxxxxxxxxxxxxxxxxx, forthcoming

B. Gaines and J. Kuklinski, "Experimental Estimation of Heterogeneous Treatment Effects Related to Self-Selection," *American Journal of Political Science*, 2011, pp. 724-736

A. Dafoe, B. Zhang, and D. Caughey, "Information Equivalence in Survey Experiments," *Political Analysis*, 2018, pp. 1-18

D. Broockman, J. Kalla, and J. Sekhorn, "The Design of Field Experiments with Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs, *Political Analysis*, 2017, pp. 435-464

D. Mook, "In Defense of External Validity," *American Psychologist*, 1983, pp. 379-387

Recommended: Other Readings on Experiments

B. Gaines and J. Kuklinski, "Treatment Effects," in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*

B. Gaines, J. Kuklinski, and P. Quirk, "The Logic of the Survey Experiment Reexamined," *Political Analysis*, 2007, pp. 1-20

J. Druckman and T. Leeper, "Learning More from Political Communication Experiments: Pretreatment and Its Effects," *American Journal of Political Science*, 2012, pp. 875-896nh-

J. Orbell and R. Dawes, "A 'Cognitive Mister' Theory of Cooperators' Advantage," *American Political Science Review,* 1991*,* pp. 515-528

J. Druckman, J. Fein, and T. Leeper, "A Source of Bias in Public Opinion Stability," *American Political Science Review*, 2012, pp. 430-454

Recommended: Differences-in-Differences

Differences-in-differences estimates depend on two assumptions: first, much can be learned about cause and effect by examining how a phenomenon of interest changes, if at all, after a real-world intervention; and, second, it is crucial to include, for purposes of comparison, units in a control condition that are as similar as possible to those in the treated condition prior to the intervention. Thus comes the term "differences-in-differences." Note that differences-in-differences analysis requires observations before and after the intervention, the more observations the better. Although differences-in-differences estimation presumably resembles experimental research, randomization is absent in the former (as it is, by definition in all quasi-experimental designs).

The first four readings focus generally on differences-in-differences logic. The remaining readings report substantive studies that use differences-in-differences designs. Card and Krueger's imaginative study is widely cited; unfortunately, the authors' conclusions are wrong (there is a lesson to be learned here). Lyall applies both matching and differences-in-differences to answer the question in the title.

J. Angrist and J-S. Pischke, *Mastering 'Metrics*, ch. 5, pp. 178-208

J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 5, pp. 221-248

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch.6

M. Bertrand, E. Duflo, and S. Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 2004, pp. 249-275

D. Card and A. Krueger, "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 1994, pp. 772-784 (and browse some of the reactions included at the end of the article)

J. Lyall, "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya," *Journal of Conflict Resolution*, 2009, pp. 331-362

Recommended: Applied Examples of Synthetic Control and DiD

D. Abadie, A. Diamond, and J. Hainmueller, "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 2014, pp. 495-510

D. Card, "The Impact of the Mariel Boatlife on the Miami Labor Market," *Industrial and Labor Relations Review*, 1990, pp. 245-257

S. Anzia and C. Berry, "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science*, 2011, pp. 478-493

W. Bullock and J. Clinton, "More a Molehill than a Mountain: The Effects of the Blanket Primary on Electoral Officials' Behavior from California," *Journal of Politics*, 2011, pp. 915-930

Recommended Reading on Qualitative and Quantitative Methods

H. Brady and D. Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, all

D. Freedman, "On Types of Scientific Enquiry: The Role of Qualitative Reasoning," in J. Box- Steffensmeier, H. Brady, and D. Collier, eds., *Oxford Handbook of Political Methodology*

R. Pahre, "Formal Theory and Case Study Methods in EU Studies," *European Union Politics*, 2004, pp. 113-146

J. Gerring, "Mere Description," *British Journal of Political Science*, 2012, pp. 721-746

L. Wedeen, "Reflections on Ethnographic Work in Political Science," 2010, pp. 255-272

Recommended: An Integration of Quantitative and Qualitative Research?

While the preceding work has emphasized the need for an integration of quantitative and qualitative research, recent efforts have offered specific ways in which this integration might occur. In the four readings, the authors propose different paths to integration.

Elizabeth Levy Paluck. "The Promising Integration of Qualitative Methods and Field Experiments". In: *The Annals of the American Academy of Political and Social Science* (2010), pp. 59–71. ISSN: 0002-7162

J. Fearon and D. Laitin, "Integrating Qualitative and Quantitative Methods," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *Oxford Handbook of Political Methodology*

E. Lieberman, "Nested Analysis as a Mixed-Method Strategy for Comparative Research," *American Political Science Review*, 2005, pp. 435-452

M. Humphreys and A. Jacobs, "Mixing Methods: A Bayesian Approach," *American Political Science Review*, 2015, pp. 653-673

J. Seawright, *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*, all

References

[1] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method". In: *American Journal of Political Science* (2014).

[2] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American Statistical Association* 105.490 (2010).

[3] Robert Adcock and David Collier. "Measurement Validity: A shared Standard for Qualitative and Measurement Validity: A shared Standard for Qualitative and Quantitative Research." In: *American Political Science Review* 95.3 (2001), pp. 529–546.

[4] Joshua D Angrist and Jörn-Steffen Pischke. *Mastering'metrics: The path from cause to effect.* Princeton University Press, 2014.

[5] Jason Barabas and Jennifer Jerit. "Are survey experiments externally valid?" In: *American Political Science Review* (2010), pp. 226–242.

[6] Howard S. Becker. *Writing for Social Scientists : How to Start and Finish Your Thesis, Book, or Article (Chicago Guides to Writing, Editin.* University Of Chicago Press, 1986. ISBN: 0226041085.

[7] Howard S. Becker. *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article.* University of Chicago Press., 1986.

[8] Taylor C Boas, F Daniel Hidalgo, and Marcus André Melo. "Norms versus action: Why voters fail to sanction malfeasance in Brazil". In: *American Journal of Political Science* 63.2 (2019), pp. 385–400.

[9] James Bogen. "Theory and Observation in Science". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2020. Metaphysics Research Lab, Stanford University, 2020.

[10] Henry E Brady. "Causation and explanation in social science". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science).* 2008.

[11] L. Breiman. "Statistical modeling: the two cultures (with comments and a rejoinder by the author)". In: *Statistical Science* 16.3 (2001), pp. 199–231.

[12] David E Broockman, Joshua L Kalla, and Jasjeet S Sekhon. "The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs". In: *Political Analysis* 25.4 (2017), pp. 435–464.

[13] Donald T Campbell. "Reforms as experiments." In: *American psychologist* 24.4 (1969), p. 409.

[14] Stephen Chaudoin, Jude Hays, and Raymond Hicks. "Do We Really Know the WTO Cures Cancer?" In: *British Journal of Political Science* 48.4 (2018), pp. 903–928.

[15] Carlos Cinelli, Jeremy Ferwerda, and Chad Hazlett. "Sensemakr: Sensitivity analysis tools for ols in r and stata". In: *Submitted to the Journal of Statistical Software* (2020).

[16] Carlos Cinelli and Chad Hazlett. "Making sense of sensitivity: Extending omitted variable bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1 (2020), pp. 39–67.

[17] J. Cornfield et al. "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions". In: *Journal of the National Cancer Institute* (1959).

[18] Vincenzo Crupi. "Confirmation". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Spring 2020. Metaphysics Research Lab, Stanford University, 2020.

[19] Allan Dafoe, Baobao Zhang, and Devin Caughey. "Information equivalence in survey experiments". In: *Political Analysis* 26.4 (2018), pp. 399–416.

[20] Gustavo Diaz, Christopher Grady, and James H Kuklinski. "Survey Experiments and the Quest for Valid Interpretation". In: *The SAGE Handbook of Research Methods in Political Science and International Relations.* SAGE Publications, 2020.

[21] Thad Dunning. *Natural Experiments in the Social Sciences: A Design-Based Approach.* First. Cambridge University Press, 2012.

[22] Zachary Elkins. "Gradations of democracy? Empirical tests of alternative conceptualizations". In: *American Journal of Political Science* (2000), pp. 293–300.

[23] P.K. Feyerabend. "How to be a good empiricist: A plea for tolerance in matters epistemological". In: *Philosophy of Science: The Delaware Seminar* 2 (1963), pp. 3–39.

[24] Paul Feyerabend. "Consolations for the Specialist". In: *Criticism and the Growth of Knowledge* 4 (1970), pp. 197–230.

[25] R.A. Fisher. *The design of experiments. 1935.* Edinburgh: Oliver and Boyd, 1935.

[26] AlexanderM Franks, Alexander D'Amour, and Avi Feller. "Flexible sensitivity analysis for observational studies without observable implications". In: *Journal of the American Statistical Association* (2019).

[27] Brian J Gaines, James H Kuklinski, and Paul J Quirk. "The logic of the survey experiment reexamined". In: *Political Analysis* 15.1 (2007), pp. 1–20.

[28] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.

[29] Alan S Gerber, Donald P Green, and Edward H Kaplan. "The illusion of learning from observational research". In: *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences* (2014). Ed. by Dawn Langan Teele, pp. 9–32.

[30] John Gerring. "Causal mechanisms: Yes, but. . . " In: *Comparative political studies* 43.11 (2010), pp. 1499–1526.

[31] Mita Giacomini. "Theory-based medicine and the role of evidence: why the emperor needs new clothes, again". In: *Perspectives in Biology and Medicine* 52.2 (2009), pp. 234–251.

[32] Gary Goertz. "Concepts, Theories and Numbers: A Checklist for Constructing, Evaluating and Using Concepts or Quantitative Measures". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. 2008.

[33] George D Gopen and Judith A Swan. "The science of scientific writing". In: *American Scientist* 78.6 (1990), pp. 550–558.

[34] Christopher Grady. *10 Things to Know About Survey Experiments*. URL: https://egap.org/resource/10-things-to-know-about-survey-experiments/ (visited on 2019).

[35] P Richard Hahn, Jared S Murray, Carlos M Carvalho, et al. "Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion)". In: *Bayesian Analysis* 15.3 (2020), pp. 965–1056.

[36] P. W. Holland. "Statistics and Causal Inference (with discussion)". In: *Journal of the American Statistical Association* 81 (1986), pp. 945–970.

[37] Carrie A Hosman, Ben B. Hansen, and Paul W. Holland. "The sensitivity of linear regression coefficients's confidence limits to the omission of a confounder". In: *The Annals of Applied Statistics* 4.2 (2010), pp. 849–870.

[38] Trevor Incerti. "Corruption information and vote share: A meta-analysis and lessons for experimental design". In: *American Political Science Review* 114.3 (2020), pp. 761–774.

[39] Simon Jackman. "Measurement". In: *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. 2008.

[40] Thomas S Kuhn. "Logic of Discovery or Psychology". In: *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Vol. 4. Cambridge University Press. 1970, p. 1.

[41] Elizabeth Levy Paluck. "The Promising Integration of Qualitative Methods and Field Experiments". In: *The Annals of the American Academy of Political and Social Science* (2010), pp. 59–71. ISSN: 0002-7162.

[42] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge U.P., 2000.

[43] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[44] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018.

[45] P.R. Rosenbaum. "Choice as an Alternative to Control in Observational Studies (with discussion)". In: *Statistical Science* 14.3 (1999), pp. 259–304.

[46] Paul R Rosenbaum. "How to see more in observational studies: Some new quasi-experimental devices". In: *Annual Review of Statistics and Its Application* 2 (Apr. 2015), pp. 21–48.

[47] Paul R Rosenbaum. "Modern algorithms for matching in observational studies". In: *Annual Review of Statistics and Its Application* 7 (2020), pp. 143–176.

[48] Paul R. Rosenbaum. *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, 2017, p. 374. ISBN: 9780674975576.

[49] Jasjeet S Sekhon and Rocio Titiunik. "When natural experiments are neither natural nor experiments". In: *American Political Science Review* 106 (2012), pp. 35–57.

[50] Tara Slough. *10 Things to Know About Measurement in Experiments*. 2020. URL: https://egap.org/resource/10-things-to-know-about-measurement-in-experiments/.

[51] Stephen Thornton. "Karl Popper". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University, 2019.

[52] J.W. Tukey. "Tightening the clinical trial". In: *Controlled Clinical Trials* 14.4 (1993), pp. 266–285. ISSN: 0197-2456.

[53] Duncan J Watts. "Should social science be more solution-oriented?" In: *Nature Human Behaviour* 1.1 (2017), p. 0015.

[54] Matthew S Winters and Rebecca Weitz-Shapiro. "Lacking information or condoning corruption: When do voters support corrupt politicians?" In: *Comparative Politics* 45.4 (2013), pp. 418–436.

[55] James Woodward. "Scientific Explanation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University, 2019.

[56] Julie Zahle. "Methodological Holism in the Social Sciences". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2016. Metaphysics Research Lab, Stanford University, 2016.

[57] José R Zubizarreta, Dylan S Small, and Paul R Rosenbaum. "Isolation in the construction of natural experiments". In: *The Annals of Applied Statistics* (2014), pp. 2096–2121.