

Syllabus: Research Design & Causality in Social Science

Fall 2017

Last updated: October 29, 2017

Course information

- Course title: Research Design & Causality in Social Science
- Course number: POLCS-UH 1210
- Credits: 4
- Term: Fall 2017 (14 weeks)
- Lecture location: C2-E051
- Lecture hours: TuTh 2:40PM - 3:55PM
- Office hours: By appointment
- This course is a methods elective for SRPP and Political Science.
- Course prerequisites: Statistics (SOCSC-UH 1010Q Statistics for the Social and Behavioral Sciences). Motivated students can take it simultaneously with this course.

Instructor information

- Instructor: Dr. Peter van der Windt
- Email: petervanderwindt@nyu.edu
- Website: www.petervanderwindt.com
- Office: Building A5, Office 147

Course description & learning objectives

This course will provide students with the “foundations” to undertake research in social science. You will learn how to identify an interesting research question. You will be introduced to different approaches that social scientists take to answer these questions. And because many of the questions we are interested in are causal (What leads to Y? What is the impact of X?), you will learn about different strategies to get at causality. We will also discuss other key issues related to good research like transparency and ethics. This class is highly recommended to students who plan to write a capstone or a research paper.

Teaching methodology

This class is hands on. I will give presentations but you will do so as well. We will make use of many examples to critically evaluate existing research. Furthermore, during the course you will create your own research design.

Relation to other courses

The course Survey Research (SRPP-AD 120) and Ethnographic Field Research (SRPP-AD 125) are about data collection. Data analysis (POLSC-AD 209) and Econometrics (ECON-AD 210) are about what to do once you have the data collected. This course is about your initial steps when

it comes to research, those things that you have to know even before you collect your data. It is recommended to take this class after Statistics, but before Data analysis or Econometrics.

Course material

We make use of two books. Additional articles will also be used, which will be shared via New-classes.

- “Social Science Methodology: A Unified Framework”. J. Gerring. 2012. NYC: Cambridge University Press.
- “Mastering Metrics: The Path from Cause to Effect”. J. Angrist and J. Pischke. 2015. Princeton: Princeton University Press.

Course grading and assignments

Your final grade consist out of five parts:

- Class participation (10%): The grade for this component is based on your preparation and participation in the classes. If you are always present in class you receive a ‘B+’. If you also participate in class this grade will go up. If you cannot be present during a class, send me an email *before* the class and include the reason. You are allowed to miss one class (or more if you have valid reasons and send me an email beforehand). After that it will lower your grade (from A to A-, from A- to B+, from B+ to B, etc.) for every missed class.
- Research presentation (10%): Depending on class size, each student will do at least one research presentation. This presentation should be around 15 minutes long and discuss the research approach of a pre-selected article. This article will then be discussed in detail together in class. These discussion classes take places after I have introduced the research approach more generally the class before.
- Mid-term exam (20%): In the middle of the semester you will make a short midterm exam in class. The exam will measure your knowledge of some key concepts that we learned in class the weeks before.
- Final exam (30%): There will be an exam at the end of the course.
- Pre-analysis plan (30%): During the course you will develop your own pre-analysis plan (PAP). The PAP will be on a topic of your own interest. The pre-analysis plan forces you to put what you learn (research question, hypotheses, data collection, etc.) into practice.

Course schedule

The next part provides the course schedule by week. The course is structured into four parts: “Introduction to research”, “Measurement and approaches to research”, “Causality”, and “Other components of good research”. Readings indicated with an “(R)” are required, those indicated with an “(O)” are optional. Readings and topics can change as we go along (among others based on your input).

Introduction to research

The first three weeks are an introduction to research. You will learn why we do research, and how not to do research. We will discuss the importance of theory, and how to move from theory to concept to measurement. We will also introduce the research design.

Week 1: Introduction to research, and research question

- What is research and why do we do it?
- Examples of bad research.
- How do I pick a research question? Should I look carefully for a “gap” in existing knowledge, or can I allow myself to be motivated by my fascination about a particular topic?
- What is a good hypothesis?
- (R) Gerring (2012), chapters 1-4
- (O) Geddes (2003), chapters 1 & 2
- (O) King et al. (1994), chapter 1

Week 2: Concepts and measurement

- What is a “theory”, and why is it necessary? What does it do for us?
- From Characteristics of theories: parsimony, falsifiability, testability.
- Testing our theories: from theory to concept to measure
- (R) Gerring (2012), chapters 5-7
- (O) Morton (1999), chapters 1, 2 & 9
- (O) Fearon (1991)

Week 3: Introduction to your research design

- The components of a good research design (research question, hypotheses, etc.)
- Publication bias and pre-analysis plans.
- (R) Gerber and Malhotra (2008)

Approaches to data collection

To test our theories empirically you need data. The next four weeks are about the different approaches social scientists take in collecting data: case studies, surveys, archival research, laboratory experiments, elite interviewing, focus groups, participant observation, etc. What are the differences between small-N and large-N research? How are quantitative and qualitative research different?

Week 4: Small-N

- What is small-N research? What are the benefits and drawbacks?
- What is a case study? How to select your case studies?
- Types: archival research, elite interviewing, etc.
- (R) Rose (1991)
- (R) Gerring (2004)
- (R) Collier and Mahoney (1996)
- (O) Gerring (2007)

- (O) George and Bennett (2004)
- (O) Geddes (1990)
- (O) Lustick (1996)
- (O) Frisch and Harris (2012)
- (O) Tansey (2007)
- (O) Leech (2002)
- (O) Bratton and Liatto-Katundu (1994)
- (O) Glaser (1996)

Week 5: Large-N

- What are the pros and cons of large-N statistical research?
- Types: surveys, experiments, etc.
- How not to ask questions
- (R) www.arabbarometer.org
- (R) Jackman (1985)
- (R) Payne (1950)
- (O) Pearce (2002)
- (O) Cohen and Arieli (2011)
- (O) Hsueh and Jensenius (2014)

Week 5: Large-N

- The difficulties of doing large-N research
- Non-coverage bias, sampling bias, non-response bias, measurement bias
- Observational studies does not mean causality
- (R) Discussion paper: Collier and Hoeffler (1998)
- (O) Windt and Humphreys (2016)
- (O) King (1986)

Week 7: Mixing methods

- The benefits of combining small-N and large-N approaches?
- And how should we do this?
- (R) Fearon and Laitin (2013)
- (R) Lieberman (2005)
- (O) Humphreys and Jacobs (2015)

Causality

It is easy to find variables that are related empirically: shoe size and height, ice cream sales and drowning, rain and the number of umbrellas on the street (have a look here for even better examples: <http://www.tylervigen.com/spurious-correlations>). It is quite another matter to assert causation. In the following five weeks we discuss approaches that allow social scientists to make causal claims.

Week 8: Causality and experiments

- What is causality?
- The fundamental problem of causal inference
- Randomizing to get at causality
- What is the difference between ‘field experiments’; ‘lab experiment’; ‘quasi-experiments’; etc.?
- (R) Angrist and Pischke (2015), chapter 1
- (R) <http://egap.org/methods-guides/10-things-you-need-know-about-causal-inference>
- (R) <http://egap.org/methods-guides/10-strategies-figuring-out-if-x-caused-y>
- (O) Gerring (2012), chapters 8-12
- (O) Mcdermott (2013)
- (O) Humphreys and Weinstein (2009)
- (O) Gerring (2005)
- (O) Gerber and Green (2012)
- (O) Mcdermott (2002)
- (O) Morton and Williams (2010)

Week 9: Threats to causal inference

- Spillovers
- Compliance problems
- Hawthorne effects
- (R) <http://egap.org/methods-guides/10-things-you-need-know-about-spillovers>
- (R) <http://egap.org/methods-guides/10-things-you-need-know-about-local-average-treatment-effect>

Week 10: Regression and matching

- (R) Angrist and Pischke (2015), chapter 2
- (R) Discussion paper: Mahal et al. (2008)
- (O) Behrman and Rosenzweig (2002)

Week 11: Instrumental variable

- (R) Angrist and Pischke (2015), chapter 3

- (R) Discussion paper: Acemoglu et al. (2001)

Week 12: Regression discontinuity

- (R) Angrist and Pischke (2015), chapter 4
- (R) Discussion paper: Butler (2009)

Week 13: Difference-in-difference

- (R) Angrist and Pischke (2015), chapter 5
- (R) Discussion paper: Menon and Van Der Meulen (2015)

Conclusion

The last week of class is reserved for course review. And students present their research designs.

Week 14:

- Design presentations: students present their research designs.
- Review and discussion.

Additional components of good research

If time permits, we can discuss additional topics related to doing good research. For example, how to practically do good research? And what are the ethical implications of doing (field) research?

Doing good research:

- Workflow: how to make sure that your work is well-organized, transparent, and reproducible.
- Replication and publication bias.
- (R) Bowers (2008)
- (R) Ioannidis (2005)
- (R) King (1995)
- (R) <http://chrisblattman.com/2015/07/23/dear-journalists-and-policymakers-what-you-need-to-know>

Ethical implications of doing research:

- Belmont report
- IRB
- (R) Humphreys (2015)
- (R) Wood (2006)
- (O) Carapico (2006)
- (O) Belmont Report, available at: <http://ohsr.od.nih.gov/guidelines/belmont.html>

Appendix: Your pre-analysis plan

During the course you will develop your own pre-analysis plan (PAP). The PAP will be on a topic of your own interest (yes, you can use a topic that you are thinking of for your capstone). The PAP addresses most of the major topics that we cover in this course. I uploaded a template of the research design to Newclasses. Please just fill out the template; do not start a new document.

Writing a PAP is not easy, but an important skill to have. It follows the same rules that you would use for writing an academic paper. Below I list several of them:

The Do's:

- Follow the guidelines indicated in the template.
- Follow what I say in class: e.g. if I ask for you to hand in “Section 1: Introduction” and points 12-14, please hand in that and only that. Nothing more. Nothing less.
- Incorporate the comments I give you.
- Use writing that people understand. No need to write more complicated than necessary. “I have set out to argue” can also be written as “I will argue”.
- Be academic in your writing. Do not use words like “fascinating”, “amazing”, etc. That’s your opinion. Readers don’t care about your opinion. The data will provide evidence.
- It also means that you have to be precise. Avoid words like “a lot” or “in the beginning” if you can be more precise.
- Use references! If you use a number somewhere (e.g. in Congo unemployment is 95%), you need to add a source. How to add references? Do it the same way as I do in this syllabus. In the main text it will be: “Edwards (2015) writes that...” and then have a section at the end with references.
- Use the same tense throughout your document. If you start writing in the present tense, do that throughout your document.
- Hypotheses should follow from the literature/motivation section. Your hypotheses should not come out of the blue. If your hypothesis is “Women in the Congo work more than men in the Congo”, your literature/motivation section should not only contain information about poverty in the Congo, but also about differences between men and women in the Congo.
- If you have many hypotheses, you will likely have many independent and dependent variables. Either write things out clearly, like “For hypothesis 1 my independent variable is X1 and my dependent variable is Y1. I’ll measure them as follows... For hypothesis 2 my independent variable is X2 and my dependent variable is Y2. I’ll measure them as follows...” If you have many hypotheses, it is probably cleanest to provide a table with the following columns: hypothesis, independent variable, dependent variable and question in the survey.
- All pages should have a page number.

The Dont's:

- Do not have hypotheses that are a combination of multiple hypotheses. E.g. “X leads to increases in Y and does not decrease Y.” These are two hypotheses. Write them out as two sentences.
- Do not make value judgements. We are academics.
- Avoid contractions: write “they’ve” as “they have”, “can’t” “as cannot”, etc.
- Avoid writing the full name of the authors, the title of their book or article, etc. Use referencing similar as how it is done in this syllabus.

- No need for a coverpage and pictures. That's for secondary school.

Examples of real pre-analysis plans can be found here (please have a look around): <http://egap.org/content/registration>. As you can see, they come in many different flavors but most of them share the main ingredients discussed below.

References

- Acemoglu, D., Johnson, S., and Robinson, J. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5):1369–1401.
- Angrist, J. D. and Pischke, J.-S. (2015). *Mastering Metrics: The Path from Cause to Effect*. Princeton University Press, Princeton.
- Behrman, J. R. and Rosenzweig, M. R. (2002). Does Increasing Women’s Schooling Raise the Schooling of the Next Generation? *American Economic Review*, (1):323–334.
- Bowers, J. (2008). Six Steps to a Better Relationship with Your Future Self. *The Political Methodologist*, 18(2):1–18.
- Bratton, M. and Liatto-Katundu, B. (1994). A Focus Group Assessment of Political Attitudes in Zambia. *African Affairs*, 93(373):535–563.
- Butler, D. M. (2009). A Regression Discontinuity Design Analysis of the Incumbency Advantage and Tenure in the U.S. House. *Electoral Studies*, 28(1):123–128.
- Carapico, S. (2006). No Easy Answers: The Ethics of Field Research in the Arab World. *PS: Political Science and Politics*, 39(3):429–431.
- Cohen, N. and Arieli, T. (2011). Field Research in Conflict Environments: Methodological Challenges and Snowball Sampling. *Journal of Peace Research*, 48(4):423–435.
- Collier, D. and Mahoney, J. (1996). Selection Bias in Qualitative Research. *World Politics*, 49(1):56–91.
- Collier, P. and Hoeffler, A. (1998). On Economic Causes of Civil War. *Oxford Economic Papers*, 50:563–573.
- Fearon, J. D. (1991). Counterfactuals and Hypothesis Testing in Political Science. *World Politics*, 43(2):169–195.
- Fearon, J. D. and Laitin, D. D. (2013). Integrating Qualitative and Quantitative Methods: Putting It Together Again. In *Oxford Handbook*, number 2011, pages 1–25.
- Frisch, S. A. and Harris, D. B. (2012). *Doing Archival Research in Political Science*. Cambria Press.
- Geddes, B. (1990). How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis*, 2:131–150.
- Geddes, B. (2003). *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. University of Michigan Press, Michigan.
- George, A. and Bennett, A. (2004). *Case Studies and Theory Development in the Social Sciences*. MIT Press, Cambridge.
- Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York City.
- Gerber, A. S. and Malhotra, N. (2008). Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, 3:313–326.
- Gerring, J. (2004). What Is a Case Study and What Is It Good for? *American Political Science Review*, 98(2):341–354.
- Gerring, J. (2005). Causation: A Unified Framework for the Social Sciences. 17(2):163–198.
- Gerring, J. (2007). *Case Study Rresearch: Principles and Practice*. Cambridge University Press, New York City.

- Gerring, J. (2012). *Social Science Methodology: A Unified Framework*. Cambridge University Press, New York City.
- Glaser, J. M. (1996). The Challenge of Campaign Watching: Seven Lessons of Participant-Observation Research. *PS: Political Science and Politics*, 29(3):533–537.
- Hsueh, R. and Jensenius, F. R. (2014). Fieldwork in Political Science: Encountering Challenges and Crafting Solutions. *PS: Political Science and Politics*, (April):391–393.
- Humphreys, M. (2015). Reflections on the Ethics of Social Experimentation. *Journal of Globalization and Development*, 6(1):87–112.
- Humphreys, M. and Jacobs, A. M. (2015). Mixing Methods: A Bayesian Approach. *American Political Science Review*, 109(4):653–673.
- Humphreys, M. and Weinstein, J. M. (2009). Field Experiments and the Political Economy of Development. *Annual Review of Political Science*, 12(1):367–378.
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings are False. *PLoS medicine*, 2(8):696–701.
- Jackman, R. W. (1985). Cross-National Statistical Research and the Study of Comparative Politics. *American Journal of Political Science*, 29(1):161–182.
- King, G. (1986). How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science*, 30(3):666–687.
- King, G. (1995). Replication, Replication. *PS: Political Science and Politics*, 28(3):444–452.
- King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. NJ: Princeton University Press.
- Leech, B. L. (2002). Interview Methods in Political Science. *PS: Political Science and Politics*, 35(4):663–664.
- Lieberman, E. S. (2005). Nested Analysis as a Mixed-Method Strategy for Comparative Research. *American Political Science Review*, 99(3):435–452.
- Lustick, I. S. (1996). History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias. *American Political Science Review*, 90(3):605–618.
- Mahal, A., Canning, D., Odumosu, K., and Okonkwo, P. (2008). Assessing the Economic Impact of HIV/AIDS on Nigerian Households: A Propensity Score Matching Approach. *AIDS*, pages S95–101.
- Mcdermott, R. (2002). Experimental Methods in Political Science. *Annual Review of Political Science*, 5:31–61.
- Mcdermott, R. (2013). The Ten Commandments of Experiments. *PS: Political Science and Politics*, pages 605–610.
- Menon, N. and Van Der Meulen, Y. (2015). War and Women’s Work: Evidence from the Conflict in Nepal. *Journal of Conflict Resolution*, 59(1):51–73.
- Morton, R. B. (1999). *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*. Cambridge University Press, New York City.
- Morton, R. B. and Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge University Press, New York City.
- Payne, S. L. (1950). Thoughts About Meaningless Questions. *Public Opinion Quarterly*, 14(4):687–696.

- Pearce, L. D. (2002). Integrating Survey and Ethnographic Methods for Systematic Anomalous Case Analysis. *Sociological Methodology*, 32:103–132.
- Rose, R. (1991). Comparing Forms of Comparative Analysis. *Political Studies*, 34:446–462.
- Tansey, O. (2007). Process Tracing and Elite Interviewing: A Case for Non-probability Sampling. *PS: Political Science and Politics*, 40(4):765–772.
- Windt, P. V. D. and Humphreys, M. (2016). Crowdsourcing in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time. *Journal of Conflict Resolution*, 60(4):748–781.
- Wood, E. J. (2006). The Ethical Challenges of Field Research in Conflict. *Qualitative Sociology*, 29:373–386.