Design & Analysis of Quasi-Experiments for Causal Inference
Educational Psychology 711-005 - Spring 2014
University of Wisconsin – Madison / Department of Educational Psychology

Lecture	TueThu 2:30-3:45 PM
Room	218 Educational Sciences
Instructor	Peter M. Steiner
Office	1061 Educational Sciences (608 262 0842)
e-mail	psteiner@wisc.edu
Office Hours	TueThu 4:00-5:00 PM, and by appointment

Course Description

Randomized experiments are frequently considered as the gold standard for inferring causal effects of a treatment, program or policy. However, randomizing subjects into a treatment and control condition is not always possible (e.g., due to ethical concerns). If a randomized experiment is not feasible we frequently rely on quasi-experimental designs for assessing causal effects of a treatment. Causal research questions like "Did NCLB increase students' achievement in reading and math?" or "Does retaining kindergartners for one year (instead of promoting them) result in negative effects on their future achievements" are typically investigated using quasi-experimental approaches. This course focuses on the design and analysis of the strongest quasi-experimental methods: regression discontinuity designs, interrupted time series designs, non-equivalent control group designs, and instrumental variable approaches. As with randomized experiments, the goal of these quasi-experimental designs is to estimate the impact of a treatment or policy on quantitative outcomes of interest. Though the focus is on causal description ("is treatment effective or not?") rather than causal explanation ("why is the treatment effective") analytic techniques for causal explanation, like structural equation modeling, will be briefly discussed.

The course starts with an introduction to the philosophy of causation and then outlines the Rubin Causal Model (RCM) in the context of randomized experiments. RCM is the currently predominant quantitative causal model in social sciences. Then we focus on four of the strongest quasi-experimental designs: regression discontinuity designs, interrupted time series designs, non-equivalent control group designs (with an emphasis on propensity score methods), and instrumental variable approaches. For each design, we discuss (i) the basic design idea for identifying the treatment effect, (ii) strategies and design elements for improving the basic design, and (iii) statistical approaches for estimating the effect. The course will close with a discussion of the empirical research on whether and under which conditions quasi-experimental methods actually work in practice.

Besides the theoretical aspects of quasi-experimental designs the course also emphasizes practical issues in implementing and analyzing quasi-experiments. For each design we will (i) analyze real data and explore and discuss different analytic strategies, and (ii) read applied studies which we critically evaluate with regard to their design and analytic approaches.

All analyses of real data will be done in R, a free language and environment for statistical computing and graphics (http://www.r-project.org/). The reason for using R instead of a different software package (e.g., STATA, SAS, or SPSS) is twofold. First, the transparency and flexibility of R, including its powerful graphics, enables us to better understand and individually modify analytic procedures (that would not be possible with "canned" software packages or procedures). Second, not all techniques required for the course are available in other software packages.

Texts & other materials

Required textbook: Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company. This book covers the design aspect of quasi-experiments and discuses a lot of design elements that are very useful for practical research. However, since it does not cover the analysis of quasi-experiments we will rely on papers and chapters of other books for the analysis part of the course (see the weekly schedule for details). You are expected to read the required texts *before* each class.

Extensive lecture notes and data sets will be available online. The statistical software package R can be downloaded from http://www.r-project.org/. You need to go to a CRAN server—click on the CRAN link on the left menu bar.

Prerequisites

Since the course focuses on the *design* & *analysis* of quasi-experiments, you should be familiar with (i) basic probability theory, (ii) the basic ideas of randomized experiments, (iii) regression techniques (multiple regression/ANCOVA and logistic regression), and (iv) the statistical computing package R (we are doing all analyses in R).

Knowledge of the basic multiple regression techniques is quintessential since the analysis of all quasi-experimental designs involves standard and more advanced regression techniques (e.g., non-parametric regression or two-stage least square regression). A full comprehension of quasi-experimental techniques is impossible if you do not have a basic understanding of regression analysis (including dummy coding, interaction effects, model selection criteria like AIC, logistic regression).

I use R in class for illustrating analytic techniques. However, you are not required to use R for the assignments. Nonetheless, I strongly encourage you to use R for this course because some procedures we are going to use are not (yet) available in other statistical software packages. Even if you are not familiar with R, doing the assignments in R is not too difficult since I provide corresponding R code in class. Moreover, only half of the assignments will require the analysis of real data. The other half of assignments is more design-oriented (and does not involve statistical analyses).

If you are not familiar with R you can work through an introductory textbook on R (e.g., Peter Daalgard (2004): Introductory Statistics with R. Springer). There is also lots of introductory material on the Comprehensive R Archive Network (CRAN) homepages (e.g., http://cran.cnr.berkeley.edu/other-docs.html).

Requirements & Grading

(1) You are required to attend classes and actively participate in discussions.
(2) There will be 8 assignments. Half of the assignments require you to analyse real datasets (datasets will be provided online), the other half consists of reviewing published papers with regard to their design and analytic strategies used. For both types of assignment, analyses and reviews, you need to write short reports (maximum of three pages, excluding appendices). Assignments need to be handed in as hard copies at the due date (electronic submissions are only possible in exceptional cases).

Since each assignment is worth 5 points the maximum number of points attainable is 40. Grading is as follows (assuming regular class attendance):

A	35-40pts
AB	30-34pts
B	25-29pts
BC	20-24pts
C	15-19pts
D	10-14pts
F	0-9pts

Analyses Using real data sets (which I will provide), you are required to run similar analyses as demonstrated in class. This typically requires only slight changes of the code I provide in class. The emphasis is on the appropriate analysis of quasi-experimental data, in particular, the estimation of treatment effects and testing the model specifications. In writing up your results you should briefly address the following points:

(a) research question & quasi-experimental design used
(b) description of analytic method and model(s) estimated (= model equations)
(c) presentation of results
(d) limitations (statistical conclusion validity / assumptions)
(e) conclusion

Maximum of three pages (excluding Appendices of plots, tables, code of statistical analysis, etc).

Reviews Reviews of papers should address both the design and analysis of the corresponding quasi-experiment. You should discuss strengths and weaknesses of the implemented approach and make suggestions for improvements, i.e., which design elements could have been used to strengthen the quasi-experimental analysis (you may make suggestions even if you suspect the required data are not available). In reviewing the papers you should carefully think about whether the assumptions required by the quasi-experimental design as well as the statistical approach to estimate the treatment effect are (likely to be) met. In particular, address potential threats to validity and how the authors could have ruled them out by using additional design elements. For this purpose you might find SCC's list of potential threats to internal and statistical conclusion validity (p45 & p55) as well as the list of design elements helpful (p157). Finally, assess whether the authors' causal claims and conclusions drawn are warranted. Each review is limited to a maximum of three pages.

Grading 5 = Excellent understanding of the assignment. All of the required elements are present and correctly interpreted. The write-up/review demonstrates excellent depth of understanding with respect to the link between theory (design & analysis), results, and interpretation. Write-up is in an appropriate and clear style with minor typographical errors.

4 = Good understanding of the assignment. All of the required elements are present and correctly interpreted. Minor lack of understanding or in the depth of the analysis/review. Write-up is in an appropriate and clear style with minor typographical errors.

3 = Adequate understanding of the assignment. All the required components are present but without sufficient depth of understanding with respect to the interpretation of results. Write-up is in appropriate style but lacks some clarity or contains some grammatical and typographical errors.

2 = Inadequate understanding of the design or statistical procedures required for analyzing the problem. Some main elements of the assignments are missing (e.g., interpretation of important results, or discussion of main strengths or weaknesses of a paper). Severe problems in understanding the link between theory, results, and interpretation. Write-up is not very clear or contains numerous typographical and grammatical errors.

1 = No understanding of the nature of the assignment. Incorrect analyses or interpretations of results, or lack of understanding of the papers to be reviewed. Write-up is in an unclear style or contains numerous typographical, logical, and grammatical errors.

0 = Assignment is not (timely) handed in.

Course schedule	Topics, Readings & Assignments [⇒ required reading; + optional reading; SCC: Shadish, Cook & Campbell, 2002]
W1 (Jan 21 & 23)	**Introduction to Causal Inference & Quasi-Experimentation: General Introduction** ⇒ SCC: Chapter 1 ⇒ Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology, 21*, 291-313.
W2 (Jan 28 & 30)	**Introduction to Causal Inference & Quasi-Experimentation: Philosophy of Causation** ⇒ SCC: Chapters 2 & 3 ⇒ Kaplan, D. (2009). Causal inference in non-experimental educational policy research. In D. N. Plank, W. E. Schmidt, & G. Sykes (Eds.), *AERA Handbook on Education Policy Research*. Washington, D. C.: AERA.
W3 (Feb 4 & 6)	**Randomized Experiments & Rubin Causal Model: Basic Design** ⇒ SCC: Chapter 8 ⇒ Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945-970. ⇒ Steiner, P. M., Kim, Y., Hall, C., & Su, D. (2014). Graphical models for quasi-experimental designs. *Working Paper.* + Bloom, H. S. (2006). The Core Analytics of Randomized Experiments for Social Research. *MDRC working paper*. + West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings. Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). Cambridge, UK: Cambridge University Press.
W4 (Feb 11 & 13) *Review 1 (13)*	**Randomized Experiments: Attrition & Non-compliance; Instrumental Variable Estimation** ⇒ SCC: Chapters 9 & 10 ⇒ Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics, 114*(2), 497-532.
W5 (Feb 18 & 20)	**Regression Discontinuity Designs: Sharp RD Design** ⇒ SCC: Chapter 7 ⇒ Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281-355. + Bloom, H. S. (2012). Modern regression discontinuity analysis. *MDRC working paper*.

W6 (Feb 25 & 27)
Analysis 1 (Feb 27)

Regression Discontinuity Designs: Fuzzy RD Design

⇒ Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122-154.

W7 (Mar 4)
Review 2 (4)

Regression Discontinuity Designs: Design Elements & Multivariate RD

+ Wong, V. C., Steiner, P. M., & Cook, T. D. (2012). Analyzing Regression-Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods.

W8 (Mar 11 & 13)

Interrupted Time Series Designs: Basic Design & Design Elements

⇒ SCC: Chapter 6

+ Bloom, H. S. (1999). Estimating Program Impacts on Student Achievement Using "Short" Interrupted Time Series. *MDRC working paper*.
+ Wagner, A. K., Soumerai, S. B., Ross-Degnan, D., & Zhang, F. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27*, 299–309.

W9 (Mar 18 & 20)

Spring recess

W10 (Mar 25 & 27)
Analysis 2 (25)
Review 3 (27)

Interrupted Time Series Designs: Analysis

⇒ Dee, T., & Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management, 30*(3), 418–446.
⇒ Wong, M., Cook, T. D., & Steiner, P. M. (2012). No Child Left Behind: An interim evaluation using two interrupted time series each with its own non-equivalent comparison series. *IPR Working Paper*, Northwestern University.

W11 (Apr 1 & 3)

Non-Equivalent Control Group Designs: Basic Design

⇒ SCC: Chapters 4 & 5

W12 (Apr 8 & 10)

Non-Equivalent Control Group Designs: Matching & Propensity Score Techniques

⇒ Schafer, J. L., & Kang, J. (2008). Average causal effects from non-randomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279-313.
⇒ Steiner, P. M., & Cook, D. (2013). *Matching and Propensity Scores*. In T. D. Little (ed.), The Oxford Handbook of Quantitative Methods.

+ Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics, 86*(1), 4-29.
+ Rosenbaum, P. R., & Rubin, D. B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

W13 (Apr 15 & 17)
Analysis 3 (15)

Non-Equivalent Control Group Designs: Design Elements

Review 4 (17) ⇒ Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*, 901-910.

W13 (Apr 22 & 24) **Instrumental Variable Design Revisited**

Analysis 4 (22)

⇒ Morgan, S. L., & Winship C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge: Cambridge University Press. Only Chapter 7.
⇒ Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. Quarterly Journal of Economics, 118(4), 495-532.
+ Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association 87:328-336.
+ Dee, T. S. (2004). Are there civic returns to education? Journal of Public Economics, 88(9-10), 1697-720.

W14 (Apr 29 & May 1) **Empirical Evaluation of Quasi-Experimental Methods: Within-Study Comparisons & Meta-Analyses**

⇒ Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management, 27*(4), 724-750.
⇒ Cook, T. D., Steiner, P. M., & Pohl, S. (2009). Assessing how bias reduction is influenced by covariate choice, unreliability and data analytic mode: An analysis of different kinds of within-study comparisons in different substantive domains. *Multivariate Behavioral Research, 44,* 828-847.

+ Shadish, W. S., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (under review). A Randomized Experiment Comparing Random to Cutoff-Based Assignment. Psychological Methods.
+ Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. *Journal of the American Statistical Association, 103*, 1334-1343.
+ Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 15(3), 250-267.
+ Cook, T. D., & Wong, V. C. (2008). Empirical tests of the validity of the regression-discontinuity design. *Annales d'Economie et de Statistique, 91/92*, 127-150.

W15 (Mai 6 & 8) **Empirical Evaluation of Quasi-Experimental Methods: Within-Study Comparisons & Meta-Analyses**

⇒ Peikes, D.N., Moreno, L. & Orzol, S.M. (2008). Propensity score matching: A note of caution for evaluators of social programs. *The American Statistician, 62*, 222-231

Further Readings
General Papers

+ Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5-86.
+ Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology, 35*(1), 1-98. [with a discussion by Michael Sobel]
+ Pearl, J. (2010). The foundations of causal inference. *Sociological Methodology, 40*(1), 75-149.
+ Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, *2*, 808-840.
+ Sobel, M. E. (1996). An introduction to causal inference. *Sociological Methods and Research 24*(3), 353-379.
+ West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings. Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). Cambridge, UK: Cambridge University Press.
+ Winship, C., & Morgan, W. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659-707.

Useful Books

Design & Analysis of Quasi-Experiments

+ Angrist, J. D., & Pischke, J.-S. (2009). Mostly Harmless Econometrics. An Empiricist’s Companion. Princeton: Princeton University Press.
+ Berk, R. A. (2004). Regression Analysis. A Constructive Critique. Thousand Oaks: Sage Publications.
+ Freedman, D. A. (2005). Statistical Models. Theory and Practice. Cambridge: Cambridge University Press.
+ Freedman, D. A. (2010). Statistical Models and Causal Inference. Cambridge: Cambridge University Press. [Collection of papers]
+ Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge. Cambridge University Press.
+ Guo, S., & Fraser, M. W. (2010). Propensity score analysis. Statistical Methods and Applications. Thousand Oaks: Sage.
+ Morgan, S. L., & Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge: Cambridge University Press.
+ Murnane, R. J., & Willett, J. B. (2011). Methods Matter. Improving Causal Inference in Educational and Social Science Research. Oxford: Oxford University Press.
+ Pearl, J. (2009). Causality: Models, Reasoning, and Inference (2nd ed.). Cambridge: Cambridge University Press.
+ Rosenbaum, P. R. (2002). Observational Studies (2nd Ed.). New York: Springer.
+ Rosenbaum, P. R. (2009). Design of Observational Studies. New York: Springer.
+ Rubin, D. B. (2006). Matched Sampling for Causal Effects. Cambridge: Cambridge University Press. [Collection of papers]

Philosophy on Causation

+ Cartwright, N. (2007). Hunting Causes and Using Them. Approaches in Philosophy and Economics. Cambridge: Cambridge University Press.
+ Collins, J., Hall, N., & Paul L. A. (2004). Causation and Counterfactuals. MIT Press.
+ Mackie, J. L. (1980). The Cement of the Universe. A Study of Causation. Oxford: Oxford University Press.
+ Lewis, D. K. (2001). Counterfactuals. 2nd Edition. Wiley-Blackwell.
+ Woodward, J. (2003). Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford University Press.