Causal Inference for the Social Sciences

Jake Bowers [*] Thomas (Tom) Leavitt [†] Seongyoon Kim [‡] Sally Sharif [§]

ICPSR Session 1 (June 20–July 7, 2023)

Abstract

This course introduces methods and concepts used to infer causal effects from comparisons of intervention and control groups. We'll use the potential outcomes framework of causality to show how a study's research design provides a foundation for estimation and testing. We focus, first, on properties of estimators and tests in randomized experiments, e.g., unbiasedness, consistency, controlled error rates. We then turn to research designs that are either partially controlled (e.g., experiments with noncompliance and/or attrition) or uncontrolled (e.g., observational studies). For observational studies, we focus primarily on matching methods implemented via `optmatch` and related packages in `R`. Finally, we turn to sensitivity analysis — namely, how to assess how inferences would change should certain assumptions about the research design be false. Examples throughout the course are drawn from economics, political science, public health, and sociology.

The course assumes familiarity with linear algebra and strong knowledge of statistical concepts, such as sampling distributions, statistical inference, and hypothesis testing. The course will rely on `R` for computation.

The course meets in CCCB 3420 from 10:00 AM to 12:45 PM US Eastern Time, Monday through Friday, from June 20 through July 7. There will be no class on June 19, the first day of the course. Tom will be the primary instructor for the first week of the course. Jake will be the primary instructor for weeks 2 and 3.

Seongyoon Kim and Sally Sharif will be the TAs for the duration of the course.

You can sign up for office hours with Jake at https://calendly.com/jakebowers/icpsr-office-hours. By default they are on Zoom, but he can meet in-person on request in Helen Newberry (rm. 414) or at a local cafe. Tom's, Seongyoon's and Sally's office hours are listed on the course's Canvas site. Tom's office is in Helen Newberry (rm. 414); Seongyoon's and Sally's offices are in Helen Newberry (rms. 412 and 417, respectively).

[*] Political Science and Statistics, University of Illinois @ Urbana-Champaign; jwbowers@illinois.edu

[†] Marxe School of Public and International Affairs at Baruch College, City University of New York (CUNY); thomas.leavitt@baruch.cuny.edu

[‡] Economics, University of Michigan – Ann Arbor; syoonkim@umich.edu

[§] School for International Studies, Simon Fraser University; s_sharif@sfu.ca

Overview

We may all warn our freshmen that association is not causation, but inferring causation has always been a central aim both for statisticians and for their collaborators. Until recently, however, inference of causation from statistical evidence depended on murky, scarcely attainable requirements; in practice, the weight of casual arguments was largely determined by the scientific authority of the people making them.

Requirements for causal inference become more clear when they are framed in terms of *potential outcomes*. This was first done by Neyman, who in the 1920s used potential outcomes to model agricultural experiments. Fisher independently proposed a related but distinct, ultimately more influential, analysis of experiments in 1935, and a rich strain of causal analysis developed among his intellectual progeny. It clarified the differing requirements for causal inference with experiments and with observational data, isolating the distinct contributions required of the statistician and of his disciplinary collaborators; generated more satisfying methods with which to address potential confounding due to measured variables; qualitatively and quantitatively advanced our grasp of unmeasured confounding and its potential ramifications; furnished statistical methods with which to eke more out of the strongest study designs, under fewer assumptions; and articulated principles with which to understand study designs as a spectrum, rather than a dichotomy between “good" experiments and “bad" observational studies. Understanding the methods and outlook of the school founded by Fisher’s student W. G. Cochran will be the central task of this course.

The course begins by applying the Fisher and Neyman-Rubin approaches to statistical inference for counterfactual causal efffects to randomized experiments, touching on considerations specific to clustered treatment assignment, “small” sample sizes and treatment effect heterogeneity. The next segment addresses conceptual and methodogical challenges of applying the same models to analysis of non-experimental data. This course segment covers ignorability, selection, “common support,” covariate balance, paired comparisons, optimal matching and propensity scores. With these foundations in place, the course then turns to sensitivity analysis — namely, how to make principled assessments of how inferences would change should crucial assumptions be false. Over the course’s three weeks, the course becomes progressively less conceptual and more applied with increasing emphasis on computing strategies in R.

Administrative

Textbooks

The main texts for the course are

Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press

Paul R Rosenbaum (2010). *Design of Observational Studies*. New York, NY: Springer

Paul R Rosenbaum (2002b). *Observational Studies*. Second. New York, NY: Springer

These three textbooks are presented in varying difficulty and we will draw from all three. Although we won’t follow these books closely, their goals and methods align with the course’s, and they will be useful as references and supplements.

Other texts that we draw on include

Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton

Guido W Imbens and Donald B Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press

Other readings will be assigned and distributed electronically.

If you're new to R, we suggest getting a hold of:

John Fox (2016). *Applied Regression Analysis and Generalized Linear Models.* 3rd. Los Angeles, CA: SAGE Publications

Hadley Wickham and Garrett Grolemund (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* First. Sebastopol, CA: O'Reilly Media

R software will be required for several specific segments of the course. With some independent effort, students not familiar with R in advance should be able to learn enough R during the course to complete these assignments. We also recommend some work with R — for example, via working through some online R courses — before the course for students who have never used it before.

Assignments

Assignments are due each Friday, at the beginning of class. Parts of the assignment will be given at the beginning of the week, but other parts will be given during class, over the course of the week. Late homework will not be accepted without cause (or prior arrangement with the teaching assistant).

Participation is expected. It can take various forms:

1. Doing in-class exercises and discussing them with your peers;
2. From time to time, making a clarification or raising a clarifying questions;
3. Contributing to in-class discussions;
4. Drop by one of the professor's office hours to share a point that you *and at least one classmate* would like to have clarified or amplified, or to point out a connection to your field;

If you are taking the course for a grade, make a point of doing at least one of 3 and 4.

Course Schedule

The course schedule is below. In the Course content section, we provide more extensive readings on each topic in the schedule, as well as additional “special topics.” The last two days of the course are reserved for special topics chosen by students.

Date	Instructor	Topic	Required readings	Application
Tues, June 20	Tom	Introduction: Randomized experiments and potential outcomes	Holland (1986) Kinder and Palfrey (1993, Section 1.2) Fisher (1935, Introduction) Rosenbaum (2017, Chapter 2)	Fisher (1935)
Wed, June 21	Tom	Randomized experiments: Fisherian exact tests	Rosenbaum (2017, Chapter 3)	Arceneaux (2005)
Thurs, June 22	Tom	Randomized experiments: Neymanian estimation and inference	Gerber and Green (2012, Chapter 2) Aronow and Middleton (2013) Middleton and Aronow (2015) Gerber and Green (2012, pp. 51 – 61)	
Fri, June 23	Tom	Randomized experiments: Covariance adjustment & regression	Gerber and Green (2012, Chapter 4) Rosenbaum (2002) Lin (2013)	
Mon, June 26 (HW 1 due)	Jake	Randomized experiments: Noncompliance and attrition	Gerber and Green (2012, Chapters 5 – 6) Rosenbaum (1996) Rosenbaum (2010, Section 5.3)	Albertson and Lawrence (2009)
Tues, June 27	Jake	Observational studies: Introduction	Bind and Rubin (2019) Gelman and Hill (2006, Sections 9.0 – 9.2) Berk (2010)	Cerdá et al (2012)
Wed, June 28	Jake	Introduction to matching	Rosenbaum (2017, pp. 65 – 90); Hansen and Bowers (2008)	
Thurs, June 29	Jake	Nuts & bolts of matching	Rosenbaum (2010, Chapters 7 – 8) Rosenbaum (2017, Chapter 11)	
Fri, June 30 (HW 2 due)	Jake	Multivariate matching with scores	Rosenbaum (2017, pp. 90 – 96) Hansen (2011)	
Mon, July 3	Jake	Outcome analysis with Stratified Designs	Gerber and Green (2021, pp. 71 – 79)	
Tues, July 4	Jake	Matching with Multi-valued "Treatments"	See § 7.1 esp. Rabb et al (2022)	
Wed, July 5	Jake	Fisherian sensitivity analysis	Rosenbaum (2017, Chapter 9) Gastwirth et al (2000) Rosenbaum (2015) Rosenbaum (2018)	
Thurs, July 6	Jake	Natural Experiments and Discontinuities	See § 7.2	
Fri, July 7 (HW 3 due)	Tom	Difference in Differences	See § 7.5	

Course content

1 Potential outcomes and random assignment

Required

Paul W Holland (1986). “Statistics and Causal Inference”. In: *Journal of the American Statistical Association* 81.396, pp. 945–960, Sections 1 – 4. (The article that brought the “Rubin Causal Model” to statisticians’ attention.)

Section 1.2, “Experimentation defined,” of Donald R Kinder and Thomas R Palfrey (1993). “On Behalf of an Experimental Political Science”. In: *Experimental Foundations of Political Science*. Ed. by Donald R Kinder and Thomas R Palfrey. Michigan Studies in Political Analysis. Ann Arbor, MI: University of Michigan Press. Chap. 1, pp. 1–39. (Particularly pp. 5 – 10.)

Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press, Chapter 2.

Chapter 1, “Introduction,” of Ronald Aylmer Fisher (1935). *The Design of Experiments*. Edinburgh, SCT: Oliver and Boyd and pages 131 – 135 of Joan Fisher Box (1978). *R. A. Fisher, the Life of a Scientist*. New York, NY: Wiley for historical context.

Recommended

Jake Bowers and Thomas Leavitt (2020). “Causality and Design-Based Inference”. In: *The SAGE Handbook of Research Methods in Political Science and International Relations*. Ed. by Luigi Curini and Robert Franzese. Vol. 2. Thousand Oaks, CA: SAGE Publications. Chap. 41, pp. 769–804

2 Random assignment as a basis for inference

Application

Kevin Arceneaux (2005). “Using Cluster Randomized Field Experiments to Study Voting Behavior”. In: *The Annals of the American Academy of Political and Social Science* 601.1, pp. 169–179

2.1 Inference for causal effects: the Fisherian tradition

Required

Chapter 3 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press.

Recommended

Pages 27 – 49 of Paul R Rosenbaum (2002b). *Observational Studies.* Second. New York, NY: Springer

Chapter 2 of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer.

2.2 Inference for causal effects: the Neyman tradition

2.2.1 Estimation of average causal effects

Required

Chapter 2 and pp. 51 – 61 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton

Peter M Aronow and Joel A Middleton (2013). "A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments". In: *Journal of Causal Inference* 1.1, pp. 135–154

Joel A Middleton and Peter M Aronow (2015). "Unbiased Estimation of the Average Treatment Effect in Cluster-Randomized Experiments". In: *Statistics, Politics and Policy* 6.1-2, pp. 39–75

2.2.2 Variance estimation and hypothesis testing

Required

Chapter 3 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton.

Endnote spanning pages A-32 and 33, David A Freedman, Robert Pisani, and Roger Purves (1998). *Statistics.* 3rd. New York, NY: W. W. Norton & Company. (This can be read as a précis of: Jerzy Splawa-Neyman, D M Dabrowska, T P Speed, et al. (1990). "On the application of probability theory to agricultural experiments. Essay on principles. Section 9". In: *Statistical Science* 5.4, pp. 465–472.)

Chapter 6, pp. 87 – 98 of Guido W Imbens and Donald B Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press

Recommended

Chapter 6 of Thad Dunning (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York, NY: Cambridge University Press

Xinran Li and Peng Ding (2017). "General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference". In: *Journal of the American Statistical Association* 112.520, pp. 1759–1769

Peng Ding (2017). "A Paradox from Randomization-Based Causal Inference". In: *Statistical Science* 32.3, pp. 331–345

Peter M Aronow, Donald P Green, Donald KK Lee, et al. (2014). "Sharp Bounds on the Variance in Randomized Experiments". In: *The Annals of Statistics* 42.3, pp. 850–871

2.3 Covariance adjustment in randomized experiments

Required

Chapter 4 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W.W. Norton.

Paul R Rosenbaum (2002a). "Covariance Adjustment in Randomized Experiments and Observational Studies". In: *Statistical Science* 17.3, pp. 286–327.

Winston Lin (2013). "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique". In: *The Annals of Applied Statistics* 7.1, pp. 295–318

Recommended

Luke W Miratrix, Jasjeet S Sekhon, and Bin Yu (2013). "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.2, pp. 369–396

David A Freedman (2008b). "On Regression Adjustments to Experimental Data". In: *Advances in Applied Mathematics* 40.2, pp. 180–193

David A Freedman (2008c). "Randomization Does Not Justify Logistic Regression". In: *Statistical Science* 23.2, pp. 237–249

David A Freedman (2008a). "On Regression Adjustments in Experiments with Several Treatments". In: *The Annals of Applied Statistics* 2.1, pp. 176–196

Cyrus Samii and Peter M Aronow (2012). "On Equivalencies between Design-based and Regression-based Variance Estimators for Randomized Experiments". In: *Statistics & Probability Letters* 82.2, pp. 365–370

Peter M Aronow and Cyrus Samii (2016). "Does Regression Produce Representative Estimates of Causal Effects?" In: *American Journal of Political Science* 60.1, pp. 250–267

Alberto Abadie, Susan Athey, et al. (2020). "Sampling-Based versus Design-Based Uncertainty in Regression Analysis". In: *Econometrica* 88.1, pp. 265–296

3 Noncompliance and Attrition

Applications

Bethany Albertson and Adria Lawrence (2009). "After the Credits Roll: The Long-Term Effects of Educational Television on Public Knowledge and Attitudes". In: *American Politics Research* 37.2, pp. 275–300

3.1 Noncompliance and instrumental variables

Required

Chapters 5 and 6 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton.

Section 5.3, "Instruments," of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Paul R Rosenbaum (1996). "Identification of Causal Effects Using Instrumental Variables: Comment". In: *Journal of the American Statistical Association* 91.434, pp. 465–468

Recommended

Section 2.3 of Paul R Rosenbaum (2002a). "Covariance Adjustment in Randomized Experiments and Observational Studies". In: *Statistical Science* 17.3, pp. 286–327

Joshua D Angrist, Guido W Imbens, and Donald B Rubin (1996). "Identification of Causal Effects Using Instrumental Variables". In: *Journal of the American Statistical Association* 91.434, pp. 444–455

Guido W Imbens and Paul R Rosenbaum (2005). "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.1, pp. 109–126

Hyunseung Kang, Laura Peck, and Luke Keele (2018). "Inference for Instrumental Variables: A Randomization Inference Approach". In: *Journal of the Royal Statistical Society. Series A: Statistics in Society* 181.4, pp. 1231–1254

Ben B Hansen and Jake Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". In: *Statistical Science* 23.2, pp. 219–236

3.2 Attrition, or missing outcomes

Recommended

Chapter 7 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton

David S Lee (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects". In: *The Review of Economic Studies* 76.3, pp. 1071–1102

Peter M Aronow, Jonathon Baron, and Lauren Pinson (2019). "A Note on Dropping Experimental Subjects who Fail a Manipulation Check". In: *Political Analysis* 27.4, pp. 572–589

Joel L Horowitz and Charles F Manski (2000). "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data". In: *Journal of the American Statistical Association* 95.449, pp. 77–84

Alexander Coppock et al. (2017). "Combining Double Sampling and Bounds to Address Nonignorable Missing Outcomes in Randomized Experiments". In: *Political Analysis* 25.2, pp. 188–206

4 Observational Studies

4.1 Introduction: "Controlling for" in observational studies

Required

Marie-Abele C Bind and Donald B Rubin (2019). "Bridging Observational Studies and Randomized Experiments by Embedding the Former in the Latter". In: *Statistical Methods in Medical Research* 28.7, pp. 1958–1978

Sections 9.0 – 9.2 (especially discussion of interpolation and extrapolation) of Andrew Gelman and Jennifer Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York, NY: Cambridge University Press

Richard Berk (2010). "What You Can and Can't Properly Do with Regression". In: *Journal of Quantitative Criminology* 26.4, pp. 481–487

Recommended

Chapter 5 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press

William G Cochran (1965). "The Planning of Observational Studies of Human Populations". In: *Journal of the Royal Statistical Society. Series A (General)* 128.2, pp. 234–266

On the problem of kitchen sink regressions, Christopher H Achen (2002). "Toward a New Political Methodology: Microfoundations and ART". in: *Annual Review of Political Science* 5, pp. 423–450

Chapters 11 and 19 (on overly influential points) of John Fox (2016). *Applied Regression Analysis and Generalized Linear Models*. 3rd. Los Angeles, CA: SAGE Publications

4.2 Matching: An introduction

Application

Magdalena Cerdá et al. (2012). “Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia”. In: *American Journal of Epidemiology* 175.10, pp. 1045–1053

Required

Pages 65 – 90 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press

Paul R Rosenbaum (2020). “Modern Algorithms for Matching in Observational Studies”. In: *Annual Review of Statistics and Its Application* 7.1, pp. 143–176

Recommended

Chapter 3 of Paul R Rosenbaum (2002b). *Observational Studies*. Second. New York, NY: Springer, specifically Sections 3.1 – 3.2 and 3.4 – 3.5.

Paul R Rosenbaum (2001b). “Observational Studies: Overview”. In: *International Encyclopedia of the Social & Behavioral Sciences*. Ed. by Neil J Smelser and Paul B Baltes. Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 10808–10815

Robert Bifulco (2012). “Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison”. In: *Journal of Policy Analysis and Management* 31.3, pp. 729–751

Kevin Arceneaux (2010). “A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark”. In: *Sociological Methods & Research* 39.2, pp. 256–282

4.3 Nuts & bolts of matching

Required

Chapters 7 – 8 of Paul R Rosenbaum (2010). *Design of Observational Studies*. New York, NY: Springer,

Chapter 11 Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press,

4.4 Propensity scores methods

Required

Pages 90 – 96 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press

Ben B Hansen (2011). "Propensity Score Matching to Extract Latent Experiments from Nonexperimental Data: A Case Study". In: *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland*. Ed. by Neil J. Dorans and Sandip Sinharay. Vol. 202. Lecture Notes in Statistics. New York, NY: Springer. Chap. 9, pp. 149–181

Recommended

Donald B Rubin (1979). "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies". In: *Journal of the American Statistical Association* 74.366a, pp. 318–328

James M Robins, Miguel Ángel Hernán, and Babette Brumback (2000). "Marginal Structural Models and Causal Inference in Epidemiology". In: *Epidemiology* 11.5, pp. 550–560

Daniel E Ho et al. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis* 15.3, pp. 199–236

Chapters 13 of Paul R Rosenbaum (2010). *Design of Observational Studies*. New York, NY: Springer

Paul R Rosenbaum and Donald B Rubin (1985). "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score". In: *The American Statistician* 39.1, pp. 33–38

4.5 Covariate balance and outcome analysis after matching

Required

Ben B Hansen and Jake Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". In: *Statistical Science* 23.2, pp. 219–236

Pages 71 – 79 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W.W. Norton

Recommended

Colin B Fogarty (2018). "On Mitigating the Analytical Limitations of Finely Stratified Experiments". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.5, pp. 1035–1056

Nicole E Pashley and Luke W Miratrix (2020). "Insights on Variance Estimation for Blocked and Matched Pairs Designs". In: *Journal of Educational and Behavioral Statistics*

Kosuke Imai (2008). "Variance Identification and Efficiency Analysis in Randomized Experiments under the Matched-Pair Design". In: *Statistics in Medicine* 27.24, pp. 4857–4873

Kosuke Imai, Gary King, and Elizabeth A Stuart (2008). "Misunderstandings between Experimentalists and Observationalists about Causal Inference". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.2, pp. 481–502

5 Sensitivity analysis

Application

Magdalena Cerdá et al. (2012). "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". In: *American Journal of Epidemiology* 175.10, pp. 1045–1053

5.1 Sensitivity analysis for sharp nulls

Required

Chapter 11 Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press,

Recommended

Chapter 4 of Paul R Rosenbaum (2002b). *Observational Studies.* Second. New York, NY: Springer

Paul R Rosenbaum (2018). "Sensitivity Analysis for Stratified Comparisons in an Observational Study of the Effect of Smoking on Homocysteine Levels". In: *Annals of Applied Statistics* 12.4, pp. 2312–2334

Paul R Rosenbaum and Abba M Krieger (1990). "Sensitivity of Two-Sample Permutation Inferences in Observational Studies". In: *Journal of the American Statistical Association* 85.410, pp. 493–498

Ben B Hansen, Paul R Rosenbaum, and Dylan S Small (2014). "Clustered Treatment Assignments and Sensitivity to Unmeasured Biases in Observational Studies". In: *Journal of the American Statistical Association* 109.505, pp. 133–144

Jesse Y Hsu and Dylan S Small (2013). "Calibrating Sensitivity Analyses to Observed Covariates in Observational Studies". In: *Biometrics* 69.4, pp. 803–811

5.2 Sensitivity analysis for weak nulls

Required

Colin B Fogarty (2023). "Testing Weak Nulls in Matched Observational Studies". In: *Biometrics*

Recommended

Colin B Fogarty (2020). "Studentized Sensitivity Analysis for the Sample Average Treatment Effect in Paired Observational Studies". In: *Journal of the American Statistical Association* 115.531, pp. 1518–1530

Colin B Fogarty et al. (2017). "Randomization Inference and Sensitivity Analysis for Composite Null Hypotheses With Binary Outcomes in Matched Observational Studies". In: *Journal of the American Statistical Association* 112.517, pp. 321–331

6 Design sensitivity

Required

Paul R Rosenbaum (2004). "Design Sensitivity in Observational Studies". In: *Biometrika* 91.1, pp. 153–164

Chapter 10 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press

7 Additional topics

7.1 Nonbipartite matching

Chapter 11 of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Bo Lu, Elaine Zanutto, et al. (2001). "Matching with Doses in an Observational Study of a Media Campaign against Drug Abuse". In: *Journal of the American Statistical Association* 96.456, pp. 1245–1253

Chapter 11 of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Bo Lu, Robert Greevy, et al. (2011). "Optimal Nonbipartite Matching and Its Statistical Applications". In: *The American Statistician* 65.1, pp. 21–30

Bo Lu (2005). "Propensity Score Matching with Time-Dependent Covariates". In: *Biometrics* 61.3, pp. 721–728

Mike Baiocchi et al. (2010). "Building a Stronger Instrument in an Observational Study of Perinatal Care for Premature Infants". In: *Journal of the American Statistical Association* 105.492, pp. 1285–1296

José R Zubizarreta, Magdalena Cerdá, and Paul R Rosenbaum (2013). "Effect of the 2010 Chilean Earthquake on Posttraumatic Stress Reducing Sensitivity to Unmeasured Bias Through Study Design". In: *Epidemiology* 24.1, pp. 79–87

Nathaniel Rabb et al. (July 2022). "The influence of social norms varies with "others" groups: Evidence from COVID-19 vaccination intentions". In: *Proceedings of the National Academy of Sciences* 119.29. DOI: https://doi.org/10.1073/pnas.2118770119

7.2 Regression discontinuity designs and "natural experiments"

Devin Caughey and Jasjeet S Sekhon (2011). "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008". In: *Political Analysis* 19.4, pp. 385–408

Matias D Cattaneo, Rocío Titiunik, and Gonzalo Vazquez-Bare (2020). "The Regression Discontinuity Design". In: *Sage Handbook of Research Methods in Political Science & International Relations.* Ed. by Luigi Curini and Robert J Franzese Jr. Washington, D.C.: Sage Publications

Adam Sales and Ben B Hansen (2020). "Limitless Regression Discontinuity". In: *Journal of Educational and Behavioral Statistics* 45.2, pp. 143–174

Luke Keele, Rocío Titiunik, and José R Zubizarreta (2015). "Enhancing a Geographic Regression Discontinuity Design through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.1, pp. 223–239

Jasjeet Sekhon and Rocío Titiunik (2017). "On Interpreting the Regression Discontinuity Design as a Local Experiment". In: *Regression Discontinuity Designs: Theory and Applications.* Ed. by Matias D. Cattaneo and Juan Carlos Escanciano. Vol. 38. Advances in Econometrics. Bingley, UK: Emerald Group Publishing. Chap. 1

Jasjeet S Sekhon and Rocío Titiunik (2016). "Understanding Regression Discontinuity Designs As Observational Studies". In: *Observational Studies* 2, pp. 174–182

Ben B Hansen and Adam Sales (2015). "Comment on Cochran's "Observational Studies"". In: *Observational Studies*, pp. 184–193

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design". In: *Econometrica* 69.1, pp. 201–209

Guido W Imbens and Thomas Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". In: *Journal of Econometrics* 142.2, pp. 615–635

David S Lee (2008). "Randomized Experiments from Non-Random Selection in US House Elections". In: *Journal of Econometrics* 142.2, pp. 675–697

Andrew Gelman and Guido W Imbens (2019). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs". In: *Journal of Business & Economic Statistics* 37.3, pp. 447–456

Justin McCrary (2008). "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test". In: *Journal of Econometrics* 142.2, pp. 698–714

Chapter 6 of Joshua D Angrist and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press

7.3 Interference

Paul R Rosenbaum (2007b). "Interference Between Units in Randomized Experiments". In: *Journal of the American Statistical Association* 102.477, pp. 191–200

Jake Bowers, Mark Fredrickson, and Costas Panagopoulos (2013). "Reasoning about Interference Between Units: A General Framework". In: *Political Analysis* 21.1, pp. 97–124

Peter M Aronow and Cyrus Samii (2017). "Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment". In: *Annals of Applied Statistics* 11.4, pp. 1912–1947

Charles F Manski (2013). "Identification of Treatment Response with Social Interactions". In: *The Econometrics Journal* 16.1, S1–S23

Susan Athey, Dean Eckles, and Guido W Imbens (2018). "Exact p-Values for Network Interference". In: *Journal of the American Statistical Association* 113.521, pp. 230–240

7.4 Factorial and complex experiments

Tirthankar Dasgupta, Natesh S Pillai, and Donald B Rubin (2015). "Causal Inference from 2^K Factorial Designs by using Potential Outcomes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.4, pp. 727–753

Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto (2014). "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments". In: *Political Analysis* 22.1, pp. 1–30

Naoki Egami and Kosuke Imai (2019). "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis". In: *Journal of the American Statistical Association* 114.526, pp. 529–540

Alan S Gerber, Donald P Green, et al. (2010). "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments". In: *Political Analysis* 18.3, pp. 297–315

7.5 Difference-in-Differences

Pages 162 – 167 of Paul R Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press

Section 4.1 of Alan S Gerber and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton

Chapter 5 of Joshua D Angrist and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press

Michael Lechner (2011). "The Estimation of Causal Effects by Difference-in-Difference Methods". In: *Foundations and Trends in Econometrics* 4.3, pp. 165–224

Charles F Manski and John V Pepper (2018). "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions". In: *The Review of Economics and Statistics* 100.2, pp. 232–244

Peng Ding and Fan Li (2019). "A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment". In: *Political Analysis* 27.4, pp. 605–615

Kosuke Imai and In Song Kim (2021). "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data". In: *Political Analysis* 29.3, pp. 405–415

7.6 Special topics in matching: Multilevel, risk set, cardinality, generalized full matching and fine balance

José R Zubizarreta and Luke Keele (2017). "Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the Effectiveness of Private Schools Under a Large-Scale Voucher System". In: *Journal of the American Statistical Association* 112.518, pp. 547–560

Shu Yang et al. (2016). "Propensity Score Matching and Subclassification in Observational Studies with Multi-level Treatments". In: *Biometrics* 72.4, pp. 1055–1065

Samuel D Pimentel et al. (2018). "Optimal Multilevel Matching Using Network Flows: An Application to a Summer Reading Intervention". In: *Annals of Applied Statistics* 12.3, pp. 1479–1505

Chapter 12 of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Yunfei Paul Li, Kathleen J Propert, and Paul R Rosenbaum (2001). "Balanced Risk Set Matching". In: *Journal of the American Statistical Association* 96.455, pp. 870–882

Chapter 10 of Paul R Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Cinar Kilcioglu and José R Zubizarreta (2016). "Maximizing the Information Content of a Balanced Matched Sample in a Study of the Economic Performance of Green Buildings". In: *The Annals of Applied Statistics* 10.4, pp. 1997–2020

Fredrik Sävje, Michael J Higgins, and Jasjeet S Sekhon (2021). "Generalized Full Matching". In: *Political Analysis*

7.7 Attributable effects and inference of random causal quantities

Paul R Rosenbaum (2001a). "Effects Attributable to Treatment: Inference in Experiments and Observational Studies with a Discrete Pivot". In: *Biometrika* 88.1, pp. 219–231

Paul R Rosenbaum (2003). "Exact Confidence Intervals for Nonconstant Effects by Inverting the Signed Rank Test". In: *The American Statistician* 57.2, pp. 132–138

Ben B Hansen and Jake Bowers (2009). "Attributing Effects to a Cluster-Randomized Get-Out-the-Vote Campaign". In: *Journal of the American Statistical Association* 104.487, pp. 873–885

Jasjeet S Sekhon and Yotam Shem-Tov (2020). "Inference on a New Class of Sample Average Treatment Effects". In: *Journal of the American Statistical Association* 116.534, pp. 798–804

Luke Keele, Dylan Small, and Richard Grieve (2017). "Randomization-based Instrumental Variables Methods for Binary Outcomes with an Application to the 'IMPROVE' Trial". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180.2, pp. 569–586

Paul R Rosenbaum (2007a). "Confidence intervals for uncommon but dramatic responses to treatment". In: *Biometrics* 63.4, pp. 1164–1171

7.8 Regression-based sensitivity analysis

Carrie A Hosman, Ben B Hansen, and Paul W Holland (2010). "The Sensitivity of Linear Regression Coefficients' Confidence Limits to the Omission of a Confounder". In: *The Annals of Applied Statistics* 4.2, pp. 849–870

Carlos Cinelli and Chad Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67

Guido W Imbens (2003). "Sensitivity to Exogeneity Assumptions in Program Evaluation". In: *The American Economic Review* 93.2, pp. 126–132

Emily Oster (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204

7.9 External validity

Holger L Kern et al. (2016). "Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations". In: *Journal of Research on Educational Effectiveness* 9.1, pp. 103–127

Luke W Miratrix, Jasjeet S Sekhon, Alexander G Theodoridis, et al. (2018). "Worth Weighting? How to Think About and Use Weights in Survey Experiments". In: *Political Analysis* 26.3, pp. 275–291

Magdalena Bennett, Juan Pablo Vielma, and José R Zubizarreta (2020). "Building Representative Matched Samples With Multi-Valued Treatments in Large Observational Studies". In: *Journal of Computational and Graphical Statistics* 29.4, pp. 744–757

Jeffrey H Silber et al. (2014). "Template Matching for Auditing Hospital Cost and Quality". In: *Health Services Research* 48.5, pp. 1446–1474

Daniel Westreich et al. (2019). "Target Validity and the Hierarchy of Study Designs". In: *American Journal of Epidemiology* 188.2, pp. 438–443

7.10 Weighting methods

Ambarish Chattopadhyay, Christopher H Hase, and José R Zubizarreta (2020). "Balancing Versus Modeling Approaches to Weighting in Practice". In: *Statistics in Medicine* 39.24, pp. 3227–3254

Kosuke Imai and Marc Ratkovic (2014). "Covariate Balancing Propensity Score". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1, pp. 243–263

José R Zubizarreta (2015). "Stable Weights that Balance Covariates for Estimation with Incomplete Outcome Data". In: *Journal of the American Statistical Association* 110.511, pp. 910–922

Raymond K W Wong and Kwun Chuen Gary Chan (2018). "Kernel-based Covariate Functional Balancing for Observational Studies". In: *Biometrika* 105.1, pp. 199–213

Zhiqiang Tan (2020). "Regularized Calibrated Estimation of Propensity Scores with Model Misspecification and High-dimensional Data". In: *Biometrika* 107.1, pp. 137–158

Yixin Wang and José R Zubizarreta (2020). "Minimal Dispersion Approximately Balancing Weights: Asymptotic Properties and Practical Considerations". In: *Biometrika* 107.1, pp. 93–105

Jens Hainmueller (2012). "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." In: *Political Analysis* 20.1, pp. 25–46

7.11 Synthetic control

Alberto Abadie, Alexis Diamond, and Jens Hainmueller (2010). "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, pp. 493–505

Alberto Abadie, Alexis Diamond, and Jens Hainmueller (2012). "Comparative Politics and the Synthetic Control Method". In: *American Journal of Political Science* 59.2, pp. 495–510

Eli Ben-Michael, Avi Feller, and Jesse Rothstein (2021). "The Augmented Synthetic Control Method". In: *Journal of the American Statistical Association*

7.12 Bayesian causal inference

Donald B Rubin (1978). "Bayesian Inference for Causal Effects: The Role of Randomization". In: *The Annals of Statistics* 6.1, pp. 34–58

Chapter 8 of Guido W Imbens and Donald B Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press

leavitt2023

Guido W Imbens and Donald B Rubin (1997). "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance". In: *The Annals of Statistics* 25.1, pp. 305–327

Luke Keele and Kevin M Quinn (2017). "Bayesian Sensitivity Analysis for Causal Effects from 2×2 Tables in the Presence of Unmeasured Confounding with Application to Presidential Campaign Visits". In: *The Annals of Applied Statistics* 11.4, pp. 1974–1997

Peng Ding and Luke Miratrix (2019). "Model-Free Causal Inference of Binary Experimental Data". In: *Scandinavian Journal of Statistics* 46.1, pp. 200–214

References

Abadie, Alberto, Susan Athey, et al. (2020). “Sampling-Based versus Design-Based Uncertainty in Regression Analysis”. In: *Econometrica* 88.1, pp. 265–296.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). “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, pp. 493–505.

— (2012). “Comparative Politics and the Synthetic Control Method”. In: *American Journal of Political Science* 59.2, pp. 495–510.

Achen, Christopher H (2002). “Toward a New Political Methodology: Microfoundations and ART”. In: *Annual Review of Political Science* 5, pp. 423–450.

Albertson, Bethany and Adria Lawrence (2009). “After the Credits Roll: The Long-Term Effects of Educational Television on Public Knowledge and Attitudes”. In: *American Politics Research* 37.2, pp. 275–300.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996). “Identification of Causal Effects Using Instrumental Variables”. In: *Journal of the American Statistical Association* 91.434, pp. 444–455.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion.* Princeton, NJ: Princeton University Press.

Arceneaux, Kevin (2005). “Using Cluster Randomized Field Experiments to Study Voting Behavior”. In: *The Annals of the American Academy of Political and Social Science* 601.1, pp. 169–179.

— (2010). “A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark”. In: *Sociological Methods & Research* 39.2, pp. 256–282.

Aronow, Peter M, Jonathon Baron, and Lauren Pinson (2019). “A Note on Dropping Experimental Subjects who Fail a Manipulation Check”. In: *Political Analysis* 27.4, pp. 572–589.

Aronow, Peter M, Donald P Green, Donald KK Lee, et al. (2014). “Sharp Bounds on the Variance in Randomized Experiments”. In: *The Annals of Statistics* 42.3, pp. 850–871.

Aronow, Peter M and Joel A Middleton (2013). “A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments”. In: *Journal of Causal Inference* 1.1, pp. 135–154.

Aronow, Peter M and Cyrus Samii (2016). “Does Regression Produce Representative Estimates of Causal Effects?” In: *American Journal of Political Science* 60.1, pp. 250–267.

— (2017). “Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment”. In: *Annals of Applied Statistics* 11.4, pp. 1912–1947.

Athey, Susan, Dean Eckles, and Guido W Imbens (2018). “Exact p-Values for Network Interference”. In: *Journal of the American Statistical Association* 113.521, pp. 230–240.

Baiocchi, Mike et al. (2010). “Building a Stronger Instrument in an Observational Study of Perinatal Care for Premature Infants”. In: *Journal of the American Statistical Association* 105.492, pp. 1285–1296.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2021). “The Augmented Synthetic Control Method”. In: *Journal of the American Statistical Association.*

Bennett, Magdalena, Juan Pablo Vielma, and José R Zubizarreta (2020). "Building Representative Matched Samples With Multi-Valued Treatments in Large Observational Studies". In: *Journal of Computational and Graphical Statistics* 29.4, pp. 744–757.

Berk, Richard (2010). "What You Can and Can't Properly Do with Regression". In: *Journal of Quantitative Criminology* 26.4, pp. 481–487.

Bifulco, Robert (2012). "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison". In: *Journal of Policy Analysis and Management* 31.3, pp. 729–751.

Bind, Marie-Abele C and Donald B Rubin (2019). "Bridging Observational Studies and Randomized Experiments by Embedding the Former in the Latter". In: *Statistical Methods in Medical Research* 28.7, pp. 1958–1978.

Bowers, Jake, Mark Fredrickson, and Costas Panagopoulos (2013). "Reasoning about Interference Between Units: A General Framework". In: *Political Analysis* 21.1, pp. 97–124.

Bowers, Jake and Thomas Leavitt (2020). "Causality and Design-Based Inference". In: *The SAGE Handbook of Research Methods in Political Science and International Relations*. Ed. by Luigi Curini and Robert Franzese. Vol. 2. Thousand Oaks, CA: SAGE Publications. Chap. 41, pp. 769–804.

Box, Joan Fisher (1978). *R. A. Fisher, the Life of a Scientist*. New York, NY: Wiley.

Cattaneo, Matias D, Rocío Titiunik, and Gonzalo Vazquez-Bare (2020). "The Regression Discontinuity Design". In: *Sage Handbook of Research Methods in Political Science & International Relations*. Ed. by Luigi Curini and Robert J Franzese Jr. Washington, D.C.: Sage Publications.

Caughey, Devin and Jasjeet S Sekhon (2011). "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008". In: *Political Analysis* 19.4, pp. 385–408.

Cerdá, Magdalena et al. (2012). "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". In: *American Journal of Epidemiology* 175.10, pp. 1045–1053.

Chattopadhyay, Ambarish, Christopher H Hase, and José R Zubizarreta (2020). "Balancing Versus Modeling Approaches to Weighting in Practice". In: *Statistics in Medicine* 39.24, pp. 3227–3254.

Cinelli, Carlos and Chad Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67.

Cochran, William G (1965). "The Planning of Observational Studies of Human Populations". In: *Journal of the Royal Statistical Society. Series A (General)* 128.2, pp. 234–266.

Coppock, Alexander et al. (2017). "Combining Double Sampling and Bounds to Address Nonignorable Missing Outcomes in Randomized Experiments". In: *Political Analysis* 25.2, pp. 188–206.

Dasgupta, Tirthankar, Natesh S Pillai, and Donald B Rubin (2015). "Causal Inference from 2^K Factorial Designs by using Potential Outcomes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.4, pp. 727–753.

Ding, Peng (2017). "A Paradox from Randomization-Based Causal Inference". In: *Statistical Science* 32.3, pp. 331–345.

Ding, Peng and Fan Li (2019). "A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment". In: *Political Analysis* 27.4, pp. 605–615.

Ding, Peng and Luke Miratrix (2019). "Model-Free Causal Inference of Binary Experimental Data". In: *Scandinavian Journal of Statistics* 46.1, pp. 200–214.

Dunning, Thad (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York, NY: Cambridge University Press.

Egami, Naoki and Kosuke Imai (2019). "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis". In: *Journal of the American Statistical Association* 114.526, pp. 529–540.

Fisher, Ronald Aylmer (1935). *The Design of Experiments*. Edinburgh, SCT: Oliver and Boyd.

Fogarty, Colin B (2018). "On Mitigating the Analytical Limitations of Finely Stratified Experiments". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.5, pp. 1035–1056.

— (2020). "Studentized Sensitivity Analysis for the Sample Average Treatment Effect in Paired Observational Studies". In: *Journal of the American Statistical Association* 115.531, pp. 1518–1530.

— (2023). "Testing Weak Nulls in Matched Observational Studies". In: *Biometrics*.

Fogarty, Colin B et al. (2017). "Randomization Inference and Sensitivity Analysis for Composite Null Hypotheses With Binary Outcomes in Matched Observational Studies". In: *Journal of the American Statistical Association* 112.517, pp. 321–331.

Fox, John (2016). *Applied Regression Analysis and Generalized Linear Models*. 3rd. Los Angeles, CA: SAGE Publications.

Freedman, David A (2008a). "On Regression Adjustments in Experiments with Several Treatments". In: *The Annals of Applied Statistics* 2.1, pp. 176–196.

— (2008b). "On Regression Adjustments to Experimental Data". In: *Advances in Applied Mathematics* 40.2, pp. 180–193.

— (2008c). "Randomization Does Not Justify Logistic Regression". In: *Statistical Science* 23.2, pp. 237–249.

Freedman, David A, Robert Pisani, and Roger Purves (1998). *Statistics*. 3rd. New York, NY: W. W. Norton & Company.

Gelman, Andrew and Jennifer Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.

Gelman, Andrew and Guido W Imbens (2019). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs". In: *Journal of Business & Economic Statistics* 37.3, pp. 447–456.

Gerber, Alan S and Donald P Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W.W. Norton.

Gerber, Alan S, Donald P Green, et al. (2010). "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments". In: *Political Analysis* 18.3, pp. 297–315.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design". In: *Econometrica* 69.1, pp. 201–209.

Hainmueller, Jens (2012). "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." In: *Political Analysis* 20.1, pp. 25–46.

Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto (2014). "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments". In: *Political Analysis* 22.1, pp. 1–30.

Hansen, Ben B (2011). "Propensity Score Matching to Extract Latent Experiments from Nonexperimental Data: A Case Study". In: *Looking Back: Proceedings of a Conference in Honor of Paul*

W. Holland. Ed. by Neil J. Dorans and Sandip Sinharay. Vol. 202. Lecture Notes in Statistics. New York, NY: Springer. Chap. 9, pp. 149–181.
Hansen, Ben B and Jake Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". In: *Statistical Science* 23.2, pp. 219–236.
— (2009). "Attributing Effects to a Cluster-Randomized Get-Out-the-Vote Campaign". In: *Journal of the American Statistical Association* 104.487, pp. 873–885.
Hansen, Ben B, Paul R Rosenbaum, and Dylan S Small (2014). "Clustered Treatment Assignments and Sensitivity to Unmeasured Biases in Observational Studies". In: *Journal of the American Statistical Association* 109.505, pp. 133–144.
Hansen, Ben B and Adam Sales (2015). "Comment on Cochran's "Observational Studies"". In: *Observational Studies*, pp. 184–193.
Ho, Daniel E et al. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis* 15.3, pp. 199–236.
Holland, Paul W (1986). "Statistics and Causal Inference". In: *Journal of the American Statistical Association* 81.396, pp. 945–960.
Horowitz, Joel L and Charles F Manski (2000). "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data". In: *Journal of the American Statistical Association* 95.449, pp. 77–84.
Hosman, Carrie A, Ben B Hansen, and Paul W Holland (2010). "The Sensitivity of Linear Regression Coefficients' Confidence Limits to the Omission of a Confounder". In: *The Annals of Applied Statistics* 4.2, pp. 849–870.
Hsu, Jesse Y and Dylan S Small (2013). "Calibrating Sensitivity Analyses to Observed Covariates in Observational Studies". In: *Biometrics* 69.4, pp. 803–811.
Imai, Kosuke (2008). "Variance Identification and Efficiency Analysis in Randomized Experiments under the Matched-Pair Design". In: *Statistics in Medicine* 27.24, pp. 4857–4873.
Imai, Kosuke and In Song Kim (2021). "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data". In: *Political Analysis* 29.3, pp. 405–415.
Imai, Kosuke, Gary King, and Elizabeth A Stuart (2008). "Misunderstandings between Experimentalists and Observationalists about Causal Inference". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.2, pp. 481–502.
Imai, Kosuke and Marc Ratkovic (2014). "Covariate Balancing Propensity Score". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1, pp. 243–263.
Imbens, Guido W (2003). "Sensitivity to Exogeneity Assumptions in Program Evaluation". In: *The American Economic Review* 93.2, pp. 126–132.
Imbens, Guido W and Thomas Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". In: *Journal of Econometrics* 142.2, pp. 615–635.
Imbens, Guido W and Paul R Rosenbaum (2005). "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.1, pp. 109–126.
Imbens, Guido W and Donald B Rubin (1997). "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance". In: *The Annals of Statistics* 25.1, pp. 305–327.
— (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press.

Kang, Hyunseung, Laura Peck, and Luke Keele (2018). “Inference for Instrumental Variables: A Randomization Inference Approach”. In: *Journal of the Royal Statistical Society. Series A: Statistics in Society* 181.4, pp. 1231–1254.
Keele, Luke and Kevin M Quinn (2017). “Bayesian Sensitivity Analysis for Causal Effects from 2×2 Tables in the Presence of Unmeasured Confounding with Application to Presidential Campaign Visits”. In: *The Annals of Applied Statistics* 11.4, pp. 1974–1997.
Keele, Luke, Dylan Small, and Richard Grieve (2017). “Randomization-based Instrumental Variables Methods for Binary Outcomes with an Application to the ‘IMPROVE’ Trial”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180.2, pp. 569–586.
Keele, Luke, Rocío Titiunik, and José R Zubizarreta (2015). “Enhancing a Geographic Regression Discontinuity Design through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.1, pp. 223–239.
Kern, Holger L et al. (2016). “Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations”. In: *Journal of Research on Educational Effectiveness* 9.1, pp. 103–127.
Kilcioglu, Cinar and José R Zubizarreta (2016). “Maximizing the Information Content of a Balanced Matched Sample in a Study of the Economic Performance of Green Buildings”. In: *The Annals of Applied Statistics* 10.4, pp. 1997–2020.
Kinder, Donald R and Thomas R Palfrey (1993). “On Behalf of an Experimental Political Science”. In: *Experimental Foundations of Political Science*. Ed. by Donald R Kinder and Thomas R Palfrey. Michigan Studies in Political Analysis. Ann Arbor, MI: University of Michigan Press. Chap. 1, pp. 1–39.
Lechner, Michael (2011). “The Estimation of Causal Effects by Difference-in-Difference Methods”. In: *Foundations and Trends in Econometrics* 4.3, pp. 165–224.
Lee, David S (2008). “Randomized Experiments from Non-Random Selection in US House Elections”. In: *Journal of Econometrics* 142.2, pp. 675–697.
— (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects”. In: *The Review of Economic Studies* 76.3, pp. 1071–1102.
Li, Xinran and Peng Ding (2017). “General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference”. In: *Journal of the American Statistical Association* 112.520, pp. 1759–1769.
Li, Yunfei Paul, Kathleen J Propert, and Paul R Rosenbaum (2001). “Balanced Risk Set Matching”. In: *Journal of the American Statistical Association* 96.455, pp. 870–882.
Lin, Winston (2013). “Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique”. In: *The Annals of Applied Statistics* 7.1, pp. 295–318.
Lu, Bo (2005). “Propensity Score Matching with Time-Dependent Covariates”. In: *Biometrics* 61.3, pp. 721–728.
Lu, Bo, Robert Greevy, et al. (2011). “Optimal Nonbipartite Matching and Its Statistical Applications”. In: *The American Statistician* 65.1, pp. 21–30.
Lu, Bo, Elaine Zanutto, et al. (2001). “Matching with Doses in an Observational Study of a Media Campaign against Drug Abuse”. In: *Journal of the American Statistical Association* 96.456, pp. 1245–1253.
Manski, Charles F (2013). “Identification of Treatment Response with Social Interactions”. In: *The Econometrics Journal* 16.1, S1–S23.

Manski, Charles F and John V Pepper (2018). “How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions”. In: *The Review of Economics and Statistics* 100.2, pp. 232–244.
McCrary, Justin (2008). “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test”. In: *Journal of Econometrics* 142.2, pp. 698–714.
Middleton, Joel A and Peter M Aronow (2015). “Unbiased Estimation of the Average Treatment Effect in Cluster-Randomized Experiments”. In: *Statistics, Politics and Policy* 6.1-2, pp. 39–75.
Miratrix, Luke W, Jasjeet S Sekhon, Alexander G Theodoridis, et al. (2018). “Worth Weighting? How to Think About and Use Weights in Survey Experiments”. In: *Political Analysis* 26.3, pp. 275–291.
Miratrix, Luke W, Jasjeet S Sekhon, and Bin Yu (2013). “Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.2, pp. 369–396.
Oster, Emily (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence”. In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.
Pashley, Nicole E and Luke W Miratrix (2020). “Insights on Variance Estimation for Blocked and Matched Pairs Designs”. In: *Journal of Educational and Behavioral Statistics*.
Pimentel, Samuel D et al. (2018). “Optimal Multilevel Matching Using Network Flows: An Application to a Summer Reading Intervention”. In: *Annals of Applied Statistics* 12.3, pp. 1479–1505.
Rabb, Nathaniel et al. (July 2022). “The influence of social norms varies with “others” groups: Evidence from COVID-19 vaccination intentions”. In: *Proceedings of the National Academy of Sciences* 119.29. DOI: `https://doi.org/10.1073/pnas.2118770119`.
Robins, James M, Miguel Ángel Hernán, and Babette Brumback (2000). “Marginal Structural Models and Causal Inference in Epidemiology”. In: *Epidemiology* 11.5, pp. 550–560.
Rosenbaum, Paul R (1996). “Identification of Causal Effects Using Instrumental Variables: Comment”. In: *Journal of the American Statistical Association* 91.434, pp. 465–468.
— (2001a). “Effects Attributable to Treatment: Inference in Experiments and Observational Studies with a Discrete Pivot”. In: *Biometrika* 88.1, pp. 219–231.
— (2001b). “Observational Studies: Overview”. In: *International Encyclopedia of the Social & Behavioral Sciences*. Ed. by Neil J Smelser and Paul B Baltes. Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 10808–10815.
— (2002a). “Covariance Adjustment in Randomized Experiments and Observational Studies”. In: *Statistical Science* 17.3, pp. 286–327.
— (2002b). *Observational Studies*. Second. New York, NY: Springer.
— (2003). “Exact Confidence Intervals for Nonconstant Effects by Inverting the Signed Rank Test”. In: *The American Statistician* 57.2, pp. 132–138.
— (2004). “Design Sensitivity in Observational Studies”. In: *Biometrika* 91.1, pp. 153–164.
— (2007a). “Confidence intervals for uncommon but dramatic responses to treatment”. In: *Biometrics* 63.4, pp. 1164–1171.
— (2007b). “Interference Between Units in Randomized Experiments”. In: *Journal of the American Statistical Association* 102.477, pp. 191–200.
— (2010). *Design of Observational Studies*. New York, NY: Springer.

Rosenbaum, Paul R (2017). *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press.

— (2018). "Sensitivity Analysis for Stratified Comparisons in an Observational Study of the Effect of Smoking on Homocysteine Levels". In: *Annals of Applied Statistics* 12.4, pp. 2312–2334.

— (2020). "Modern Algorithms for Matching in Observational Studies". In: *Annual Review of Statistics and Its Application* 7.1, pp. 143–176.

Rosenbaum, Paul R and Abba M Krieger (1990). "Sensitivity of Two-Sample Permutation Inferences in Observational Studies". In: *Journal of the American Statistical Association* 85.410, pp. 493–498.

Rosenbaum, Paul R and Donald B Rubin (1985). "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score". In: *The American Statistician* 39.1, pp. 33–38.

Rubin, Donald B (1978). "Bayesian Inference for Causal Effects: The Role of Randomization". In: *The Annals of Statistics* 6.1, pp. 34–58.

— (1979). "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies". In: *Journal of the American Statistical Association* 74.366a, pp. 318–328.

Sales, Adam and Ben B Hansen (2020). "Limitless Regression Discontinuity". In: *Journal of Educational and Behavioral Statistics* 45.2, pp. 143–174.

Samii, Cyrus and Peter M Aronow (2012). "On Equivalencies between Design-based and Regression-based Variance Estimators for Randomized Experiments". In: *Statistics & Probability Letters* 82.2, pp. 365–370.

Sävje, Fredrik, Michael J Higgins, and Jasjeet S Sekhon (2021). "Generalized Full Matching". In: *Political Analysis*.

Sekhon, Jasjeet and Rocío Titiunik (2017). "On Interpreting the Regression Discontinuity Design as a Local Experiment". In: *Regression Discontinuity Designs: Theory and Applications*. Ed. by Matias D. Cattaneo and Juan Carlos Escanciano. Vol. 38. Advances in Econometrics. Bingley, UK: Emerald Group Publishing. Chap. 1.

Sekhon, Jasjeet S and Yotam Shem-Tov (2020). "Inference on a New Class of Sample Average Treatment Effects". In: *Journal of the American Statistical Association* 116.534, pp. 798–804.

Sekhon, Jasjeet S and Rocío Titiunik (2016). "Understanding Regression Discontinuity Designs As Observational Studies". In: *Observational Studies* 2, pp. 174–182.

Silber, Jeffrey H et al. (2014). "Template Matching for Auditing Hospital Cost and Quality". In: *Health Services Research* 48.5, pp. 1446–1474.

Splawa-Neyman, Jerzy, D M Dabrowska, T P Speed, et al. (1990). "On the application of probability theory to agricultural experiments. Essay on principles. Section 9". In: *Statistical Science* 5.4, pp. 465–472.

Tan, Zhiqiang (2020). "Regularized Calibrated Estimation of Propensity Scores with Model Misspecification and High-dimensional Data". In: *Biometrika* 107.1, pp. 137–158.

Wang, Yixin and José R Zubizarreta (2020). "Minimal Dispersion Approximately Balancing Weights: Asymptotic Properties and Practical Considerations". In: *Biometrika* 107.1, pp. 93–105.

Westreich, Daniel et al. (2019). "Target Validity and the Hierarchy of Study Designs". In: *American Journal of Epidemiology* 188.2, pp. 438–443.

Wickham, Hadley and Garrett Grolemund (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. First. Sebastopol, CA: O'Reilly Media.

Wong, Raymond K W and Kwun Chuen Gary Chan (2018). “Kernel-based Covariate Functional Balancing for Observational Studies”. In: *Biometrika* 105.1, pp. 199–213.

Yang, Shu et al. (2016). “Propensity Score Matching and Subclassification in Observational Studies with Multi-level Treatments”. In: *Biometrics* 72.4, pp. 1055–1065.

Zubizarreta, José R (2015). “Stable Weights that Balance Covariates for Estimation with Incomplete Outcome Data”. In: *Journal of the American Statistical Association* 110.511, pp. 910–922.

Zubizarreta, José R, Magdalena Cerdá, and Paul R Rosenbaum (2013). “Effect of the 2010 Chilean Earthquake on Posttraumatic Stress Reducing Sensitivity to Unmeasured Bias Through Study Design”. In: *Epidemiology* 24.1, pp. 79–87.

Zubizarreta, José R and Luke Keele (2017). “Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the Effectiveness of Private Schools Under a Large-Scale Voucher System”. In: *Journal of the American Statistical Association* 112.518, pp. 547–560.