Causal Inference for the Social Sciences

Jake Bowers[1] Thomas (Tom) Leavitt[2] Ben Hansen[3]

[Session Dates Redacted]

[1]Departments of Political Science and Statistics, University of Illinois at Urbana–Champaign; [Email Redacted]

[2]Marxe School of Public and International Affairs, Baruch College, City University of New York (CUNY); [Email Redacted]

[3]Department of Statistics, University of Michigan; [Email Redacted]

Class meets in [Classroom Redacted].

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.

We assume familiarity with linear algebra and strong knowledge of statistical concepts, such as sampling distributions, statistical inference, and hypothesis testing. Demonstrations, examples and assignments make extensive use of R.

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 effects to randomized experiments, touching on considerations specific to clustered treatment assignment, "small" sample sizes and treatment effect heterogeneity. The next segment addresses conceptual and methodological 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

Peng Ding (2024). *A First Course in Causal Inference.* Chapman and Hall/CRC

Paul R. Rosenbaum (2025). *An Introduction to the Theory of Observational Studies.* Springer Texts in Statistics. New York, NY: Springer

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 a 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 question;
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 schedule appears below; the Course content section gives fuller readings for each topic, plus the menu of student-chosen special topics for the final day.

Table 1: Course schedule by date, instructor, topic, readings, and application

Date	Instructor	Topic	Required readings	Application
Mon, June 15	Tom	Introduction: Random allocation, potential outcomes, and Fisher's exact test	Holland (1986, § 1–4) Rosenbaum (2017, Chapter 2)	Fisher (1935)
Tues, June 16	Tom	The sharp framework: Fisherian inference	Rosenbaum (2017, Chapter 3) Fisher (1935, § 5–10)	Arceneaux (2005)
Wed, June 17	Tom	Weak framework I: Estimating average effects and the variance of the difference-in-means	Gerber and Green (2012, Ch. 2 and pp. 51–61) Aronow and Middleton (2013) Middleton and Aronow (2015) Freedman et al. (1998) Imbens and Rubin (2015, pp. 87–98)	
Thurs, June 18	Tom	Weak framework II: Conservative variance estimation, hypothesis testing, and covariance adjustment	Gerber and Green (2012, Ch. 3–4) Rosenbaum (2002); Lin (2013)	
Fri, June 19 (HW 1 due)	–	No Class: Juneteenth U.S. Holiday		
Mon, June 22	Tom	Noncompliance and attrition	Gerber and Green (2012, Ch. 5–7) Rosenbaum (1996) Rosenbaum (2010, § 5.3)	Albertson and Lawrence (2009)
Tues, June 23	Tom	Observational studies and as-if randomization	Leavitt and Miratrix (2026, § 1) Rosenbaum (2017, Chapter 5) Gelman and Hill (2006, § 9.0–9.2) Berk (2010)	Gilligan and Sergenti (2008); Cerdá et al. (2012)
Wed, June 24	Tom	Building a matched design: Distances and forming matches	Leavitt and Miratrix (2026, § 2.1–2.2.4) Rosenbaum (2017, pp. 65–96) Rosenbaum (2010, Ch. 7–8) Hansen (2011)	
Thurs, June 25	Tom	Evaluating a matched design: Balance and effective sample size	Leavitt and Miratrix (2026, § 2.2.5) Hansen and Bowers (2008)	
Fri, June 26 (HW 2 due)	Tom	Inference under as-if randomization I: The sharp framework	Leavitt and Miratrix (2026, § 2.3–2.3.1) Rosenbaum (2017, Chapter 3)	
Mon, June 29	Jake	Inference under as-if randomization II: The weak framework	Leavitt and Miratrix (2026, § 2.3.2) Gerber and Green (2012, pp. 71–79) Fogarty (2018); Pashley and Miratrix (2020)	
Tues, June 30	Jake	Sensitivity analysis I: The sharp framework	Leavitt and Miratrix (2026, § 2.4–2.4.1) Rosenbaum (2017, Chapter 9) Rosenbaum (2018)	
Wed, July 1	Jake	Sensitivity analysis II: The weak framework	Leavitt and Miratrix (2026, § 2.4.2) Fogarty (2023)	
Thurs, July 2 (HW 3 due)	Jake	Full pipeline end-to-end and/or special topics (students' choice)	Leavitt and Miratrix (2026)	

Course content

1 Randomized experiments

1.1 Introduction: Random allocation, potential outcomes and Fisher's exact test

Required

Paul W. Holland (1986). "Statistics and Causal Inference". In: *Journal of the American Statistical Association* 81.396, pp. 945–960, Sections 1 – 4.

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

Recommended

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.)

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.

Jersey Neyman (1923). "Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes". In: *Roczniki Nauk Rolniczych* 10, pp. 1–51

Donald B. Rubin (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66.5, p. 688

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

1.2 The sharp framework: Fisherian inference for causal effects

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

Required

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

Sections 5 – 10 (pp. 11 – 19) of Ronald Aylmer Fisher (1935). *The Design of Experiments.* Edinburgh, SCT: Oliver and Boyd.

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.

Devin Caughey, Allan Dafoe, Xinran Li, and Luke W. Miratrix (2023). "Randomisation Inference Beyond the Sharp Null: Bounded Null Hypotheses and Quantiles of Individual Treatment Effects". In: *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 85.5, pp. 1471–1491

David Kim, Yongchang Su, Jake Bowers, and Xinran Li (2026). "Randomization Tests for Distributions of Individual Treatment Effects via Combined Rank Statistics". ArXiv preprint, arXiv:2605.08027

Andrew Gelman (2003). "A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-fit Testing". In: *International Statistical Review* 71.2, pp. 369–382

1.3 The weak framework I: Estimating average effects and the variance of the difference-in-means

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

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: Jersey Neyman (1923). "Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes". In: *Roczniki Nauk Rolniczych* 10, pp. 1–51.)

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

1.4 The weak framework II: Conservative 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.

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

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

James M. Robins (1988). "Confidence Intervals for Causal Parameters". In: *Statistics in Medicine* 7.7, pp. 773–785

Christopher Harshaw, Joel A. Middleton, and Fredrik Sävje (2026). "Optimized Variance Estimation under Interference and Complex Experimental Designs". In: *Journal of the American Statistical Association*

1.5 Covariance adjustment (sharp and weak frameworks)

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, Guido W. Imbens, and Jeffrey M. Wooldridge (2020). “Sampling-Based versus Design-Based Uncertainty in Regression Analysis”. In: *Econometrica* 88.1, pp. 265–296

Colin B. Fogarty (2018b). “Regression-assisted Inference for the Average Treatment Effect in Paired Experiments”. In: *Biometrika* 105.4, pp. 994–1000

Kevin Guo and Guillaume W. Basse (2023). “The Generalized Oaxaca-Blinder Estimator”. In: *Journal of the American Statistical Association* 118.541, pp. 524–536

Peter L. Cohen and Colin B. Fogarty (2023). “No-Harm Calibration for Generalized Oaxaca-Blinder Estimators”. In: *Biometrika* 111.1, pp. 331–338

Haoge Chang, Joel A. Middleton, and P. M. Aronow (2024). “Exact Bias Correction for Linear Adjustment of Randomized Controlled Trials”. In: *Econometrica* 92, pp. 1503–1519

1.6 Noncompliance and attrition

Application

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

1.6.1 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

P. M. Aronow, Haoge Chang, and Patrick Lopatto (2025). "Randomization-Based Confidence Sets for the Local Average Treatment Effect". In: *Biometrika.* In press

Nicole E. Pashley (2022). "Note on the Delta Method for Finite Population Inference with Applications to Causal Inference". In: *Statistics & Probability Letters* 188, p. 109540

Nicole E. Pashley, Luke Keele, and Luke W. Miratrix (2024). "Improving Instrumental Variable Estimators with Poststratification". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 188.3, pp. 765–790

Matthew Blackwell and Nicole E. Pashley (2021). "Noncompliance and Instrumental Variables for 2^K Factorial Experiments". In: *Journal of the American Statistical Association* 118.542, pp. 1102–1114

1.6.2 Attrition, or missing outcomes

Required

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

Recommended

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, Alan S. Gerber, Donald P. Green, and Holger L. Kern (2017). "Combining Double Sampling and Bounds to Address Nonignorable Missing Outcomes in Randomized Experiments". In: *Political Analysis* 25.2, pp. 188–206

Xinran Li, Peizan Sheng, and Zeyang Yu (2025). "Randomization Inference with Sample Attrition". ArXiv preprint, arXiv:2507.00795

Haoge Chang and Zeyang Yu (2026). "Randomization Inference For the Always-Reporter Average Treatment Effect". ArXiv preprint, arXiv:2603.24970

2 Observational studies

2.1 Observational studies and as-if randomization

Required

Section 1, "Design-Based Foundations of the Matching Pipeline," of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

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

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

Recommended

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

Donald B. Rubin (1977). "Assignment to Treatment Group on the Basis of a Covariate". In: *Journal of Educational Statistics* 2.1, pp. 1–26

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

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

2.2 Building a matched design: Measuring similarity and forming matches

Application

Michael J. Gilligan and Ernest J. Sergenti (2008). "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference". In: *Quarterly Journal of Political Science* 3.2, pp. 89–122

Magdalena Cerdá, Jeffrey D. Morenoff, Ben B. Hansen, Kimberly J. Tessari Hicks, Luis F. Duque, Alexandra Restrepo, and Ana V. Diez-Roux (2012). "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". In: *American Journal of Epidemiology* 175.10, pp. 1045–1053

Required

Sections 2.1 – 2.2.4 (running example through forming the matches) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

Pages 65 – 96 of Paul R. Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press

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

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

Ben B. Hansen (2004). "Full Matching in an Observational Study of Coaching for the SAT". in: *Journal of the American Statistical Association* 99.467, pp. 609–618

Ben B. Hansen and Stephanie Olsen Klopfer (2006). "Optimal Full Matching and Related Designs via Network Flows". In: *Journal of Computational and Graphical Statistics* 15.3, pp. 609–627

Xing Sam Gu and Paul R. Rosenbaum (1993). "Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms". In: *Journal of Computational and Graphical Statistics* 2.4, pp. 405–420

Donald B. Rubin and Richard P. Waterman (2006). "Estimating the Causal Effects of Marketing Interventions Using Propensity Score Methodology". In: *Statistical Science* 21.2, pp. 206–222

Ben B. Hansen (2008b). "The Prognostic Analogue of the Propensity Score". In: *Biometrika* 95.2, pp. 481–488

Adam C. Sales, Ben B. Hansen, and Brian Rowan (2018). "Rebar: Reinforcing a Matching Estimator With Predictions From High-Dimensional Covariates". In: *Journal of Educational and Behavioral Statistics* 43.1, pp. 3–31

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

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

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, Kosuke Imai, Gary King, and Elizabeth A. Stuart (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis* 15.3, pp. 199–236

Chapter 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

2.3 Evaluating a matched design: Balance and effective sample size

Required

Section 2.2.5 (balance and effective sample size) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

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

Recommended

Peter C. Austin (2009). "Balance Diagnostics for Comapring the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples". In: *Statistics in Medicine* 28.25, pp. 3083–3107

Elizabeth A. Stuart (2010). "Matching Methods for Causal Inference: A Review and a Look Forward". In: *Statistical Science* 25.1, p. 1

Donald B. Rubin (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation". In: *Health Services and Outcomes Research Methodology* 2.3, pp. 169–188

Peter C. Austin (2008). "A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003". In: *Statistics in Medicine* 27.12, pp. 2037–2049

Ben B. Hansen (2008a). "The Essential Role of Balance Tests in Propensity-Matched Observational Studies: Comments on 'A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003' by Peter Austin". In: *Statistics in Medicine* 27.12, pp. 2050–2054

Sture Holm (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70

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

2.4 Inference under as-if randomization I: The sharp framework

Required

Sections 2.3 – 2.3.1 (inference under the sharp framework) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

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

Recommended

Nicole E. Pashley, Guillaume W. Basse, and Luke W. Miratrix (2021). "Conditional as-if analyses in randomized experiments". In: *Journal of Causal Inference* 9.1

2.5 Inference under as-if randomization II: The weak framework

Required

Section 2.3.2 (inference under the weak framework) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

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

Colin B. Fogarty (2018a). "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*

Recommended

Fan Li, Kari Lock Morgan, and Alan M. Zaslavsky (2018). "Balancing Covariates via Propensity Score Weighting". In: *Journal of the American Statistical Association* 113.521, pp. 390–400

Nicole E. Pashley, Guillaume W. Basse, and Luke W. Miratrix (2021). "Conditional as-if analyses in randomized experiments". In: *Journal of Causal Inference* 9.1

3 Sensitivity analysis

Application

Michael J. Gilligan and Ernest J. Sergenti (2008). "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference". In: *Quarterly Journal of Political Science* 3.2, pp. 89–122

Magdalena Cerdá, Jeffrey D. Morenoff, Ben B. Hansen, Kimberly J. Tessari Hicks, Luis F. Duque, Alexandra Restrepo, and Ana V. Diez-Roux (2012). "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". In: *American Journal of Epidemiology* 175.10, pp. 1045–1053

3.1 Sensitivity analysis I: The sharp framework

Required

Sections 2.4 – 2.4.1 (sensitivity analysis under the sharp framework) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

Chapter 9 of Paul R. Rosenbaum (2017). *Observation and Experiment: An Introduction to Causal Inference.* Cambridge, MA: Harvard University Press

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

Recommended

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

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

Joseph L. Gastwirth, Abba M. Krieger, and Paul R. Rosenbaum (2000). "Asymptotic separability in sensitivity analysis". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 545–555

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

3.2 Sensitivity analysis II: The weak framework

Required

Section 2.4.2 (sensitivity analysis under the weak framework) of Thomas Leavitt and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*

Colin B. Fogarty (2023). "Testing Weak Nulls in Matched Observational Studies". In: *Biometrics* 79.3, pp. 2196–2207

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, Pixu Shi, Mark E. Mikkelsen, and Dylan S. Small (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

4 Additional topics

4.1 Design sensitivity

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

Ruth Heller, Paul R. Rosenbaum, and Dylan S. Small (2009). "Split Samples and Design Sensitivity in Observational Studies". In: *Journal of the American Statistical Association* 104.487, pp. 1090–1101

Jesse Y. Hsu, Dylan S. Small, and Paul R. Rosenbaum (2013). "Effect Modification and Design Sensitivity in Observational Studies". In: *Journal of the American Statistical Association* 108.501, pp. 135–148

Dylan S. Small, Jing Cheng, M. Elizabeth Halloran, and Paul R. Rosenbaum (2013). "Case Definition and Design Sensitivity". In: *Journal of the American Statistical Association* 108.504, pp. 1457–1468

4.2 Special topics in matching: nonbipartite, multilevel, risk-set, cardinality, coarsened exact, template, exterior, generalized full, and fine balance

Bo Lu, Elaine Zanutto, Robert Hornik, and Paul R. Rosenbaum (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

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

Pages 207 – 221 of Paul R. Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

Jeffrey H. Silber, Paul R. Rosenbaum, Richard N. Ross, Justin M. Ludwig, Wei Wang, Bijan A. Niknam, Nabanita Mukherjee, Philip A. Saynisch, Orit Even-Shoshan, Rachel R. Kelz, and Lee A. Fleisher (2014). "Template Matching for Auditing Hospital Cost and Quality". In: *Health Services Research* 48.5, pp. 1446–1474

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

Samuel D. Pimentel, Lindsay C. Page, Matthew Lenard, and Luke Keele (2018). "Optimal Multilevel Matching Using Network Flows: An Application to a Summer Reading Intervention". In: *Annals of Applied Statistics* 12.3, pp. 1479–1505

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

José R. Zubizarreta, Ricardo D. Paredes, and Paul R. Rosenbaum (2014). "Matching for balance, pairing for heterogeneity in an observational study of the effectiveness of for-profit and not-for-profit high schools in Chile". In: *The Annals of Applied Statistics* 8.1, pp. 204–231

Stefano M. Iacus, Gary King, and Giuseppe Porro (2012). "Causal Inference without Balance Checking: Coarsened Exact Matching". In: *Political Analysis* 20.1, pp. 1–24

Paul R. Rosenbaum and Jeffrey H. Silber (2013). "Using the Exterior Match to Compare Two Entwined Matched Control Groups". In: *The American Statistician* 63.2, pp. 67–75

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

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

Shoshana R. Daniel, Katrina Armstrong, Jeffrey H. Silber, and Paul R. Rosenbaum (2008). "An Algorithm for Optimal Tapered Matching, With Application to Disparities in Survival". In: *Journal of Computational and Graphical Statistics* 17.4, pp. 914–924

Ruoqi Yu, Dylan S. Small, and Paul R. Rosenbaum (2021). "The Information in Covariate Imbalance in Studies of Hormone Replacement Therapy". In: *The Annals of Applied Statistics* 15.4, pp. 2023–2042

Paul R. Rosenbaum, Richard N. Ross, and Jeffrey H. Silber (2007). "Minimum Distance Matched Sampling With Fine Balance in an Observational Study of Treatment for Ovarian Cancer". In: *Journal of the American Statistical Association* 102.477, pp. 75–83

Dan Yang, Dylan S. Small, Jeffrey H. Silber, and Paul R. Rosenbaum (2012). "Optimal Matching with Minimal Deviation from Fine Balance in a Study of Obesity and Surgical Outcomes". In: *Biometrics* 68.2, pp. 628–636

Ruoqi Yu (2023). "How Well Can Fine Balance Work for Covariate Balancing". In: *Biometrics* 79.3, pp. 2346–2356

Samuel D. Pimentel, Rachel R. Kelz, Jeffrey H. Silber, and Paul R. Rosenbaum (2015). "Large, Sparse Optimal Matching With Refined Covariate Balance in an Observational Study of the Health Outcomes Produced by New Surgeons". In: *Journal of the American Statistical Association* 110.510, pp. 515–527

Samuel D. Pimentel, Frank Yoon, and Luke Keele (2015). "Variable-ratio Matching with Fine Balance in a Study of the Peer Health Exchange". In: *Statistics in Medicine* 34.30, pp. 4070–4082

4.3 Extensions to covariate balance testing

Johann Gagnon-Bartsch and Yotam Shem-Tov (2019). "The Classification Permutation Test: A Flexible Approach to Testing for Covariate Imbalance in Observational Studies". In: *The Annals of Applied Statistics* 13.3, pp. 1464–1483

Zach Branson (2021). "Randomization Tests to Assess Covariate Balance When Designing and Analyzing Matched Datasets". In: *Observational Studies* 7.2, pp. 1–36

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

4.4 Residual imbalance on observed covariates

Paul R. Rosenbaum (1988). "Sensitivity Analysis for Matching with Multiple Controls". In: *Biometrika* 75.3, pp. 577–581

Samuel D. Pimentel and Yaxuan Huang (2024). "Covariate-adaptive Randomization Inference in Matched Designs". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.5, pp. 1312–1338

Kan Chen, Siyu Heng, Qi Long, and Bo Zhang (2023). "Testing Biased Randomization Assumptions and Quantifying Imperfect Matching and Residual Confounding in Matched Observational Studies". In: *Journal of Computational and Graphical Statistics* 32.2, pp. 528–538

Siyu Heng, Yanxin Shen, and Pengyun Wang (2025). "Reconciling Overt Bias and Hidden Bias in Sensitivity Analysis for Matched Observational Studies". ArXiv preprint, arXiv:2311.11216

Jianan Zhu, Jeffrey Zhang, Zijian Guo, and Siyu Heng (2025). "Randomization-Based

Inference for Average Treatment Effects in Inexactly Matched Observational Studies". ArXiv preprint, arXiv:2308.02005

4.5 Defining an interpretable study population

Mikhail Traskin and Dylan S. Small (2011). "Defining the Study Population for an Observational Study to Ensure Sufficient Overlap: A Tree Approach". In: *Statistics in Biosciences* 3, pp. 94–118

Colin B. Fogarty, Mark E. Mikkelsen, David F. Gaieski, and Dylan S. Small (2016). "Discrete Optimization for Interpretable Study Populations and Randomization Inference in an Observational Study of Severe Sepsis Mortality". In: *Journal of the American Statistical Association* 111.514, pp. 447–458

Richard K. Crump, V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik (2009). "Dealing with limited overlap in estimation of average treatment effects". In: *Biometrika* 96.1, pp. 187–199

Gary King and Langche Zeng (2006). "The Dangers of Extreme Counterfactuals". In: *Political Analysis* 14.2, pp. 131–159

4.6 Rerandomization

Kari Lock Morgan and Donald B. Rubin (2012). "Rerandomization to Improve Covariate Balance in Experiments". In: *Annals of Statistics* 40.2, pp. 1263–1282

Kari Lock Morgan and Donald B. Rubin (2015). "Rerandomization to Balance Tiers of Covariates". In: *Journal of the American Statistical Association* 110.512, pp. 1412–1421

Xinran Li, Peng Ding, and Donald B. Rubin (2018). "Asymptotic Theory of Rerandomization in Treatment–Control Experiments". In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9157–9162

Xinran Li and Peng Ding (2020). "Rerandomization and Regression Adjustment". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1, pp. 241–268

Zach Branson, Xinran Li, and Peng Ding (2023). "Power and Sample Size Calculations for Rerandomization". In: *Biometrika* 111.1, pp. 355–363

Xinhe Wang, Tingyu Wang, and Hanzhong Liu (2021). "Rerandomization in Stratified Randomized Experiments". In: *Journal of the American Statistical Association* 118.542, pp. 1295–1304

Per Johansson, Donald B. Rubin, and Mårten Schultzberg (2021). "On Optimal Rerandomization Designs". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.2, pp. 395–403

Xin Lu, Tianle Liu, Hanzhong Liu, and Peng Ding (2023). "Design-Based Theory for Cluster Rerandomization". In: *Biometrika* 110.2, pp. 467–483

Max Cytrynbaum (2024). “Finely Stratified Rerandomization Designs”. ArXiv preprint, arXiv:2407.03279

Antônio Carlos Herling Ribeiro Junior and Zach Branson (2025). “Does Rerandomization Help Beyond Covariate Adjustment? A Review and Guide for Theory and Practice”. ArXiv preprint, arXiv:2512.05290

4.7 Randomization versus optimal design

Jack Kiefer (1959). “Optimum Experimental Designs”. In: *Journal of the Royal Statistical Society Series B: Methodological* 21.2, pp. 272–319

Valerii V. Fedorov (1972). *Theory of Optimal Experiments*. New York: Academic Press

David A. Harville (1975). “Experimental Randomization: Who Needs It?” In: *The American Statistician* 29.1, pp. 27–31

Debabrata Basu (1980). “Randomization Analysis of Experimental Data: The Fisher Randomization Test — Rejoinder”. In: *Journal of the American Statistical Association* 75.371, pp. 593–595

Chien-Fu Wu (1981). “On the Robustness and Efficiency of Some Randomized Designs”. In: *The Annals of Statistics* 9.6, pp. 1168–1177

Stephen Senn (1994). “Fisher’s Game with the Devil”. In: *Statistics in Medicine* 13.3, pp. 217–230

Stephen Senn (2013). “Seven Myths of Randomisation in Clinical Trials”. In: *Statistics in Medicine* 32.9, pp. 1439–1450

Dimitris Bertsimas, Mac Johnson, and Nathan Kallus (2015). “The Power of Optimization over Randomization in Designing Experiments Involving Small Samples”. In: *Operations Research* 63.4, pp. 868–876

Maximilian Kasy (2016). “Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead”. In: *Political Analysis* 24.3, pp. 324–338

Abhijit Banerjee, Sylvain Chassang, and Erik Snowberg (2017). “Decision Theoretic Approaches to Experiment Design and External Validity”. In: *Handbook of Field Experiments*. Ed. by Esther Duflo and Abhijit Banerjee. Vol. 1. Amsterdam, NL: North-Holland. Chap. 4, pp. 141–174

Nathan Kallus (2018). “Optimal A Priori Balance in the Design of Controlled Experiments”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.1, pp. 85–112

Abhijit Banerjee, Sylvain Chassang, Sergio Montero, and Erik Snowberg (2020). “A Theory of Experimenters: Robustness, Randomization, and Balance”. In: *American Economic Review* 110.4, pp. 1206–1230

Nathan Kallus (2021). “On the Optimality of Randomization in Experimental Design: How to Randomize for Minimax Variance and Design-Based Inference”. In: *Jour-*

nal of the Royal Statistical Society Series B: Statistical Methodology 83.2, pp. 404–409

Adam Kapelner, Abba M. Krieger, Michael Sklar, Uri Shalit, and David Azriel (2021). "Harmonizing Optimized Designs with Classic Randomization in Experiments". In: *The American Statistician* 75.2, pp. 195–206

Adam Kapelner, Abba M. Krieger, Michael Sklar, and David Azriel (2022). "Optimal Rerandomization Designs via a Criterion that Provides Insurance against Failed Experiments". In: *Journal of Statistical Planning and Inference* 219, pp. 63–84

Yuehao Bai (2023). "Why Randomize? Minimax Optimality under Permutation Invariance". In: *Journal of Econometrics* 232.2, pp. 565–575

Christopher Harshaw, Fredrik Sävje, Daniel A. Spielman, and Peng Zhang (2024). "Balancing Covariates in Randomized Experiments with the Gram–Schmidt Walk Design". In: *Journal of the American Statistical Association* 119.548, pp. 2934–2946

Marco Martinez and David Teira (2024). "Why Experimental Balance Is Still a Reason to Randomize". In: *The British Journal for the Philosophy of Science* 75.2, pp. 519–535

4.8 Quasi-experimental devices

Donald T. Campbell (1957). "Factors Relevant to the Validity of Experiments in Social Settings". In: *Psychological Bulletin* 54.4, pp. 297–312

Donald T. Campbell and Julian C. Stanley (1963). *Experimental and Quasi-Experimental Designs for Research.* Boston, MA: Houghton Mifflin Company

William R. Shadish, Thomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton Mifflin Company

Section IV of Paul R. Rosenbaum (2025). *An Introduction to the Theory of Observational Studies.* Springer Texts in Statistics. New York, NY: Springer

4.8.1 Difference-in-Differences

Dae Woong Ham and Luke W. Miratrix (2024). "Benefits and Costs of Matching Prior to a Difference in Differences Analysis When Parallel Trends Does Not Hold". In: *Annals of Applied Statistics* 18.3, pp. 2096–2122

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

Brantly Callaway and Pedro H. C. Sant’Anna (2021). “Difference-in-Differences with Multiple Time Periods”. In: *Journal of Econometrics* 225.2, pp. 200–230

Ashesh Rambachan and Jonathan Roth (2023). “A More Credible Approach to Parallel Trends”. In: *Review of Economic Studies* 90.5, pp. 2555–2591

Jonathan Roth (2022). “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends”. In: *American Economic Review: Insights* 4.3, pp. 305–322

Jonathan Roth and Pedro H. C. Sant’Anna (2023). “When Is Parallel Trends Sensitive to Functional Form?” In: *Econometrica* 91.2, pp. 737–747

Jonathan Roth, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe (2023). “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature”. In: *Journal of Econometrics* 235.2, pp. 2218–2244

Andrew Baker, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant’Anna (2026). “Difference-in-Differences Designs: A Practitioner’s Guide”. In: *Journal of Economic Literature* 64.2, pp. 498–557

Jonas M. Mikhaeil and Christopher Harshaw (2025). “Valid Inference when Testing Violations of Parallel Trends for Difference-in-Differences”. ArXiv preprint, arXiv:2510.26470

Thomas Leavitt and Laura A. Hatfield (2025). “Averaged Prediction Models (APM): Identifying Causal Effects in Controlled Pre-Post Settings with Application to Gun Policy”. In: *Annals of Applied Statistics* 19.3

Thomas Leavitt (2026). “Beyond Pre-Trends: A Discordance-Based Sensitivity Analysis for Difference-in-Differences”. Accepted at *Observational Studies*

4.8.2 Regression discontinuity designs

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 C. 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 S. 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

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

4.9 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

Fredrik Sävje, P. M. Aronow, and Michael G. Hudgens (2021). "Average Treatment Effects in the Presence of Unknown Interference". In: *Annals of Statistics* 49.2, pp. 673–701

Ye Wang, Cyrus Samii, Haoge Chang, and P. M. Aronow (2025). "Design-Based Inference for Spatial Experiments with Interference". In: *Annals of Applied Statistics* 19.1, pp. 744–768

Vardis Kandiros, Charilaos Pipis, Constantinos Daskalakis, and Christopher Harshaw (2024). "The Conflict Graph Design: Estimating Causal Effects under Arbitrary Neighborhood Interference". ArXiv preprint, arXiv:2411.10908

Jake Bowers, Bruce A. Desmarais, Mark Frederickson, Nahomi Ichino, Hsuan-Wei Lee, and Simi Wang (2018). "Models, Methods and Network Topology: Experimental Design for the Study of Interference". In: *Social Networks* 54, pp. 196–208

Christopher Harshaw, Joel A. Middleton, and Fredrik Sävje (2026). "Optimized Variance Estimation under Interference and Complex Experimental Designs". In: *Journal of the American Statistical Association*

4.10 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

Christopher Harshaw, Joel A. Middleton, and Fredrik Sävje (2026). "Optimized Variance Estimation under Interference and Complex Experimental Designs". In: *Journal of the American Statistical Association*

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, Edward H. Kaplan, and Holger L. Kern (2010). "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments". In: *Political Analysis* 18.3, pp. 297–315

Matthew Blackwell and Nicole E. Pashley (2026). "Bounds on Causal Effects in 2^K Factorial Experiments with Noncompliance". In: *Biometrika* 113.1

Max Goplerud, Kosuke Imai, and Nicole E. Pashley (2025). "Estimating Heterogeneous Causal Effects of High-Dimensional Treatments: Application to Conjoint Analysis". In: *Annals of Applied Statistics* 19.2

4.11 Joint inference of sharp and weak causal hypotheses

Eun Yi Chung and Joseph P. Romano (2013). "Exact and Asymptotically Robust Permutation Tests". In: *The Annals of Statistics* 41.2, pp. 484–507

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

Jason Wu and Peng Ding (2021). "Randomization Tests for Weak Null Hypotheses". In: *Journal of the American Statistical Association* 116.536, pp. 1898–1913

Peter L. Cohen and Colin B. Fogarty (2022). "Gaussian Prepivoting for Finite Population Causal Inference". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84.2, pp. 295–320

4.12 Attributable effects, the ATT, and alternative effect models

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 S. 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

Paul R. Rosenbaum (1999). "Reduced Sensitivity to Hidden Bias at Upper Quantiles in Observational Studies with Dilated Treatment Effects". In: *Biometrics* 55.2, pp. 560–564

Pages 46 – 49 of Paul R. Rosenbaum (2010). *Design of Observational Studies.* New York, NY: Springer

4.13 Sensitivity analysis: regression-based approaches and recent extensions

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

Dongxiao Wu and Xinran Li (2025). "Sensitivity Analysis for Quantiles of Hidden Biases in Matched Observational Studies". In: *Journal of the American Statistical Association* 120.551, pp. 1657–1668

Xinran Li (2025). "Sensitivity Analysis for Observational Studies with Flexible Matched Designs". In: *Biometrika* 112.4

Colin B. Fogarty (2025). "Tilted Sensitivity Analysis in Matched Observational Studies". ArXiv preprint, arXiv:2503.09736

William Bekerman, Anurag Mehta, Rebecca E. Hasson, Leah E. Robinson, Dylan S. Small, and Colin B. Fogarty (2026). "Powerful Multivariate Sensitivity Analysis via Sample Splitting in an Observational Study of the Effects of Poverty on Cardiovascular Disease Risk Factors". ArXiv preprint, arXiv:2606.04416

Mengqi Lin, Colin B. Fogarty, and Gongjun Xu (2026). "Stochastic Sensitivity Analysis for Matched Observational Studies". ArXiv preprint, arXiv:2606.05120

4.14 External validity

Holger L. Kern, Elizabeth A. Stuart, Jennifer Hill, and Donald P. Green (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, and Luis F. Campos (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, Paul R. Rosenbaum, Richard N. Ross, Justin M. Ludwig, Wei Wang, Bijan A. Niknam, Nabanita Mukherjee, Philip A. Saynisch, Orit Even-Shoshan, Rachel R. Kelz, and Lee A. Fleisher (2014). "Template Matching for Auditing Hospital Cost and Quality". In: *Health Services Research* 48.5, pp. 1446–1474

Daniel Westreich, Jessie K. Edwards, Catherine R. Lesko, Stephen R. Cole, and Elizabeth A. Stuart (2019). "Target Validity and the Hierarchy of Study Designs". In: *American Journal of Epidemiology* 188.2, pp. 438–443

Naoki Egami and Erin Hartman (2022). "Elements of External Validity: Framework, Design, and Analysis". In: *American Political Science Review* 117.3, pp. 1070–1088

4.15 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

4.16 Synthetic control

Alberto Abadie and Javier Gardeazabal (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country”. In: *The American Economic Review* 93.1, pp. 113–132

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

Alberto Abadie (2021). “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects”. In: *Journal of Economic Literature* 59.2, pp. 391–425

Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021). “Synthetic Difference-in-Differences”. In: *American Economic Review* 111.12, pp. 4088–4118

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

4.17 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

Thomas Leavitt (2023). “Randomization-based, Bayesian Inference of Causal Effects”. In: *Journal of Causal Inference* 11.1, pp. 1–25

Thomas Leavitt (2024). “Fisher Meets Bayes: The Value of Randomisation for Bayesian Inference of Causal Effects”. In: *International Statistical Review*

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

Zhaoyang Liu, Tingxuan Han, Donald B. Rubin, and Ke Deng (2025). "A Bayesian Criterion for Rerandomization". In: *Journal of the American Statistical Association* 120.552, pp. 2809–2821

References

Abadie, Alberto (2021). “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects”. In: *Journal of Economic Literature* 59.2, pp. 391–425.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (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.

Abadie, Alberto and Javier Gardeazabal (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country”. In: *The American Economic Review* 93.1, pp. 113–132.

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.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021). “Synthetic Difference-in-Differences”. In: *American Economic Review* 111.12, pp. 4088–4118.

Aronow, P. M., Haoge Chang, and Patrick Lopatto (2025). “Randomization-Based Confidence Sets for the Local Average Treatment Effect”. In: *Biometrika.* In press.

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, and Donald K. K. Lee (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.

Aronow, Peter M. 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.

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.

Austin, Peter C. (2008). “A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003”. In: *Statistics in Medicine* 27.12, pp. 2037–2049.

— (2009). “Balance Diagnostics for Comapring the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples”. In: *Statistics in Medicine* 28.25, pp. 3083–3107.

Bai, Yuehao (2023). “Why Randomize? Minimax Optimality under Permutation Invariance”. In: *Journal of Econometrics* 232.2, pp. 565–575.

Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant’Anna (2026). “Difference-in-Differences Designs: A Practitioner’s Guide”. In: *Journal of Economic Literature* 64.2, pp. 498–557.

Banerjee, Abhijit, Sylvain Chassang, Sergio Montero, and Erik Snowberg (2020). “A Theory of Experimenters: Robustness, Randomization, and Balance”. In: *American Economic Review* 110.4, pp. 1206–1230.

Banerjee, Abhijit, Sylvain Chassang, and Erik Snowberg (2017). “Decision Theoretic Approaches to Experiment Design and External Validity”. In: *Handbook of Field Experiments*. Ed. by Esther Duflo and Abhijit Banerjee. Vol. 1. Amsterdam, NL: North-Holland. Chap. 4, pp. 141–174.

Basu, Debabrata (1980). “Randomization Analysis of Experimental Data: The Fisher Randomization Test — Rejoinder”. In: *Journal of the American Statistical Association* 75.371, pp. 593–595.

Bekerman, William, Anurag Mehta, Rebecca E. Hasson, Leah E. Robinson, Dylan S. Small, and Colin B. Fogarty (2026). “Powerful Multivariate Sensitivity Analysis via Sample Splitting in an Observational Study of the Effects of Poverty on Cardiovascular Disease Risk Factors”. ArXiv preprint, arXiv:2606.04416.

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 A. (2010). “What You Can and Can’t Properly Do with Regression”. In: *Journal of Quantitative Criminology* 26.4, pp. 481–487.

Bertsimas, Dimitris, Mac Johnson, and Nathan Kallus (2015). “The Power of Optimization over Randomization in Designing Experiments Involving Small Samples”. In: *Operations Research* 63.4, pp. 868–876.

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.

Blackwell, Matthew and Nicole E. Pashley (2021). "Noncompliance and Instrumental Variables for 2^K Factorial Experiments". In: *Journal of the American Statistical Association* 118.542, pp. 1102–1114.
— (2026). "Bounds on Causal Effects in 2^K Factorial Experiments with Noncompliance". In: *Biometrika* 113.1.
Bowers, Jake, Bruce A. Desmarais, Mark Frederickson, Nahomi Ichino, Hsuan-Wei Lee, and Simi Wang (2018). "Models, Methods and Network Topology: Experimental Design for the Study of Interference". In: *Social Networks* 54, pp. 196–208.
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.
Branson, Zach (2021). "Randomization Tests to Assess Covariate Balance When Designing and Analyzing Matched Datasets". In: *Observational Studies* 7.2, pp. 1–36.
Branson, Zach, Xinran Li, and Peng Ding (2023). "Power and Sample Size Calculations for Rerandomization". In: *Biometrika* 111.1, pp. 355–363.
Callaway, Brantly and Pedro H. C. Sant'Anna (2021). "Difference-in-Differences with Multiple Time Periods". In: *Journal of Econometrics* 225.2, pp. 200–230.
Campbell, Donald T. (1957). "Factors Relevant to the Validity of Experiments in Social Settings". In: *Psychological Bulletin* 54.4, pp. 297–312.
Campbell, Donald T. and Julian C. Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.
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, Allan Dafoe, Xinran Li, and Luke W. Miratrix (2023). "Randomisation Inference Beyond the Sharp Null: Bounded Null Hypotheses and Quantiles of Individual Treatment Effects". In: *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 85.5, pp. 1471–1491.
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, Jeffrey D. Morenoff, Ben B. Hansen, Kimberly J. Tessari Hicks, Luis F. Duque, Alexandra Restrepo, and Ana V. Diez-Roux (2012). "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". In: *American Journal of Epidemiology* 175.10, pp. 1045–1053.
Chang, Haoge, Joel A. Middleton, and P. M. Aronow (2024). "Exact Bias Correction for Linear Adjustment of Randomized Controlled Trials". In: *Econometrica* 92, pp. 1503–1519.
Chang, Haoge and Zeyang Yu (2026). "Randomization Inference For the Always-Reporter Average Treatment Effect". ArXiv preprint, arXiv:2603.24970.
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.

Chen, Kan, Siyu Heng, Qi Long, and Bo Zhang (2023). "Testing Biased Randomization Assumptions and Quantifying Imperfect Matching and Residual Confounding in Matched Observational Studies". In: *Journal of Computational and Graphical Statistics* 32.2, pp. 528–538.

Chung, Eun Yi and Joseph P. Romano (2013). "Exact and Asymptotically Robust Permutation Tests". In: *The Annals of Statistics* 41.2, pp. 484–507.

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.

Cohen, Peter L. and Colin B. Fogarty (2022). "Gaussian Prepivoting for Finite Population Causal Inference". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84.2, pp. 295–320.

— (2023). "No-Harm Calibration for Generalized Oaxaca-Blinder Estimators". In: *Biometrika* 111.1, pp. 331–338.

Coppock, Alexander, Alan S. Gerber, Donald P. Green, and Holger L. Kern (2017). "Combining Double Sampling and Bounds to Address Nonignorable Missing Outcomes in Randomized Experiments". In: *Political Analysis* 25.2, pp. 188–206.

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik (2009). "Dealing with limited overlap in estimation of average treatment effects". In: *Biometrika* 96.1, pp. 187–199.

Cytrynbaum, Max (2024). "Finely Stratified Rerandomization Designs". ArXiv preprint, arXiv:2407.03279.

Daniel, Shoshana R., Katrina Armstrong, Jeffrey H. Silber, and Paul R. Rosenbaum (2008). "An Algorithm for Optimal Tapered Matching, With Application to Disparities in Survival". In: *Journal of Computational and Graphical Statistics* 17.4, pp. 914–924.

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

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

— (2024). *A First Course in Causal Inference*. Chapman and Hall/CRC.

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 W. 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 Erin Hartman (2022). "Elements of External Validity: Framework, Design, and Analysis". In: *American Political Science Review* 117.3, pp. 1070–1088.

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.

Fedorov, Valerii V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.

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

Fogarty, Colin B. (2018a). “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.
— (2018b). “Regression-assisted Inference for the Average Treatment Effect in Paired Experiments”. In: *Biometrika* 105.4, pp. 994–1000.
— (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* 79.3, pp. 2196–2207.
— (2025). “Tilted Sensitivity Analysis in Matched Observational Studies”. ArXiv preprint, arXiv:2503.09736.
Fogarty, Colin B., Mark E. Mikkelsen, David F. Gaieski, and Dylan S. Small (2016). “Discrete Optimization for Interpretable Study Populations and Randomization Inference in an Observational Study of Severe Sepsis Mortality”. In: *Journal of the American Statistical Association* 111.514, pp. 447–458.
Fogarty, Colin B., Pixu Shi, Mark E. Mikkelsen, and Dylan S. Small (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.
Gagnon-Bartsch, Johann and Yotam Shem-Tov (2019). “The Classification Permutation Test: A Flexible Approach to Testing for Covariate Imbalance in Observational Studies”. In: *The Annals of Applied Statistics* 13.3, pp. 1464–1483.
Gastwirth, Joseph L., Abba M. Krieger, and Paul R. Rosenbaum (2000). “Asymptotic separability in sensitivity analysis”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 545–555.
Gelman, Andrew (2003). “A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-fit Testing”. In: *International Statistical Review* 71.2, pp. 369–382.
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, Edward H. Kaplan, and Holger L. Kern (2010). “Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments”. In: *Political Analysis* 18.3, pp. 297–315.

Gilligan, Michael J. and Ernest J. Sergenti (2008). “Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference”. In: *Quarterly Journal of Political Science* 3.2, pp. 89–122.

Goplerud, Max, Kosuke Imai, and Nicole E. Pashley (2025). “Estimating Heterogeneous Causal Effects of High-Dimensional Treatments: Application to Conjoint Analysis”. In: *Annals of Applied Statistics* 19.2.

Gu, Xing Sam and Paul R. Rosenbaum (1993). “Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms”. In: *Journal of Computational and Graphical Statistics* 2.4, pp. 405–420.

Guo, Kevin and Guillaume W. Basse (2023). “The Generalized Oaxaca-Blinder Estimator”. In: *Journal of the American Statistical Association* 118.541, pp. 524–536.

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.

Ham, Dae Woong and Luke W. Miratrix (2024). “Benefits and Costs of Matching Prior to a Difference in Differences Analysis When Parallel Trends Does Not Hold”. In: *Annals of Applied Statistics* 18.3, pp. 2096–2122.

Hansen, Ben B. (2004). “Full Matching in an Observational Study of Coaching for the SAT”. In: *Journal of the American Statistical Association* 99.467, pp. 609–618.

— (2008a). “The Essential Role of Balance Tests in Propensity-Matched Observational Studies: Comments on ‘A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003’ by Peter Austin”. In: *Statistics in Medicine* 27.12, pp. 2050–2054.

— (2008b). “The Prognostic Analogue of the Propensity Score”. In: *Biometrika* 95.2, pp. 481–488.

— (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. and Stephanie Olsen Klopfer (2006). “Optimal Full Matching and Related Designs via Network Flows”. In: *Journal of Computational and Graphical Statistics* 15.3, pp. 609–627.

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 C. Sales (2015). "Comment on Cochran's "Observational Studies"". In: *Observational Studies*, pp. 184–193.
Harshaw, Christopher, Joel A. Middleton, and Fredrik Sävje (2026). "Optimized Variance Estimation under Interference and Complex Experimental Designs". In: *Journal of the American Statistical Association*.
Harshaw, Christopher, Fredrik Sävje, Daniel A. Spielman, and Peng Zhang (2024). "Balancing Covariates in Randomized Experiments with the Gram–Schmidt Walk Design". In: *Journal of the American Statistical Association* 119.548, pp. 2934–2946.
Harville, David A. (1975). "Experimental Randomization: Who Needs It?" In: *The American Statistician* 29.1, pp. 27–31.
Heller, Ruth, Paul R. Rosenbaum, and Dylan S. Small (2009). "Split Samples and Design Sensitivity in Observational Studies". In: *Journal of the American Statistical Association* 104.487, pp. 1090–1101.
Heng, Siyu, Yanxin Shen, and Pengyun Wang (2025). "Reconciling Overt Bias and Hidden Bias in Sensitivity Analysis for Matched Observational Studies". ArXiv preprint, arXiv:2311.11216.
Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart (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.
Holm, Sture (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70.
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.
Hsu, Jesse Y., Dylan S. Small, and Paul R. Rosenbaum (2013). "Effect Modification and Design Sensitivity in Observational Studies". In: *Journal of the American Statistical Association* 108.501, pp. 135–148.
Iacus, Stefano M., Gary King, and Giuseppe Porro (2012). "Causal Inference without Balance Checking: Coarsened Exact Matching". In: *Political Analysis* 20.1, pp. 1–24.
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.
Johansson, Per, Donald B. Rubin, and Mårten Schultzberg (2021). “On Optimal Rerandomization Designs”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.2, pp. 395–403.
Kallus, Nathan (2018). “Optimal A Priori Balance in the Design of Controlled Experiments”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.1, pp. 85–112.
— (2021). “On the Optimality of Randomization in Experimental Design: How to Randomize for Minimax Variance and Design-Based Inference”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.2, pp. 404–409.
Kandiros, Vardis, Charilaos Pipis, Constantinos Daskalakis, and Christopher Harshaw (2024). “The Conflict Graph Design: Estimating Causal Effects under Arbitrary Neighborhood Interference”. ArXiv preprint, arXiv:2411.10908.
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.
Kapelner, Adam, Abba M. Krieger, Michael Sklar, and David Azriel (2022). “Optimal Rerandomization Designs via a Criterion that Provides Insurance against Failed Experiments”. In: *Journal of Statistical Planning and Inference* 219, pp. 63–84.
Kapelner, Adam, Abba M. Krieger, Michael Sklar, Uri Shalit, and David Azriel (2021). “Harmonizing Optimized Designs with Classic Randomization in Experiments”. In: *The American Statistician* 75.2, pp. 195–206.
Kasy, Maximilian (2016). “Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead”. In: *Political Analysis* 24.3, pp. 324–338.
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 S. 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., Elizabeth A. Stuart, Jennifer Hill, and Donald P. Green (2016). "Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations". In: *Journal of Research on Educational Effectiveness* 9.1, pp. 103–127.
Kiefer, Jack (1959). "Optimum Experimental Designs". In: *Journal of the Royal Statistical Society Series B: Methodological* 21.2, pp. 272–319.
Kim, David, Yongchang Su, Jake Bowers, and Xinran Li (2026). "Randomization Tests for Distributions of Individual Treatment Effects via Combined Rank Statistics". ArXiv preprint, arXiv:2605.08027.
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.
King, Gary and Langche Zeng (2006). "The Dangers of Extreme Counterfactuals". In: *Political Analysis* 14.2, pp. 131–159.
Leavitt, Thomas (2023). "Randomization-based, Bayesian Inference of Causal Effects". In: *Journal of Causal Inference* 11.1, pp. 1–25.
— (2024). "Fisher Meets Bayes: The Value of Randomisation for Bayesian Inference of Causal Effects". In: *International Statistical Review*.
— (2026). "Beyond Pre-Trends: A Discordance-Based Sensitivity Analysis for Difference-in-Differences". Accepted at *Observational Studies*.
Leavitt, Thomas and Laura A. Hatfield (2025). "Averaged Prediction Models (APM): Identifying Causal Effects in Controlled Pre-Post Settings with Application to Gun Policy". In: *Annals of Applied Statistics* 19.3.
Leavitt, Thomas and Luke W. Miratrix (2026). "Building a Design-Based Matching Pipeline: From Principles to Practical Implementation in R". Accepted at *Observational Studies*.
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, Fan, Kari Lock Morgan, and Alan M. Zaslavsky (2018). "Balancing Covariates via Propensity Score Weighting". In: *Journal of the American Statistical Association* 113.521, pp. 390–400.
Li, Xinran (2025). "Sensitivity Analysis for Observational Studies with Flexible Matched Designs". In: *Biometrika* 112.4.
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.
— (2020). "Rerandomization and Regression Adjustment". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1, pp. 241–268.
Li, Xinran, Peng Ding, and Donald B. Rubin (2018). "Asymptotic Theory of Rerandomization in Treatment–Control Experiments". In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9157–9162.
Li, Xinran, Peizan Sheng, and Zeyang Yu (2025). "Randomization Inference with Sample Attrition". ArXiv preprint, arXiv:2507.00795.

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, Mengqi, Colin B. Fogarty, and Gongjun Xu (2026). "Stochastic Sensitivity Analysis for Matched Observational Studies". ArXiv preprint, arXiv:2606.05120.
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.
Liu, Zhaoyang, Tingxuan Han, Donald B. Rubin, and Ke Deng (2025). "A Bayesian Criterion for Rerandomization". In: *Journal of the American Statistical Association* 120.552, pp. 2809–2821.
Lu, Bo, Robert Greevy, Xinyi Xu, and Cole Beck (2011). "Optimal Nonbipartite Matching and Its Statistical Applications". In: *The American Statistician* 65.1, pp. 21–30.
Lu, Bo, Elaine Zanutto, Robert Hornik, and Paul R. Rosenbaum (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.
Lu, Xin, Tianle Liu, Hanzhong Liu, and Peng Ding (2023). "Design-Based Theory for Cluster Rerandomization". In: *Biometrika* 110.2, pp. 467–483.
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.
Martinez, Marco and David Teira (2024). "Why Experimental Balance Is Still a Reason to Randomize". In: *The British Journal for the Philosophy of Science* 75.2, pp. 519–535.
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.
Mikhaeil, Jonas M. and Christopher Harshaw (2025). "Valid Inference when Testing Violations of Parallel Trends for Difference-in-Differences". ArXiv preprint, arXiv:2510.26470.
Miratrix, Luke W., Jasjeet S. Sekhon, Alexander G. Theodoridis, and Luis F. Campos (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.
Morgan, Kari Lock and Donald B. Rubin (2012). "Rerandomization to Improve Covariate Balance in Experiments". In: *Annals of Statistics* 40.2, pp. 1263–1282.
— (2015). "Rerandomization to Balance Tiers of Covariates". In: *Journal of the American Statistical Association* 110.512, pp. 1412–1421.
Neyman, Jersey (1923). "Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes". In: *Roczniki Nauk Rolniczych* 10, pp. 1–51.
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. (2022). "Note on the Delta Method for Finite Population Inference with Applications to Causal Inference". In: *Statistics & Probability Letters* 188, p. 109540.

Pashley, Nicole E., Guillaume W. Basse, and Luke W. Miratrix (2021). “Conditional as-if analyses in randomized experiments”. In: *Journal of Causal Inference* 9.1.
Pashley, Nicole E., Luke Keele, and Luke W. Miratrix (2024). “Improving Instrumental Variable Estimators with Poststratification”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 188.3, pp. 765–790.
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. and Yaxuan Huang (2024). “Covariate-adaptive Randomization Inference in Matched Designs”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.5, pp. 1312–1338.
Pimentel, Samuel D., Rachel R. Kelz, Jeffrey H. Silber, and Paul R. Rosenbaum (2015). “Large, Sparse Optimal Matching With Refined Covariate Balance in an Observational Study of the Health Outcomes Produced by New Surgeons”. In: *Journal of the American Statistical Association* 110.510, pp. 515–527.
Pimentel, Samuel D., Lindsay C. Page, Matthew Lenard, and Luke Keele (2018). “Optimal Multilevel Matching Using Network Flows: An Application to a Summer Reading Intervention”. In: *Annals of Applied Statistics* 12.3, pp. 1479–1505.
Pimentel, Samuel D., Frank Yoon, and Luke Keele (2015). “Variable-ratio Matching with Fine Balance in a Study of the Peer Health Exchange”. In: *Statistics in Medicine* 34.30, pp. 4070–4082.
Rambachan, Ashesh and Jonathan Roth (2023). “A More Credible Approach to Parallel Trends”. In: *Review of Economic Studies* 90.5, pp. 2555–2591.
Ribeiro Junior, Antônio Carlos Herling and Zach Branson (2025). “Does Rerandomization Help Beyond Covariate Adjustment? A Review and Guide for Theory and Practice”. ArXiv preprint, arXiv:2512.05290.
Robins, James M. (1988). “Confidence Intervals for Causal Parameters”. In: *Statistics in Medicine* 7.7, pp. 773–785.
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. (1988). “Sensitivity Analysis for Matching with Multiple Controls”. In: *Biometrika* 75.3, pp. 577–581.
— (1996). “Identification of Causal Effects Using Instrumental Variables: Comment”. In: *Journal of the American Statistical Association* 91.434, pp. 465–468.
— (1999). “Reduced Sensitivity to Hidden Bias at Upper Quantiles in Observational Studies with Dilated Treatment Effects”. In: *Biometrics* 55.2, pp. 560–564.
— (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.

Rosenbaum, Paul R. (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.
— (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.
— (2025). *An Introduction to the Theory of Observational Studies*. Springer Texts in Statistics. New York, NY: Springer.
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., Richard N. Ross, and Jeffrey H. Silber (2007). "Minimum Distance Matched Sampling With Fine Balance in an Observational Study of Treatment for Ovarian Cancer". In: *Journal of the American Statistical Association* 102.477, pp. 75–83.
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.
Rosenbaum, Paul R. and Jeffrey H. Silber (2013). "Using the Exterior Match to Compare Two Entwined Matched Control Groups". In: *The American Statistician* 63.2, pp. 67–75.
Roth, Jonathan (2022). "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends". In: *American Economic Review: Insights* 4.3, pp. 305–322.
Roth, Jonathan and Pedro H. C. Sant'Anna (2023). "When Is Parallel Trends Sensitive to Functional Form?" In: *Econometrica* 91.2, pp. 737–747.
Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe (2023). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature". In: *Journal of Econometrics* 235.2, pp. 2218–2244.
Rubin, Donald B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66.5, p. 688.
— (1977). "Assignment to Treatment Group on the Basis of a Covariate". In: *Journal of Educational Statistics* 2.1, pp. 1–26.
— (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.
— (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation". In: *Health Services and Outcomes Research Methodology* 2.3, pp. 169–188.
Rubin, Donald B. and Richard P. Waterman (2006). "Estimating the Causal Effects of Marketing Interventions Using Propensity Score Methodology". In: *Statistical Science* 21.2, pp. 206–222.

Sales, Adam C. and Ben B. Hansen (2020). “Limitless Regression Discontinuity”. In: *Journal of Educational and Behavioral Statistics* 45.2, pp. 143–174.
Sales, Adam C., Ben B. Hansen, and Brian Rowan (2018). “Rebar: Reinforcing a Matching Estimator With Predictions From High-Dimensional Covariates”. In: *Journal of Educational and Behavioral Statistics* 43.1, pp. 3–31.
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, P. M. Aronow, and Michael G. Hudgens (2021). “Average Treatment Effects in the Presence of Unknown Interference”. In: *Annals of Statistics* 49.2, pp. 673–701.
Sävje, Fredrik, Michael J. Higgins, and Jasjeet S. Sekhon (2021). “Generalized Full Matching”. In: *Political Analysis*.
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.
— (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.
Senn, Stephen (1994). “Fisher’s Game with the Devil”. In: *Statistics in Medicine* 13.3, pp. 217–230.
— (2013). “Seven Myths of Randomisation in Clinical Trials”. In: *Statistics in Medicine* 32.9, pp. 1439–1450.
Shadish, William R., Thomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.
Silber, Jeffrey H., Paul R. Rosenbaum, Richard N. Ross, Justin M. Ludwig, Wei Wang, Bijan A. Niknam, Nabanita Mukherjee, Philip A. Saynisch, Orit Even-Shoshan, Rachel R. Kelz, and Lee A. Fleisher (2014). “Template Matching for Auditing Hospital Cost and Quality”. In: *Health Services Research* 48.5, pp. 1446–1474.
Small, Dylan S., Jing Cheng, M. Elizabeth Halloran, and Paul R. Rosenbaum (2013). “Case Definition and Design Sensitivity”. In: *Journal of the American Statistical Association* 108.504, pp. 1457–1468.
Stuart, Elizabeth A. (2010). “Matching Methods for Causal Inference: A Review and a Look Forward”. In: *Statistical Science* 25.1, p. 1.
Tan, Zhiqiang (2020). “Regularized Calibrated Estimation of Propensity Scores with Model Misspecification and High-dimensional Data”. In: *Biometrika* 107.1, pp. 137–158.
Traskin, Mikhail and Dylan S. Small (2011). “Defining the Study Population for an Observational Study to Ensure Sufficient Overlap: A Tree Approach”. In: *Statistics in Biosciences* 3, pp. 94–118.
Wang, Xinhe, Tingyu Wang, and Hanzhong Liu (2021). “Rerandomization in Stratified Randomized Experiments”. In: *Journal of the American Statistical Association* 118.542, pp. 1295–1304.
Wang, Ye, Cyrus Samii, Haoge Chang, and P. M. Aronow (2025). “Design-Based Inference for Spatial Experiments with Interference”. In: *Annals of Applied Statistics* 19.1, pp. 744–768.

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, Jessie K. Edwards, Catherine R. Lesko, Stephen R. Cole, and Elizabeth A. Stuart (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.

Wu, Chien-Fu (1981). "On the Robustness and Efficiency of Some Randomized Designs". In: *The Annals of Statistics* 9.6, pp. 1168–1177.

Wu, Dongxiao and Xinran Li (2025). "Sensitivity Analysis for Quantiles of Hidden Biases in Matched Observational Studies". In: *Journal of the American Statistical Association* 120.551, pp. 1657–1668.

Wu, Jason and Peng Ding (2021). "Randomization Tests for Weak Null Hypotheses". In: *Journal of the American Statistical Association* 116.536, pp. 1898–1913.

Yang, Dan, Dylan S. Small, Jeffrey H. Silber, and Paul R. Rosenbaum (2012). "Optimal Matching with Minimal Deviation from Fine Balance in a Study of Obesity and Surgical Outcomes". In: *Biometrics* 68.2, pp. 628–636.

Yu, Ruoqi (2023). "How Well Can Fine Balance Work for Covariate Balancing". In: *Biometrics* 79.3, pp. 2346–2356.

Yu, Ruoqi, Dylan S. Small, and Paul R. Rosenbaum (2021). "The Information in Covariate Imbalance in Studies of Hormone Replacement Therapy". In: *The Annals of Applied Statistics* 15.4, pp. 2023–2042.

Zhu, Jianan, Jeffrey Zhang, Zijian Guo, and Siyu Heng (2025). "Randomization-Based Inference for Average Treatment Effects in Inexactly Matched Observational Studies". ArXiv preprint, arXiv:2308.02005.

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. 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.

Zubizarreta, José R., Ricardo D. Paredes, and Paul R. Rosenbaum (2014). "Matching for balance, pairing for heterogeneity in an observational study of the effectiveness of for-profit and not-for-profit high schools in Chile". In: *The Annals of Applied Statistics* 8.1, pp. 204–231.