Matching for Adjustment and Causal Inference
Escuela de Invierno en Métodos y Análisis de Datos UCU-DCSP

Jake Bowers
jwbowers@illinois.edu
http://jakebowers.org/

Invierno 2023

1 Overview

Esta clase es una introducción al ajuste estadístico utilizando la estratificación emparejada en el estilo iniciado por Rubin y Rosenbaum y actualmente en rápido desarrollo en las ciencias sociales y las disciplinas estadísticas. Una motivación importante para el emparejamiento es aproximarse a un diseño experimental o al menos hacer un diseño de investigación en un contexto no aleatorio que pueda compararse con algún punto de referencia aleatorio. Y, dado que tal motivación surge del deseo de hacer declaraciones transparentes y defendibles sobre las relaciones causales, presentaremos como formalizaciones de tales ideas la concepción contrafactual de la inferencia causal y la idea de "resultados potenciales". También dedicaremos un tiempo a la inferencia estadística (estimación, prueba de hipótesis, creación de intervalos de confianza) después de la creación de un diseño estratificado. Finalmente, abordaremos algunas de las preguntas que son temas de investigación actuales en esta área: ¿Cuándo y cómo se puede afirmar que se ha ajustado "suficientemente"? ¿Cómo podemos abordar las preocupaciones sobre los sesgos no observados incluso si nos hemos ajustado a lo que observamos?

Dado que los métodos de emparejamiento se están desarrollando rápidamente en la literatura metodológica, aquí nos centraremos en la forma más simple y antigua: la posestratificación. Los conceptos generales y el flujo de trabajo deben ser transportables a métodos más sofisticados de ajuste combinado.

This class is an introduction to statistical adjustment using matched stratification in the style pioneered by Rubin and Rosenbaum and currently in rapid development across the social science and statistical disciplines. An important motivation for matching is to approximate an experimental design or at least make an research design in a non-randomized context that can be compared to some randomized benchmark. And, since such a motivation arises from a desire to make transparent and defensible statements about causal relations, we will introduce the counterfactual conception of causal inference and the potential outcome formalization of these ideas. We will also spend some time on statistical inference (estimation, hypothesis testing, confidence interval creation) after the creation of a matched design. Finally, we will grapple with some of the questions that are current research topics in this area: When and how one can claim to have adjusted "enough"? How can we engage with concerns about unobserved confounds even if we have adjusted for what we observe?

Since methods of matching are rapidly developing in the methodology literature, we will here focus on the simplest and oldest form: post-stratification. The general concepts and work-flow should be transportable to more sophisticated methods of matched adjustment.

Goals and Expectations

This course aims to help you think about statistical adjustment using stratification and matching as compared to statistical adjustment using the linear model directly (adjustment by parametric model or "residualization").

The course ought to give you opportunities to practice producing matched designs for your data and to ask questions that puzzle you as you do this work.

The point of the course is to position you to do the future learning that is at the core of your work as an academic analyzing data.

This course does not delve deeply into the theories of causal inference, statistical inference, or algorithms at the heart of these methods of statistical adjustment. Rather, through practice using tools, I hope that your curiosity is awakened and you begin to read more broadly and understand more deeply on your own.

Expectations I assume some previous engagement with high school mathematics and probability and statistical computing in the R statistical computing environment. If you have not used R, you are welcome to take the class, but I encourage you to get a little experience with R before the first class session. Feel free to email me to ask for advice about how to practice with R before the class begins. You can see some basic guidance about R and RStudio in Spanish here and in English here.

Participation We will be doing hands-on work. I plan to lecture relatively little and instead will hope to pose problems of statistical theory, research design, and data for you to solve at your computers. I anticipate that you'll work in small groups, asking me and/or the group questions as you proceed. I will break away to draw on the board or demonstrate on my own computer now and then to clarify points or help you around particularly difficult tasks.

Computing We will be using R in class so those of you with laptops available should bring them. Of course, I will not tolerate the use of computers for anything other than class related work during active class time. Please install R (`http://www.r-project.org`) on your computers before the first class session. You may prefer to use R in the context of the Rstudio IDE (`http://www.rstudio.com/`).

Computing is an essential part of modern statistical data analysis — both for turning data into information and for conveying that information persuasively (and thus transparently and reliably) to the scholarly community. In this course we will pay attention to computing, with special emphasis on understanding what is going on behind the scenes. You will be writing your own routines for a few simple and common procedures.

Books

We will use the Rosenbaum book as our primary source. The other books are useful for further study.

Required Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer (pdf free to download from some university ip addresses or via university library springerlink subscriptions: `https://link.springer.com/book/10.1007/978-3-030-46405-9`)

Recommended Paul R. Rosenbaum (2017). *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, p. 374. ISBN: 9780674975576 (This book is particularly easy to read. Particularly Chapter 3 on Statistical Inference and Chapter 11 on Matching Techniques.)

Howard S. Becker (1986). *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. University of Chicago Press. (On writing.)

Richard Berk (2004). *Regression Analysis: A Constructive Critique*. Sage (For clarity about what a linear model does well and poorly.)

A. Gelman and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press (particularly chapters 9,10 and 23 see `http://www.stat.columbia.edu/~gelman/arm/`).

Stephen L. Morgan and Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press. ISBN: 0521671930 (See http://www.wjh.harvard.edu/~cwinship/cfa.html for some links and background reading)

Paul R. Rosenbaum (2002b). *Observational Studies*. Second. Springer-Verlag (see http://www-stat.wharton.upenn.edu/~rosenbap/index.html for lots of papers and presentations).

Donald B. Rubin (2006). *Matched sampling for causal effects*. Cambridge; New York: Cambridge University Press (for early papers on matching)

Guido W Imbens and Donald B Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press (for discussion of statistical inference for causal effects)

Scott Cunningham (2021). *Causal inference: The mixtape*. Yale university press (see the online version https://mixtape.scunning.com/)

II Schedule

Note: This schedule is preliminary and subject to change. I anticipate mixing group discussions of your questions from the readings with in-class work using your own laptops or those provided by the school to get practice creating, analyzing, and assessing matched designs.

1— Experiments, Potential Outcomes, and Treatment Effects

Questions and Reading What is the point of experiments? What are the key characteristics of experiments? Why are experiments special? Why are we talking about randomized experiments if this class is about matching in observational studies; in studies were we have not randomized?

D.R. Kinder and T.R. Palfrey (1993). "On behalf of an experimental political science". In: *Experimental foundations of political science*, pp. 1–39

Alan S Gerber and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton, Chap 1

How can we bolster the interpretability of our comparisons if we do not have an experiment?

Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 1

What do we mean by "causal inference"? Can we convince ourselves that experiments have special advantages like "unbiased estimation of averages of potential outcomes"?

Alan S Gerber and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton, Chap 2

Jake Bowers, Maarten Voors, and Nahomi Ichino (2021). *The Theory and Practice of Field Experiments: An Introduction from the EGAP Learning Days*. Open Source Textbook. Evidence in Governance and Politics. URL: https://egap.github.io/theory_and_practice_of_field_experiments/, Modulo 3: Inferencia Causal / Module 3: Causal Inference(edición en español or english version).

Extra Reading A. Gelman and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chap 9.0 – 9.3 (On potential outcomes and causal inference)

J.D. Angrist and J.S. Pischke (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton Univ Pr. ISBN: 0691120358, Chap 2

P. W. Holland (1986). "Statistics and Causal Inference (with discussion)". In: *Journal of the American Statistical Association* 81, pp. 945–970 (on the Counterfactual/Manipulationist conception of causality)

Henry E. Brady (2008). "Causation and explanation in social science". In: ed. by Janet M Box-Steffensmeier, Henry E Brady, and David Collier. Oxford University Press, pp. 217–270 (for a survey of other major conceptions of causality from the perspective of applied social science)

2— Adjustment by Simple and Complex, Algorithmic Stratification

Given the problems that arise from the use of the linear model, how can we use research design to approximate the randomized experiment? How would we assess whether a matched design adjusted enough or not? What does it mean to adjust enough? What actual steps do we take in order to flexibly express our ideas about what it means for "like to be compared with like."

Questions and Reading Why not use the linear model for adjustment? How do we know when we have adjusted enough to make a strong case for clear, interpretable comparisons?

Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 6

A. Gelman and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chap 9.5–9.6

What is the basic intuition behind post-stratified designs and modern optimal versions of post-stratification that we call "matched designs"? What is "balance"? How do we use R to create and evaluate matched designs?

Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 7–9, 13 (Especially Chap 7, Chap 8.6, Chap 9)

Extra Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 3

Ben B. Hansen (Sept. 2004). "Full Matching in an Observational Study of Coaching for the SAT". in: *Journal of the American Statistical Association* 99, p. 609

Ben B Hansen (2011). "Propensity score matching to extract latent experiments from nonexperimental data: A case study". In: *Looking Back*. Springer, pp. 149–181 for an example walk-through of a matched analysis including a discussion of missing data on covariates.

Daniel Ho et al. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis* 15, pp. 199–236 [esp. their discussion of model sensitivity, for example their Fig 2]

Four different ideas about balance testing:

1. Require strict common support Imai, Gary King, and Stuart 2008 (see also literature on coarsened exact matching)
2. Minimize imbalance measured multiple ways J. Sekhon 2007a[1];
3. Compare to equivalent randomized design and test across many covariates B.B. Hansen and J. Bowers 2008 B.B. Hansen 2008 or for a less mathematical version of the same argument Jake Bowers 2011, §3;
4. Specify a range for imbalance and test Hartman and Hidalgo 2018

3— Statistical Inference for Matched/Post-stratified Designs

Questions and Reading Given a matched design, how can we produce tests of substantively meaningful hypotheses about the unobserved comparisons of potential outcomes that we call "causal effects?"

Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 2

How can we produce confidence intervals for an estimate of an average treatment effect?

Thad Dunning (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press, Chap 6.1 and Appendix 6.1

G. Imbens and D. Rubin (2009). "Causal Inference in Statistics". Unpublished book manuscript. Forthcoming at Cambridge University Press., Chap 17

[1] http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf

Extra Richard Berk (2004). *Regression Analysis: A Constructive Critique*. Sage, Chap 4 on general requirements for statistical inference (i.e. what does it mean to do statistical inference at all, what are we inferring to?)

Winston Lin (Sept. 2011). "Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique". Unpublished manuscript provides some useful proofs supporting the idea that linear regression with "robust" HC2 standard errors provides a useful large-sample way to do statistical inference about average treatment effects.

L.W. Miratrix, J.S. Sekhon, and B. Yu (2012). "Adjusting treatment effect estimates by post-stratification in randomized experiments". In: *JR Stat. Soc. Ser. B. Stat. Methodol. To appear* teach us about statistical inference for average treatment effects when matching after experimental outcomes have been collected.

David A Freedman (2008a). "Randomization does not justify logistic regression". In: *Statistical Science* 23.2, pp. 237–249; David A. Freedman (2008b). "On regression adjustments to experimental data". In: *Advances in Applied Mathematics* 40.2, pp. 180–193; David A. Freedman (2007). "On regression adjustments in experiments with several treatments". In: *Annals of Applied Statistics (To Appear)*; David A. Freedman (2006). "On the So-called "Huber Sandwich Estimator" and "Robust Standard Errors"". In: *The American Statistician* 60.4, pp. 299–302 Suggesting that even the large sample statistical inference from using linear regression in randomized experiments is biased. Also arguing that the Huber-White standard errors are not a good idea.

Paul R. Rosenbaum (2002a). "Covariance adjustment in randomized experiments and observational studies". In: *Statistical Science* 17.3, pp. 286–327 and Jake Bowers and Costas Panagopoulos (May 2011). "Fisher's randomization mode of statistical inference, then and now." Unpublished manuscript showing how covariance adjustment is compatible with Fisher's randomization inference (and thus can be unproblematic after matching).

Peter Schochet (2009). "Is regression adjustment supported by the Neyman model for causal inference". In: *Journal of Statistical Planning and Inference*; Donald P. Green (July 2009). "Regression Adjustments to Experimental Data: Do David Freedman's Concerns Apply to Political Science?" Unpublished Manuscript Suggesting that in large samples these biases worried about by Freedman ought not to worry us.

G. Imbens and D. Rubin (2009). "Causal Inference in Statistics". Unpublished book manuscript. Forthcoming at Cambridge University Press., Chap 6–8 Suggesting, similarly to Green and Schochet, that regression is fine for statistical inference in experiments (and further suggesting the use of the Huber-White robust standard errors).

Alberto Abadie and Guido Imbens (2004). "On the Failure of the Bootstrap for Matching Estimators". In: *NBER, Unpublished Manuscript* suggesting that the bootstrap is not a good approach with matched designs.

For advanced reading on the latest in statistical theory for statistical inference for "matching estimators" (which include but are not restricted to post-stratified studies) see:

Ben Hansen (2009). *Propensity score matching to recover latent experiments: diagnostics and asymptotics*. Tech. rep. 486. Statistics Department, University of Michigan for theory using randomization-inference.

A. Abadie and G. Imbens (2009). "Matching on the estimated propensity score". Unpublished Manuscript for a large-sample, Normal theory approach.

4— Sensitivity Analysis: An observational study is not a randomized experiment.

Questions and Reading We addressed concerns about variables that we do observe using matching. Although randomization addresses concerns about all background covariates (observed and unobserved), any non-randomized study may be criticized on the basis that it does not adjust for variables that were not observed.

Since an observational study (no matter how well matched) cannot adjust for unobserved confounders, how can we address concerns about such unobserved variables?

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

Carrie A Hosman, Ben B. Hansen, and Paul W. Holland (2010). "The sensitivity of linear regression coefficients's 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

Extra Paul R. Rosenbaum (2002b). *Observational Studies*. Second. Springer-Verlag, Chap 4

Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 3, 14

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

5— *—Other Topics

If we move quickly, or if the class has some organized preferences, we could change the syllabus to discussion some of these ideas.

Non-bipartite matching – Advances in Multivariate Matching: Beyond Binary Treatment Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 11; B. Lu et al. (2011). "Optimal nonbipartite matching and its statistical applications". In: *The American Statistician* 65.1, pp. 21–30. ISSN: 0003-1305; Nathaniel Rabb et al. (July 2022). "The influence of social norms varies with "others" groups: Evidence from COVID-19 vaccination intentions". In: *Proceedings of the National Academy of Sciences* 119.29. DOI: https://doi.org/10.1073/pnas.2118770119 (an example using non-bipartite matching in a survey with application to COVID attitudes)

Longitudinal Matching – Advances in Multivariate Matching: Matching with Longitudinal Data Paul R. Rosenbaum (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer, Chap 12

Multilevel Matching 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 et al. (2018). "Optimal multilevel matching using network flows: An application to a summer reading intervention". In: *The Annals of Applied Statistics* 12.3, pp. 1479–1505

6— Stuff that was painfully left out but which is important

Here are just a few extra citations to launch self-study of aspects of matching which we did not cover in our class.

For example: Genetic Matching Diamond and J. S. Sekhon 2013; J. Sekhon 2007b or Coarsened Exact Matching Iacus, G. King, and Porro 2009; Iacus, G. King, and Porro 2011 or Balance Optimization Subset Selection Nikolaev et al. 2012 or more fine-tuned versions of the optimal post-stratification that we consider in this class (for example, J. Zubizarreta 2012, Sävje, Higgins, and J. S. Sekhon 2021). Nor did we have time to engage with many other applied and theoretical topics in causal inference for observational studies such as the work establishing causal interpretation of the propensity score (cited in the Rosenbaum textbook), or the alternative approaches to causal inference based on weighting by functions of the propensity score such as those arising from work by Jamie Robins Glynn and Quinn 2010, let alone alternative conceptualizations of causal relations such at those developed by Judea Pearl Pearl 2000 or the work on estimation by Heckman or bounding causal inferences by Manski.

III References

References

Abadie, A. and G. Imbens (2009). “Matching on the estimated propensity score”. Unpublished Manuscript.

Abadie, Alberto and Guido Imbens (2004). “On the Failure of the Bootstrap for Matching Estimators”. In: *NBER, Unpublished Manuscript*.

Angrist, J.D. and J.S. Pischke (2009). *Mostly harmless econometrics: an empiricist’s companion*. Princeton Univ Pr. ISBN: 0691120358.

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

Berk, Richard (2004). *Regression Analysis: A Constructive Critique*. Sage.

Bowers, Jake (2011). “Making Effects Manifest in Randomized Experiments”. In: *Cambridge Handbook of Experimental Political Science*. Ed. by James N. Druckman et al. New York, NY: Cambridge University Press. Chap. 32.

Bowers, Jake and Costas Panagopoulos (May 2011). “Fisher’s randomization mode of statistical inference, then and now.” Unpublished manuscript.

Bowers, Jake, Maarten Voors, and Nahomi Ichino (2021). *The Theory and Practice of Field Experiments: An Introduction from the EGAP Learning Days*. Open Source Textbook. Evidence in Governance and Politics. URL: https://egap.github.io/theory_and_practice_of_field_experiments/.

Brady, Henry E. (2008). “Causation and explanation in social science”. In: ed. by Janet M Box-Steffensmeier, Henry E Brady, and David Collier. Oxford University Press, pp. 217–270.

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.

Cornfield, J. et al. (1959). “Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions”. In: *Journal of the National Cancer Institute*.

Cunningham, Scott (2021). *Causal inference: The mixtape*. Yale university press.

Diamond, Alexis and Jasjeet S Sekhon (2013). “Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies”. In: *Review of Economics and Statistics* 95.3, pp. 932–945.

Dunning, Thad (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

Freedman, David A (2008a). “Randomization does not justify logistic regression”. In: *Statistical Science* 23.2, pp. 237–249.

– (2006). “On the So-called “Huber Sandwich Estimator” and “Robust Standard Errors””. In: *The American Statistician* 60.4, pp. 299–302.

– (2007). “On regression adjustments in experiments with several treatments”. In: *Annals of Applied Statistics (To Appear)*.

– (2008b). “On regression adjustments to experimental data”. In: *Advances in Applied Mathematics* 40.2, pp. 180–193.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gerber, Alan S and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Glynn, Adam N and Kevin M Quinn (2010). “An introduction to the augmented inverse propensity weighted estimator”. In: *Political Analysis* 18.1, pp. 36–56.

Green, Donald P. (July 2009). "Regression Adjustments to Experimental Data: Do David Freedman's Concerns Apply to Political Science?" Unpublished Manuscript.

Hansen, B.B. (May 2008). "Comment: The essential role of balance tests in propensity-matched observational studies". In: *Statistics in Medicine* 27.12.

Hansen, B.B. and J. Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". In: *Statistical Science* 23, p. 219.

Hansen, Ben (2009). *Propensity score matching to recover latent experiments: diagnostics and asymptotics*. Tech. rep. 486. Statistics Department, University of Michigan.

Hansen, Ben B (2011). "Propensity score matching to extract latent experiments from nonexperimental data: A case study". In: *Looking Back*. Springer, pp. 149–181.

– (Sept. 2004). "Full Matching in an Observational Study of Coaching for the SAT". In: *Journal of the American Statistical Association* 99, p. 609.

Hartman, Erin and F Daniel Hidalgo (2018). "An equivalence approach to balance and placebo tests". In: *American Journal of Political Science* 62.4, pp. 1000–1013.

Ho, Daniel et al. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis* 15, pp. 199–236.

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

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

Iacus, S.M., G. King, and G. Porro (2009). "Causal inference without balance checking: Coarsened exact matching". In: *Retrieved September* 13, p. 2010.

– (2011). "Multivariate matching methods that are monotonic imbalance bounding". In: *Journal of the American Statistical Association* 106.493, pp. 345–361.

Imai, Kosuke, Gary King, and Elizabeth Stuart (2008). "Misunderstandings among Experimentalists and Observationalists about Causal Inference". In: *Journal of the Royal Statistical Society, Series A* 171.2, pp. 1–22.

Imbens, G. and D. Rubin (2009). "Causal Inference in Statistics". Unpublished book manuscript. Forthcoming at Cambridge University Press.

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

Imbens, Guido W and Donald B Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kinder, D.R. and T.R. Palfrey (1993). "On behalf of an experimental political science". In: *Experimental foundations of political science*, pp. 1–39.

Lin, Winston (Sept. 2011). "Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique". Unpublished manuscript.

Lu, B. et al. (2011). "Optimal nonbipartite matching and its statistical applications". In: *The American Statistician* 65.1, pp. 21–30. ISSN: 0003-1305.

Miratrix, L.W., J.S. Sekhon, and B. Yu (2012). "Adjusting treatment effect estimates by post-stratification in randomized experiments". In: *JR Stat. Soc. Ser. B. Stat. Methodol. To appear*.

Morgan, Stephen L. and Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press. ISBN: 0521671930.

Nikolaev, Alexander G. et al. (2012). *Balance Optimization Subset Selection (BOSS): An Alternative Approach for Causal Inference with Observational Data*. Tech. rep. University of Buffalo.

Pearl, Judea (2000). *Causality : Models, Reasoning, and Inference*. Reprinted with corrections. Cambridge University Press. ISBN: 0521773628.

Pimentel, Samuel D et al. (2018). "Optimal multilevel matching using network flows: An application to a summer reading intervention". In: *The Annals of Applied Statistics* 12.3, pp. 1479–1505.

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

Rosenbaum, Paul R. (2002a). “Covariance adjustment in randomized experiments and observational studies”. In: *Statistical Science* 17.3, pp. 286–327.
– (2002b). *Observational Studies*. Second. Springer-Verlag.
– (2017). *Observation and experiment : an introduction to causal inference*. Cambridge, MA: Harvard University Press, p. 374. ISBN: 9780674975576.
– (2020). *Design of Observational Studies*. 2nd. Springer Series in Statistics. Springer.
Rubin, Donald B. (2006). *Matched sampling for causal effects*. Cambridge; New York: Cambridge University Press.
Sävje, Fredrik, Michael J Higgins, and Jasjeet S Sekhon (2021). “Generalized full matching”. In: *Political Analysis* 29.4, pp. 423–447.
Schochet, Peter (2009). “Is regression adjustment supported by the Neyman model for causal inference”. In: *Journal of Statistical Planning and Inference*.
Sekhon, J.S. (2007a). “Alternative balance metrics for bias reduction in matching methods for causal inference”. In: *Survey Research Center, University of California, Berkeley*.
– (2007b). “Multivariate and propensity score matching software with automated balance optimization: The matching package for r”. In: *Journal of Statistical Software* 10.2, pp. 1–51.
Zubizarreta, J.R. (2012). “Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery”. In: *Journal of the American Statistical Association, Forthcoming*.
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.