References

Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, Vol. 97(457): 284–292.

Abadie, A. (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, Vol. 113(2): 231–263.

Abadie, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, Vol. 72(1): 1–19.

Abadie, A., J. Angrist, and G. Imbens, (2002), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, Vol. 70(1): 91–117.

Abadie, A., A. Diamond, and J. Hainmueller, (2010), "Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, Vol. 105(490): 493–505.

Abadie, A., D. Drukker, H. Herr, and G. Imbens, (2003), "Implementing Matching Estimators for Average Treatment Effects in STATA," *The STATA Journal*, Vol. 4(3): 290–311.

Abadie, A., and G. Imbens, (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, Vol. 74(1): 235–267.

Abadie, A., and G. Imbens, (2008), "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, Vol. 76(6): 1537–1557.

Abadie, A., and G. Imbens, (2009), "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business and Economic Statistics*, Vol. 29(1): 1–11.

Abadie, A., and G. Imbens, (2010), "Estimation of the Conditional Variance in Paired Experiments," *Annales d'Economie et de Statistique*, Vol. 91: 175–187.

Abadie, A., and G. Imbens, (2011), "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business and Economic Statistics*, Vol. 29(1): 1–11.

Abadie, A., and G. Imbens, (2012), "Matching on the Estimated Propensity Score," National Bureau of Economic Research Working paper 15301.

Abbring, J., and G. van den Berg, (2003), "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, Vol. 71(5): 1491–1517.

Althauser, R., and D. Rubin, (1970), "The Computerized Construction of a Matched Sample," *The American Journal of Sociology*, Vol. 76(2): 325–346.

Altman, D. (1991), *Practical Statistics for Medical Research*, Chapman and Hall/CRC.

Angrist, J. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, Vol. 80: 313–335.

Angrist, J. (1998), "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, Vol. 66(2): 249–288.

Angrist, J. D., and J. Hahn, (2004) "When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects," *Review of Economics and Statistics*, Vol. 86(1): 58–72.

Angrist, J., G. Imbens, and D. Rubin, (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, Vol. 91: 444–472.

Angrist, J., and A. Krueger, (1999), "Does Compulsory Schooling Affect Schooling and Earnings," *Quarterly Journal of Economics*, Vol. CVI(4): 979–1014.

Angrist, J. D., and A. B. Krueger, (2000), "Empirical Strategies in Labor Economics," in A. Ashenfelter and D. Card, eds. *Handbook of Labor Economics*, vol. 3. Elsevier Science.

Angrist, J. D., and G. M. Kuersteiner, (2011), "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score," *Review of Economics and Statistics*, Vol. 93(3): 725–747.

Angrist, J., and V. Lavy, (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, Vol. CXIV: 1243.

Angrist, J., and S. Pischke, (2008), *Mostly Harmless Econometrics: An Empiricists' Companion*, Princeton University Press.

Anscombe, F. J. (1948), "The Validity of Comparative Experiments," *Journal of the Royal Statistical Society, Series A*, Vol. 61: 181–211.

Ashenfelter, O. (1978), "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, Vol. 60: 47–57.

Ashenfelter, O., and D. Card, (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, Vol. 67: 648–660.

Athey, S., and G. Imbens, (2006), "Identification and Inference in Nonlinear Difference-In-Differences Models," *Econometrica*, Vol. 74(2): 431-497.

Athey, S., and G. Imbens, (2014), "Supervised Learning Methods for Causal Effects" Unpublished Manuscript.

Athey, S., and S. Stern, (1998), "An Empirical Framework for Testing Theories About Complementarity in Organizational Design," NBER working paper 6600.

Austin, P. (2008), "A Critical Appraisal of Propensity-Score Matching in the Medical Literature Between 1996 and 2003," *Statistics in Medicine*, Vol. 27: 2037–2049.

Baiocchi, M., D. Small, S. Lorch, and P. Rosenbaum, (2010), "Building a Stronger Instrument in an Observational Study of Perinatal Care for Premature Infants," *The Journal of the American Statistical Association*, Vol. 105(492): 1285–1296.

Baker, S. (2000), "Analyzing a Randomized Cancer Prevention Trial with a Missing Binary Outcome, an Auxiliary Variable, and All-or-None Compliance," *The Journal of the American Statistical Association*, Vol. 95(449): 43–50.

Ball, S., G. Bogatz, D. Rubin, and A. Beaton, (1973), "Reading with Television: An Evaluation of The Electric Company. A Report to the Children's Television Workshop," Vol 1 and 2, Educational Testing Service, Princeton NJ.

Barnard, J., J. Du, J. Hill, and D. Rubin, (1998), "A Broader Template for Analyzing Broken Randomized Experiments," *Sociological Methods & Research*, Vol. 27: 285–317.

Barnow, B. S., G. G. Cain, and A. S. Goldberger, (1980), "Issues in the Analysis of Selectivity Bias," in E. Stromsdofer and G. Farkas, eds. *Evaluation Studies*, vol. 5, Sage.

Becker, S., and A. Ichino, (2002), "Estimation of Average Treatment Effects Based on Propensity Scores," *The Stata Journal*, Vol. 2(4): 358–377.

Beebee, H., C. Hitchcock, and P. Menzies, (2009), *The Oxford Handbook of Causation*, Oxford University Press.

Belloni, A., V. Chernozhukov, and C. Hansen, (2014), "Inference on Treatment Effects After Selection Amongst High-Dimensional Controls," *The Review of Economic Studies*, Vol. 81(2): 608–650.

Bertrand, M., and S. Mullainathan, (2004), "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, Vol. 94(4): 991–1013.

Bitler, M., J. Gelbach, and H. Hoynes, (2002), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, Vol. 96(4): 988-1012.

Björklund, A., and R. Moffitt, (1987), "The Estimation of Wage Gains and Welfare Gains in Self–Selection Models," *Review of Economics and Statistics*, Vol. LXIX: 42–49.

Black, S. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, Vol. CXIV: 577.

Bloom, H. (1984), "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, Vol. 8: 225–246.

Blundell, R., and M. Costa-Dias, (2000), "Evaluation Methods for Non-Experimental Data," *Fiscal Studies*, Vol. 21(4): 427–468.

Blundell, R., and M. Costa-Dias, (2002), "Alternative Approaches to Evaluation in Empirical Microeconomics," *Portuguese Economic Journal*, Vol. 1(1): 91–115.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir, (2007), "Changes in the Distribution of Male and Female Wages Accounting for the Employment Composition," *Econometrica*, Vol. 75(2): 323–363.

Box, G., S. Hunter, and W. Hunter, (2005), *Statistics for Experimenters: Design, Innovation and Discovery*, Wiley.

Box, G., and G. Tiao, (1973), *Bayesian Inference in Statistical Analysis*, Addison Wesley.

Breiman, L., and P. Spector, (1992), "Submodel Selection and Evaluation in Regression: The x-Random Case," *International Statistical Review*, Vol. 60: 291–319.

Brillinger, D. R., L. V. Jones, and J. W. Tukey, (1978), "Report of the statistical task force for the weather modification advisory board." *The Management of Western Resources, Vol. II: The Role of Statistics on Weather Resources Management.* Stock No. 003-018-00091-1, Government Printing Office, Washington, DC.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng, (2011), *Handbook of Markov Chain Monte Carlo*, Chapman and Hall.

Bühlman, P., and S. van der Geer, (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Verlag.

Busso, M., J. DiNardo, and J. McCrary, (2009), "New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators," Unpublished Working Paper.

Caliendo, M. (2006), *Microeconometric Evaluation of Labour Market Policies*, Springer Verlag.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in Christofides, E. K. Grant, and R. Swidinsky, ed. *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, University of Toronto Press.

Card, D., and A. Krueger, (1994), "Minimum Wages and Employment: A Case Study of the Fast-food Industry in New Jersey and Pennsylvania," *American Economic Review*, Vol. 84(4): 772–784.

Card, D., and D. Sullivan, (1988), "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, Vol. 56(3), 497–530.

Chernozhukov, V., and C. Hansen, (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, Vol. 73(1): 245–261.

Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Schanzenbach, and D. Yagan, (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, Vol. 126(4): 1593–1660.

Clogg, C., D. Rubin, N. Schenker, B. Schultz, and L. Weidman, (1991), "Multiple Imputation of Industry and Occupation Codes in Census Public-Use Samples Using Bayesian Logistic Regression,", *Journal of the American Statistical Association*, Vol. 86(413): 68–78.

Cochran, W. G. (1965), "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society, Series A (General)*, Vol. 128(2): 234–266.

Cochran, W. G. (1968) "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, Vol. 24: 295–314.

Cochran, W. G. (1977), *Sampling Techniques*, Wiley.

Cochran, W. G., and G. Cox, (1957), *Experimental Design*, Wiley Classics Library.

Cochran, W. G., and D. Rubin, (1973), "Controlling Bias in Observational Studies: A Review," *Sankhya*, Vol. 35: 417–46.

Cook, T. (2008), "'Waiting for Life to Arrive': A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics," *Journal of Econometrics*, Vol. 142(2): 636–654.

Cook, T., and D. DeMets, (2008), *Introduction to Statistical Methods for Clinical Trials*, Chapman and Hall/CRC.

Cornfield et al. (1959), "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions," *Journal of the National Cancer Institute*, Vol. 22: 173–203.

Cox, D. (1956), "A Note on Weighted Randomization," *The Annals of Mathematical Statistics*, Vol. 27(4): 1144-1151.

Cox, D. (1958), *Planning of Experiments*, Wiley Classics Library.

Cox, D. (1992), "Causality: Some Statistical Aspects," *Journal of the Royal Statistical Society*, *Series A*, Vol. 155: 291–301.

Cox, D., and P. McCullagh, (1982), "Some Aspects of Covariance," (with discussion). *Biometrics*, Vol. 38: 541–561.

Cox, D., and N. Reid, (2000), *The Theory of the Design of Experiments*, Chapman and Hall/CRC.

Crump, R., V. J. Hotz, G. Imbens, and O. Mitnik, (2008), "Nonparametric Tests for Treatment Effect Heterogeneity," *Review of Economics and Statistics*, Vol. 90(3): 389–405.

Crump, R., V. J. Hotz, G. Imbens, and O. Mitnik, (2009), "Dealing with Limited Overlap in Estimation of Average Treatment Effects," *Biometrika*, Vol. 96: 187–99.

Cuzick, J., R. Edwards, and N. Segnan, (1997), "Adjusting for Non-Compliance and Contamination in Randomized Clinical Trials," *Statistics in Medicine,* Vol. 16: 1017–1039.

Darwin, C. (1876), *The Effects of Cross- and Self-Fertilisation in the Vegetable Kingdom*, John Murry.

Davies, O. (1954), *The Design and Analysis of Industrial Experiments*, Oliver and Boyd.

Dawid, P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society*, *Series B*, Vol. 41(1): 1–31.

Dawid, P. (2000), "Causal Inference Without Counterfactuals," *Journal of the American Statistical Association*, Vol. 95(450): 407–424.

Deaton, A. (2010), "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, Vol. 48(2): 424–455.

Dehejia, R. (2002), "Was There a Riverside Miracle? A Hierarchical Framework for Evaluating Programs with Grouped Data," *Journal of Business and Economic Statistics*, Vol. 21(1): 1–11.

Dehejia, R. (2005a), "Practical Propensity Score Matching: A Reply to Smith and Todd," *Journal of Econometrics*, Vol. 125: 355–364.

Dehejia, R. (2005b), "Program Evaluation as a Decision Problem," *Journal of Econometrics*, Vol. 125: 141–173.

Dehejia, R., and S. Wahba, (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, Vol. 94: 1053–1062.

Dehejia, R., and S. Wahba, (2002), "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, Vol. 84(1): 151–161.

Diaconis, P. (1976), "Finite Forms of de Finetti's Theorem on Exchangeability," Technical Report 84, Department of Statistics, Stanford University.

Diamond, A., and J. Sekhon, (2013), "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," *Review of Economics and Statistics*, Vol. 95(3): 932–945.

Diehr, P., D. Martin, T. Koepsell, and A. Cheadle, (1995), "Breaking the Matches in a Paired *t*-Test for Community Interventions When the Number of Pairs is Small," *Statistics in Medicine*, Vol. 14: 1491–1504.

Donner, A. (1987), "Statistical Methodology for Paired Cluster Designs," *American Journal of Epidemiology,* Vol. 126(5), 972–979.

Du, J. (1998) "Valid Inferences After Propensity Score Subclassification Using Maximum Number of Subclasses as Building Blocks," PhD Thesis, Department of Statistics, Harvard University.

Duflo, E., R. Hanna, and S. Ryan, (2012), "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, Vol. 102(4): 1241–1278.

Efron, B., and D. Feldman, (1992), "Compliance as an Explanatory Variable in Clinical Trials," *Journal of the American Statistical Association*, Vol. 86(413): 9–17.

Efron, B., and R. Tibshirani, (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Engle, R., D. Hendry, and J.-F. Richard, (1974), "Exogeneity," *Econometrica*, Vol. 51(2): 277–304.

Espindle, L. (2004), "Improving Confidence Coverage for the Estimate of the Treatment Effect in a Subclassification Setting," Undergraduate Thesis, Department of Statistics, Harvard University.

de Finetti, B. (1964), "Foresight: Its Logical Laws, Its Subjective Sources," in Kyburg and Smokler, eds. *Studies in Subjective Probability*, Wiley.

de Finetti, B. (1992), *Theory of Probability: A Critical Introductory Treatment*, Vol. 1 & 2, Wiley Series in Probability & Mathematical Statistics.

Feller, W. (1965), *An Introduction to Probability and its Applications*, Vol. 1, John Wiley and Sons, New York City.

Firpo, S. (2003), "Efficient Semiparametric Estimation of Quantile Treatment Effects," PhD Thesis, Chapter 2, Department of Economics, University of California, Berkeley.

Firpo, S. (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects,"*Econometrica*, Vol. 75(1): 259-276.

Fisher, L., D. Dixon, J. Herson, R. Frankowski, M. Hearron, and K. Peace, (1990), "Intention to Treat in Clinical Trials," in Peace, ed. *Statistical Issues in Drug Research and Development*, Marcel Dekker, Inc.

Fisher, R. A. (1918), "The Causes of Human Variability," *Eugenics Review*, Vol. 10: 213–220.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, 1st ed, Oliver and Boyd.

Fisher, R. A. (1935), *Design of Experiments*, Oliver and Boyd.

Fisher, R., and W. MacKenzie, (1923), "Studies in Crop Vacation. II. The Manurial Response of Different Potato Varieties," *Journal of Agricultural Science,* Vol. 13: 311–320.

Fisher, L. et al. (1990), "Intention-to-Treat in Clinical Trials," in Peace ed., *Statistical Issues in Drug Research and Development*, Marcel Dekker.

Fraker, T., and R. Maynard, (1987), "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs," *Journal of Human Resources*, Vol. 22(2): 194–227.

Frangakis, C., and D. Rubin, (2002), "Principal Stratification," *Biometrics*, Vol. 58(1): 21–29.

Freedman, D. A. (2006), "Statistical Models for Causation: What Inferential Leverage Do They Provide," *Evaluation Review* , Vol. 30(6): 691–713.

Freedman, D. A. (2008a), "On Regression Adjustments to Experimental Data, " *Advances in Applied Mathematics* , Vol. 30(6): 180–193.

Freedman, D. A. (2008b), "On Regression Adjustments in Experiments with Several Treatments," *Annals of Applied Statistics*, Vol. 2: 176–196.

Freedman, D. A. (2009), in D. Collier, J. S. Sekhon, and P. B. Stark, eds. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*, Cambridge University Press.

Freedman, D. A., R. Pisani, and R. Purves, (1978), *Statistics*, Norton.

Friedlander, D., and J. Gueron, (1995), "Are High-Cost Services More Effective Than Low-Cost Services," in C. Manski and I. Garfinkel, eds. *Evaluating Welfare and Training Programs*, Harvard University Press, pp. 143–198.

Friedlander, D., and P. Robins, (1995), "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods," *American Economic Review*, Vol. 85: 923–937.

Frölich, M. (2000), "Treatment Evaluation: Matching versus Local Polynomial Regression," Discussion paper 2000-17, Department of Economics, University of St. Gallen.

Frölich, M. (2004a), "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *The Review of Economics and Statistics*, Vol. 86(1): 77–90.

Frölich, M. (2004b), "A Note on the Role of the Propensity Score for Estimating Average Treatment Effects," *Econometric Reviews*, Vol. 23(2): 167–174.

Frumento, P., F. Mealli, B. Pacini, and D. Rubin, (2012), "Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data," *Journal of the American Statistical Association*, No. 498: 450–466.

Gail, M. H., S. Mark, R. Carroll, S. Green, and D. Pee, (1996), "On Design Considerations and Randomization-based Inference for Community Intervention Trials," *Statistics in Medicine*, Vol. 15: 1069–1092.

Gail, M. H., W. Tian, and S. Piantadosi, (1988), "Tests for No Treatment Effect in Randomized Clinical Trials," *Biometrika*, Vol. 75(3): 57–64.

Gail, M. H., S. Wieand, and S. Piantadosi, (1984), "Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates," *Biometrika*, Vol. 71(3): 431–444.

Gelman, A., J. Carlin, H. Stern, and D. Rubin, (1995), *Bayesian Data Analysis*, Chapman and Hall.

Gelman, A., and J. Hill, (2006), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Gill, R., and J. Robins, (2001), "Causal Inference for Complex Longitudinal Data: The Continuous Case," *Annals of Statistics*, Vol. 29(6): 1785–1811.

Goldberger, A. (1991), *A Course in Econometrics*, Harvard University Press.

Graham, B. (2008), "Identifying Social Interactions through Conditional Variance Restrictions," *Econometrica*, Vol. 76(3): 643–660.

Granger, C. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, Vol. 37(3): 424–438.

Greene, W. (2011), *Econometric Analysis*, 7th Edition, Prentice Hall.

Gu, X., and P. Rosenbaum, (1993), "Comparison of Multivariate Matching Methods: Structures, Distances and Algorithms," *Journal of Computational and Graphical Statistics*, Vol. 2: 405–420.

Guo, S., and M. Fraser, (2010), *Propensity Score Analysis*, Sage Publications.

Gutman, R., and D. Rubin, (2014), "Robust Estimation of Causal Effects of Binary Treatments in Unconfounded Studies with Dichotomous Outcomes," *Statistics in Medicine*, forthcoming.

Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, Vol. 11(1):1–12.

Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Econometrica*, Vol. 11.

Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, Vol. 66(2): 315–331.

Hahn, J., P. Todd, and W. VanderKlaauw, (2000), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, Vol. 69(1): 201–209.

Hainmueller, J. (2012), "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, Vol. 20: 25–46.

Ham, J., and R. Lalonde, (1996), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, Vol. 64: 1.

Hansen, B. (2007), "Optmatch: Flexible, Optimal Matching for Observational Studies," *R News*, Vol. 7(2): 18–24.

Hansen, B. (2008), "The Prognostic Analogue of the Propensity Score," *Biometrika*, Vol. 95(2): 481–488.

Hansen, B., and S. Klopfer, (2006), "Optimal Full Matching and Related Designs via Network Flows," *Journal of Computational and Graphical Statistics*, Vol. 15(3): 609–627.

Hartigan, J. (1983), *Bayes Theory*, Springer Verlag.

Hartshorne, C., and P. Weiss (Eds.), (1931), *Collected Papers of Charles Sanders Peirce* (Vol. 1), Harvard University Press.

Hearst, N., Newman, T., and S. Hulley, (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment," *New England Journal of Medicine*, Vol. 314 (March 6): 620–624.

Heckman, J., and J. Hotz, (1989), "Alternative Methods for Evaluating the Impact of Training Programs" (with discussion), *Journal of the American Statistical Association*, Vol. 84(804): 862–874.

Heckman, J., H. Ichimura, and P. Todd, (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, Vol. 64: 605–654.

Heckman, J., H. Ichimura, and P. Todd, (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, Vol. 65: 261–294.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, Vol. 66: 1017–1098.

Heckman, J., R. Lalonde, and J. Smith, (2000), "The Economics and Econometrics of Active Labor Markets Programs," in Ashenfelter and Card, eds. *Handbook of Labor Economics*, vol. 3, Elsevier Science.

Heckman, J., and R. Robb, (1984), "Alternative Methods for Evaluating the Impact of Interventions," in Heckman and Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge University Press.

Heckman, J., and E. Vytlacil, (2007a), "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in J. Heckman and E. Leamer, eds. *Handbook of Econometrics*, vol. 6B, Chapter 70, 4779-4874, Elsevier Science.

Heckman, J., and E. Vytlacil, (2007b), "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in J. Heckman and E. Leamer, eds. Handbook of Econometrics, vol. 6B, Chapter 71, 4875-5143, Elsevier Science.

Heitjan, D., and R. Little, (1991), "Multiple Imputation for the Fatal Accident Reporting System," *Applied Statistics*, Vol. 40: 13–29.

Hendry, D., and M. Morgan, (1995), *The Foundations of Econometric Analysis*, Cambridge University Press.

Hewitt, E., and L. Savage, (1955), "Symmetric Measures on Cartesian Products,"*Transactions of the American Mathematical Society*, Vol. 80: 470-501.

Hinkelmann, K., and O. Kempthorne, (2005), *Design and Analysis of Experiments*, Vol. 2, Advance Experimental Design, Wiley.

Hinkelmann, K., and O. Kempthorne, (2008), *Design and Analysis of Experiments*, Vol. 1, Introduction to Experimental Design, Wiley.

Hirano, K., and G. Imbens, (2001), "Estimation of Causal Effects Using Propensity Score Weighting: An Application of Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, Vol. 2: 259–278.

Hirano, K., and G. Imbens, (2004), "The Propensity Score with Continuous Treatments," in Gelman and Meng, eds. *Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives*, Wiley.

Hirano, K., G. Imbens, and G. Ridder, (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, Vol. 71(4): 1161–1189.

Hirano, K., G. Imbens, D. Rubin, and A. Zhou, (2000), "Estimating the Effect of Flu Shots in a Randomized Encouragement Design," *Biostatistics*, Vol. 1(1): 69–88.

Ho, D., and K. Imai, (2006), "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election," *Journal of the American Statistical Association*, Vol. 101(476): 888–900.

Ho, D., K. Imai, G. King, and E. Stuart, (2007), "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, Vol. 81: 945–970.

Hodges, J. L., and Lehmann, E., (1970), *Basic Concepts of Probability and Statistics*, 2nd ed., Holden-Day.

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

Holland, P. (1988), "Causal Inference, Path Analysis, and Recursive Structural Equations Models," (with discussion), *Sociological Methodology*, Vol. 18: 449–484.

Holland, P., and D. Rubin, (1982), "Introduction: Research on Test Equating Sponsored by Educational Testing Service, 1978–1980," in *Test Equating*, Academic Press, Inc. pp. 1–6.

Holland, P., and D. Rubin, (1983), "On Lord's Paradox," in Wainer and Messick, eds. *Principles of Modern Psychological Measurement: A Festschrift for Frederick Lord*, Erlbaum, pp. 3–25.

Hood, W., and T. Koopmans, (1953), *Studies in Econometric Method*, Wiley, New York.

Horowitz, J. (2002), "The Bootstrap," in Heckman and Leamer, eds. *Handbook of Econometrics*, Vol. 5, Elsevier.

Horvitz, D., and D. Thompson, (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, Vol. 47: 663–685.

Hotz, V. J., G. Imbens, and J. Klerman, (2001), "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program," *Journal of Labor Economics,* Vol. 24(3): 521–566.

Hotz, J., G. Imbens, and J. Mortimer, (2005), "Predicting the Efficacy of Future Training Programs Using Past Experiences," *Journal of Econometrics*, Vol. 125: 241–270.

Huber, M., M. Lechner, and C. Wunsch, (2012), "The Performance of Estimators Based on the Propensity Score," *Journal of Econometrics*, Vol. 175(1): 1–21.

Imai, K. (2008), Variance Identification and Efficiency Analysis in Randomized Experiments under the Matched-Pair Design." *Statistics in Medicine*, Vol. 27(24) (October): 4857–4873.

Imai, K., and D. van Dyk, (2004), "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association*, Vol. 99: 854–866.

Imai, K., G. King, and E. A. Stuart, (2008), "Misunderstandings among Experimentalists and Observationalists about Causal Inference," *Journal of the Royal Statistical Society*, *Series A (Statistics in Society)*, Vol. 171(2): 481-502.

Imbens, G. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika*, Vol. 87(3): 706–710.

Imbens, G. (2003), "Sensitivity to Exogeneity Assumptions in Program Evaluation," *American Economic Review*, Papers and Proceedings.

Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, Vol. 86(1): 1–29.

Imbens, G. (2010), "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, Vol. 48(2): 399–423.

Imbens, G. (2011), "On the Finite Sample Benefits of Stratification, Blocking and Pairing in Randomized Experiments," Unpublished Manuscript.

Imbens, G. (2014), "Instrumental Variables: An Econometrician's Perspective," *Statistical Science*, Vol. 29(3): 375–379.

Imbens, G., (2015), "Matching Methods in Practice: Three Examples," forthcoming, *Journal of Human Resources.*

Imbens, G., and J. Angrist, (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, Vol. 61(2): 467–476.

Imbens, G., and K. Kalyanaraman, (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator Review of Economic Studies," *Review of Economic Studies*, Vol. 79(3): 933–959.

Imbens, G., and T. Lemieux, (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, Vol. 142(2): 615–635.

Imbens, G., and P. Rosenbaum, (2005), "Randomization Inference with an Instrumental Variable," *Journal of the Royal Statistical Society*, *Series A*, Vol. 168(1): 109–126.

Imbens, G., and D. Rubin, (1997a), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economic Studies*, Vol. 64(3): 555–574.

Imbens, G., and D. Rubin, (1997b), "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance," *Annals of Statistics*, Vol. 25(1): 305–327.

Imbens, G., and J. Wooldridge, (2009), "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, Vol. 47(1): 1–81.

Jin, H., and D. B. Rubin, (2008), "Principal Stratification for Causal Inference with Extended Partial Compliance: Application to Efron-Feldman Data," *Journal of the American Statistical Association*, Vol. 103: 101–111.

Kane, T., and C. Rouse, (1995), "Labor-Market Returns to Two- and Four- Year College," *American Economic Review*, Vol. 85(3): 600–614.

Kang, J., and J. Schafer, (2007), "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data," *Statistical Science*, Vol. 22(4): 523–539.

Kempthorne, O. (1952), *The Design and Analysis of Experiments*, Robert Krieger Publishing Company.

Kempthorne, O. (1955), "The Randomization Theory of Experimental Evidence," *Journal of the American Statistical Association*, Vol. 50927(1): 946–967.

Ketel, N., E. Leuven, H. Oosterbeek, and B. VanderKlaauw, (2013), "The Returns to Medical School in a Regulated Labor Market: Evidence from Admission Lotteries," Unpublished Manuscript.

Koch, G., C. Tangen, J. W Jung, and I. Amara, (1998), "Issues for Covariance Analysis of Dichotomous and Ordered Categorical Data from Randomized Clinical Trials and Non-Parametric Strategies for Addressing Them," *Statistics in Medicine*, Vol. 17: 1863–1892.

Koopmans, T., (1950), *Statistical Inference in Dynamic Economic Models*, Wiley, New York.

Krueger, A. (1999), "Experimental Estimates of Education Production Functions Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, Vol. 114(2): 497-532.

Lalonde, R. J., (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, Vol. 76: 604–620.

Lancaster, T. (2004), *An Introduction to Modern Bayesian Econometrics*, Blackwell Publishing.

Leamer, E. (1988), "Discussion on Marini, Singer, Glymour, Scheines, Spirtes, and Holland," *Sociological Methodology*, Vol. 18: 485-493.

Lechner, M. (1999), "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany After Unification," *Journal of Business and Economic Statistics*, Vol. 17(1): 74–90.

Lechner, M. (2001), "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption," in Lechner and Pfeiffer, eds. *Econometric Evaluations of Active Labor Market Policies in Europe*, Heidelberg.

Lechner, M. (2002), "Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies," *Review of Economics and Statistics*, Vol. 84(2): 205–220.

Lechner, M. (2008), "A Note on the Common Support Problem in Applied Evaluation Studies," *Annales d'Economie et de Statistique*, Vol. 91-92: 217–234.

Lee, D. (2008), "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, Vol. 142(2): 675–697.

Lee, D., and T. Lemieux, (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, Vol. 48(2): 281–355.

Lee, M.-J. (2005), *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press.

Lehman, E. (1974), *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day.

Lesaffre, E., and S. Senn, (2003), "A Note on Non-Parametric ANCOVA for Covariate Adjustment in Randomized Clinical Trials," *Statistics in Medicine*, Vol. 22: 3583–3596.

Lin, W. (2012), "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *Annals of Applied Statistics*.

Lindley, D. V., and N. R. Novick, (1981), "The Role of Exchangeability in Inference," *Annals of Statistics*, Vol. 9: 45–58.

Little, R., and D. Rubin, (2002), *Statistical Analysis with Missing Data*, Wiley.

Lord, F. (1967), "A Paradox in the Interpretation of Group Comparisons," *Psychological Bulletin*, Vol. 68: 304–305.

Lui, K.-J. (2011), *Binary Data Analysis of Randomized Clinical Trials with Noncompliance*, Wiley, Statistics in Practice.

Lynn, H., and C. McCulloch, (1992), "When Does It Pay to Break the Matches for Analysis of a Matched-pair Design," *Biometrics*, Vol. 48: 397–409.

McCarthy, M. D. (1939), "On the Application of the z-Test to Randomized Blocks," *Annals of Mathematical Statistics*, Vol. 10: 337.

McClellan, M., and J. P. Newhouse, (1994), "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality," *Journal of the American Medical Association*, Vol. 272(11): 859–866.

McDonald, C., S. Hiu, and W. Tierney, (1992), "Effects of Computer Reminders for Influenza Vaccination on Morbidity During Influenza Epidemics," *MD Computing*, Vol. 9: 304–312.

McNamee, R. (2009), "Intention to Treat, Per Protocol, as Treated and Instrumental Variable Estimators Given Non-Compliance and Effect Heterogeneity," *Statistics in Medicine*, Vol. 28: 2639–2652.

Mann, H. B., and D. R. Whitney, (1947), "On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other," *Annals of Mathematical Statistics*, Vol. 18(1): 50–60.

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, Vol. 80: 319–323.

Manski, C. (1996), "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *The Journal of Human Resources*, Vol. 31(4): 709–773.

Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer-Verlag.

Manski, C. (2013), *Public Policy in an Uncertain World*, Harvard University Press.

Manski, C., G. Sandefur, S. McLanahan, and D. Powers, (1992), "Alternative Estimates of the Effect of Family Structure During Adolescence on High School," *Journal of the American Statistical Association*, Vol. 87(417): 25–37.

Marini, M., and B. Singer, (1988), "Causality in the Social Sciences," *Sociological Methodology*, Vol. 18: 347–409.

Mealli, F., and D. Rubin, (2002a), "Assumptions When Analyzing Randomized Experiments with Noncompliance and Missing Outcomes," *Health Services Outcome Research Methodology*, Vol. 3: 225–232.

Mealli, F., and D. Rubin, (2002b), "Discussion of Estimation of Intervention Effects with Noncompliance: Alternative Model Specification by Booil Jo," *Journal of Educational and Behavioral Statistics*, Vol. 27: 411–415.

Meier, P. (1991), "Compliance as an Explanatory Variable in Clinical Trials: Comment," *Journal of the American Statistical Association*, Vol. 86(413):19–22.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, and M. Van der Laan, (1991), "Promoting Transparency in Social Science Research," *Science*, Vol. 343(6166): 30–31.

Mill, J. S. (1973), *A system of logic* , In Collected Works of John Stuart Mill, University of Toronto Press.

Miratrix, L., J. Sekhon, and B. Yu, (2013), "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments," *Journal of the Royal Statistical Society*, Series B, 75, 369–396.

Morgan, K., and D. Rubin, (2012), "Rerandomization to Improve Covariate Balance in Experiments," *Annals of Statistics*, Vol. 40(2): 1263–1282.

Morgan, S. (2013), *Handbook of Causal Analysis for Social Research*, Springer.

Morgan, S., and C. Winship, (2007), *Counterfactuals and Causal Inference*, Cambridge University Press.

Morris, C., and J. Hill, (2000), "The Health Insurance Experiment: Design Using the Finite Selection Model," *Public Policy and Statistics: Case Studies from RAND 2953*. Springer, New York.

Morton, R., and K. Williams, (2010), *Experimental Political Science and the Study of Causality*, Cambridge University Press.

Mosteller, F. (1995), "The Tennessee Study of Class Size in the Early School Grades," *The Future of Children: Critical Issues for Children and Youths*, V(1995): 113–127.

Murnane, R., and J. Willett, (2011), *Methods Matter: Improving Causal Inference in Educational and Social Science Research*, Oxford University Press.

Murphy, D., and L. Cluff, (1990), "SUPPORT: Study to understand prognoses and preferences for outcomes and risks of treatments: study design," *Journal of Clinical Epidemiology*, Vol. 43: 1S–123S.

Neyman, J. (1923, 1990), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," translated in *Statistical Science*, (with discussion), Vol. 5(4): 465–480, 1990.

Neyman, J. (1934), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, Vol. 97(4): 558-625.

Neyman, J. with the cooperation of K. Iwaskiewicz and St. Kolodziejczyk, (1935), "Statistical Problems in Agricultural Experimentation" (with discussion), Supplement, *Journal of the Royal Statistical Society*, *Series B*, Vol. 2: 107–180.

Pattanayak, C., D. Rubin, and E. Zell, (2011), "Propensity Score Methods for Creating Covariate Balance in Observational Studies," *Review of Experimental Cardiology*, Vol. 64(10): 897-903.

Paul, I., J. Beiler, A. McMonagle, M. Shaffer, L. Duda, and C. Berlin, (2007), "Effect of Honey, Dextromerhorphan, and No Treatment on Nocturnal Cough and Sleep Quality for Coughing Childre and Their Parents," *Archives of Pediatric and Adolescent Medicine*, Vol. 161(12): 1140–1146.

Pearl, J. (1995), "Causal Diagrams for Empirical Research," *Biometrika*, Vol. 82: 669–688.

Pearl, J. (2000, 2009), *Causality: Models, Reasoning and Inference*, Cambridge University Press.

Peirce, C., and J. Jastrow, (1885), "On Small Differences in Sensation," *Memoirs of the National Academy of Sciences*, Vol. 3: 73–83.

Peters, C., and W. van Vorhis, (1941), *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill.

Politis, D., and J. Romano, (1999), *Subsampling*, Springer Verlag.

Porter, J. (2003), "Estimation in the Regression Discontinuity Model," Unpublished Manuscript, Harvard University.

Powers, D., and S. Swinton, (1984), "Effects of Self-Study for Coachable Test Item Types," *Journal of Educational Measurement*, Vol. 76: 266–278.

Quade, D. (1982), "Nonparametric Analysis of Covariance by Matching," *Biometrics*, Vol. 38: 597–611.

Reid, C. (1998), *Neyman from Life*, Springer.

Reinisch, J., S. Sanders, E. Mortensen, and D. Rubin, (1995), "In Utero Exposure to Phenobarbital and Intelligence Deficits in Adult Men," *The Journal of the American Medical Association*, Vol. 274(19): 1518-1525.

Robert, C. (1994), *The Bayesian Choice*, Springer Verlag.

Robert, C., and G. Casella, (2004), *Monte Carlo Statistical Methods*, Springer Verlag.

Robins, J. (1986), "A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods - Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, Vol. 7: 1393–1512.

Robins, J., and Y. Ritov, (1997), "Towards a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-parametric Models," *Statistics in Medicine*, Vol. 16: 285–319.

Robins, J. M., and A. Rotnitzky, (1995), "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, Vol. 90: 122–129.

Robins, J. M., Rotnitzky, A., and Zhao, L.-P., (1995), "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, Vol. 90: 106–121.

Romer, C. D., and D. H. Romer, (2004), "A New Measure of Monetary Shocks: Derivation and Implications," *The American Economic Review*, Vol. 94(4): 1055–1084.

Rosenbaum, P. (1984a), "Conditional Permutation Tests and the Propensity Score in Observational Studies," *Journal of the American Statistical Association*, Vol. 79: 565–574.

Rosenbaum, P. (1984b), "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society*, *Series A*, Vol. 147: 656–666.

Rosenbaum, P. (1987), "The Role of a Second Control Group in an Observational Study," *Statistical Science*, (with discussion), Vol. 2(3), 292–316.

Rosenbaum, P. (1988), "Permutation Tests for Matched Pairs," *Applied Statistics*, Vol. 37: 401–411.

Rosenbaum, P. (1989a), "Optimal Matching in Observational Studies," *Journal of the American Statistical Association*, 84, 1024–1032.

Rosenbaum, P. (1989b), "On Permutation Tests for Hidden Biases in Observational Studies: An Application of Holley's Inequality to the Savage Lattice," *Annals of Statistics*, Vol. 17: 643–653.

Rosenbaum, P. (1995, 2002), *Observational Studies*, Springer Verlag.

Rosenbaum, P. (2009), *Design of Observational Studies*, Springer Verlag.

Rosenbaum, P. (2002), "Covariance Adjustment in Randomized Experiments and Observational Studies," *Statistical Science*, Vol. 17(3): 286–304.

Rosenbaum, P., and D. Rubin, (1983a), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70: 41–55.

Rosenbaum, P., and D. Rubin, (1983b), "Assessing the Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," *Journal of the Royal Statistical Society*, *Series B*, Vol. 45: 212–218.

Rosenbaum, P., and D. Rubin, (1984), "Reducing the Bias in Observational Studies Using Sub-classification on the Propensity Score," *Journal of the American Statistical Association*, Vol. 79: 516–524.

Rosenbaum, P., and D. Rubin, (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician*, Vol. 39: 33–38.

Rubin, D. B. (1973a), "Matching to Remove Bias in Observational Studies," *Biometrics*, Vol. 29: 159–183.

Rubin, D. B. (1973b), "The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies," *Biometrics*, Vol. 29: 185–203.

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, Vol. 66: 688–701.

Rubin, D. B. (1975), "Bayesian Inference for Causality: The Importance of Randomization," *Proceedings of the Social Statistics Section of the American Statistical Association*, 233–239.

Rubin, D. B. (1976a), "Multivariate Matching Methods That Are Equal Percent Bias Reducing, I: Some Examples," *Biometrics*, Vol. 32(1): 109–120.

Rubin, D. B. (1976b), "Multivariate Matching Methods That Are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes," *Biometrics*, Vol. 32(1): 121–132.

Rubin, D. B. (1976c), "Inference and Missing Data," *Biometrika*, (with discussion and reply), Vol. 63(3): 581–592.

Rubin, D. B. (1977), "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, Vol. 2(1): 1–26.

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, Vol. 6: 34–58.

Rubin, D. B. (1979), "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, Vol. 74: 318–328.

Rubin, D. B. (1980a), "Discussion of "Randomization Analysis of Experimental Data in the Fisher Randomization Test" by Basu," *The Journal of the American Statistical Association*, Vol. 75(371): 591–593.

Rubin, D. B. (1980b), "Bias Reduction Using Mahalanobis' Metric Matching," *Biometrics*, Vol. 36(2): 293–298.

Rubin, D. B. (1986a), "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, Vol. 81(396): 961–962.

Rubin, D. B. (1986b), "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations," *Journal of Business and Economic Statistics*, Vol. 4(1): 87–94.

Rubin, D. B. (1990a), "Formal Modes of Statistical Inference for Causal Effects," *Journal of Statistical Planning and Inference*, Vol. 25: 279–292.

Rubin, D. B. (1990b), "Comment on Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, Vol. 5(4): 472–480.

Rubin, D. B. (1998), "More Powerful Randomization-Based p-Values in Double-Blind Trials with Non-Compliance," *Statistics in Medicine*, Vol. 17: 371–385.

Rubin, D. B. (2001), "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services & Outcomes Research Methodology,* Vol. 2: 169–188.

Rubin, D. B. (2004), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," Fisher Lecture, *The Journal of the American Statistical Association*, Vol. 100(469): 322–331.

Rubin, D. B. (2006), *Matched Sampling for Causal Effects*, Cambridge University Press.

Rubin, D. B. (2007), "The Design versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials," *Statistics in Medicine*, Vol. 26(1): 20–30.

Rubin, D. B. (2008), "The Design and Analysis of Gold Standard Randomized Experiments. Comment on 'Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random to Nonrandom Assignment' by Shadish, Clark, and Steiner," *Journal of the American Statistical Association*, Vol. 103: 1350–1353.

Rubin, D. B. (2010), "Reflections Stimulated by the Comments of Shadish (2010) and West and Thoemmes (2010)," *Psychological Methods*, Vol. 15(1): 38–46.

Rubin, D. B. (2012), "Analyses That Inform Policy Decisions," *Biometrics*, Vol. 68: 671–775.

Rubin, D., and E. Stuart, (2006), "Affinely Invariant Matching Methods with Discriminant Mixtures of Ellipsoidally Symmetric Distributions," *Annals of Statistics*, Vol. 34(4): 1814–1826.

Rubin, D. B., and N. Thomas, (1992a), "Affinely Invariant Matching Methods with Ellipsoidal Distributions," *Annals of Statistics*, Vol. 20(2): 1079–1093.

Rubin, D. B., and N. Thomas, (1992b), "Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Distributions," *Biometrika*, Vol. 79(4): 797–809.

Rubin, D. B., and N. Thomas, (1996), "Matching Using Estimated Propensity Scores: Relating Theory to Practice," *Biometrics* 52: 249-264.

Rubin, D. B., and N. Thomas, (2000), "Combining Propensity Score Matching with Additional Adjustment for Prognostic Covariates," *Journal of the American Statistical Association*, Vol. 95(450): 573–585.

Rubin, D. B., X. Wang, L. Yin, and E. Zell, (2010), "Bayesian Causal Inference: Approaches to Estimating the Effect of Treating Hospital Type on Cancer Survival in Sweden Using Principal Stratification," in A. O'Hagen and M. West, eds. *The Handbook of Applied Bayesian Analysis*, Chapter 24, pp. 679–706.

Rubin, D. B., and E. Zell, (2010), "Dealing with Noncompliance and Missing Outcomes in a Randomized Trial using Bayesian Technology: Prevention of Perinatal Sepsis Clinical Trial, Soweto, South Africa," *Statistical Methodology*, Vol. 7(3): 338–350.

Sabbaghi, A., and D. Rubin, (2014), "Comments on the Neyman-Fisher Controversy and Its Consequences," *Statistical Science*, Vol. 29(2): 267–284.

Samii, C., and P. Aronow, (2012), "Equivalencies Between Design-Based and Regression-Based Variance Estimators for Randomized Experiments," *Statistics and Probability Letters*, Vol. 82: 365–370.

Sekhon, J. (2004–2013), "Matching: Multivariate and Propensity Score Matching with Balance Optimization," http://sekhon.berkeley.edu/matching, http://cran.r-project.org/package=Matching.

Senn, S. (1994), "Testing for Baseline Balance in Clinical Trials," *Statistics in Medicine*, Vol. 13: 1715–1726.

Shadish, W., T. Campbell, and T. Cook, (2002), *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, Houghton and Mifflin.

Sheiner, L., and D. Rubin, (1995), "Intention-to-treat Analysis and the Goals of Clinical Trial," *Clinical Pharmacology and Therapeutics*, Vol. 57: 6–15.

Shipley, M., P. Smith, and M. Dramaix, (1989), "Calculation of Power for Matched Pair Studies when Randomization is by Group," *International Journal of Epidemiology*, Vol. 18(2): 457–461.

Shu, Y., G. Imbens, Z. Cui, D. F., and Z. Kadziola, (2013), "Propensity Score Matching and Subclassification with Multivalued Treatments," Unpublished Manuscript.

Sianesi, B. (2001), "Psmatch: Propensity Score Matching in STATA," University College London, and Institute for Fiscal Studies.

Sims, C. (1972), "Money, Income and Causality," *American Economic Review*, Vol. 62(4): 540-552.

Smith, J. A., and P. E. Todd, (2001), "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *American Economic Review*, Papers and Proceedings, Vol. 91: 112–118.

Smith, J. A., and P. E. Todd, (2005), "Does Matching Address LaLonde's Critique of Nonexperimental Estimators," *Journal of Econometrics*, Vol. 125(12): 305–353.

Snedecor, G., and W. Cochran, (1967, 1989), *Statistical Methods*, Iowa State University Press.

Sommer, A., I. Tarwotjo, E. Djunaedi, K. West, A. Loeden, R. Tilden, and L. Mele, (1986), "Impact of Vitamin A Supplementation on Child Mortality: A Randomized Controlled Community Trial," *Lancet*, Vol. 1: 1169–1173.

Sommer, A., and S. Zeger, (1991), "On Estimating Efficacy from Clinical Trials," *Statistics in Medicine*, Vol. 10: 45–52.

Stigler, S. (1986), *American Contributions to Mathematical Statistics in the Nineteenth Century*, Arno Press.

Stock, J., and F. Trebbi, (2003),"Who Invented Instrumental Variable Regression?" *Journal of Economic Perspectives*, Vol. 17: 177–194.

"Student" (1923), "On Testing Varieties of Cereals," *Biometrika*, Vol. 15: 271–293.

Tanner, M. (1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer Verlag.

Tanner, M., and W. Wong, (1987),"The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, Vol. 82(398): 528–540.

Thistlewaite, D., and D. Campbell, (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology*, Vol. 51: 309–317.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, *Series B* (Methodological), Vol. 58(1): 267–288.

Tinbergen, J. (1930), "Bestimmung und Deutung von Angebotskurven: in Beispiel," *Zeitschrift für Nationalökonomie*, 669–679.

Torgerson, D., and M. Roland, (1998), "Understanding Controlled Trials: What Is Zelen's Design?" *BMJ*, Vol. 316: 606.

Van Der Klaauw, W. (2002), "A Regression–discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment," *International Economic Review*, Vol. 43(4): 1249–1287.

Van Der Laan, M., and J. Robins, (2003), *Unified Methods for Censored Longitudinal Data and Causality*, Springer Verlag.

Van Der Vaart, A., (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Victora, C., J.-P. Habicht, and J. Bryce, (2004), "Evidence-Based Public Health: Moving Beyond Randomized Trials," *American Journal of Public Health*, Vol. 94(3): 400–405.

Waernbaum, I. (2010), "Model Misspecification and Robustness in Causal Inference: Comparing Matching with Doubly Robust Estimation," *Statistics in Medicine*, Vol. 31(15): 1572–1581.

Welch, B. (1937), "On the z Test in Randomized Blocks and Latin Squares," *Biometrika*, Vol. 29: 21–52.

Wilcoxon, F. (1945), "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, Vol. 1(6): 80–83.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, MIT Press.

Wright, P. (1928), *The Tariff on Animal and Vegetable Oils*, Macmillan.

Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*, Vol. 20: 257-285.

Wright, S. (1923), "The Theory of Path Coefficients: A Reply to Niles' Criticism," *Genetics*, Vol. 8: 239–255.

Wu, J., and Hamada, M. (2009), *Experiments, Planning, Analysis and Optimization*, Wiley Series in Probability and Statistics.

Yang, S., G. Imbens, Z. Cui, D. Faries, and Z. Kadziola, (2014) "Propensity Score Matching and Subclassification with Multi-level Treatments," unpublished manuscript.

Yule, G. N. (1897), "On the Theory of Correlation," *Journal of the Royal Statistical Society*, 812–854.

Zelen, M. (1979), "A New Design for Randomized Clinical Trials," *New England Journal of Medicine*, Vol. 300: 1242–1245.

Zelen, M. (1990), "Randomized Consent Designs for Clinical Trials: An Update," *Statistics in Medicine*, Vol. 9: 645–656.

Zhao, Z. (2004), "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics and an Application," *Review of Economics and Statistics*, Vol. 86(1): 91–107.

Zhang, J., D. Rubin, and F. Mealli, (2009), "Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification," *Journal of the American Statistical Association*, Vol. 104(485): 166–176.