

# Quasi-Experimentation Workshops – August 2012

*Sponsored by the Institute of Education Sciences, U.S. Department of Education  
and the Institute for Policy Research, Northwestern University*

---

## Workshop Reference List

- Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, 4, 290-311.
- Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *The Review of Economics and Statistics*, 86(1), 180-194.
- Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J., & Hsuing, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Evaluation Review*, 22, 207-244.
- Ayllon, T., Layman, D., & Kandel, H. J. (1975). A behavioral-educational alternative to drug control of hyperactive children. *Journal of Applied Behavior Analysis*, 8(2), 137.
- Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114, 533-576.
- Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The State Journal*, 2(4), 358-377. Berk et al., in press;
- Black, D. C., Galdo, J., & Smith, J. A. (2007). Evaluating the worker profiling and reemployment services system using a regression discontinuity approach, *American Economic Review*, 97(2), 104-107.
- Blandford, B. J., & Lloyd, J. W. (1987). Effects of a self-instructional procedure on handwriting. *Journal of learning disabilities*, 20(6), 342.
- Bloom, H. S., Michalopoulos, C., & Hill, C. J. (2005). Using experiments to assess nonexperimental comparison-group methods for measuring program effects. In H. S. Bloom (Ed.), *Learning more from social experiments* (pp. 173-235). New York, NY: Russell Sage Foundation.
- Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Buddelmeyer, H., & Skoufias, E. (2004). *An evaluation of the performance of regression discontinuity design on PROGRESA*. (World Bank Policy Research Working Paper No. 3386; IZA Discussion Paper No. 827). Retrieved from [http://ssrn.com/abstract\\_434600](http://ssrn.com/abstract_434600)
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand-McNally.
- Cook, R. D. & Weisberg, S. (1999). *Applied regression including computing and graphics*. New York, NY: Wiley-Interscience.

- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand-McNally.
- Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which observational studies produce the same results as experiments. *Journal of Policy Analysis and Management*, 27(4), 724-750.
- Dee, T., & Jacob, B. 2009. The impact of No Child Left Behind on student achievement. Working Paper 15531, National Bureau of Economic Research.
- Diaz, J. J. & Handa, S. (2006). An assessment of propensity score matching as a nonexperimental impact estimator: Evidence from Mexico's PROGRESA Program, *Journal of Human Resources*, 41(2), 319-345.
- Gill, B., Lockwood, J. R., Martorell, F., Setodji, C. M., & Booker, K. (2007). State and local implementation of the No Child Left Behind Act. Washington, DC: Policy and Program Studies Service, Office of Planning, Evaluation, and Policy Development, U.S. Department of Education.
- Glazerman, S., Levy, D. M., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy*, 589, 63-93.
- Goldberger, A. S. (1972b). Selection bias in evaluating treatment effects: The case of interaction. Institute for Research on Poverty, University of Wisconsin-Madison.
- Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis*, 17(4), 400.
- Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, 15(3), 609-627.
- Hedges, L. V., and E. Hedberg. (2007). Interclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis* 29(1), 60-87.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8)
- Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, 27(3), 205-224.
- Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101, 901-910.
- Imbens, G., & Kalyanaraman, K. (2009). Optimal bandwidth choice for the regression discontinuity estimator. Working Paper 14726. National Bureau of Economic Research.
- Jacob, B., & Lefgren, L. (2004a). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, 39, 50-79.
- Jacob, B., & Lefgren, L. (2004b). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, LXXXVI, 226-244.

- Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going. Working Paper 9703, National Bureau of Economic Research.
- Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4), 523-539.
- Kosanke, J., & Bergstralh, E. (2004). gmatch: Match 1 or more controls to cases using the GREEDY algorithm. <http://mayoresearch.mayo.edu/mayo/research/biostat/upload/gmatch.sas>
- Leuven, E., & Sianesi, B. (2003). psmatch2. *STATA module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, version, 1(3)*.
- Ludwig, J., & Miller, D. L. (2005). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. Working Paper 11702, National Bureau of Economic Research.
- Martorell, F. (2005). *Does failing a high school graduation exam matter?* Mimeo, RAND.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- McKilip, J. (1992). Research without control groups: A control construct design. In F. B. Bryant, J. Edwards, R. S. Tindale, E. J. Posavac, L. Heath & E. Henderson (Eds.), *Methodological issues in applied psychology* (pp. 159–175). New York, NY: Plenum.
- Minton, J. H. (1975). The impact of "Sesame Street" on reading readiness of kindergarten children. *Sociology of Education*, 48, 141-151.
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.
- Nefdt, N., Koegel, R., Singer, G., & Gerber, M. (2010). The use of a self-directed learning program to provide introductory training in pivotal response treatment to parents of children with autism. *Journal of Positive Behavior Interventions*, 12(1), 23-32.
- Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*.
- Parker, E. B., Cook, T. D., & Campbell, D. T. (1966). *Effects of introducing TV on library circulation*. Unpublished manuscript, Stanford, CA.
- Parsons, L. S. (2001). Reducing bias in a propensity score matched-pair sample using greedy matching techniques. SAS Institute Inc., *Proceedings of the Twenty-Sixth Annual SAS® Users Group International Conference*, Paper 214-26. Cary, NC: SAS Institute Inc., URL <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>.
- Pohl, S., Steiner, P. M., Eisermann, J., Soellner, R., & Cook, T. D. (2009). Unbiased causal inference from an observational study: Results of a within-study comparison. *Educational Evaluation and Policy Analysis*, 31(4), 463–479.
- R Development Core Team (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

- Reynolds, K. D., & West, S. G. (1987). A multiplist strategy for strengthening nonequivalent control group designs. *Evaluation Review*, 11, 691-714.
- Roifman, C. M., Levison, H., & Gelfand, E. W. (1987). High-dose versus low-dose intravenous immunoglobulin in hypogammaglobulinaemia and chronic lung disease. *The Lancet*, 329(8541), 1075-1077.
- Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.
- Schafer, J. L., & Kang, J. (2008). Average causal effects from non-randomized studies: A practical guide and simulated example. *Psychological Methods*, 13(4), 279-313.
- Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software*, 42 (7).
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. *Journal of the American Statistical Association*, 103, 1334-1343.
- Shadish, W., Cook, T. D., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal influence*. Boston, MA: Houghton, Mifflin and Company.
- Shadish, W.R., Galindo, R., Wong, V.C., Steiner, P.M., Cook, T.D. (2011). A randomized experiment comparing random to cutoff-based assignment. *Psychological Methods*, 16(2), 179-191
- Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics*, 36(2), 213
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*. 15(3), 250-67.
- Sween, J. A. (1971). *The experimental regression design: An inquiry into the feasibility of nonrandom treatment allocation*. (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.
- Trochim, W. M. K., & Spiegelman, C. (1980). The relative assignment variable approach to selection bias in pretest-posttest designs. *Proceedings of the Social Statistical Section. American Statistical Association*. 376-381.
- van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review*, 43(4), 1249-1287.
- Wilde, E. T., & Hollister, R. (2007). How close is close enough? Testing on experimental estimates of impact against experimental estimates of impact with education test scores as outcomes. *Journal of Policy Analysis and Management*, 26, 455-477.

- Wing, C. & Cook, T.D. (in press) The Moving to Opportunity for Fair Housing Demonstration: How Relevant is it as an Evaluation of its Health Consequences of the Federal Section 8 Housing Program? To appear in summer 2012 special issue of *Cityscape*.
- Wong, V. C., Steiner, P. M., & Cook, T. D. (2012). Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods. *Journal of Educational and Behavioral Statistics*.
- Wong, V. C., Cook, T. D., Barnett, S. W., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.
- Wood, S. N. (2006). *Generalized Additive Models. An Introduction with R*. Boca Raton: Chapman & Hall/CRC.