



Actualidades en Psicología

ISSN: 0258-6444

actualidades.psicologia@ucr.ac.cr

Instituto de Investigaciones Psicológicas

Costa Rica

Pavlovich Panko, Pavel; Curtis, Jacob D.; Gorrall, Brittany K.; Little, Todd Daniel
Alternatives to Randomized Control Trials: A Review of Three Quasi-experimental
Designs for Causal Inference

Actualidades en Psicología, vol. 29, núm. 119, 2015, pp. 19-27

Instituto de Investigaciones Psicológicas

San José, Costa Rica

Available in: <http://www.redalyc.org/articulo.oa?id=133242591004>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

Alternatives to Randomized Control Trials: A Review of Three Quasi-experimental Designs for Causal Inference

Alternativas a las Pruebas Controladas Aleatorizadas: una revisión de tres diseños cuasi experimentales para la inferencia causal

Pavel Pavolovich Panko¹

Jacob D. Curtis²

Brittany K. Gorrall³

Todd Daniel Little⁴

Texas Tech University, United States

Abstract. The Randomized Control Trial (RCT) design is typically seen as the gold standard in psychological research. As it is not always possible to conform to RCT specifications, many studies are conducted in the quasi-experimental framework. Although quasi-experimental designs are considered less preferable to RCTs, with guidance they can produce inferences which are just as valid. In this paper, the authors present 3 quasi-experimental designs which are viable alternatives to RCT designs. These designs are Regression Point Displacement (RPD), Regression Discontinuity (RD), and Propensity Score Matching (PSM). Additionally, the authors outline several notable methodological improvements to use with these designs.

Keywords. Psychometrics, Quasi-Experimental Design, Regression Point Displacement, Regression Discontinuity, Propensity Score Matching.

Resumen. Los diseños de Pruebas Controladas Aleatorizadas (PCA) son típicamente vistas como el mejor diseño en la investigación en psicología. Como tal, no es siempre posible cumplir con las especificaciones de las PCA y por ello muchos estudios son realizados en un marco cuasi experimental. Aunque los diseños cuasi experimentales son considerados menos convenientes que los diseños PCA, con directrices estos pueden producir inferencias igualmente válidas. En este artículo presentamos tres diseños cuasi experimentales que son formas alternativas a los diseños PCA. Estos diseños son Regresión de Punto de Desplazamiento (RPD), Regresión Discontinua (RD), Pareamiento por Puntaje de Propensión (PPP). Adicionalmente, describimos varias mejoras metodológicas para usar con este tipo de diseños.

Palabras clave. Psicometría, Diseños cuasi experimentales, Regresión de Punto de Desplazamiento, Regresión Discontinua, Pareamiento por Puntaje de Propensión.

¹Pavel Pavolovich Panko. Educational Psychology - Research, Evaluation, Measurement, and Statistics (REMS) Concentration, Institute for Measurement, Methodology, Analysis & Policy (IMMAP), Texas Tech University, United States. Postal Address: Department TTU-EDUCATION, Texas Tech University - National Wind Institute, 1009 Canton, Ave, Room Number 211, Lubbock, TX 79409, United States. Email: pavel.panko@ttu.edu

²Jacob D. Curtis. Institute for Measurement, Methodology, Analysis & Policy (IMMAP), Texas Tech University, United States. Email: jacob.curtis@ttu.edu

³Brittany K. Gorrall. Institute for Measurement, Methodology, Analysis & Policy (IMMAP), Texas Tech University, United States. Email: britt.gorrall@ttu.edu

⁴Todd Daniel Little. Institute for Measurement, Methodology, Analysis, and Policy (IMMAP), Texas Tech University, United States. Email: yhat@ttu.edu



Introduction

Randomized controlled trial (RCT) designs are typically seen as the pinnacle in experimental research because they eliminate selection bias in assigning treatment (Shultz & Grimes, 2002). RCT designs, however, are sometimes not practical due to a lack of resources or inability to exercise full control over study conditions. Additionally, ethical reasons prohibit implementing random assignment when there are groups that require treatment due to higher need. In these instances, designs that are more quasi-experimental in nature are more appropriate.

In this paper, the authors outline three possible quasi-experimental designs that are robust to violations of standard RCT practice. The authors start with the regression point displacement (RPD) design, which is suitable in cases where there is a minimum of one treatment unit. Next, the authors discuss the Regression Discontinuity (RD) design, which utilizes a "cut point" to determine treatment assignment, allowing those most in need of a treatment to receive it. Finally, the authors present Propensity Score Matching (PSM), which matches control and treatment groups based on covariates that reflect the potential selection process.

The purpose of this paper is to give an introduction of each of the three quasi-experimental designs. For an in-depth discussion on each design, please refer to the included references. In addition, the authors discuss novel techniques to improve upon these designs. These techniques address the limitations often inherent in quasi-experimental designs. As well, illustrative examples are provided in each section.

Regression Point Displacement Design

Regression Point Displacement is a research design applicable in quasi-experimental situations such as pilot studies or exploratory causal inferences. The method of analysis for this design is a special case of linear regression where the post-test of an outcome measure is regressed on to its own pre-test to determine the degree of predictability. Treatment effectiveness is estimated by comparing a vertical displacement of the treatment unit(s) on the posttest against the regression

trend of the control group (Linden et al., 2006; Trochim & Campbell, 1996; 1999). If the treatment did have an effect, the treatment group would be significantly displaced from the control group regression line. In this case, the treatment condition would be evaluated for whether it is statistically different from the control.

A regression equation in the form of Linden et al. (2006) can be represented in the following way:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i \quad (1)$$

where Y_i is the score of individual i on outcome Y , β_0 is the intercept coefficient, β_1 is the pretest coefficient, X_i is the pretest score, β_2 is the coefficient for the difference due to treatment, Z_i is the dummy-coded variable indicating whether the individual received treatment ($Z_i = 1$) or not ($Z_i = 0$), and e_i the individual error term. If the p value for β_2 is significant, the treatment had an effect. This effect can be visually observed by plotting a regression line and inspecting whether or not the treatment condition is out of the confidence interval of the trend for the control groups.

RPD designs have several unique features (Trochim & Campbell, 1996; 1999). First, it requires a minimum of only one treatment unit (Trochim, 2006). Because of this minimum requirement, however, the data may be highly variable, so it is a good idea to use aggregated units (e.g. schools) due to their greater tendency toward centrality when compared with persons as the individual units. Second, this design is applicable in contexts where randomization is not possible, such as pilot studies (Linden et al., 2006) or after a particular group receives treatment a priori. Third, RPD designs avoid regression artifacts with the use of an observed regression line (Trochim & Campbell, 1996; 1999). Lastly, it is possible to add covariates to explain baseline differences between the treatment and control units (Trochim & Campbell, 1996). The effect of the covariates can be interpreted visually by using residual differences between pre and post-tests. By regressing the pretest and the posttest on the covariate, a plot with more than one predictor using the resulting residuals can be created. The residuals of the regression on the covariate should be saved for

both pre-test and post-test and used in the regression equation just as before. In this way, the residuals are representative of the pretest and the posttest with the influence of the covariate taken out.

As an example, the regression point displacement design was used to estimate the effect of a behavioral treatment on twenty-four schools. One of the schools was selected to receive the treatment. The pre and posttest outcomes were operationalized

by the number of disciplinary events for their respective years.

Figure 1 demonstrates that the treatment school was displaced by 1384 disciplinary class removals from the trend – this residual value provides a tangible effect size estimate that has real and direct interpretation. In other words, this large number can be interpreted as a real difference in removals between the trend of the control schools and the treatment school. The p value

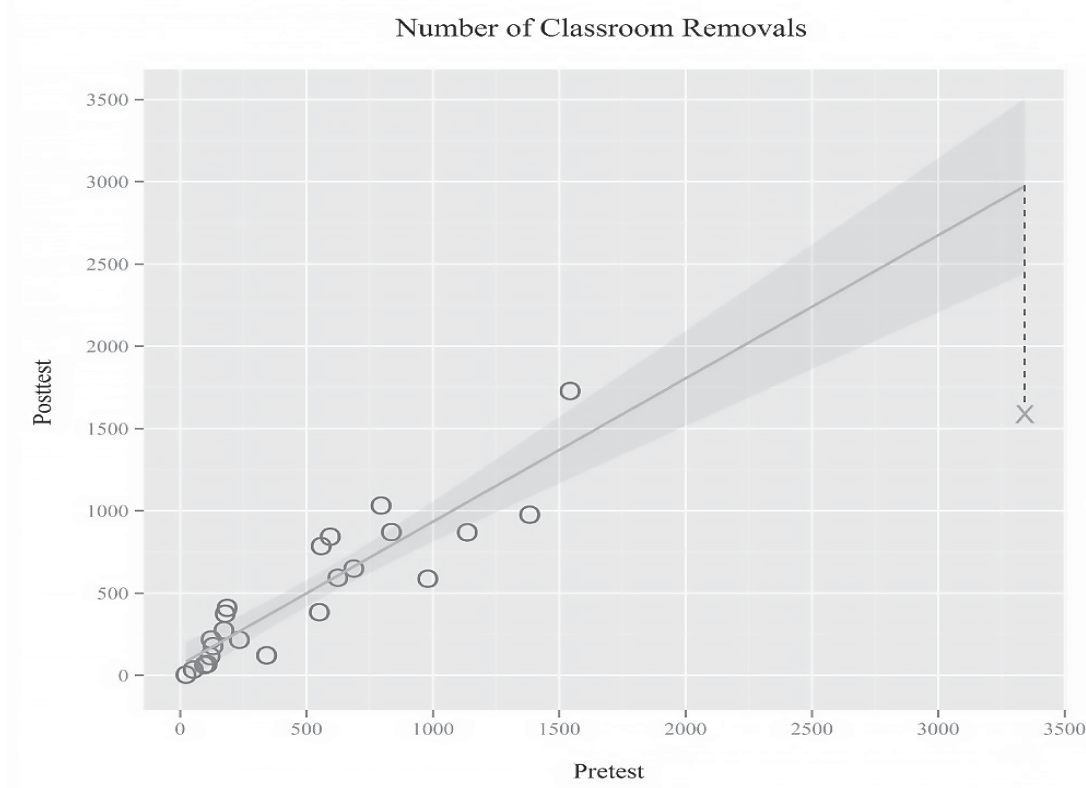


Figure 1. Displacement of the Treatment School (x) from the control group regression line.

Table 1

Regression Model Statistics

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	63.20	58.66	1.08	0.29
Disciplinary Removal (pretest)	0.87	0.09	9.82	0.00
Treatment School	-1384.2	315.13	-4.39	0.00

indicates that the displacement of the treatment unit was significant.

Regression point displacement designs also have inherent limitations. If the treatment unit is not randomly selected, the design will have the same selection bias problems as other non-RCT designs (Linden et al., 2006). Due to this limitation, it is possible that the treatment unit may not generalize to the population of interest. On the other hand, the treatment unit can be thoroughly scrutinized prior to treatment. As a result, prior knowledge and prudent selection of the context of the treatment, mitigates these issues particularly in sight of the benefits. The RPD design studies are inexpensive and perfectly suited for exploratory and pilot study frameworks (Linden et al., 2006) as well as circumscribed contexts such as program evaluations. That is, a single program can be evaluated by selecting a number of control programs and using the RPD design to evaluate the selected unit.

Regression Discontinuity.

The Regression Discontinuity (RD) design is a quasi-experimental technique that determines the effectiveness of a treatment based on the linear discontinuity between two groups. In RD designs, a cut point on an assignment variable determines whether individuals are assigned to a treatment condition or a control (comparison) condition (Shadish, Cook, & Campbell, 2002). The cut point should be a specific value on the assignment variable decided a priori. In order to make a causal conclusion about the effectiveness of a treatment or intervention, the change in the mean-level or slope-angle of the outcome variable is analyzed (see Greenwood & Little, 2007).

Figure 2 illustrates a hypothetical example of an RD design that is depicting the effect of a program intended to increase math test scores. In the RD design, the y-axis represents the outcome variable, in this case math test scores, and the x-axis represents the screening measure. In Figure 2, the trend for the control group, called the counterfactual regression line shows what the regression line would be if the treatment had no

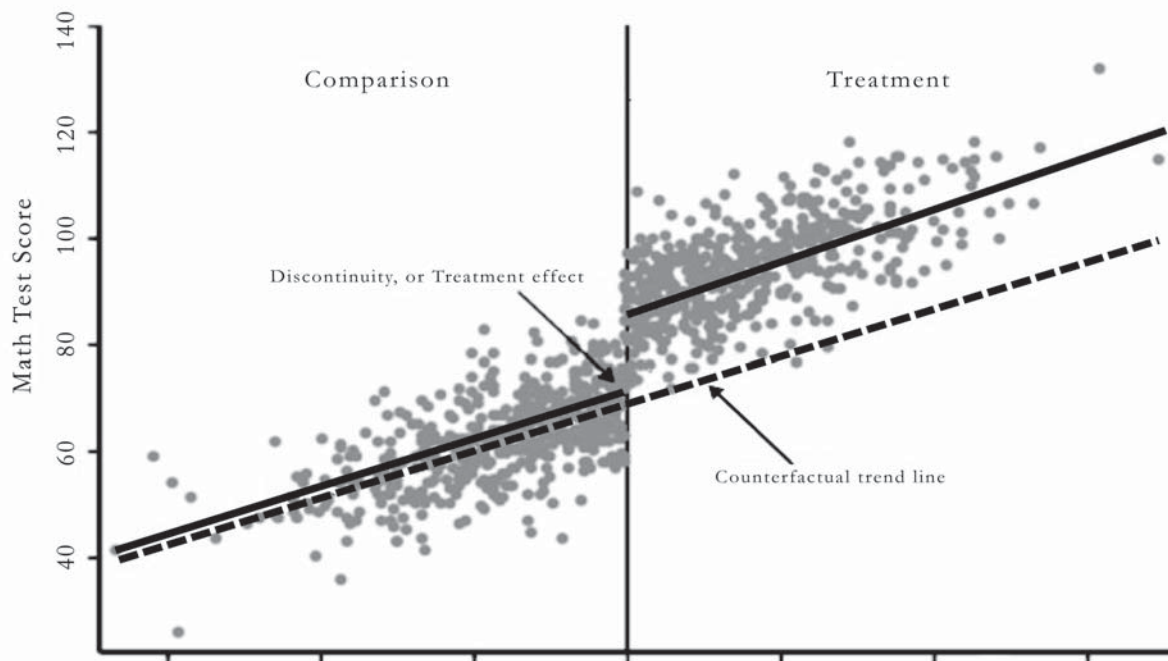


Figure 2. Hypothetical results of a treatment designed to increase math test scores. The discontinuity in the solid line indicates a treatment effect.

effect. The counterfactual line is usually smooth across the cut point, as seen in Figure 2. A discontinuity in the actual regression line indicates a treatment effect, with the size of the discontinuity providing a measure of the magnitude of the treatment effect on the outcome variable (Braden & Bryant, 1990). To see the basic form of the regression discontinuity technique, refer to Campbell, 1984; Shadish, Cook & Campbell, 2002. Also, refer to Moss, Yeaton & Floyd (2014) for discussion on polynomial and nonlinear forms.

RD designs have three main limitations. First, RD designs are dependent on statistical modeling assumptions. Participants must be grouped solely by the cut point criterion (Trochim, 1984; 2006). Second, it may not be appropriate to extrapolate the results to all the participants as only the scores immediately before and after the cut point are used to calculate the treatment effect. This limitation means that if the treatment had a differential effect on participants away from the cut point, the design would not capture it (Angrist & Rokkanen, 2012; Battistin & Rettore, 2008). Third, traditional RD designs also have low statistical power (Pellegrini, Terribile, Tarola, Muccigrosso, & Busillo, 2013).

To remedy these limitations, Wing and Cook (2013) propose the addition of a pretest comparison group. The reasoning for using pretest scores is to provide information about the relationship between the cut point and outcome prior to treatment. The first advantage of this approach is that the differences between pre and post measures will give an indication of bias in assignment, thereby attenuating the limitation of controlled assignment. Second, the treatment effect can be generalized beyond the cut point to include all individuals in the treatment group. This extended generalizability is so because adding a pretest allows for extrapolation beyond the cut point in the posttest period. Third, the inclusion of the pretest strengthens the predictive power of RD, making it comparable in power to an RCT. The addition of a comparison function gives the RD design all the benefits of an RCT design but is coupled with the dissonance reduction that serving

the neediest provides.

The pretest RD design equation from Wing and Cook (2013) is defined by the following:

$$Y(1)_{it} = Pre_{it} \theta_p + g(A_i) + e_{it} \quad (2)$$

The variable $Y(1)_{it}$ represents the outcome for the treatment group at time t . Conversely, if 0 was in place of 1, it would be the outcome of the untreated group. Pre_{it} is a dummy variable identifying observations during a pretest period where the treatment has yet to be implemented. The θ_p parameter is a fixed difference of conditional mean outcomes across pretest and posttest periods. An unknown smoothing function is represented by the $g(A_i)$, and it is assumed to be constant across the pre- and posttest (for further discussion of smoothing parameters see Peng, 1999).

Wing and Cook (2013) used the data from the Cash and Counseling Demonstration RCT (Dale & Brown, 2007) to test the efficacy of a pre-post RD design. In the original study, disabled Medicaid beneficiaries were randomly assigned to obtain two types of healthcare services to examine the differences on a variety of health, social, and economic outcomes. In the subsequent analysis, Wing and Cook used baseline age as the assignment variable to reexamine the outcomes in an RD framework. The researchers identified three age cut points (i.e., 35, 50, and 70) for the treatment assignment. Additionally, the pretest was used to estimate the average treatment effect for everyone older than the cut point in the pretest RD design.

For each age cut point, Wing and Cook compared the outcomes within the RD design as well as between the RD and RCT models. They found that the pre-post RD design leads to unbiased estimates of the treatment effects both at the cut point and beyond the cut point. Also, adding the pretest helped to obtain more precise parameter estimates than traditional posttest-only RD designs. Therefore, the results from the within-study comparisons showed that the pretest helped to improve the standard RD design method by approximating the same causal estimates of an RCT design. This example demonstrates that the pre-post

Regression Discontinuity design is a useful alternative to and can rival the performance of RCT designs.

Propensity Score Matching.

Propensity Score Matching (PSM) is a quasi-experimental technique first published by Rosenbaum and Rubin (1983). Propensity score matching attempts to rectify selection bias that can occur when random assignment is not possible by creating two groups that are statistically equivalent based on a set of important characteristics (e.g., age, gender, ethnicity, personality, health status, IQ, experience, etc) that are relevant to the study at hand. Here, each participant gets a score on their likelihood (propensity) to be assigned to the treatment group based on the characteristics that drive selection (termed, covariates). A treatment participant is matched to a corresponding control participant based on the similarity of their respective propensity score. That is, the control participants included in the analysis are those who match treatment participants on the potential confounding selection variables; in this way, selection bias is controlled.

Before propensity scores can be estimated, the likely selection covariates must be identified. Most researchers include all variables that could potentially correlate with the selection influences impacting treatment and outcome (Coffman, 2012; Cuong, 2013; Lanza, Coffman, & Xu, 2013; Stuart et al., 2013), regardless of the magnitude of correlation (Rubin, 1997).

In practice, propensity scores are typically estimated using logistic (e.g., Lanza, Moore, & Butera, 2013), probit (e.g., Lalani et al., 2010), or multiple binomial logistic regression models (e.g., Slade et al., 2008) in which the group membership is the dependent variable predicted by the selection variables in the dataset (Caliendo & Kopeinig, 2008; Lanza et al., 2013). The logistic regression model, as proposed by Cox (1970), has been the most commonly employed technique in propensity score calculations (Rosenbaum & Rubin, 1985). The probability score, a decimal value ranging from 0 to 1, is retained and used to match participants from the treatment and

control groups.

Once the propensity scores have been estimated, each participant from the treatment condition is matched with a participant from the control condition. As mentioned, the matching of these participants is based upon the similarity of their propensity scores. Matching participants from the treatment condition with similar participants from the control condition can be completed utilizing the nearest neighbor, caliper, stratification, and kernaling techniques (e.g., Austin, 2011). Of these methods, differences exist in the number of participants from the control group who are matched to treatment participants and whether or not control participants can be matched more than once (Coca-Perraillon, 2006).

The nearest neighbor and caliper techniques are among the most popular (Coca-Perraillon, 2006). The treatment and control groups are randomly sorted for both methods. Then, the first treatment participant is matched without replacement with the control participant who has the closest propensity score. The algorithm moves down the list of all the treatment participants and repeats the process until all the treatment participants are matched with a control counterpart. If any control participants are left over, they are discarded (Coca-Perraillon, 2006). The difference in the techniques is that with caliper matching, treatment participants are only used if there is a control participant within a specified range. Thus, in this technique, unlikely matches are avoided (Coca-Perraillon, 2006).

The optimal full matching technique (Hansen, 2004) improves on these popular techniques in two ways. First, it creates closer matches than the previous techniques – with caliper and nearest neighbor, a match is made independently of the other pairs. On the other hand, optimal full matching always creates matches with the smallest possible average propensity score differences between matched treatment and control participants by taking into account all the other matches. Second, optimal full matching allows for all control participants to be used (Hansen, 2004). After matching, the participants in the treatment and control

groups are assumed to have the same likelihood of being in the treatment group. The treatment effect is calculated and is now an unbiased estimator of the treatment effect.

Although the use of PSM is relatively new, there are well-explicated applications in many published manuscripts. One such example is a recent manuscript published by Lanza et al. (2013) in which they sought to examine the benefit of attending Head Start on children's reading ability over parental pre-school care.

Utilizing the Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K; Institute of Education Sciences, 2009), a nationally representative, longitudinal dataset, Lanza et al. (2013) examined the causal effect of Head Start instruction on reading development, comparing it to parental care during preschool years. Given that the ECLS-K is a dataset comprised of observational (e.g., non-experimental) data, they were unable to randomly assign students to a Head Start or parental condition. Instead, they utilized Head Start enrollment as the marker for those who were a part of the treatment condition. Additionally, they selected over 20 covariates to include in the prediction of Head Start enrollment. The selection of these covariates was comprehensive because they wanted to account for all of the possible variation in attending Head Start.

Lanza et al. (2013) fit a logistic regression to the data, with the covariates as predictors and Head Start enrollment as the dependent variable, to estimate the propensity scores. They then matched the participants from the treatment group with similar participants from the control group using the optimal matching algorithm. Using this method the researchers obtained pairs with optimally close propensity scores. After examining the quality and sensitivity of the matches, they examined the causal inference hypothesis.

Lanza et al. (2013) reported that children who stayed at home during the pre-school years had higher reading scores upon entering kindergarten than children who attended Head Start. While one may intuitively think that early intervention through preschool should

increase achievement in kindergarten; they noted that due to potential confounding variables, this relationship would not be as clear. Controlling for the influence of confounding variables, such as the child's gender, ethnicity, and maternal education, they found that there was not much difference between the two groups. This result demonstrates that Propensity Score Matching is a useful technique when selection bias is a concern.

Conclusion

Data often do not meet the necessities of a truly experimental randomized-control trial. Specifically, random assignment may not have been employed for a number of reasons. In these cases, researchers still have the ability to make conclusive inferences using the designs that the authors have discussed in this article.

The authors began with Regression Point Displacement, which is most useful when either one or a small number of treatment conditions are present for comparison. In this design, the vertical displacement of the treatments unit from the control trend is used to infer significance of the treatment effect. Next, the authors discussed the Regression Discontinuity design which assigns participants to treatment and control conditions based on a just and defensible cut point on an assignment variable and subsequently measure the discontinuity of the treatment and control trends. The inference of this design becomes much stronger when utilizing the pre-post framework outlined by Wing and Cook (2013), making RD comparable to an RCT. Lastly, the authors discussed Propensity Score Matching, which pairs control and treatment participants on the similarity of their scores to account for selection bias. Although there are several methods within PSM, the authors most strongly recommend using optimal full matching because it creates the most likely matches available.

This paper demonstrates that although RCT designs are the gold standard in the social sciences and beyond, there are alternative designs that can be just as valid and reliable in a quasi-experimental framework.

Consequently, even if a potential study is limited in the total number of participants, the ability to randomly assign treatment, or in the number of treatment units, there are methods that can be employed to make the causal inferences perfectly viable.

References

- Angrist, J., & Rokkanen, M. (2012). *Wanna get away? RD identification away from the cutoff*. Working Paper # 18662, National Bureau of Economic Research, Retrieved from <http://www.nber.org/papers/w18662>.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46, 399-424.
- Battistin, E., & Rettore, E. (2008). Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. *Journal of Econometrics*, 142, 715-730.
- Braden, J. P. & Bryant, T. J. (1990). Regression discontinuity designs: Applications for school psychology. *School Psychology Review*, 19(2), 232-239.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22, 31-72.
- Coca-Perraillon, M. (2006). *Matching with propensity scores to reduce bias in observational studies*. Retrieved from <http://www.nesug.org/proceedings/nesug06/an/da13.pdf>
- Coffman, D. (2012). *Methodology workshop: Propensity score methods for estimating causality in the absence of random assignment: Applications for child care policy research*. Presented at the annual meeting of the Child Care Policy Research Consortium, Bethesda, MD.
- Cox, D.R. (1970). *The analysis of binary data*. London, UK: Methuen.
- Cuong, N. V. (2013). Which covariates should be controlled in propensity score matching? Evidence from a simulation study. *Statistica Neerlandica*, 67, 169-180.
- Dale, S. B., & Brown, R. S. (2007). How does cash and counseling affect costs? *Health Services Research*, 42, 488-509.
- Greenwood, C. R., & Little, T. D. (2007). *Use of regression discontinuity designs in special education research*. Paper commissioned as one in a series of NCSER, IES papers devoted to special education research methodology topics. Hyattsville, MD: Optimal Solutions Group, LLC.
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99, 609-618.
- Institute of Education Sciences. (2009). Combined user's manual for the ECLS-K eighth-grade and K-8 full sample data files and electronic codebooks. Washington, D.C.: National Center for Education Statistics, U.S. Department of Education.
- Lalani, T., Cabell, C. H., Benjamin, D. K., Lasca, O., Naber, C., Fowler, V. G., & Wang, A. (2010). Analysis of the impact of early surgery on in-hospital mortality of native valve endocarditis use of propensity score and instrumental variable methods to adjust for treatment-selection bias. *Circulation*, 121, 1005-1013.
- Lanza, S. T., Coffman, D. L., & Xu, S. (2013) Causal inference in latent class analysis, structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 20, 361-383.
- Lanza, S. T., Moore, J. E., & Butera, N. M. (2013). Drawing causal inferences using propensity scores: A practical guide for community psychologists. *American Journal of Community Psychology*, 52, 380-392.
- Linden, A., Trochim, W. K., & Adams, J. L. (2006). Evaluating program effectiveness using the regression point displacement design. *Evaluation & the Health Professions*, 29, 407-423.
- Moss, B. G., Yeaton, W. H., & Lloyd, J. E. (2014). Evaluating the effectiveness of developmental

- mathematics by embedding a randomized experiment within a regression discontinuity design. *Educational Evaluation and Policy Analysis*, 36(2), 170-185.
- Pellegrini, G., Terribile, F., Tarola, O., Muccigrosso, T., & Busillo, F. (2013). Measuring the effects of European regional policy on economic growth: A regression discontinuity approach. *Papers in Regional Science*, 92, 217-233.
- Rosenbaum, P. and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 4155.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.
- Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127, 757-763.
- Schulz, K. F., & Grimes, D. A. (2002). Allocation concealment in randomized trials: Defending against deciphering. *The Lancet*, 359, 614-618.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Slade, E.P., Stuart, E.A., Salkever, D.S., Karakus, M., Green, K.M., & Lalongo, N. (2008). Impacts of age of onset of substance use disorders on risk of adult incarceration among disadvantaged urban youth: A propensity score matching approach. *Drug and Alcohol Dependence*, 95, 1-13.
- Stuart, E. A., DuGoff, E., Abrams, M., Salkever, D., & Steinwachs, D. (2013). Estimating causal effects in observational studies using electronic health data: challenges and (some) solutions. *Generating Evidence & Methods to improve patient outcomes (eGEMS)*, 1, 1-10.
- Trochim, W.M.K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage.
- Trochim, W.M.K. (2006). The Research Methods Knowledge Base, 2nd Ed. Retrieved from <http://www.socialresearchmethods.net/kb/>
- Trochim, W.M.K., & Campbell, D. T. (1996). The regression point displacement design for evaluating community-based pilot programs and demonstration projects. Unpublished manuscript. Retrieved from <http://www.socialresearchmethods.net/research/RPD/RPD.pdf>
- Trochim, W.M.K., & Campbell, D.T. (1999). Design for community-based demonstration projects. In D.T. and M.J. Russo (Eds.), *Social experimentation*. Thousand Oaks, CA: Sage.
- Wing, C., & Cook, T.D. (2013). Strengthening the regression-discontinuity design using additional design elements: A within-study comparison. *Journal of Policy Analysis and Management*, 32, 853-877.

Received: June 14th, 2015

Accepted: September 9th, 2015