Regression Discontinuity Design

Joseph C Cappelleri, Pfizer Inc, Groton, CT, USA
William M Trochim, Cornell University, Ithaca, NY, USA

This article is a revision of the previous edition article by W. M. Trochim, volume 19, pp. 12939–12945, © 2001, Elsevier Ltd.

Abstract

The regression-discontinuity (RD) research design assigns participants to treatment groups solely on the basis of a pretreatment cutoff score, allowing the relative effect of treatment to be studied in participants who most need or deserve a particular treatment. In this article a brief history of the design is given, the basic RD design structure is described, design considerations and variations are highlighted (on cutoff selection, assignment variations, multiple cutoff points, multiple assignment measures, treatment variations, and posttreatment measurement variations, internal validity, measurement error, statistical power), and recent methodological developments are presented (on nonparametric statistical analysis, the 'fuzzy' RD design, and design variations).

Definition

The regression-discontinuity (RD) research design is a quasi-experimental method that can be used to assess the effects of a treatment or intervention. Unique to the RD design is that participants are assigned to groups solely on the basis of a pretreatment cutoff score. The name 'regression-discontinuity' comes from the fact that a treatment effect appears as a 'jump' or discontinuity at the cutoff point in the regression function linking the assignment variable to the outcome. In its simplest form, the design has a pretest or pretreatment (the assignment variable) measure, two groups (those scoring above and below the cutoff), and a posttest or posttreatment (the outcome) measure. More complex variations are also possible.

To many, the RD design initially seems counterintuitive. Using a cutoff to assign participants to the program (or treatment) group and comparison (or control) group creates a pretreatment group nonequivalence that seems like it should lead to selection bias. While the design does induce initial nonequivalence, it does not necessarily lead to biased treatment effect estimates. Instead of assuming pretreatment equivalence (in measured and unmeasured means and variances) as the randomized experiment does, the RD design assumes the two groups are equivalent in their pre–post regression functions in the absence of a treatment effect. This assumption can be tested.

The major attraction of RD designs is that they can be used to estimate the effects of treatments given to those who most need or deserve them – provided that need or merit is determined using a qualification score on the pretreatment assignment variable (i.e., the cutoff) to experimental conditions, and nothing else. Because the design does not require that some needy or deserving individuals get assigned to a no-treatment or comparison group, in some settings it may have ethical advantages over experiments for assessing treatment effects.

History

In a 2008 special issue on RD methodology in the *Journal of Econometrics*, Cook (2008) provided a detailed and comprehensive history of the RD design in three academic disciplines – psychology, statistics, and economics – which covers the reinvention of the design across these disciplines and the differential waxing and waning by discipline (Cook, 2008). Applications of the RD design have also expanded to the health sciences (Trochim, 1990; Trochim and Cappelleri, 1992; Cappelleri and Trochim, 1994, 1995; Finkelstein et al., 1996a, 1996b; Linden et al., 2006; Zuckerman et al., 2006). Many noteworthy contributions using the RD design have been made in the 1960s (Thistlethwaite and Campbell, 1960), 1970s (Goldberger, 1972; Wilder, 1972; Tallmadge and Horst, 1976; Rubin, 1977), 1980s (Berk and Rauma, 1983), 1990s (Cappelleri et al., 1991; Trochim et al., 1991; Mark and Mellor, 1991; Cappelleri et al., 1994; Reichardt et al., 1995; Aiken et al., 1998; Angrist and Lavy, 1999; Berk and de Leeuw, 1999), and 2000s (Van Der Klaauw, 2002; DiNardo and Lee, 2004; Card and Shore-Sheppard, 2004; Jacob and Lefgren, 2004; Ludwig and Miller, 2007; Matsudaira, 2008; Wong et al., 2008; Van Der Klaauw, 2008; Lalive, 2008; Lee, 2008; Gamse et al., 2008).

A growing number of books have featured the RD design. It was first discussed by Campbell and Stanley (1963) and in greater detail in books on social experimentation (Riecken et al., 1974; Bennett and Lumsdaine, 1975). Later the RD design was extensively discussed in texts on quasi-experimentation (Cook and Campbell, 1979; Shadish et al., 2002) and on evaluation (Judd and Kenny, 1981; Mohr, 1995). Trochim (1984) wrote the first and, to date, only book devoted exclusively to the method. With the dawning of the twenty-first century several books discussed the RD design, presenting updates on its methodology and applications, including (but not limited to) books by Fleiss et al. (2003), Senn (2007), Angrist and Pischke (2009), and Murnane and Willett (2011).

The Basic RD Design Structure

Figure 1 shows an example of the RD design using simulated data to depict the case of a compensatory treatment – perhaps a reading program designed to help students who initially score poorly (defined as below the pretest cutoff) on a standardized reading measure. The figure is a standard bivariate plot of

 http://dx.doi.org/10.1016/B978-0-08-097086-8.44049-3

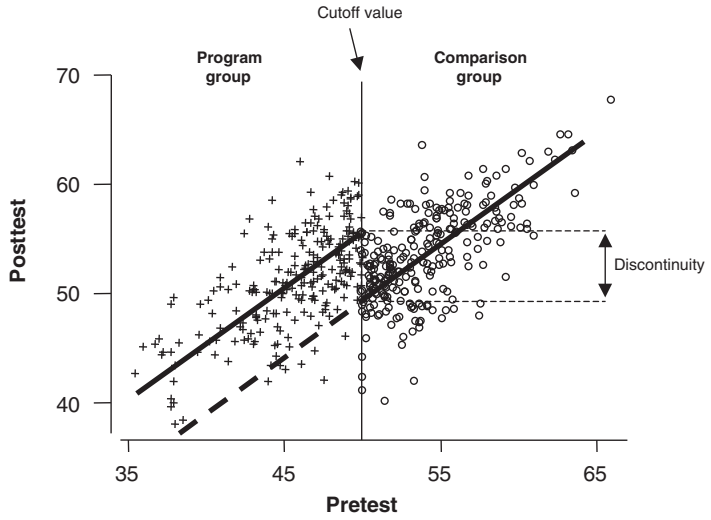

Figure 1 Basic regression-discontinuity design for a compensatory treatment showing a positive effect.

posttest and pretest for 500 simulated participants. Data points for the treatment recipients are represented with '+,' while the comparison cases are 'o.' The cutoff value in this example occurs at the middle of the pretest distribution (pretest = 50) and is indicated by the vertical line. All students scoring below 50 are defined as needing compensatory training and get the treatment, while the remaining (i.e., those to the right of the cutoff) are assigned to the comparison (control) group.

The solid lines represent the straight-line regression of posttest on pretest for each group. In the absence of a treatment effect, the critical assumption is that the regression line in the comparison group would continue to the left of the cutoff into the treatment group region, as indicated by the heavy dashed line. That is, if the reading program does not work, then the observed treatment group regression line would be where the dashed line is. The fact that the observed treatment group line is displaced from this 'expected' line suggests a treatment effect – treatment group participants scored higher on the posttest (i.e., vertically) than would have been predicted from the comparison group pre–post relationship. The size of the treatment main effect is estimated as the vertical jump or *discontinuity* between the *regression* lines at the cutoff, hence the name *regression-discontinuity*. Assuming that higher scores on both measures indicate better performance, one can conclude from the figure that the treatment improved participant performance by the amount of the vertical discontinuity.

In experimental designs random assignment assures that the treatment and comparison groups are initially probabilistically equivalent. Posttreatment differences can be then attributed to the intervention. In the RD design one obviously does not expect this kind of equivalence. Instead, it is assumed that the *relationship* between the assignment variable and the outcome is equivalent for the two groups – that is, the same continuous regression line describes both groups. Thus, interpretation of the RD design depends mostly on two factors: (1) that no alternative cause could induce a discontinuity at the cutoff and (2) that we can perfectly model the pre–post relationship. These issues, particularly the second one, constitute the major problems in the statistical analysis of the RD design (Trochim, 1984, 1990; Winship and Morgan, 1999).

Design Considerations and Variations

Selection of the Cutoff

The RD design requires that the assignment and outcome measures be of sufficient quality of measurement to warrant an appropriate regression analysis. The choice of cutoff value can be made solely on the basis of available resources. For instance, if a program can handle only 25 persons and 70 apply, then a cutoff point can be selected that distinguishes the 25 most 'needy' persons from the rest. The cutoff can also be chosen on substantive grounds. If the assignment measure is an indication of severity of illness measured on a one to seven scale, relevant experts might contend that all those scoring an average of five or more are ill enough to deserve treatment. Therefore, a theoretically justified cutoff of five could be used.

Assignment Variations

A 'sharp' RD design has its pretreatment cutoff value followed perfectly, without exception. While the use of a pretreatment cutoff distinguishes the RD design, it is often difficult to implement. In principle it does not allow room for professional judgment or discretion. In fact, the design does not preclude incorporating such judgment so long as it can be quantified or explicitly accounted for when specifying the cutoff value.

Violation of the assignment rule leads to 'misassignment' and to what is often termed the 'fuzzy' RD design. Typically, misassignment is most prevalent for 'close call' cases that fall near the cutoff. For instance, teachers or parents may successfully argue that students who just miss the cutoff for a novel compensatory tutoring program should be assigned to it based on grounds other than their assignment score. In the fuzzy RD case, standard analyses will yield biased estimates of treatment effect. While statistical procedures for dealing with

misassignment have been suggested (Trochim, 1984), misassignment is better avoided if possible.

Multiple Cutoff Points

RD designs are not limited to a single cutoff value. In the absence of treatment the assumption is that the pre–post relationship can be described as a single continuous regression line extending over the entire range of the assignment scores. Even when multiple cutoff points are used, the comparison group's pre–post relationship is still used as the counterfactual for projecting where other groups' regression lines should be if there is no treatment effect.

The simplest multiple cutoff case involves two cutoff scores. In this variant, persons scoring above the higher cutoff might be assigned to one treatment, those scoring below the lower cutoff to another, and those scoring within the cutoff interval might be controls. The assignment of treatments to each part of the assignment distribution can be made in alternative ways, including by random assignment. For instance, individuals within the interval between the two cutoffs could be randomly assigned to conditions, while those scoring below the lower cutoff and above the upper cutoff are analyzed as from an RD design (Boruch, 1973). Trochim and Cappelleri (1992) discuss four other ways of coupling random assignment to RD design features.

Multiple cutoff points are also involved in designs where the treatment is implemented a series of times (i.e., in phases), perhaps beginning with the most needy before successively giving it to those less needy. In the first phase, one uses a single cutoff value, assigning those most in need to the treatment and then measuring posttreatment performance. Then, the cutoff value would be moved to include the next most needy group on the assignment measure. Eventually, everybody could receive the treatment, but in groups ordered by need and phase. This is particularly useful in organizations seeking to 'roll out' a treatment over time, but eventually serving all.

Multiple Assignment Measures

The requirement of strict adherence to the cutoff criterion entails a single quantitative indicator. When this does not capture the degree of pretreatment need well, Trochim (1990) discusses two multivariate strategies. The first involves the use of several separate measures, each having its own cutoff value. This is most feasible when a prospective participant must meet the different cutoff criteria on measure 1 'and' measure 2 'and' measure 3, and so on. Less feasible is an 'or' rule when assignment depends on meeting a fixed number of criteria from a larger set – say, 5 of 10. The second involves assignment according to a new variable that is a composite of several individual measures. Using multiple measures this way can help incorporate more judgment into a quantitative assignment strategy.

Treatment Variations

When treatment comparisons are absolute, the control (or comparison) group receives no formal or placebo treatment. When they are relative, the comparison group receives an alternative treatment, very often the current standard treatment to which a new one is to be compared.

An advantage of the RD design is that it enables assignment of progressively riskier treatments to those in greater need. Assume three treatment programs: the standard plus two increasingly riskier experimental ones. It is then possible to have two cutoff points and to assign the riskiest treatment to the neediest patients, the least risky (i.e., standard) treatment to the least needy, and the moderately risky treatment to those falling between the two cutoffs. (If no a priori basis exists for making risk judgments, an alternative would be to use a single cutoff and to assign those least in need to standard treatment while randomly assigning those on the other side of the cutoff into one of the two experimental conditions.) RD designs are flexible for examining treatment variations.

Posttreatment Measurement Variations

In RD designs, pre and post measures do not have to be on the same type of measure. For instance, one might assign persons to a health education treatment on the basis of household income but then examine its effects on attitudes toward health; here the assignment variable is an economic indicator and the outcome is a cognitive one.

The RD design is not restricted to a single outcome, and for each posttreatment measure there can be a separate RD design and analysis. However, it will often be useful to create an aggregate outcome or outcomes. For instance, we may have many knowledge items that are scaled to give a total (aggregate) score, with subtest scores for mathematical reasoning, computational skills, verbal reasoning, analogies, and so on. We might then conduct separate analyses for both the total score and each subtest.

Outcome measures need not be continuous normal variables. While such distributions facilitate RD analysis, they are not necessary for it. For example, Berk and Rauma (1983) evaluated the effects of a California law that extended unemployment benefits to released prisoners who had not previously been eligible. In order to qualify, prisoners had to earn at least $1500 working in the prison over the 12 months prior to release. Would extending these benefits reduce subsequent recidivism? A dichotomous outcome was created with 1 indicating recidivism and 0, staying out of prison. To analyze this outcome, Berk and Rauma (1983) relied on a linear random utility model related to the binary logit model.

Internal Validity

Internal validity refers to the degree to which causal inference is reasonable because alternative explanations for an apparent treatment effect can plausibly be ruled out. The RD design is very strong with respect to internal validity because the process of selection into treatments is fully known and only factors that would serendipitously induce a discontinuity in the pre–post relationship at the cutoff can be considered threats. However, validity also depends on how well the analyst can model the true pre–post relationship so that underlying nonlinear relationships do not masquerade as discontinuities at the cutoff point.

The RD design does not yield biased estimates of the treatment effect when the assignment variable is measured with

error (Cappelleri et al., 1991; Trochim et al., 1991). The reason such error does not induce bias here, although it does in other quasi-experiments, is that assignment to treatment is by the fallible assignment variable and not by some latent underlying selection process, as happens in quasi-experiments.

Statistical Power

It takes approximately 2.75 times as many participants in an RD design to achieve statistical power comparable to a simple randomized experiment, assuming equal numbers assigned to the treatment and control groups (Goldberger, 1972). When RD is combined with random assignment within an interval around this cutoff, 2.48 times the number of respondents are needed if 20% of participants are randomized, 1.96 times with 40% randomization, 1.46 with 60%, and 1.14 with 80% (Cappelleri and Trochim, 1994, 1995, 2010).

Parametric Analysis

Assumptions of RD Analysis

Five central assumptions must be made in order for responsible analysis of data from an RD design:

1. The cutoff criterion: The cutoff criterion must be followed without exception. Otherwise, unless some valid adjustment is made, a selection threat arises, and estimates of the treatment effect are likely to be biased.
2. The pre–post distribution: The true pre–post distribution must be known and correctly specified as linear, polynomial, logarithmic, or the like. Reichardt et al. (1995) point out the difficulties in specifying the appropriate model, especially when some curvilinearity exists in the pre-post relationship.
3. Comparison group pretest variance: There must be a sufficient range of pretest values in the comparison group to enable adequate estimation of the pre–post regression line for that group. Variability in the treatment group is also desirable, although not strictly required because one can project the comparison group line even to a single treatment group point.
4. Continuous pretest distribution: Both groups must come from a single continuous pretest distribution. In some cases one might find intact groups that serendipitously divide on a measure so as to imply some cutoff (e.g., patients from two different geographic locations). However, such naturally discontinuous groups would confound geography with treatment and should not be used.
5. Treatment implementation: It is assumed that the treatment is uniformly delivered to all recipients – that is, they all receive the same dosage, length of stay, amount of training, or whatever. If not, it is necessary to model explicitly the treatment as implemented, thus complicating the analysis considerably.

A Model for the Basic RD Design

The model presented here is for the basic RD design. Given a pretreatment assignment measure, x_i, and a posttreatment measure, y_i, the model for the expected value (E) of y_i can be stated as follows:

$$E(y_i) = \beta_0 + \beta_1 z_i + \beta_2 \tilde{x}_i + \beta_3 \tilde{x}_i z_i + \cdots + \beta_{n-1} \tilde{x}_i^s + \beta_n \tilde{x}_i^s z_i$$

where

$\tilde{x}_i$ = pretreatment measure for individual i minus the cutoff value (i.e., $\tilde{x}_i = x_i - x_0$)
y_i = posttreatment measure for individual i
z_i = assignment variable (1 if treatment participant; 0 if comparison participant)
s = the degree of the polynomial for the associated $\tilde{x}_i$
β_0 = parameter for comparison group intercept at cutoff x_0
β_1 = treatment effect parameter
β_2 = linear slope parameter
β_3 = linear interaction parameter
β_{n-1} = parameter for the sth-order polynomial
β_n = parameter for the sth-order polynomial interaction term

The major null hypothesis of interest H_0: $\beta_1 = 0$ is tested against the alternative H_1: $\beta_1 \neq 0$. The model estimates both main and interaction effects at the cutoff point in order to examine changes in means and slopes. It postulates a polynomial pre–post relationship. It also requires subtracting the cutoff score from each pretest score. The term $\tilde{x}_i$ has a superscript tilde to indicate this transformation of the pretest x_i.

Finally, the model allows for any order of polynomial function (although certain restrictions are made in specifying the function as described below). Thus, the true pre–post relationship can in theory be linear, quadratic, cubic, quartic, and so on.

In the simplest model with no interaction term, which includes only treatment and the pretest (pretreatment) assignment covariate as predictors, the treatment effect (β_1) is constant across pretest assignment scores and, therefore, the same at the cutoff score as any other pretest score on the assignment continuum. In a model with a linear treatment-by-pretest interaction, the treatment effect is no longer constant across the (pretest) assignment continuum but depends linearly on the value of the pretest assignment score, allowing for the population slopes of the pre–post relationship to differ on opposite sides of the discontinuity.

Model Specification

The key analytic problem is correctly specifying the model for the data – in this example, as a polynomial model. No simple or mechanical way exists to determine definitively the appropriate model. As is the case with any judicious statistical modeling, the RD analysis requires judgment and discretion, along with conducting multiple analyses based on different assumptions about the true pre–post relationship.

Steps in the Analysis

Recall what data the basic RD design provides on each unit. There is a pretest assignment value x_i in the model. Knowing this and the cutoff value allows creating a new variable z_i, which is equal to 1 in the treatment group and 0 if not. Finally, there is a posttest (or posttreatment) score labeled y_i in the

model. Given the variables x_i, z_i, and y_i, the steps to be followed in the polynomial analysis are as follows:

1. Transform the pretest

The analysis begins by subtracting the cutoff value from each pretest score thus creating the term $\tilde{x}_i$ as in the model. This sets the intercept equal to the cutoff so that estimates of effect are made at the cutoff rather than at $x_i = 0$.

2. Examine the relationship visually

It is important to determine whether there is any visually discernible discontinuity at the cutoff. It could be a change in level, slope, or both. If a discontinuity is visually clear at the cutoff one should not be satisfied with analytic results that indicate no effect. However, if no discontinuity is apparent, variability in the data may be masking an effect, and one must attend carefully to the analytic results.

The second thing to look for is the degree of polynomial that may be required as indicated by the overall slope of the distribution, but particularly in the comparison group part. A good approach is to count the number of flexion points (how often the distribution 'flexes' or 'bends'). A linear distribution implies no flexion points, while a single flexion point would indicate a quadratic function, and so on, such that the estimated polynomial is equal to one order less than the total number of flexion points (i.e., with two flexion points one would hypothesize a cubic or third-order polynomial). This information is used to specify the initial model.

3. Create higher order terms and interactions

Depending on the number of flexion points, create transformations of the transformed assignment variable $\tilde{x}_i$. The rule of thumb is to go two orders of polynomial higher than indicated by the number of flexion points. Thus, if the bivariate relationship appeared linear (i.e., no flexion points), one would want to create transformations up to a second-order $(0 + 2)$ polynomial. The first-order polynomial already exists in the model $(\tilde{x}_i)$, and one would only have to create the second-order polynomial by squaring $\tilde{x}_i$ to obtain $\tilde{x}_i^2$. For each transformation of $\tilde{x}_i$ one also creates the interaction term by multiplying the polynomial by z_i. In this example there would be two interaction terms: $\tilde{x}_i z_i$ and $\tilde{x}_i^2 z_i$. If there seems to be two flexion points in the bivariate distribution, one would then create transformations up to the fourth $(2 + 2)$ power and their interactions. This rule of thumb errs toward overestimating the true polynomial function needed, for reasons outlined in Trochim (1984).

4. Estimate the initial model

The true analysis can now begin. One simply regresses the posttest scores, y_i, on $\tilde{x}_i$, z_i and all higher order transformations and interactions created in step 3 above. The regression coefficient associated with the z_i term (i.e., the group membership variable) is the estimate of the main effect of the treatment. If there is a vertical discontinuity at the cutoff it will be estimated by this coefficient. One can test the significance of the coefficient (or any other) by constructing a standard t-test. If the analyst correctly overestimated the polynomial function required to model the distribution at step 3, then the treatment effect estimate will be unbiased. However, by initially including terms that may not be needed in the true model, the estimate is likely to be inefficient, that is, standard error terms will be inflated and statistical significance underestimated. However, if the coefficient for the effect is highly significant in this initial model, it would be reasonable to conclude there is an effect. Interaction effects can also be examined. A linear interaction is implied by a significant coefficient for the $\tilde{x}_i z_i$ term.

5. Refine the model

The procedure described thus far is conservative, designed to reduce the chances of a biased treatment effect estimate even at the risk of increasing the error of the estimate (Trochim, 1984). Therefore, on the basis of the results of step 4 one might wish to attempt to remove apparently unnecessary terms and reestimate the treatment effect with greater efficiency. This is tricky and should be approached with caution lest it introduces bias. One should certainly examine the regression results from step 4, noting the degree to which the overall model fits the data, the presence of any insignificant coefficients, and the pattern of residuals. One might examine the highest order term in the current model and its interaction. If both coefficients are nonsignificant, and the goodness-of-fit measures and pattern of residuals indicate a good fit, one might then drop these two terms and reestimate the model. One would repeat this procedure until (1) all the coefficients are significant, (2) the goodness-of-fit measure drops appreciably, or (3) the pattern of residuals indicates a poor fit.

Analyses with Multiple Cutoff Points

The basic model described above can be applied directly when multiple cutoff points are used and random assignment is followed within cutoff intervals. If assignment within intervals is nonrandom, one must also address the potential for selection bias in the analysis. When multiple cutoffs are used to distinguish separate treatments (i.e., multiple treatments are not assigned within the same cutoff interval), one would have to construct multiple treatment assignment variables for the analytic model (e.g., z_1, z_2, z_3) and all necessary interaction terms. Clearly, as more cutoffs and groups are added, model specification becomes more complex.

Analyses with Multiple Assignment Measures

Imagine having multiple measures, each with its own cutoff. Each assignment measure is then transformed by having its own cutoff value subtracted from it to create $\tilde{x}_i$. The analysis would then include all transformed assignment measures, group membership, higher order terms, and interactions. For the simple first-order (or linear) case, one could use the following model:

$$E(y_i) = \beta_0 + \beta_1 z_i + \beta_2 \tilde{x}_{1i} + \beta_3 \tilde{x}_{2i} + \beta_4 \tilde{x}_{3i} + \beta_5 \tilde{x}_{1i} z_i + \beta_6 \tilde{x}_{2i} z_i + \beta_7 \tilde{x}_{3i} z_i$$

This model does not include any two-way assignment variable multiplicative terms (e.g., $\tilde{x}_{1i} x_{2i}$ or $\tilde{x}_{1i} x_{2i} z_i$) or any three-way assignment variable multiplicative terms (e.g.,

$\tilde{x}_{1i}x_{2i}x_{3i}$ or $\tilde{x}_{1i}x_{2i}x_{3i}z_i$), but such an assumption may be reasonable where primary interest is in estimating β_1, the treatment effect at the cutoff, and not interactions. Clearly, however, the use of multiple assignment variables with higher order polynomial models will quickly lead to an unwieldy analysis.

The situation is simpler when multiple assignment variables are rescaled to reflect the number of assignment variables on which a person meets the criterion (e.g., on 5 of 10). This number then becomes the x_i variable, and one would conduct the basic analysis described earlier. When multiple assignment variables are combined into a single index, the analysis is also a straightforward application of the basic procedures described initially.

Recent Methodological Developments

Nonparametric Statistical Analysis

As noted, the most serious limitation of RD designs is the possible sensitivity of the functional form between the posttest outcome and pretest assignment covariate. Traditionally, parametric models with polynomial regression have been used to analyze RD design. If the extrapolation beyond the cutoff score is not adequate, then what masquerades as a jump due to treatment might simply be a nonlinear relationship showing no real treatment effect.

To reduce the likelihood of such mistakes, economists have proposed and implemented nonparametric approaches that look only at the data in the neighborhood around the cutoff score (Hahn et al., 2001; Ludwig and Miller, 2007; Imbens and Lemieux, 2008; Angrist and Pischke, 2009). Comparisons of average outcome scores in small-enough neighborhoods to the left and right of the cutoff can be estimated in a way that does not depend on the correct functional form. Instead of providing a constant treatment effect generalizable across all values of the pretest assignment covariate (in the case of a model with no pretest-treatment interaction), as occurs with the parametric approach, the average treatment effect in the nonparametric approach is restricted to (and conditional on) the cutoff score.

The nonparametric approach to RD requires good estimates of the mean posttreatment scores in small neighborhoods to the immediate right and left of the cutoff. However, such estimation can be tricky. One notably conspicuous problem is the sparse or limited data in a small neighborhood of the cutoff means. Local linear regression, a nonparametric smoothing technique, has been proposed to address the typically relatively scant data around the cutoff means (Hahn et al., 2001; Imbens and Lemieux, 2008; Angrist and Pischke, 2009).

Nonparametric RD estimation is not without its shortcomings (Bloom, 2012). The method requires very large samples to provide an adequate number of observations in the two bandwidths adjacent to the cutoff point. Another limitation is the potential sensitivity of nonparametric estimations to the choice of bandwidth. A trade-off is needed between, on the one hand, introducing bias from a bandwidth that is too wide and, on the other hand, losing precision from a bandwidth that is too narrow. The most widely used empirical approach for making such a trade-off is cross-validation (Imbens and Lemieux, 2008).

Fuzzy RD Design

As noted previously, a violation of the assignment rule to treatment and comparison groups results in a fuzzy RD design. Here external factors result in individuals being placed in one group when they should have been placed in the other group according to their pretreatment assignment score. A conventional analysis of a fuzzy RD design, which does not account for this misassignment, would result in a biased treatment effect. Under certain circumstances, a potential solution to obtain an unbiased treatment effect in the fuzzy RD design is to apply the technique of instrumental variables (Angrist and Pischke, 2009; Murnane and Willett, 2011). Here the pretreatment (or pretest) assignment measure can serve as a credible instrumental variable for actual group membership if this pretreatment assignment measure predicts actual group membership and, at the same time, is not directly related to the posttreatment measure.

A two-stage least squares regression method (a parametric approach based on ordinary least squares) to instrumental variable estimation has been proposed and applied to analyze the fuzzy RD design (Jacob and Lefgren, 2004; Angrist and Pischke, 2009; Murnane and Willett, 2011). The nonparametric version of fuzzy RD consists of instrumental variable estimation in a small neighborhood around the discontinuity. Hahn et al. (2001) developed a nonparametric procedure for instrumental variables using local linear regression around the cutoff, which was applied to estimate the effect of class size on childrens' test scores (Angrist and Lavy, 1999).

In research related to instrumental variables, Battistin and Retorre (2008) and Bloom (2012) provide additional insights and expanded paradigms for fuzzy RD designs. They provide conditions that are required to make fuzzy RD designs valid in identifying average treatment effects.

Design Variations

Building on previous research on logistical and analytical issues of RD designs with varying amounts of randomization (Trochim and Cappelleri, 1992; Cappelleri and Trochim, 1994, 1995, 2010; Cappelleri et al., 1994), Mandell (2008) provides extended formulations and insights into this hybrid design in which cutoff-based assignment and random assignment are combined depending on values of the pretreatment assignment indicator to experimental conditions. Different hybrid designs can have varying degrees of randomization and discontinuity. They entail smaller sacrifices in statistical efficiency than the basic or pure RD design (which has no randomization embedded within it) and are also likely, in certain circumstances, to be judged considerably fairer than a true randomized experiment and possibly even a pure RD design.

Another variation is the clustered RD design where groups (rather than individuals) are assigned to an intervention. Considerably more methodological work has been performed on the design and analysis of RD designs when treatments are applied to individuals (Shadish et al., 2002;

Imbens and Lemieux, 2008) than to groups. More recently, however, increased methodological research on RD designs has been applied in the context of community or group-level interventions. The clustered RD design, for example, was the primary design to evaluate the federal education program mandated in the No Child Left Behind Act of 2001 (Gamse et al., 2008). Motivated by 'No Child Left Behind,' Schochet (2009) examined the statistical power under such clustered RD designs (without randomization) using techniques from the causal inference and hierarchical linear modeling literature. The main conclusion is that three to four times larger samples are typically required under the clustered RD design than under the clustered randomized design to achieve estimates of effects with the same level of precision.

Pennell et al. (2011) noted that designs in education are often quite different from designs used in public health and investigated the design and analysis of clustered RD designs in which a varying proportion of groups is randomized. These researchers extend upon previous work to consider analysis, statistical power, and sample size implications of RD designs in community-based intervention studies in public health. In doing so they examined the power of these designs as a function of the intraclass correlation, number of groups, and number of members per group and compared the results with the traditional group randomized trial.

Variations on the RD design can also be distinguished by the nature of the assignment. Research on the RD design has typically focused on applications with a single assignment variable. In many settings, however, it may be more appropriate to impose cutoffs on several assignment variables in defining a set of different treatments. For instance, cutoffs on two assignment variables in a basic RD design would result in two dimensions with four treatment regions. By using regression to model the response surface in each region, and obtaining predicted values along each of the discontinuity edges, Papay et al. (2011) show how to generalize the basic RD design to include multiple assignment variables simultaneously and to estimate the effects of several treatments.

Conclusions

From a methodological point of view, inferences drawn from a well-implemented RD design are considered comparable in internal validity to conclusions from randomized experiments. However, the lower statistical efficiency of RD designs, and the resulting sample size demands, may limit the design's utility. From an ethical perspective, RD designs are compatible with the goal of getting the treatment to those most deserving or in need. It is not necessary to deny the treatment from potentially deserving recipients simply for the sake of a scientific test. From an administrative viewpoint, the RD design may be directly usable when allocation formulas are the basis for assigning treatment or treatments.

With all these considerations in mind, the RD design must be used judiciously. In general, the randomized experiment is still the method of first choice when assessing causal hypotheses (Cappelleri and Trochim, 2010). However, where they are ruled out as impractical, the RD design should be considered as a practical high-quality alternative.

See also: Causal Counterfactuals in Social Science Research; Comparative Studies: Method and Design; Control Variables in Research; External Validity; Internal Validity; Nonequivalent Group Designs.

Bibliography

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.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.

Angrist, J.D., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 533–575.

Battistin, E., Retorre, E., 2008. Ineligibles and eligibles non-participants as a double comparison group in regression discontinuity designs. Journal of Econometrics 142, 715–730.

Bennett, C.A., Lumsdaine, A.A. (Eds.), 1975. Evaluation and Experiment. Academic Press, New York, NY.

Berk, R.A., Rauma, D., 1983. Capitalizing on nonrandom assignment to treatment: a regression discontinuity of a crime control treatment. Journal of the American Statistical Association 78, 21–28.

Berk, R.A., de Leeuw, J., 1999. An evaluation of California's inmate classification system using a generalized regression discontinuity design. Journal of the American Statistical Association 94, 1045–1052.

Bloom, H.S., 2012. Modern regression discontinuity analysis. Journal of Research on Educational Effectiveness 5, 43–82.

Boruch, R.F., 1973. Regression-Discontinuity Designs Revisited. Unpublished manuscript, Northwestern University.

Campbell, D.T., Stanley, J.C., 1963. Experimental and quasi-experimental designs for research on teaching. In: Gage, N.L. (Ed.), Handbook of Research on Teaching. Rand McNally, Chicago, IL (Also published as Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally, 1966.).

Cappelleri, J.C., Trochim, W.M.K., 2010. Cutoff designs. In: Chow, S.-C. (Ed.), Encyclopedia of Biopharmaceutical Statistics, third ed. revised and expanded. Informa Healthcare, New York, NY, pp. 386–391.

Cappelleri, J.C., Trochim, W.M.K., 1995. Ethical and scientific features of cutoff-based designs of clinical trials: a simulation study. Medical Decision Making 15, 387–394.

Cappelleri, J.C., Darlington, R.B., Trochim, W.M.K., 1994. Power analysis of cutoff-based randomized clinical trials. Evaluation Review 18, 141–152.

Cappelleri, J.C., Trochim, W.M.K., 1994. An illustrative statistical analysis of cutoff-based randomized clinical trials. Journal of Clinical Epidemiology 47, 261–270.

Cappelleri, J.C., Trochim, W.M.K., Stanley, T.D., Reichardt, C.S., 1991. Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: I. The case of no interaction. Evaluation Review 15, 395–419.

Card, D., Shore-Sheppard, L.D., 2004. Using discontinuous eligibility rules to identify the effects of the federal Medicaid expansions on low-income children. Review of Economics and Statistics 86, 752–766.

Cook, T.D., 2008. "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. Journal of Econometrics 142, 636–654.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis for Field Settings. Rand McNally, Chicago, IL.

DiNardo, J., Lee, D.S., 2004. Economic impacts of new unionization on private sector employers: 1984–2001. Quarterly Journal of Economics 119, 1383–1441.

Finkelstein, M.O., Levin, B., Robbins, H., 1996a. Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. American Journal of Public Health 86, 691–695.

Finkelstein, M.O., Levin, B., Robbins, H., 1996b. Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. American Journal of Public Health 86, 696–705.

Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical Methods for Rates and Proportions, third ed. John Wiley & Sons, New York, NY.

Gamse, B.C., Jacob, R.T., Horst, M., Boulay, B., Unlu, F., 2008. Reading First Impact Study Final Report (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldberger, A.S., 1972. Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations. Discussion Papers, 123–72. Institute for Research on Poverty, University of Wisconsin, Madison, WI.

Hahn, J., Todd, T., Van Der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrics 69, 201–209.

Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. Journal of Econometrics 142, 625–635.

Jacob, B.A., Lefgren, L., 2004. Remedial education and student achievement: a regression-discontinuity analysis. Review of Economics and Statistics 86, 226–244.

Judd, C.M., Kenny, D.A., 1981. Estimating the Effects of Social Interventions. Cambridge University Press, New York, NY.

Lalive, R., 2008. How do extended benefits affect unemployment duration? a regression-discontinuity approach. Journal of Econometrics 142, 785–806.

Lee, D.S., 2008. Randomized experiments from non-random selection in U.S. House elections. Journal of Econometrics 142, 675–697.

Linden, A., Adams, J.L., Roberts, N., 2006. Evaluating disease management programme effectiveness: an introduction to the regression discontinuity design. Journal of Evaluation in Clinical Practice 12, 124–131.

Ludwig, J., Miller, D.L., 2007. Does head start improve children's life chances? evidence from a regression discontinuity design. Quarterly Journal of Economics 122, 159–208.

Mandell, M.B., 2008. Having one's cake and eating it, too: combining true experiments with regression discontinuity designs. Evaluation Review 32, 415–434.

Mark, M.M., Mellor, S., 1991. Effect of self-relevance of an event on hindsight bias: the foreseeability of a layoff. Journal of Applied Psychology 76, 569–577.

Matsudaira, J.D., 2008. Mandatory summer school and student achievement. Journal of Econometrics 142, 829–850.

Mohr, L.B., 1995. Impact Analysis for Program Evaluation, second ed. Sage Publications, Thousands Oaks, CA.

Murnane, R.J., Willett, J.B., 2011. Methods Matter: Improving Causal Inference in Educational and Social Science Research. Oxford University Press, New York, NY.

Papay, J.P., Willett, J.B., Murnane, R.J., 2011. Extending the regression-discontinuity approach to multiple assignment variables. Journal of Econometrics 161, 203–207.

Pennell, M.L., Hade, E.M., Murray, D.M., Rhoda, D.A., 2011. Cutoff designs for community-based intervention studies. Statistics in Medicine 30, 1865–1882.

Reichardt, C.S., Trochim, W.M.K., Cappelleri, J.C., 1995. Reports of the death of regression-discontinuity analysis are greatly exaggerated. Evaluation Review 19, 39–63.

Riecken, H.W., Boruch, R., Campbell, D.T., Caplan, N., Glenman, T.K., et al., 1974. Social Experimentation: A Method for Planning and Evaluating Social Intervention. Academic Press, New York, NY.

Rubin, D.B., 1977. Assignment to treatment on the basis of a covariate. Journal of Educational Statistics 2, 1–26.

Schochet, P.Z., 2009. Statistical power of regression discontinuity designs in education evaluations. Journal of Educational and Behavioral Statistics 34, 238–266.

Senn, S.J., 2007. Statistical Issues in Drug Development, second ed. John Wiley & Sons, Chichester, UK.

Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin Company, Boston, MA.

Tallmadge, G.K., Horst, D.P., 1976. A Procedural Guide for Validating Achievement Gains in Educational Projects. Office of Education, Washington, DC.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-discontinuity analysis: an alternative to the *ex post facto* experiment. Journal of Educational Psychology 51, 309–317.

Trochim, W.M.K., Cappelleri, J.C., 1992. Cutoff assignment strategies for enhancing randomized clinical trials. Controlled Clinical Trials 13, 190–212.

Trochim, W.M.K., Cappelleri, J.C., Reichardt, C.S., 1991. Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: II. When an interaction effect is present. Evaluation Review 15, 571–604.

Trochim, W.M.K., 1990. Regression-discontinuity design in health evaluation. In: Sechrest, L., Perrin, E., Bunker, J. (Eds.), Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. U.S. Department of Health & Human Services, Rockville, MD, pp. 119–139.

Trochim, W.M.K., 1984. Research Design for Program Evaluation: The Regression-Discontinuity Approach. Sage, Beverly, CA.

Van Der Klaauw, W., 2008. Breaking the link between poverty and low student achievement: an evaluation of title I. Journal of Econometrics 142, 731–756.

Van Der Klaauw, W., 2002. Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach. International Economic Review 43, 1249–1287.

Wilder, C.S., 1972. Physician Visits, Volume, and Interval since Last Visit, U.S., 1969. National Center for Health Statistics, Rockville, MD (Series 10, No. 75; DHEW Pub. No. HSM 72-1064).

Winship, C., Morgan, S.L., 1999. The estimation of causal effects from observational data. Annual Review of Sociology 25, 659–707.

Wong, V.C., Cook, T.D., Barnett, W.S., Jung, K., 2008. An effectiveness-based evaluation of five state pre-kindergarten programs. Journal of Policy Analysis and Management 27, 122–154.

Zuckerman, I.H., Lee, E., Wutoh, A.K., Zue, Z., Stuart, B., 2006. Application of regression-discontinuity analysis in pharmaceutical health services research. Health Services Research 41, 550–563.