# On Interpreting the Regression Discontinuity Design as a Local Experiment*


Jasjeet S. Sekhon†
Robson Professor
Political Science and Statistics
UC Berkeley

Rocío Titiunik‡
James Orin Murfin Associate Professor
Political Science
University of Michigan


October 2, 2016


**Abstract**

We discuss the two most popular frameworks for identification, estimation and inference in regression discontinuity (RD) designs: the continuity-based framework, where the conditional expectations of the potential outcomes are assumed to be continuous functions of the score at the cutoff, and the local randomization framework, where the treatment assignment is assumed to be as good as randomized in a neighborhood around the cutoff. Using various examples, we show that (i) assuming random assignment of the RD running variable in a neighborhood of the cutoff neither implies that the potential outcomes and the treatment are statistically independent nor that the potential outcomes are unrelated to the running variable in this neighborhood, and (ii) assuming local independence between the potential outcomes and the treatment does not imply the exclusion restriction that the score affects the outcomes only through the treatment indicator. Our discussion highlights key distinctions between "locally randomized" RD designs and real experiments, including that statistical independence and random assignment are conceptually different in RD contexts, and that the RD treatment assignment rule places no restrictions on how the score and potential outcomes are related. Our findings imply that the methods for RD estimation, inference and falsification used in practice will necessarily be different (both in formal properties and in interpretation) according to which of the two frameworks is invoked.



*We are indebted to Matias Cattaneo, Anthony Fowler, Kellie Ottoboni, Nicolás Idrobo Rincón, Kosuke Imai, Joshua Kalla, Fredrik Sävje, Yotam Shem-Tov, and three anonymous reviewers for helpful comments and discussion. Sekhon gratefully acknowledges support from the Office of Naval Research (N00014-15-1-2367) and Titiunik gratefully acknowledges financial support from the National Science Foundation (SES 1357561).



†UC Berkeley, Travers Department of Political Science, 210 Barrows Hall #1950, Berkeley, CA 94720-1950. Email: <sekhon@berkeley.edu>, Web: `http://sekhon.berkeley.edu`.

‡Department of Political Science, 505 South State St., 5700 Haven Hall, University of Michigan, Ann Arbor, MI 48109. Email: <titiunik@umich.edu>, Web: `http://www.umich.edu/~titiunik`.

The regression discontinuity (RD) design is a research strategy based on three main components—a score or "running variable," a cutoff, and a treatment. Its basic characteristic is that the treatment is assigned based on a known rule: all units receive a score value, and the treatment is offered to those units whose score is above a cutoff and not offered to those units whose score is below it (or viceversa).

The RD design has been available since the 1960s (Thistlethwaite and Campbell 1960), but its popularity has grown particularly fast in recent years. In the last decade, an increasing number of empirical researchers across the social and biomedical sciences has turned to the RD design to estimate causal effects of treatments that are not, and often cannot be, randomly assigned. This growth in RD empirical applications has occurred in parallel with a rapid development of methodological tools for estimation, inference and interpretation of RD effects. See Cook (2008), Imbens and Lemieux (2008) and Lee and Lemieux (2010) for early reviews, and the introduction to this volume by Cattaneo and Escanciano for a comprehensive list of recent references.

The recent popularity of the RD design was in part sparked by the work of Hahn, Todd, and van der Klaauw (2001), who translated the design into the Neyman-Rubin potential outcomes framework (Holland 1986) and offered minimal conditions for non-parametric identification of average effects. These authors showed that, when all units comply with their assigned treatment, the average effect of the treatment at the cutoff can be identified under the assumption that the conditional expectations of the potential outcomes given the score are continuous (and other mild regularity conditions). This emphasis on continuity conditions for identification was a departure from other "quasi-experimental" research designs, which are typically based on independence or mean independence assumptions.

In the Neyman-Rubin framework, randomized experiments are seen as the gold standard, and quasi-experimental designs are broadly characterized in terms of the assumptions under which the (non-random) treatment assignment mechanism is as good as randomized. For example, in observational studies based on the unconfoundedness or "selection-on-observables" assumption, the treatment is as good as randomly assigned, albeit with unknown distribution, after conditioning on observable covariates (Imbens and Rubin 2015). Similarly, instrumental variables designs can be seen as randomized experiments with imperfect compliance, and difference-in-difference designs compare treatment and control groups that, on average—and except for time-invariant characteristics that can be removed by differencing—differ only on treatment status. In all these cases, a treatment and a control group are well defined and, under certain assumptions, can be compared to identify the average treatment effect for the population of interest (or a subpopulation thereof).

In this context, the results derived by Hahn, Todd, and van der Klaauw (2001) set the RD design apart. In contrast to most other quasi-experimental designs, RD identification was established only for the average treatment effect at a single point: the cutoff. This implied that, unlike differences-in-differences or selection-on-observable designs, the RD design could not be accurately described by appealing to a comparison between a treatment and a control group. Given the RD assignment rule, under perfect compliance it is impossible for treated and control units to have the same score value. Moreover, when the running variable is continuous, the probability of seeing an observation with a score value exactly equal to the cutoff is zero. Thus, in the continuity-based RD setup, identification of the average treatment effect at the cutoff necessarily relies on extrapolation, as there are neither treated nor control observations with score values exactly equal to the cutoff.

The influential contribution by Lee (2008) changed the way RD was perceived, and aligned the interpretation of RD designs with experimental (and the other quasi-experimental) research designs. Lee (2008) argued that in a RD setting where the score can be influenced by the subjects' choices and unobservable characteristics, treatment status can be interpreted to be as good as randomized in a local neighborhood of the cutoff as long as subjects lack the ability to precisely determine the value of the score they receive—i.e., as long as their score contains a random chance component. Lee's framework captured the original ideas in the seminal article by Thistlethwaite and Campbell, who called a hypothetical experiment where the treatment is randomly assigned near the cutoff an "experiment for which the regression-discontinuity analysis may be regarded as a substitute" (Thistlethwaite and Campbell 1960, p. 310).

The interpretation of RD designs as local experiments developed by Lee (2008) has been very influential, both conceptually and practically. Among other things, it established the need to provide falsification tests based on predetermined covariates just as one would do in the analysis of experiments, a practice that has now been widely adopted and has increased the credibility of countless RD applications (e.g., Caughey and Sekhon 2011; Eggers et al. 2014; Hyytinen et al. 2015). Moreover, it provided an intuitive interpretation of the RD parameter that allowed researchers to think about treatment and control groups instead of an effect at a single point—the cutoff—where there are effectively no observations.

The claim that the RD treatment assignment rule is "as good as randomized" in a neighborhood of the cutoff can be interpreted in at least two ways. In one interpretation, it means that there should be no treatment effect on predetermined covariates at the cutoff, and that the validity of the underlying RD assumptions can be evaluated by testing the null hypothesis that the RD treatment effect is zero on predetermined covariates. In another interpretation, it means that the treatment is (as good as) randomly assigned near the cutoff, and estimation and inference for treatment effects (and covariate balance tests) can be carried out using the same tools used in experimental analysis. The first interpretation does not imply the second because one can test for covariate balance (at the cutoff) under the usual continuity assumptions. While the first interpretation has resulted in an increased and much needed focus on credibility and falsification, the second has been the source of considerable confusion. Our goal is to discuss the source of such confusion in detail. In doing so, we clarify crucial conceptual distinctions within the local experiment RD framework, which in turn elucidate the differences and similarities between this framework and the more standard continuity-based approach.

We explore both the relationship between continuity and local randomization assumptions in RD designs, and the implications of adopting a RD framework based on an explicit local randomization assumption. Our paper builds on prior studies that have considered this issue. Hahn, Todd, and van der Klaauw (2001) first invoked a local randomization assumption for identification of RD effects under noncompliance. More recently, Cattaneo, Frandsen, and Titiunik (2015) formalized an analogous assumption using a Fisherian, randomization-based RD framework, which was extended by Cattaneo, Titiunik, and Vazquez-Bare (2016a,b). Various super-population versions of the local randomization RD assumption were also proposed by Keele, Titiunik, and Zubizarreta (2015) and Angrist and Rokkanen (2015), and more recently de la Cuesta and Imai (2016) discussed the relationship between local randomization and continuity RD assumptions.

We make two main arguments. First, we show that the usual RD continuity assumptions are not sufficient for a literal local randomization interpretation of RD designs—that is, not sufficient to ensure that, near the cutoff, the potential outcomes are independent of the treatment assignment and unrelated to the running variable. This has been argued previously (Cattaneo, Frandsen, and Titiunik 2015; de la Cuesta and Imai 2016); we simply provide a stylized example to further illustrate the main issues. Second, contrary to common practice, we show that the assumption that the treatment is randomly assigned among units in a neighborhood of the cutoff *cannot* be used to justify analyzing and interpreting RD designs as actual experiments. The restrictions imposed by the treatment assignment rule in a sharp RD design rule out a local experiment where one randomly assigns treatment status without changing the units' score values. Instead, one could assume that score values are randomly assigned in a neighborhood of the cutoff, a randomization model that is consistent with the RD assignment mechanism. However, even under this model, interpreting and analyzing the RD design as an experiment will generally result in invalid inferences because the score can affect the potential outcomes directly in addition to through treatment status.

In particular, we show that (i) assuming random assignment of the RD running variable in a neighborhood of the cutoff does not imply that the potential outcomes and the treatment assignment are statistically independent or that the potential outcomes are unrelated to the running variable in this neighborhood, and (ii) assuming local independence between the potential outcomes and the treatment assignment does not imply the exclusion restriction that the score affects the outcomes only via the treatment assignment indicator. Our discussion makes clear that the RD treatment assignment rule need not place any restrictions on the ways in which the score influences the potential outcomes and shows that, in local randomization RD settings, statistical independence and random assignment are conceptually different.

The discussion that follows focuses on the sharp RD design, in which units' compliance with treatment assignment is perfect, and the probability of receiving treatment changes from zero to one at the cutoff. We do not explicitly discuss fuzzy RD settings, where compliance is imperfect and the decision to take treatment is endogenous, because most of our arguments and conclusions apply equally to both sharp and fuzzy RD designs. When our discussion must be modified for the fuzzy case, we note it explicitly. Throughout, we refer to the standard RD design based on continuity identification assumptions as the *continuity-based* RD design; and to the RD design based on a local randomization assumption as the *local randomization* or *randomization-based* RD design.

# 1 The Continuity-Based RD Framework

In this and the subsequent sections we assume that we have a random sample $\{Y_i(1), Y_i(0), X_i\}_{i=1}^n$, where $X_i$ is the score on the basis of which a binary treatment $T_i$ is assigned according to the rule $T_i = \mathbb{1}(X_i \geq c)$—or $T_i = \mathbb{1}(X_i \leq c)$ depending on the example—for a known constant $c$, $Y_i(1)$ is the potential outcome under the treatment condition, and $Y_i(0)$ the potential outcome under the untreated or control condition. For every unit $i$ we observe either $Y_i(0)$ or $Y_i(1)$, so the observed sample is $\{Y_i, X_i\}_{i=1}^n$ where $Y_i := T_i Y_i(1) + (1 - T_i)\, Y_i(0)$. We assume throughout that all moments we employ exist, and that the density of $X_i$ is positive and continuous at the cutoff or in the intervals we consider.

The continuity-based framework is based on the identification conditions and estimation methods first proposed by Hahn, Todd, and van der Klaauw (2001). The authors proposed the following assumption:

**Assumption 1: Continuity.** The regression functions $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ are continuous in $x$ at the cutoff $c$.

Under this assumption, they showed that

$$\tau_{\mathtt{CB}}^{\mathtt{RD}} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]. \tag{1}$$

Thus, when the focus is on mean effects, the target parameter in the continuity-based RD framework is $\tau_{\mathtt{CB}}^{\mathtt{RD}}$—the average treatment effect at cutoff. The identification result in equation (1) states that, under continuity, this parameter—which depends on potential outcomes that are fundamentally unobservable, can be expressed as the difference between the right and left limits of the average observed outcomes at the cutoff. Estimation is therefore concerned with constructing appropriate estimators for $\lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]$ and $\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]$.

The most commonly used approach to estimate these limits is to rely on local polynomial methods, fitting two polynomials of the observed outcome on the score—one for observations above the cutoff, the other for observations below it— using only observations in a neighborhood of the cutoff, with kernel weights assigning higher weights to observations closer to the cutoff. Because these fitted polynomials are approximations to the unknown regression functions $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ near the cutoff, the choice of neighborhood, commonly known as *bandwidth*, is crucial. Given the order of the polynomial—typically one—the bandwidth controls the quality of the approximation, with smaller bandwidths reducing the bias of the approximation and larger bandwidths reducing its variance—see Cattaneo and Vazquez-Bare (2016) for an overview of RD neighborhood selection.

Although the technical details of local polynomial estimation and inference are outside the scope of our discussion, we highlight several issues that are central to distinguishing the continuity-based approach from the local randomization approach we discuss next. In the continuity-based RD design:[1]

- The target parameter, $\tau_{\texttt{CB}}^{\texttt{RD}}$, is an average effect at a single point, rather than in an interval.
- The functional form of the regression functions $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ is unknown, and is locally approximated by a polynomial for estimation and inference.
- In general, the polynomial approximation will be imperfect. The approximation error is controlled by the bandwidth sequence, $h_n$: the smaller $h_n$, the smaller the error.
- Local polynomial estimation and inference are based on large-sample results that require the bandwidth to shrink to zero as the sample size increases. For example, consistent estimation of $\tau_{\texttt{CB}}^{\texttt{RD}}$ requires $h_n \to 0$ and $nh_n \to \infty$.
- Local polynomial methods require smoothness conditions on the underlying regression functions $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ in order to control the leading biases of the local polynomial RD estimators. These smoothness conditions are stronger than the continuity assumption required for identification of $\tau_{\texttt{CB}}^{\texttt{RD}}$.
- The bandwidth $h_n$ plays no role in the identification of $\tau_{\texttt{CB}}^{\texttt{RD}}$.
- Neither the continuity assumptions required for identification nor the smoothness assumptions required for estimation and inference are implied by the RD treatment assignment rule. See Sekhon and Titiunik (2016) for further discussion.

## 1.1 Continuity Does Not Imply Local Randomization

We now consider a stylized example that shows that continuity of the potential outcomes regression functions (as in Assumption 1) does not imply that the treatment can be seen as

[1] For a general treatment of local polynomial methods, see Fan and Gijbels (1996) and Calonico, Cattaneo, and Farrell (2016) for recent higher-order results. For local polynomial methods applied specifically to the RD case, see Hahn, Todd, and van der Klaauw (2001), Porter (2003), Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014), and Calonico et al. (2016).

locally randomly assigned in a literal or precise sense. We assume that we have a sample of $n$ students who take a mathematics exam in the first quarter of the academic year. Each student receives a test score $X_i$ in the exam, which ranges from 0 to 100, and students whose grade is equal to or below 50 receive a double dose of algebra instruction ($T_i = 1$) during the second quarter—while students with $X_i > 50$ receive a single dose ($T_i = 0$). The outcome of interest is the test score $Y_i$ obtained in another mathematics exam taken at the end of the second quarter, also ranging between 0 and 100. We assume that test scores are fine enough so that they have no mass points and can be treated as continuous random variables.

To illustrate our argument, we assume that a student's expected grade in the second quarter under the control condition given her score in the first quarter, $\mathbb{E}[Y_i(0)|X_i = x]$, is simply equal to her score in the first quarter, $X_i$. We also assume that the double-dose algebra treatment effect, $\tau$, is constant for all students. The potential outcomes regression functions are therefore:

$$\begin{aligned} \mathbb{E}[Y_i(0)|X_i] &= X_i \qquad (2) \\ \mathbb{E}[Y_i(1)|X_i] &= X_i + \tau. \end{aligned}$$

This model is of course extremely simplistic, but we adopt it because it allows us to illustrate the difference between the treatment assignment mechanism and the outcome model in a straightforward way. We now assume that the initial grade $X_i$ is entirely determined by each student's fixed inherent (and unobservable) ability $a_i$—e.g. $X_i = g(a_i)$ for some strictly increasing function $g(\cdot)$. Given this setup, the assignment of treatment according to the rule $\mathbb{1}(X_i \leq 50)$ is as far from random as can be conceived, since it assigns all students of lower ability to the treatment group and all students of higher ability to the control group, inducing a complete lack of common support in the distribution of ability between the groups. Despite this severe selection into treatment based on ability, the effect of the treatment at the cutoff is readily identifiable, because the regression functions in equation (2) are continuous at the cutoff (and everywhere else).

In this context, what does it mean to say that the RD design induces as-if randomness near the cutoff? No matter how close to the threshold we get, the average ability in the control group is always higher than in the treatment group: the effect of $X$ on $\mathbb{E}[Y_i(j)|X_i]$ near the cutoff is always nonzero—$d\mathbb{E}[Y_i(j)|X_i = x]/dx = 1$ for $j = 0, 1$ and for all $x$. Thus, the continuity of the regression functions is entirely compatible with a very strong relationship between outcome and score. This means that continuity is not sufficient to guarantee the comparability of units on either side of the cutoff, even in a small neighborhood around it. In other words, for any $w > 0$, one can always conceive a data generating process such that the distortion induced by ignoring the relationship between $X$ and $Y$ in the window $[c-w, c+w]$ is arbitrarily large (a uniformity argument).

Thus, from an identification point of view, it is immediate to see that the continuity condition in Assumption 1 that ensures identification in the super population does not imply that the usual finite-sample "random assignment" identification assumptions hold. We will discuss the latter type of assumptions in detail in the following section, but we now consider one possibility in the context of the example in equation (2). Imagine that in this example we wish to invoke a mean independence assumption in a small neighborhood $W = [c-w, c+w]$ around the cutoff, with $w > 0$, such as $\mathbb{E}[Y_i(j)|T_i, X_i \in W] = \mathbb{E}[Y_i(j)|X_i \in W]$ for $j = 0, 1$; recall that in this example $T_i = \mathbb{1}(X_i \leq c)$.

Focusing, for example, on the potential outcome under control, and given equation (2), we have

$$\begin{aligned}
\mathbb{E}[Y_i(0)|T_i = 0, X_i \in W] &= \mathbb{E}\left[Y_i(0)|X_i > c, X_i \in [c-w, c+w]\right] \\
&= \mathbb{E}\left[Y_i(0)|X_i \in [c, c+w]\right] \\
&= \mathbb{E}\left[X_i|X_i \in [c, c+w]\right]
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[Y_i(0)|X_i \in W] &= \mathbb{E}\left[Y_i(0)|X_i \in [c-w, c+w]\,\right] \\ &= \mathbb{E}\left[X_i|X_i \in [c-w, c+w]\,\right]\end{aligned}$$

In general, $\mathbb{E}\left[X_i|X_i \in [c, c+w]\,\right] \neq \mathbb{E}\left[X_i|X_i \in [c-w, c+w]\,\right]$, so the mean independence assumption that is typically invoked in experiments cannot be invoked in this case. In other words, this example shows that the continuity assumption is not enough to guarantee independence between the potential outcomes and the treatment near the cutoff, and consequently cannot be used to justify analyzing a RD design as one would analyze an experiment.

From an estimation point of view, imagine that we mistakenly decided to analyze this valid continuity-based RD design as an experiment in the fixed neighborhood $[c-w, c+w]$ around the cutoff, defining the parameter of interest as average treatment effect in this window. To calculate this effect, we would simply compare the average treated-control difference in the observed outcomes in $[c-w, c+w]$. Unsurprisingly, this approach would lead to an incorrect answer, since

$$\begin{aligned}\mathbb{E}[Y_i|c-w \leq X_i \leq c] - &\mathbb{E}[Y_i|c < X_i \leq c+w] \\ &= \tau + \mathbb{E}[X_i|c-w \leq X_i \leq c] - \mathbb{E}[X_i|c < X_i \leq c+w] \\ &< \tau,\end{aligned}$$

where the last line follows from the fact that $\mathbb{E}[X_i|c-w \leq X_i \leq c] - \mathbb{E}[X_i|c < X_i \leq c+w] \in [-2w, 0)$.[2] Thus, for any fixed window $W = [c-w, c+w]$, analyzing this RD design as one would analyze an experiment will lead to a biased treatment effect estimate.

In this example, the smaller $w$, the closer the naive local randomization estimate will be to the true effect $\tau$. However, as we discuss below, in order for the local randomiza-

[2] If the treatment assignment rule were $\mathbb{1}(X_i \geq c)$, which is the more common definition of the treatment in the RD literature, we would have $\mathbb{E}[Y_i|c \leq X_i \leq c+w] - \mathbb{E}[Y_i|c-w \leq X_i < c] = \tau + \mathbb{E}[X_i|c \leq X_i \leq c+w] - \mathbb{E}[X_i|c-w \leq X_i < c]$, and $\mathbb{E}[X_i|c \leq X_i \leq c+w] - \mathbb{E}[X_i|c-w \leq X_i < c] \in (0, 2w]$.

tion RD framework to be conceptually distinct from the continuity-based framework, the neighborhood $W$ must necessarily be conceived as fixed. And, given a fixed $W$, the lack of comparability between treated and control groups cannot be eliminated by increasing the sample size—a direct consequence of that fact that such lack of comparability is an identification problem rather than an estimation one.

In contrast, in the continuity-based framework, focusing on the average treated-control outcome difference in a neighborhood of the cutoff can be understood as approximating the unknown potential outcomes regression functions with a local constant fit. Such strategy will result in a possibly large approximation error; however, the error is entirely due to the estimation strategy and will vanish asymptotically. In this fundamental sense, a RD design based on a continuity assumption cannot be interpreted literally as a local experiment. In other words, assuming the conditional regression functions $\mathbb{E}[Y_i(1)|X_i]$ and $\mathbb{E}[Y_i(0)|X_i]$ are continuous in $X_i$ at $c$ is not enough to treat the RD design as a pure experiment near the cutoff. If we are to treat the RD design as a local experiment, we must effectively fix a window width, and thus change the parameter of interest.

# 2 The Local Randomization RD Framework

The simple example above shows that continuity of the conditional regression functions is not enough to justify analyzing or interpreting the RD design as an experiment in a neighborhood of the cutoff. We now consider a local randomization RD setup where we explicitly assume that the treatment is randomly assigned for all subjects with $X_i \in [c - w, c + w]$. As we discuss, formalizing precisely the local randomization RD framework is difficult, because the notion of random assignment near the cutoff can be interpreted in different ways.

## 2.1 Local Randomization of Score or Treatment?

Intuition suggests that simply assuming that the treatment is randomly assigned in a small neighborhood around the cutoff should be enough to analyze the RD design as a local experiment. However, this is not the case for two reasons: (i) the RD treatment assignment rule places strict restrictions on the type of random assignments that are conceivable, and (ii) the assignments that are conceivable do not imply an exclusion restriction that is always true in actual experiments.

We consider two possible scenarios, according to two different interpretations of what it means for the treatment to be locally randomly assigned near the cutoff. In the first scenario, the values of $X_i$ stay constant but the treatment received is randomly changed for subjects near the cutoff. In the second scenario, the value of $X_i$ is randomly assigned for all subjects near the cutoff. As we show, this is an important distinction that must be considered when formalizing the local randomization RD assumption.

**Scenario 1: Treatment status randomly assigned for all subjects near the cutoff**

The first way to understand locally random treatment assignment in the RD context is to imagine a situation where all subjects with score in a neighborhood of the cutoff—that is, with $X_i \in [c-w, c+w]$—are randomly assigned to receive treatment or control. In the context of our education example introduced in the previous section, we could accomplish this if, for example, we randomly assigned all students who scored between 45 and 55 in the first exam to receive either a single or double dose of algebra, with every student receiving double dose with the same positive probability. Given our assumption that the test score in the first quarter is entirely determined by the students' ability, this assignment mechanism would break the relationship between ability and treatment status induced by the RD rule $\mathbb{1}(X_i \leq 50)$, because it would make receiving the treatment entirely independent of the grade obtained in the first exam.

However, this way of conceptualizing randomization implies that the score $X_i$ is unrelated to treatment status in the local neighborhood, which is incompatible with the treatment assignment rule. In particular, this assignment would imply that there would be both treated and control subjects on both sides of the cutoff in the neighborhood $[c-w, c+w]$, contradicting the RD treatment assignment rule $\mathbb{1}(X_i \leq c)$. This illustrates the general point that, in a sharp RD design, treatment status is deterministic given the score, and thus it is not possible to randomly assign the treatment without altering the values of the running variable.[3]

**Scenario 2: Score value randomly assigned for all subjects near the cutoff**

The alternative is to assume that the running variable, not treatment status, is randomly assigned near the cutoff—a manipulation that would be consistent with the sharp RD treatment rule. There are multiple ways in which one could conceive of such an experiment. For example, following our double-dose algebra example, we might believe that, even though exam performance is broadly influenced by ability, the precise grade received by each student involves some degree of randomness and arbitrariness, such that students whose grades are four points or less apart should be of comparable ability. The school may therefore implement a two-stage process, where first all exams are graded, and for those students whose grades fall between 48 and 52, the original score is replaced with a uniform random number between 48 and 52. The justification for this two-stage procedure might be that it still assigns the treatment to those students who most need it (all students who get very low grades are guaranteed to receive the treatment) but it implements a transparent and fair process to assign the treatment to those students very near the cutoff whose characteristics may be indistinguishable.

Under this setup, the test score in the first exam, $X_i$, is still broadly determined by ability—so that a student who obtains a grade of 25 is on average of lower ability than a student who obtains a grade of 90—but for two students whose grades are between 48 and 52, the second-stage ensures that they have the same ex-ante probability of receiving treatment.

[3]Note that in a fuzzy RD design, it is possible to have both treated and control observations on either side of the cutoff. However, the simple random assignment we have described would still be incompatible with a fuzzy RD assignment because it would assign treatment with the same probability on either side of the cutoff, which would violate the assumption of discontinuity of the probability of treatment assignment at the cutoff.

This hypothetical random assignment of score values implies that the average level of ability (and any other predetermined confounder) of treated students with $X_i \in [48, 50]$ is the same as for control students with $X_i \in [50, 52]$. Note that this setup also assumes that the score $X_i$ collected by the researcher is the second stage score for all students, so that the treatment rule $\mathbb{1}(X_i \leq 50)$ correctly distinguishes treated and control students. Such a two-stage rule is of course artificial, although a second-stage randomization has been used in some actual RD settings (see, e.g., Hyytinen et al. 2015).[4]

The crucial question is whether the assumption that the value of the score is (as-if) randomly assigned in a neighborhood of the cutoff—that is, the assumption that all units with score in $[c - w, c + w]$ have the same probability of receiving any score value $x$ within this neighborhood and therefore the same probability of receiving treatment or control—is enough to justify analyzing and interpreting the RD design in this neighborhood as an actual experiment. At first glance, it might appear as if it is. We are assuming that, among units in a local neighborhood of the cutoff, the value of the score is randomly assigned; since the treatment is deterministically assigned based on this score, this guarantees that, on average, there are no differences in the predetermined characteristics of units above and below the

[4]The reason we use two stages to set up a scenario where the score is randomly assigned is because we want to create a scenario where ability—a predetermined characteristic—is not systematically different between treated and control groups. Because of potential individual-level heterogeneity in ability within the window $[c-w, c+w]$, this cannot be guaranteed unless we assume a two-stage procedure or make an explicit assumption about the distribution of ability for subjects that fall in the neighborhood of the cutoff. Note also that the two-stage scenario ensures comparability of all predetermined characteristics near the cutoff, whereas a scenario based on a single stage must consider assumptions about all possible predetermined characteristics that affect the score.

To see this, consider the following alternative setup, where we assume that the score is a step function of ability, such that higher ability leads to higher grades but the relationship between both variables is constant in intervals. For example, we might imagine $X_i = f(a_i) + \varepsilon_i$, with $f(a_i) = f(\bar{a})$ for all $i$ such that $c - w \leq X_i \leq c + w$. Then,

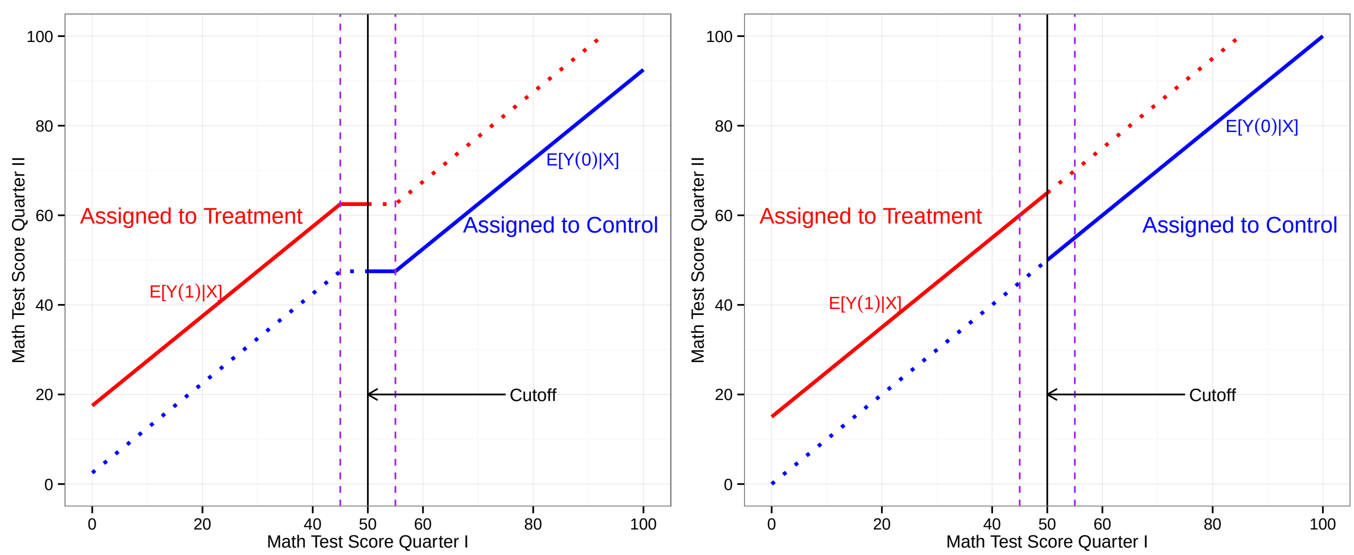
$$\begin{aligned}
\mathbb{P}\left[T_i = 1 \leq c \mid c - w \leq X_i \leq c + w\right] &= \mathbb{P}\left[X_i \leq c \mid c - w \leq X_i \leq c + w\right] \\
&= \mathbb{P}\left[f(\bar{a}) + \varepsilon_i \leq c \mid c - w \leq X_i \leq c + w\right] \\
&= F_\varepsilon\left(c - f(\bar{a}) \mid c - w \leq X_i \leq c + w\right)
\end{aligned}$$

where $F_\varepsilon(\cdot)$ is the CDF of $\varepsilon$, which is the same function for all individuals with $c - w \leq X_i \leq c + w$. Under this setup, all individuals in the local neighborhood of the cutoff have the same probability of receiving treatment and the same ability.

cutoff within this window.

However, this reasoning ignores the possibility that the score itself may affect the potential outcomes, a phenomenon we have explicitly allowed in our example in equation (2). When we introduced this equation, we motivated it by imagining that students' unobservable ability affects both the grade obtained in the first exam, $X_i$, and the grade obtained in the second exam, $Y_i$. Our new assumption that the score is unrelated to students' predetermined characteristics in the neighborhood $[48, 52]$ might seem at first to imply that the running variable must be unrelated to the regression functions in this neighborhood, as illustrated in Figure 1(a). However, if the running variable $X_i$ affects the potential outcomes through factors other than predetermined characteristics, the regression functions can follow equation (2) even when $X_i$ is randomly assigned in a window around the cutoff—this situation is illustrated in 1(b).

Figure 1: Two Scenarios with Randomly Assigned Score




(a) Test scores locally unrelated to potential outcomes

(b) Test scores locally related to potential outcomes

A more precise way to denote the potential outcomes would be $Y_i(T_i, X_i)$, where the first argument captures the effect of the treatment assignment—i.e., of being on one side of the cutoff or the other—and the second argument captures the effect of the running variable on the outcome that occurs independently of the treatment status. This "direct" effect of $X_i$ on $Y_i$ would occur in our example if, for instance, a student's test score in the first exam had a reinforcing effect. If near the cutoff students who receive lower test scores were discouraged and put less effort in the second exam than students who receive higher scores (because, for example, they feel stigmatized by being assigned to the treatment group and are frustrated by having been so close to the cutoff), $X_i$ might be positively associated with the potential outcomes even in the absence of a treatment effect. Under this assumption, the model in equation (2) and Figure 1(b) could still be true even if the score were randomly assigned in $[48, 52]$.

Another way to see the distinction is to note that any experiment can be seen as a RD design in which the score is a random number and the cutoff is chosen to ensure the desired probability of treatment assignment. In our example, if instead of having instructors grade the exams we assigned each student a uniform random number between 0 and 100, the treatment assignment rule $\mathbb{1}(X_i \leq 50)$ would effectively become a rule that assigns double-dose algebra entirely randomly among students, with each student having a 50% probability of receiving treatment. In this case, however, there would be no need to add an extra argument in the regression functions for the random number that determines treatment, as the score used to randomly assign treatment is simply a computer-generated pseudo-random number that is unrelated to the potential outcomes by construction.[5]

In contrast, in a RD design where the score is assumed to be randomly assigned in the the local neighborhood $[c - w, c + w]$ but is not an arbitrarily generated number unrelated to the phenomena under study, there is nothing preventing the value of the running variable from affecting the potential outcomes directly. In other words, local random assignment of $X_i$ does not guarantee the exclusion restriction that we typically take for granted in actual randomized experiments. This illustrates the distinction between assumptions about the assignment of the score $X_i$, and assumptions about the shape of the regression functions:

[5]Note that this statement assumes that students are never informed about the particular random "grade" that they receive, as occurs in actual experiments.

assumptions about the law of the random variable $X_i$ place no restrictions on the functions $\mathbb{E}[Y_i(1)|X_i]$, and $\mathbb{E}[Y_i(0)|X_i]$.

Note also that this phenomenon is analogous to the instrumental variables (IV) design, where random assignment of the instrument does not imply the exclusion restriction that is required to interpret the usual estimand as the average treatment effect for the compliers (Angrist, Imbens, and Rubin 1996). The parallelism arises because in both IV and RD designs, researchers are interested in the effect of the treatment on the outcome, but there is a third variable (the score in RD, the instrument in IV) that induces a change in treatment status and can also affect the potential outcomes directly. In both cases, the way in which this third variable is assigned imposes no general restrictions on its relationship with the potential outcomes.

The analogy between IV and fuzzy RD designs has been long recognized in the continuity-based framework, where the similarities arise because of imperfect compliance with the treatment assignment rule in fuzzy RD settings. However, in the context of the local randomization RD framework, similarities occur even in the sharp RD case where compliance with treatment assignment is perfect. The analogy we point out is not related to treatment compliance but rather to the possible role of the score (instrument) as a determinant of the outcome irrespective of treatment assignment and/or status.

## 2.2 Formalizing the Local Randomization RD Assumption

We now attempt to formalize a local randomization RD assumption. The first such formalization was proposed by Cattaneo, Frandsen, and Titiunik (2015), who used a Fisherian randomization-based approach in which the potential outcomes are seen as fixed as opposed to random variables. Their proposed assumption has two parts. The first part states that the conditional distribution function of the score in the finite sample is the same for all units in the neighborhood of the cutoff. The second explicitly adopts an exclusion restriction that rules out any "direct" effects of the score on the potential outcomes and states that, within the window, the fixed potential outcomes depend on the running variable solely through the treatment indicator.[6] As our discussion above illustrates, the exclusion restriction plays a crucial role in the analogy between RD designs and experiments, an issue that has not been generally recognized by scholars outside of the Fisherian framework.

Since the continuity-based RD framework has been almost exclusively developed within a random sampling framework in which the potential outcomes are random variables, and the analogy between RD and experiments is often understood in these terms, we formalize the local randomization assumption using the random sampling setting introduced in Section 1. We thus retain the random sampling framework throughout the paper. In particular, our formalization is based on statistical independence between the potential outcomes $Y_i(1)$ and $Y_i(0)$ and the binary treatment assignment indicator in a neighborhood of the cutoff, an assumption routinely invoked in experimental analysis and guaranteed by construction by the random assignment of treatment in actual randomized experiments. A condition of this kind has been invoked in local randomization RD settings by, for example, Angrist and Rokkanen (2015), de la Cuesta and Imai (2016), Keele, Titiunik, and Zubizarreta (2015).

Formally, we state this assumption as:

**Assumption 2: (Super-population) Local Independence.** There exists a neighborhood around the cutoff $c$, $W = [c - w, c + w]$, $w > 0$, such that

$Y_i(1), Y_i(0) \perp\!\!\!\perp T_i \mid X_i \in W$,

where $\perp\!\!\!\perp$ denotes statistical independence and $T_i = \mathbb{1}(X_i \geq c)$ as defined above.

Under this assumption, the average treatment effect in the window $W$ is identified by

$$\tau_{\mathtt{LR}}^{\mathtt{RD}} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|X_i \in W] = \mathbb{E}[Y_i|T_i = 1, X_i \in W] - \mathbb{E}[Y_i|T_i = 0, X_i \in W]$$

If the window $W$ is known, estimation can proceed, for example, simply by computing a

[6] Cattaneo, Titiunik, and Vazquez-Bare (2016a) relax this restriction to allow the potential outcomes functions to depend on the score, and use that information to adjust the potential outcomes and build the randomization distribution of the desired test-statistic. Their approach clarifies that one could use a finite-sample randomization-based framework to analyze a RD design when the exclusion restriction is violated, provided one specifies a particular functional form for the potential outcome function $y_i(x, t)$.

difference in means between treated and control groups for observations in this window.

Under Assumption 2, the local randomization RD framework has the following features:

- The target parameter, $\tau_{\mathtt{LR}}^{\mathtt{RD}}$, is the average treatment effect in an interval of the support of the running variable, rather than at a point.

- Because the focus is on an average over an interval rather than at a point, approximation of the functional form of the regression functions $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ is not necessary for estimation.

- Knowledge of the window $W = [c - w, c + w]$ around the cutoff is necessary for identification of $\tau_{\mathtt{LR}}^{\mathtt{RD}}$.

- Estimation and inference methods assume the window is fixed as the sample size grows.

Thus, as we have defined it, the local randomization RD setup stands in contrast to the continuity-based framework described above. We highlight two distinctions in particular.

**Remark 1** (Randomization in window versus at the cutoff)**.** Our characterization of the local randomization RD framework is explicit in stating a "random assignment" assumption that holds in an interval around the cutoff rather than at the cutoff point. This is in contrast to some formalizations of the local randomization RD framework that make assumptions at the cutoff point. The reason we focus on an interval rather than a point is because the requirement of statistical independence at a point is trivial: any random variable is independent of a constant, so the potential outcomes and the score will always be independent at the cutoff. In other words, an assumption of randomization or independence at the cutoff has no empirical or theoretical content, and thus cannot be used as a justification to analyze RD designs as experiments.

Naturally, one can still use the local randomization RD interpretation simply as an approximation device, where the window is not seen as the interval where a randomization condition holds but rather as the interval where the unknown regression functions are approximated—see Cattaneo, Frandsen, and Titiunik (2015, §6.5). But used in this heuristic way, the local randomization RD framework becomes in essence identical to the continuity-based framework: the parameter of interest becomes the treatment effect at the cutoff, and identification and estimation results are ultimately based on some kind of continuity condition (see, e.g., Lee 2008; Canay and Kamat 2016). □

**Remark 2** (The Role of Neighborhood)**.** If the local randomization RD framework is understood in terms of a fixed window as in Assumption 2 and not as the heuristic approximation discussed in Remark 1, the role of the neighborhood in this approach is conceptually different from the role of the bandwidth in the continuity-based framework. In the latter framework, the bandwidth is used to control the bias and variance of the local polynomial approximation to the unknown regression functions; this implies, among other things, that optimal and valid estimation and inference will require choosing a different bandwidth for every outcome variable or covariate analyzed. In contrast, in the local randomization RD framework based on an assumption such as Assumption 2, the window is a fundamental piece of the research design, since it is the interval where the required identification assumption holds. Consequentially, in the local randomization framework, this single window is used to perform estimation and inference for all outcomes and covariates. See Cattaneo and Vazquez-Bare (2016) for further details on the role of neighborhood selection in RD estimation and inference. □

# 3 The Difference Between Random Assignment and Independence in RD contexts

We now consider whether Assumption 2 is implied by the random assignment of the score near the cutoff and whether it implies that the exclusion restriction is satisfied in the RD context. An in depth consideration of these issues reveals subtle relationships between the concepts of random treatment assignment, statistical independence, and exclusion restriction in RD settings. Our discussion makes three main points: (i) the random assignment of the score does not imply local independence between treatment assignment and potential outcomes; (ii) the local statistical independence between treatment assignment and potential outcomes does not imply the exclusion restriction that holds by construction in experiments; and (iii) an experimental analysis is possible under local independence if the exclusion restriction fails, provided that the interpretation of the parameter is modified accordingly.

## 3.1 Random Assignment Of Score Does Not Imply Local Independence

The local independence assumption as stated in Assumption 2 is not guaranteed to hold when we randomly assign the score near the cutoff—i.e., when all subjects with $c-w \leq X_i \leq c+w$ face the same ex-ante probability of receiving every value of the score between $c - w$ and $c + w$. The reason is that, even if all subjects in the neighborhood of the cutoff have the same marginal distribution of the running variable, the potential outcomes and the running variable may be related in ways that violate the local independence assumption. If, as discussed above, higher test scores lead students to expend systematically more or less effort in future exams, the randomly assigned value of $X_i$ will affect the potential outcomes directly even inside the local neighborhood $[c-w, c+w]$, inducing a relationship between the potential outcomes $Y_i(1), Y_i(0)$ and the treatment indicator $\mathbb{1}(X_i \leq c)$ or $\mathbb{1}(X_i \geq c)$ that may violate independence.

Thus, unlike in the case of actual randomized experiments, in a local randomization RD design in the sense of Assumption 2, statistical independence and random assignment are conceptually different. The reason is that, as discussed above, the implicit randomization rule in a local randomization RD design does not and cannot manipulate treatment status directly; instead, the rule must randomly assign the score values. Therefore, if the randomly assigned score has a direct impact on the potential outcomes—so that, for example, high values of $X_i$ lead to high values of the potential outcomes—the indicators $\mathbb{1}(X_i \leq c)$ and $\mathbb{1}(X_i \geq c)$ may fail to be statistically independent of the potential outcomes.

To see this more formally, simply consider the example in equation (2) introduced above. We already showed in Section 1 that this example violates Assumption 2; now simply assume that $X_i$ is randomly assigned in a neighborhood of the cutoff, which is of course allowed by equation (2).

## 3.2 Local Independence Does Not Imply Exclusion Restriction

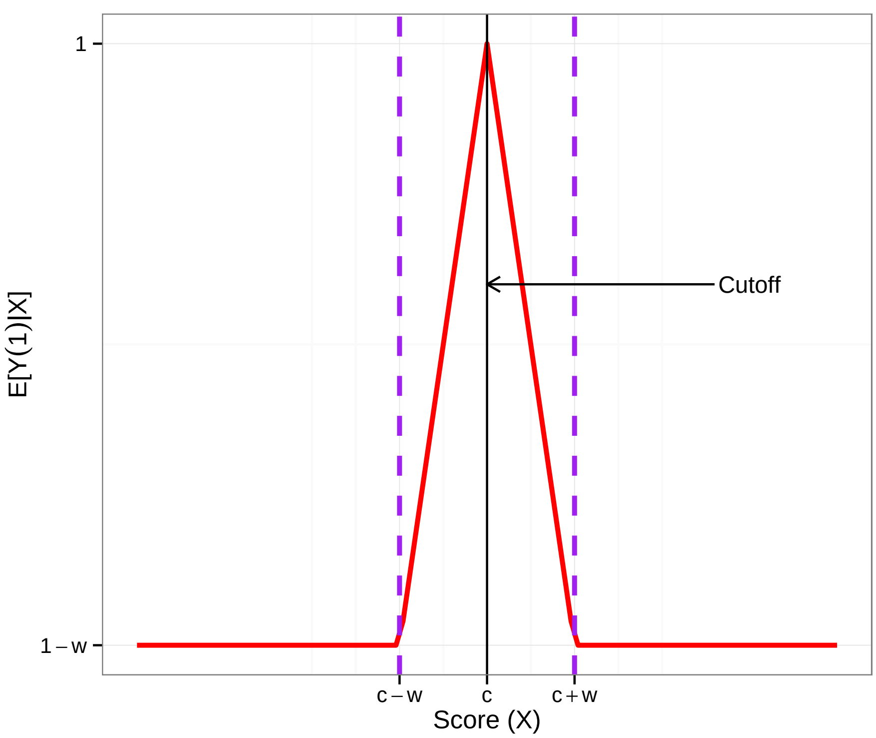
Moreover, the local independence assumption as stated in Assumption 2 is not a sufficient condition for the exclusion restriction that the score does not affect the potential outcomes except via the treatment indicator. That is, the local independence assumption does not imply that the score $X_i$ and the regression functions $\mathbb{E}[Y_i(0) \mid X_i]$, $\mathbb{E}[Y_i(1) \mid X_i]$ are unrelated in a local neighborhood of the cutoff. Graphically, this means that regression functions need not be flat in the neighborhood $[c-w, c+w]$ even when Assumption 2 holds; in other words, local independence does not necessarily imply a scenario like the one illustrated in Figure 1(a).

We briefly present an example to illustrate this point. It suffices to focus on one potential outcome, so we focus on $Y_i(1)$.

$$Y_i(1) = \mathbb{1}(X_i \in W) \cdot (1 - |X_i - c|) + \mathbb{1}(X_i \notin W) \cdot (1 - w) + \varepsilon_i \tag{3}$$

for all $i$, with $\varepsilon_i$ a random error independent of $X_i$, $c$ the cutoff, $\mathbb{1}(X_i \leq c)$ the treatment rule, and $W = [c - w, c + w]$ as before. The regression function $\mathbb{E}[Y_i(1)|X]$ implied by this model for $Y_i(1)$ is shown in Figure 2. To simplify notation, we redefine $\tilde{x} := x - c$ so that the cutoff for $\tilde{x}$ is normalized to zero. We also assume that the density of $\tilde{X}$, $f(\tilde{x})$, is symmetric around zero in $[-w, w]$.

Figure 2: Example Where Exclusion Restriction Is Violated But Local Independence Holds




We wish to show that in this setup, which clearly violates the exclusion restriction, the local independence assumption holds. Let $\tilde{W} = [-w, w]$. We want to show

$$\mathbb{P}[Y(1) \leq y, \tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}] = \mathbb{P}[Y(1) \leq y \mid \tilde{X} \in \tilde{W}] \cdot \mathbb{P}[\tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}]$$

Under the assumptions we made, we can show that

$$\begin{aligned}
\mathbb{P}[Y(1) \leq y, \tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}] &= \mathbb{P}[Y(1) \leq y \mid -w \leq \tilde{X} \leq 0] \cdot \mathbb{P}[\tilde{X} \leq 0 \mid -w \leq \tilde{X} \leq w] \\
&= \mathbb{E}[\, \mathbb{P}[Y(1) \leq y \mid \tilde{X}] \mid -w \leq \tilde{X} \leq 0] \cdot \mathbb{P}[\tilde{X} \leq 0 \mid -w \leq \tilde{X} \leq w] \\
&= \frac{\int_{-w}^{0} F_\varepsilon(y - 1 + |\tilde{x}|) \cdot f(\tilde{x}) \cdot d\tilde{x}}{\int_{-w}^{0} f(\tilde{x}) d\tilde{x}} \cdot \frac{\int_{-w}^{0} f(\tilde{x}) d\tilde{x}}{\int_{-w}^{w} f(\tilde{x}) d\tilde{x}} \\
&= \frac{\frac{1}{2}\left(\int_{-w}^{0} F_\varepsilon(y - 1 + |\tilde{x}|) \cdot f(\tilde{x}) \cdot d\tilde{x} + \int_{0}^{w} F_\varepsilon(y - 1 + |\tilde{x}|) \cdot f(\tilde{x}) \cdot d\tilde{x}\right)}{\int_{-w}^{w} f(\tilde{x}) d\tilde{x}} \\
&= \frac{\frac{1}{2}\int_{-w}^{w} F_\varepsilon(y - 1 + |\tilde{x}|) \cdot f(\tilde{x}) \cdot d\tilde{x}}{\int_{-w}^{w} f(\tilde{x}) d\tilde{x}} \\
&= \frac{1}{2}\mathbb{P}[Y(1) \leq y \,|\tilde{X} \in \tilde{W}]
\end{aligned}$$

Since we assumed that the density of $\tilde{x}$ is symmetric around zero in the window, $\mathbb{P}[\tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}] = \frac{1}{2}$. Thus, we have shown that in this example

$$\begin{aligned}
\mathbb{P}[Y(1) \leq y, \tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}] &= \frac{1}{2}\mathbb{P}[Y(1) \leq y \,|\tilde{X} \in \tilde{W}] \\
&= \mathbb{P}[\tilde{X} \leq 0 \mid \tilde{X} \in \tilde{W}] \cdot \mathbb{P}[Y(1) \leq y \mid \tilde{X} \in \tilde{W}]
\end{aligned}$$

and therefore $Y(1)$ and $\mathbb{1}(\tilde{X} \leq 0)$ are independent in $\tilde{W}$.

The most important assumptions in this example are the symmetry of the density of the running variable $X_i$ around the cutoff in the window $[c - w, c + w]$, the independence between the error term $\varepsilon_i$ and $X_i$ in this window, and the symmetry around the cutoff of the functional form that relates $Y(1)$ and $X$. The intuition behind the result is as follows: if $X$ is symmetrically distributed in $[c - w, c + w]$, the value of the random variable $Y = 1 - |X - c| + \varepsilon$ will be the same regardless of whether $X > c$ or $X < c$; being above or below the cutoff does not provide any information regarding the particular value of $Y(1)$ that will be observed. Thus the local independence condition between the potential outcome $Y_i(1)$ and the treatment indicator $\mathbb{1}(X \leq c)$ holds even though $Y(1)$ and $X$ are related.

## 3.3 Experimental Analysis Under Local Independence Will Capture 'Overall' Effect If Exclusion Restriction Fails

We have shown that local statistical independence as stated in Assumption 2 does not guarantee that the exclusion restriction holds. Although referring to the potential outcomes functions as $Y_i(1)$ and $Y_i(0)$ may give the impression that these functions are only affected by the treatment indicator but not the score itself, this notation is commonly used to broadly refer to the potential outcome under the treatment and control conditions, including everything that these conditions entail. In the context of actual randomized experiments, it is generally unnecessary to make the notation more explicit to let the potential outcomes depend on the particular randomization device used. But in contexts where the variable that induces variation in treatment assignment is an important determinant of the potential outcomes rather than an arbitrary device, generalizing the potential outcomes notation is necessary. For example, a more general potential outcomes notation has been used to allow for the direct effect of the instrument in instrumental variable setups (Angrist, Imbens, and Rubin 1996) and of the cutoff in multi-cutoff RD setups (Cattaneo et al. 2016).

Following the notation we introduced briefly in Scenario 2 in Section 2, we use let $Y_i(T_i, X_i)$ denote the potential outcomes, now explicitly acknowledging that the potential outcomes may depend on the value of the score $X_i$ directly in addition to through the treatment assignment indicator. Using this notation, Assumption 2 implies the mean independence condition $\mathbb{E}[Y_i(j, X_i)|T_i = j, X_i \in W] = \mathbb{E}[Y_i(j, X_i)|X_i \in W]$ for $j \in \{0, 1\}$. Given this generalization, we now consider whether the local randomization RD framework can be used in a context where Assumption 2 holds but the exclusion restriction fails.

Even if the exclusion restriction is violated in the sense that $Y_i(j, X_i) \neq Y_i(j, X_i')$ for

$X_i \neq X_i'$, $j \in \{0,1\}$, under Assumption 2 we have

$$\mathbb{E}[Y_i|T_i = 1, X_i \in W] = \mathbb{E}[Y_i(1, X_i)|T_i = 1, X_i \in W] = \mathbb{E}[Y_i(1, X_i)|X_i \in W]$$
$$\mathbb{E}[Y_i|T_i = 0, X_i \in W] = \mathbb{E}[Y_i(0, X_i)|T_i = 0, X_i \in W] = \mathbb{E}[Y_i(0, X_i)|X_i \in W],$$

leading to

$$\mathbb{E}[Y_i|T_i = 1, X_i \in W] - \mathbb{E}[Y_i|T_i = 0, X_i \in W] = \mathbb{E}[Y_i(1, X_i) - Y_i(0, X_i)|X_i \in W] = \tau_{\mathtt{LR}}^{\mathtt{RD}}.$$

Thus, in a broad sense, Assumption 2 justifies analyzing the RD design in the neighborhood of the cutoff as one would analyze an experiment because it allows identification of the average treatment effect $\tau_{\mathtt{LR}}^{\mathtt{RD}}$. However, if the exclusion restriction does not hold, the treated and control average outcomes within the local window will combine the effect of the treatment (e.g., a student receiving double-dose algebra or a party winning an election) on the outcome, *plus* the additional effect of the score on the outcome that would occur regardless of treatment status (e.g., students who receive higher grades may feel motivated to study more, political donors may wish to donate more money to localities where parties obtain higher vote shares). In this case, the parameter $\tau_{\mathtt{LR}}^{\mathtt{RD}}$ will not capture the (local) average effect of the treatment alone, but rather the effect of obtaining a value of the score in $[c, c+w]$ versus $[c-w, c)$, which will include, among other things, the effect of the treatment.

Using the more general notation, we can see that in a local randomization RD design where Assumption 2 holds but the exclusion restriction does not, the parameter $\tau_{\mathtt{LR}}^{\mathtt{RD}}$, unlike $\tau_{\mathtt{CB}}^{\mathtt{RD}}$, is not the average effect at any single point $x$:

$$\begin{aligned}\tau_{\mathtt{LR}}^{\mathtt{RD}} &= \mathbb{E}[Y_i(1) - Y_i(0)|X_i \in W] = \mathbb{E}[Y_i(1, X_i) - Y_i(0, X_i)|X_i \in W] \\ &\neq \mathbb{E}[Y_i(1, x) - Y_i(0, x)|X_i \in W]\end{aligned}$$

Imagine, for example, that $\mathbb{E}[Y_i(0, X_i)|X_i \in W] = \mu_0$, constant for every $i$ in the window, and $Y_i(1, X_i)$ follows equation 3. In this case,

$$\tau^{\mathtt{RD}}_{\mathtt{LR}} = \mathbb{E}[1 - |X_i - c||X_i \in W] - \mu_0 = \frac{\int_{c-w}^{c+w}(1 - |x - c|)f(x)dx}{\int_{c-w}^{c+w} f(x)dx} - \mu_0,$$

which will take any value between $1 - w - \mu_0$ and $1 - \mu_0$, depending on the density of the running variable $X$ inside the window. This example shows that, in a scenario where the exclusion restriction fails, the average effect $\tau^{\mathtt{RD}}_{\mathtt{LR}}$ will take different values according to the likelihood of observations concentrating in particular ranges of the running variable inside the window. For example, the same setup just described would lead to different values of $\tau^{\mathtt{RD}}_{\mathtt{LR}}$ if the observations within the window were uniformly distributed than if they were disproportionally concentrated near the cutoff. Thus, when the exclusion restriction fails and the regression functions are not constant in the window around the cutoff, the interpretation of the average effect $\tau^{\mathtt{RD}}_{\mathtt{LR}}$ differs from the interpretation it would have in a standard experiment.

# 4 Concluding Remarks

Our discussion highlights that the local randomization interpretation of the RD design, if taken literally, introduces conceptual distinctions that are absent in pure experimental designs. As we have shown, in the context of RD designs, the distinction between random assignment and statistical independence is consequential. This distinction is often meaningful in natural experiments (Sekhon and Titiunik 2012). In an actual experiment, random assignment of treatment leads to statistical independence between treatment status and potential outcomes, because the "score" used to randomly assign subjects is a device (likely a computer-generated pseudo-random number) that is by construction arbitrary and unrelated to the potential outcomes or any systematic characteristic of the experimental subjects. As we observed, a randomized experiment can be recast as a RD design where the score is a randomly generated number and the cutoff is chosen to ensure the desired probability of receiving treatment. Thus, seen as functions of this random number, the regression functions are guaranteed to be constant (graphically flat), because the score is a randomly generated number that is by construction unrelated to the potential outcomes.

In contrast, in a RD design, the score used to assign treatment, even if its values are randomly allocated near the cutoff, are usually important determinants of the outcome of interest. Indeed, the importance of the RD score is often what motivates using it as the basis of treatment assignment in the first place. In a context where the score is meaningfully related to the outcome (past and future test scores, past and future vote shares, etc.), random assignment of the score value contains no information about the particular form of the regression functions $\mathbb{E}[Y_i(1)|X_i]$ and $\mathbb{E}[Y_i(0)|X_i]$. This is straightforward to see once one realizes that restrictions on the randomization distribution of the score $X_i$ (and implicitly of the treatment assignment) are fundamentally different from restrictions on the shape of the regression functions—and more generally, on the conditional distribution of the potential outcomes given the score. No matter how much information we have about the way in which $X_i$ is assigned, this information will generally be insufficient to determine how $X_i$ and the potential outcomes are related.

Given the additional assumptions needed for the local randomization interpretation to hold, in most applications one should proceed using the continuity assumption alone. This is typically a plausible assumption if there are neither formal nor informal mechanisms for sorting—i.e., for units to appeal and change the score value they were originally assigned in order to receive their preferred treatment condition. However, in practice, the methods used to estimate the continuous functions at the cutoff are consequential. Estimation is delicate because the functional forms are unknown, and one is estimating the trend at an endpoint (the cutoff) of measure zero. The example in Hyytinen et al. (2015) illustrates how inferences in RD designs are sensitive to the method used to estimate the continuous functions on both sides of the cutoff even when sample sizes are not small, and the design is valid by construction. More generally, the properties of RD estimation and inference based on local polynomial methods depend crucially on the bandwidth choice; for example, the commonly used mean-squared-error optimal bandwidth, though valid for point estimation, is too large for inference and requires either undersmoothing or robust methods to yield valid confidence intervals (Calonico, Cattaneo, and Titiunik 2014). One appeal of the local randomization interpretation of RD is that it avoids these estimation and inference issues, but unfortunately, this interpretation requires additional assumptions that may not be plausible.

# References


Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91 (434): 444–455.

Angrist, Joshua D., and Miikka Rokkanen. 2015. "Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff." *Journal of the American Statistical Association* 110 (512): 1331–1344.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2016. "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference." Working paper, University of Michigan.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2016. "Regression Discontinuity Designs Using Covariates." Working paper, University of Michigan.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Canay, Ivan A, and Vishal Kamat. 2016. "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design." Working Paper*, Northwestern Univeristy*.

Cattaneo, Matias D., Brigham Frandsen, and Rocio Titiunik. 2015. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." *Journal of Causal Inference* 3 (1): 1–24.

Cattaneo, Matias D., and Gonzalo Vazquez-Bare. 2016. "The Choice of Neighborhood in Regression Discontinuity Designs." forthcoming *Observational Studies*.

Cattaneo, Matias D., Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *Journal of Politics* 78 (3): 1229-1248.

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2016a. "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality." Working paper, University of Michigan.

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2016b. "Inference in Regression Discontinuity Designs Under Local Randomization." *The Stata Journal* 16 (2): 331-367.

Caughey, Devin, and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008." *Political Analysis* 19 (4): 385-408.

Cook, Thomas D. 2008. "“Waiting for Life to Arrive”: A history of the regression-discontinuity design in Psychology, Statistics and Economics." *Journal of Econometrics* 142 (2): 636–654.

de la Cuesta, Brandon, and Kosuke Imai. 2016. "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19: 375–396.

Eggers, Andrew C., Olle Folke, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder. 2014. "On The Validity Of The Regression Discontinuity Design For Estimating Electoral Effects: New Evidence From Over 40,000 Close Races." *American Journal of Political Science*, forthcoming.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–209.

Holland, Paul W. 1986. "Statistics and causal inference." *Journal of the American statistical Association* 81 (396): 945–960.

Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen, and Janne Tukiainen. 2015. "Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes." *VATT Institute for Economic Research, Working Paper 59.*

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

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

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

Keele, Luke, Rocío Titiunik, and José R Zubizarreta. 2015. "Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178 (1): 223–239.

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

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

Porter, Jack. 2003. "Estimation in the Regression Discontinuity model." Working paper*, University of Wisconsin at Madison.*

Sekhon, Jasjeet S., and Rocío Titiunik. 2012. "When natural experiments are neither natural nor experiments." *American Political Science Review* 106 (01): 35–57.

Sekhon, Jasjeet S., and Rocío Titiunik. 2016. "Understanding Regression Discontinuity Designs As Observational Studies." forthcoming *Observational Studies*.

Thistlethwaite, Donald L., and Donald T. Campbell. 1960. "Regression-discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309–317.