WestEd Justice & Prevention Research Center

BRIEF

Regression Discontinuity Design: A Method to Rigorously Evaluate Interventions to Reduce Crime and Improve the Justice System

February 2024

Jonathan Nakamoto
Senior Research Associate

Alexis Grant
Research Associate

Trent Baskerville
Research Assistant

Anthony Petrosino
Director

Introduction

The evaluation of interventions, programs, or policies involves a multitude of methods. The expectation is that these diverse approaches, referred to as "research designs," should converge on the same conclusions. Ideally, the specific method used to assess a program would have no bearing on the outcome of the study. However, the reality is that the landscape of research is far from straightforward.

The results we observe can often be attributed to, or confounded by, the evaluation design we employ to determine the success of a treatment or intervention. A less rigorous design could result in a false positive, in which an intervention that is ineffective is incorrectly credited with a good outcome. And the converse is also true: a weak design could result in a false negative, in which the intervention is wrongly determined to have been unsuccessful.

Aims of the Brief

- Provide a brief, accessible summary of Regression Discontinuity Design (RDD) for practitioners, policymakers, researchers unfamiliar with the approach, graduate students, and others who deal with crime and justice issues in their work.
- Highlight examples that show how RDD was successfully implemented to evaluate interventions designed to reduce crime or improve the justice system.
- Identify opportunities in which RDD could be used.

Fortunately, researchers have been toiling for decades in developing and promoting methods that are more rigorous. These methods aim to increase our confidence and reduce our skepticism about observed outcomes. One of the most well-known approaches is the randomized controlled trial (RCT). In an RCT, individuals or groups (such as prison units or neighborhoods) are randomly assigned to either receive an intervention or to be part of a control group that does not receive an intervention.

A less familiar approach is the Regression Discontinuity Design (RDD). Our review of the crime and justice literature reveals that, historically, RDD has been less frequently used compared to other methods, notably the RCT. In this brief, we offer an overview of RDD, explaining what it is and underscoring why it stands as a powerful approach to evaluation. We highlight several practical instances illustrating RDD in practice. We conclude with a call to action, advocating for its increased use to answer questions related to crime reduction and justice system improvement.

How do we know a program works?

Several graduate-level research and evaluation courses devote substantial time to the examination of methods employed to assess program effectiveness and determine the level of confidence to place in these methodologies. Referred to as causal inference by researchers, this confidence pertains to evaluating whether the intervention being studied is responsible for producing the observed outcome. This becomes particularly challenging in the world of social programs, such as those aiming to reduce crime and improve justice, because of the logistical challenges related to randomly assigning study participants and withholding potentially beneficial programming from some participants.

Interventions are delivered in the real world, not within controlled laboratory environments under constant observation. In these real-world scenarios, conditions are uncontrolled, and changes may occur without continuous monitoring. Consequently, the outcomes are often not immediately observable or as dramatic as in comparison with those seen in assessing the effectiveness of using a parachute when jumping out of an airplane compared to not using one.

However, determining whether an after-school program for youths effectively reduces crime differs significantly from conducting an RCT in a laboratory. In a real-world scenario, it is impossible to monitor every young person continuously or account for all the other influences in their lives. Additionally, the difference attributed solely to the after-school program is not likely to be substantial and dramatic, even if we hope it will be. Moreover, frequently, we examine impact by collecting data on the youths' behavior (e.g., whether they were arrested or appeared in juvenile court). Voilà! The young people committed fewer offenses at the end of the program than when they started. The program worked!

But let's take a step back. Several factors could have influenced those results, known to researchers as *confounders*. One common factor is the natural maturing process of young people. What we might think is a program's impact could, in fact, be attributed to the young people's improving their behavior as they grow older.

Experts in evaluation would argue that collecting data only at the program's start and conclusion, what we often refer to as the "pretest–posttest design" or, more commonly, the "before and after study," does not yield strong causal inferences. Researchers would not be able to conclusively assert that the program directly *caused* the observed outcomes.

This leads us toward research designs aiming to produce compelling evidence linking improved outcomes to a program. As previously mentioned, researchers have developed an extensive toolbox of potential designs to help bolster our confidence and allay skepticism regarding observed results. Most of these evaluation approaches, including the use of an RCT to investigate the effectiveness of an after-school youth program, will result in developing a control or comparison group of young people who are as similar as possible to the after-school participants but do not partake in the after-school program.

However, there are many ethical and logistical issues inherent in an RCT because, in this context, some of the youths who could benefit from the program will not be able to access it. RDD emerges as an ingenious approach within the research design toolbox. RDD facilitates robust causal inference while ensuring that the most underserved still receive the program.

What is Regression Discontinuity Design?

RDD allows us to assess the impact of an intervention by assigning individuals or groups to treatment and control conditions based solely on a cutoff threshold on a numeric score. In such situations, entities scoring above the cutoff receive treatment, while those scoring below it do not.[1] These numeric scores can be derived from various types of data. Towns assigned to implement a new violence prevention initiative based on exceeding a certain violent crime rate would be one example. Every town above the rate would get the program; every town below the rate would not.

Another illustrative example pertains to the allocation of additional services to individuals at high risk and on probation, determined primarily by their scores on a risk and needs assessment instrument. Those meeting a select threshold (let's say "75 and above") qualify to receive the services, while those who score below ("74 and under") do not receive them. Furthermore, age can serve as a numeric criterion within an RDD. For instance, the legal drinking age of 21 serves as the cutoff, distinguishing between legal and underage drinkers.

You might wonder, why is the use of this numeric score crucial in bolstering the efficacy of the RDD? It ties back to the earlier discussed rationale for the comparison group, especially concerning the entities scoring just above or below the cutoff threshold.

Consider the prior scenario of using a risk and needs assessment score to allocate treatment services to high-risk persons on probation. We can assume that someone who scores a 99 and someone who scores a 43 differ significantly in their risk levels. However, the individuals scoring 74 and those scoring 75 present a more intriguing scenario. Despite nearly matching in their levels of risks and needs, and likely sharing many other characteristics, one will receive enhanced services, and the other will not.

RDD leverages this cutoff rule. Since we can assume the similarity between individuals just above and just below the cutoff, we can also confidently assume that the difference in outcomes between entities just above the cutoff

[1] In other situations, entities scoring below the cutoff may be assigned the treatment, while the control group comprises those who score above the cutoff.

compared to those just below the cutoff provides a valid estimate of the impact of the intervention. Researchers would argue that these estimates from RDD are at the high end of causal inference (there's that phrase again), enhancing our confidence in the observed results.

To further illustrate the potency of RDD, let's delve into a hypothetical example that uses a risk and needs assessment instrument within a city's implementation of a youth crime prevention initiative. The city opts to assess youths' risk of involvement in crime by administering an assessment instrument. The decision-makers stipulate that anyone scoring 75 or above is considered to be at high risk and will receive the intervention. Those scoring 74 or below will not be involved in the specialized initiative.

Minimum Conditions for RDD

- Individuals or groups (or other entities) are assigned to the intervention based on whether they are above or below a threshold (i.e., the cutoff score) on a numeric variable (i.e., the assignment variable).
- This numeric variable needs to be measured before the start of the intervention and be a continuous measure, such as a risk assessment score, crime rates, and age.
- The numeric assignment variable has at least five to seven possible scores. A risk scale with 20 intervals could be perfect. But a risk assessment scale that was simply based on three categories, such as "1—low risk," "2—moderate risk," and "3—high risk," would not work.

Even Better Conditions for RDD

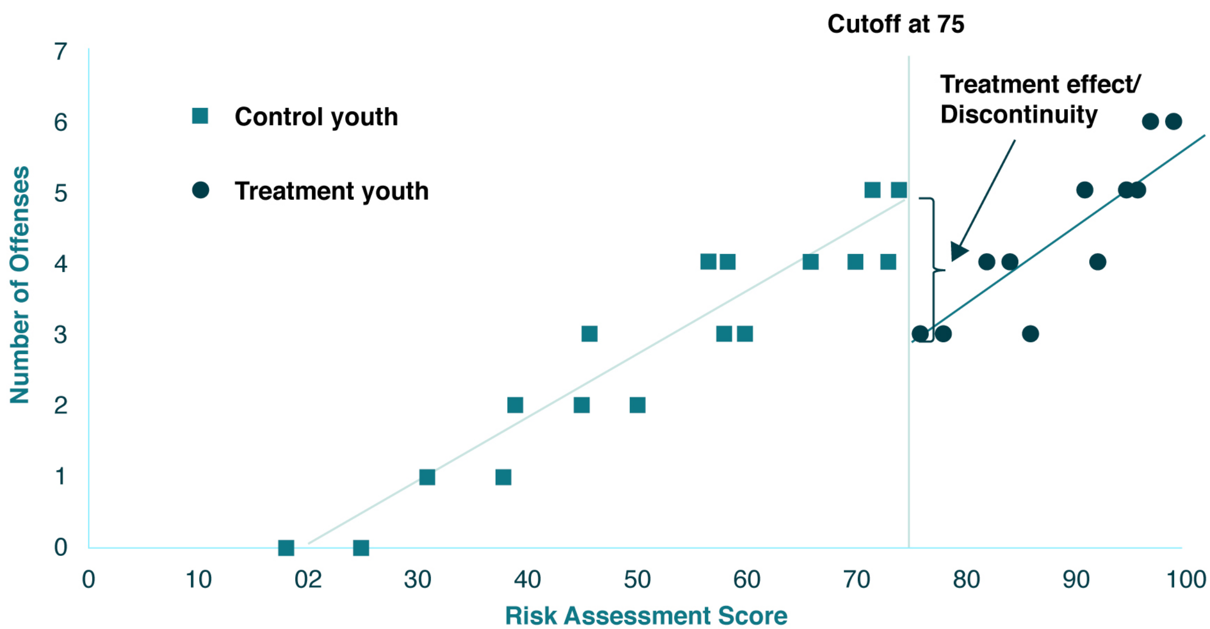
- It's not too fuzzy! It's best when there are a limited number of overrides to the cutoff (e.g., cases that receive the intervention even when their scores on the assignment variable were below the cutoff and should have resulted in them not receiving the intervention).
- The study's sample size is fairly large. You probably wouldn't want to do an RDD with 50 cases. RDD works best with large samples, especially around that cutoff score where cases are assumed to be similar.

As previously mentioned, it is widely accepted that substantial differences probably exist between youths scoring "40" (indicating low risk for crime) and those scoring "90" (indicating very high risk for crime). Any "comparison" of low- and high-scoring youths necessitates the consideration of other characteristics, such as the socioeconomic status of the youths' families, the youths' academic performance, the degree of crime and disorder in the neighborhood surrounding their schools, and so on.

However, we would not anticipate systematic differences between youths scoring a "73" or "74" on the assessment and youths scoring "75" or "76." Nonetheless, some youths will be assigned to treatment (the "75" and above crowd) while others will not (the "74" and below crowd). If there is a positive treatment impact, youths just above the cutoff should exhibit better outcomes concerning crime than youths just below the cutoff. We should see a “discontinuity” or “break” in the expected crime outcomes. In the absence of programmatic impact, there likely would be no “discontinuity” or “break.”

Graphical illustrations frequently play an important role in presenting the findings from RDD studies. In many cases, substantively important findings will be visible to the naked eye. Figure 1 suggests that the program was successful. There is a noticeable discontinuity or break, with the youths just above the cutoff committing fewer offenses at the end of 1 year compared to the youths just below the cutoff who did not receive the program.

Figure 1. Results From a Hypothetical RDD Study Examining a Youth Crime Prevention Initiative



How many technical details do you need to pay attention to when using RDD?

While RDD can involve a considerable amount of technical depth, our objective is to not go too deep into the intricacies within this brief. We also aim to avoid portraying RDD as if there are no challenges to implementing it. There are a range of technical intricacies tied to RDD planning and analysis that will require the expertise of a methodologist experienced with the design. The analyses and interpretation of RDD results can be complex. There are a range of different statistical models that researchers need to use when conducting RDD and a variety of diagnostic tests that need to be performed to determine whether the assumptions underlying RDD are met. Furthermore, decisions need to be made about how to handle overrides to the cutoff and whether to include all cases in an analysis or to limit an analysis to cases close to the cutoff. We compiled a recommended reading list of RDD resources at the end of the brief.

What about some examples of how RDD is used in practice?

Constructing a study to comprehend the impact of prison presents a significant challenge. Typically, individuals sentenced to prison commit offenses that are more serious than those who receive alternative sanctions such as probation. But Mitchell and his colleagues (2017) devised an innovative way to construct such a study using RDD. They capitalized on a large historical database in Florida that included more than 262,000 individuals convicted of felonies and their respective sentences.

Florida utilizes a system that assigns points at sentencing; these points are known as “total sentence points.” How these points are assigned in Florida is based on several factors, including the seriousness of the offense and the defendant’s prior criminal record. Cases that have more than 44 total sentence points are “scored to prison,” while cases with 44 or fewer points are allocated to probation, jail, or house arrest.

One challenge in implementing RDD is the potential for slippage, even if a score like total sentencing points is the sole determinant of whether a person gets prison or not. In Florida, judges possess considerable discretion to override this assignment. And, further complicating matters, it turns out that these overrides happen quite frequently: 13 percent of cases just below the cutoff still received prison sentences (when they should have gotten alternative sanctions), while, surprisingly, only 39 percent of cases just above the cutoff received prison sentences (meaning 61% who should have gotten prison just above the cutoff did not). Technically speaking, when there is slippage like this, researchers refer to the RDD as being “fuzzy.”

Despite this fuzziness, researchers argue that there is a sufficient sample at the cutoff to allow for valid conclusions to be drawn. Mitchell and his colleagues (2017) conducted complex analyses. Their overarching finding was significant: For cases near the cutoff, there is no evidence that prison sentences led to reductions in subsequent recidivism over a 3-year period. This finding carries immense weight, considering the potential harm of incarceration on individuals and their families, as well as its higher cost compared to alternative sanctions such as probation (Beckett & Goldberg, 2022).

There are various ways to apply RDD beyond using individual-level justice data and total sentencing points. Doleac and Sanders (2015) pursued an unconventional path, beyond a specific justice intervention, to examine the impact on crime of additional daylight due to daylight savings time (DST). Their hypothesis centered on the theory that increased “lighting” would decrease crime (Welsh & Farrington, 2008). Unlike individual-level data, their study used aggregate data from 558 counties and cities.

This analysis was conducted at the jurisdiction level, examining historical data and the crime rates in these jurisdictions during each hour of the day. The assignment variable revolved around the days before and after DST, marking the start of the “intervention” on the first day of DST. The treatment group comprised all days after DST, while the control group encompassed all days before DST.

RDD results revealed striking findings: There were 27 percent fewer robberies and 38 percent fewer rapes in the hour around sunset during DST when there was additional lighting compared to the same hour pre-DST when it was darker. Post-DST, there were declines in daily crime rates, albeit smaller (7% for robberies and 11% for rapes), but this suggests that criminal activity was not entirely displaced to darker hours later in the day.

These first two examples are retrospective. That is, researchers cleverly analyzed data that permitted them to use RDD to examine the impact of sentencing in the Florida example or DST in the second one. But what about prospective studies in which RDD is planned before the intervention takes place? We have only been able to identify a small number of published studies in which researchers worked together with policy or practice leaders in the justice system in the planning stages to implement and conduct an RDD.

One such prospective study comes from Germany (Engel et al., 2022). In Germany, the juvenile courts frequently assign youths to regular probation, which is a mild sanction with a low amount of supervision by a probation officer. However, German officials were interested in examining the impact of an intensive probation program on recidivism for high-risk juveniles convicted of a crime. This intensive program would provide a greater level of support with a focus on rehabilitation.

Engel and his colleagues (2022) worked with judges and probation staff to develop a scorecard, based on judges' ratings, to evaluate the severity of the offenses. These scores based on judges' ratings ranged from 0 to 28. This study assigned youths with scores of 13 or higher to the intensive probation program; a score of 12 or lower resulted in an individual's being assigned to regular probation.[2] The RDD analysis showed that the intensive probation program reduced recidivism by 10 percentage points at 6 months and 30 percentage points 1 to 3 years later for youths around the cutoff of 13 points on the scorecard.

What do I do now?

We believe that policymakers (e.g., agency leaders), practitioners, and researchers should consider the use of RDD because it allows for stronger conclusions to be drawn about the impact of an intervention than do many other types of research designs. RDD is particularly well-suited for many situations within the crime and justice domain because it allows intentional targeting of the individuals, areas, or entities with the highest needs to receive specific treatment.

In general, we might worry about the bias of an approach that deliberately selects and assigns individuals to treatment. However, utilizing a numeric score and threshold (i.e., the cutoff) to assign the intervention in this instance allows us to turn that bias around and exploit it in a powerful way to increase causal inference and bolster our confidence in the research findings.

We reiterate that it is not our intention to suggest in this brief that RDD is an easy process. Implementing and carrying out rigorous research designs, in general, is rarely simple, and RDD is no exception. However, our examples demonstrate that RDD can effectively operate within messy real-world settings. Even when there is not a strict adherence to the cutoff score, RDD can still produce strong causal conclusions about the effectiveness of an intervention.

2 This study is more complicated than a typical RDD study because the authors randomly assigned only half of the youths above the cutoff to receive the intensive probation. This resulted in a fuzzy RDD like the Florida study, but in this case, the fuzziness was planned from the start of the study and allowed for additional analyses based on an RCT with youths above the cutoff.

The document provides guidance on employing RDD in both prospective and retrospective studies within the context of crime and justice. It emphasizes the importance of collaboration between agency leaders, practitioners, and researchers to plan and execute RDD studies effectively.

For Prospective RDD

- Partner with researchers to discuss a prospective RDD study, especially when assignment will be based on risk or need.
- Discuss what the numeric assignment variable would be or create one (e.g., a risk assessment score).
- Identify the optimal cutoff you believe is needed for those above or below to receive the intervention.
- Work with researchers to limit fuzziness and increase the sample size.

We encourage agency leaders to collaborate proactively with researchers in planning RDD studies. An example would be when an evaluation is needed for a particular program. Let's say it is for a treatment program for high-risk people. The agency leaders and practitioners could work with researchers to identify an existing instrument for classifying risk or develop a new one. In some cases, relatively minor changes to existing practices (e.g., the development of a more formalized risk assessment system) would allow for the use of RDD. Policymakers and practitioners could also work with researchers to identify an acceptable and comfortable cutoff score.

Planning an RDD prospectively in this way can have several benefits, such as selecting the right factor to assign entities (the assignment variable) with enough variation in scores (e.g., we don't want 1-2-3 as the scale), selecting the best cutoff threshold that is not too high or low (e.g., a scale of 1–100 that assigns only those scoring over 95 to treatment), and stressing the importance of limiting fuzziness (e.g., overrides to the cutoff threshold). Although innovative statistical methods are always being developed to reduce bias, it is often said that the best statistical method is to have a strong design that limits such statistical corrections from being needed in the first place.

When prospective studies cannot be done, let's not forget that most RDD studies to date have been done retrospectively. In fact, in our early analysis of RDD studies in crime and justice, we have found that nearly all of them are retrospective. In the study we cited from Florida, researchers used existing data and were able to distinguish between entities receiving the intervention or not (i.e., the cutoff threshold) and analyze the impact for these groups on selected outcomes (crime or recidivism).

For Retrospective RDD

- Identify interventions that your agency implemented.
- Identify those in which a score-based assignment variable was used (e.g., risk assessment score, crime rates, age), and the assignment variable can be linked to the outcomes of interest.
- Make data available to allow researchers to conduct analyses.

Retrospective studies also have the advantage in that they can be cheaper than a prospective study in the field (as the data, in some sense, have already been collected), and they are unobtrusive—the researchers can do the analyses without bothering anyone outside of the research team. However, it is critical that data on the assignment variable and outcomes of interest be collected by interested parties and can be linked to the entities included in the analyses. Furthermore, interested parties need to grant researchers access to the data. But such retrospective analyses, as in the two instances described earlier, can yield important insights to guide crime and justice policy.

Conclusions

The push for evidence-based policy within crime and justice has increased the focus on the quality of evidence produced by evaluations. As a result, organizations have developed evidence clearinghouses tasked with evaluating programs based on the robustness and caliber of their conducted studies. For instance, the National Institute of Justice's CrimeSolutions initiative not only serves as a repository for programs and practices but also features an assessment process. The evaluation methodology used for interventions stands as a crucial determinant in this assessment process. Most evidence clearinghouses, like CrimeSolutions, widely acknowledge RDD as one of the most influential designs available to researchers for program evaluation.

Although the number of published studies that use RDD in the crime and justice field has grown in recent years, it is still quite low relative to how many evaluations have been published since RDD was first popularized in the 1960s. Our preliminary review of the literature found fewer than 70 published or available RDD studies. Our objective in composing this brief was to emphasize the significance of this design and advocate for its broader implementation whenever feasible, with the intent of enhancing the rigor of our evaluations of programs and policies.

Recommended Resources for Regression Discontinuity Design

Research Methodology Textbooks

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin, and Company.

Peer-Reviewed Journal Articles

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*(2), 636–654. https://doi.org/10.1016/j.jeconom.2007.05.002

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Rhodes, W., & Jalbert, S. K. (2013). Regression discontinuity design in criminal justice evaluation: An introduction and illustration. *Evaluation Review*, *37*(3-4), 239–273. https://doi.org/10.1177/0193841X14523004

Research Reports and Other Guides

Jacob, R., Zhu, P., Somers, M.-A., & Bloom, H. (2012). *A practical guide to regression discontinuity*. MDRC. https://files.eric.ed.gov/fulltext/ED565862.pdf

Schochet, P. Z. (2008). *Technical methods report: Statistical power for regression discontinuity designs in education evaluations* (NCEE 2008-4026). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED511782.pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

References

Beckett, K., & Goldberg, A. (2022). The effects of imprisonment in a time of mass incarceration. *Crime and Justice*, *51*, 349–398. https://doi.org/10.1086/721018

Doleac, J. L., & Sanders, N. J. (2015). Under the cover of darkness: How ambient light influences criminal activity. *The Review of Economics and Statistics*, *97*(5), 1093–1103. https://doi.org/10.1162/REST_a_00547

Engel, C., Goerg, S. J., & Traxler, C. (2022). Intensified support for juvenile offenders on probation: Evidence from Germany. *Journal of Empirical Legal Studies*, *19*(2), 447–490. https://doi.org/10.1111/jels.12311

Mitchell, O., Cochran, J. C., Mears, D. P., & Bales, W. D. (2017). The effectiveness of prison for reducing drug offender recidivism: A regression discontinuity analysis. *Journal of Experimental Criminology*, *13*, 1–27. https://doi.org/10.1007/s11292-017-9282-6

Welsh, B. C., & Farrington, D. P. (2008). Effects of improved street lighting on crime. *Campbell Systematic Reviews*, *4*(1), 1–51. https://doi.org/10.4073/csr.2008.13

WestEd.org

Suggested citation: Nakamoto, J., Grant, A., Baskerville, T., & Petrosino, A. (2024). *Regression discontinuity design: A method to rigorously evaluate interventions to reduce crime and improve the justice system*. WestEd.

The authors would like to thank **Trevor Fronius** and **Sean Tanner** of WestEd for reviewing drafts of this brief.