CHAPTER 5

Regression Discontinuity Design

Social programs often use an index to decide who is eligible to enroll in the program and who is not. For example, antipoverty programs are typically targeted to poor households, which are identified by a poverty score or index. The poverty score can be based on a proxy means formula that measures a set of basic household assets. Households with low scores are classified as poor, and households with higher scores are considered relatively well-off. The program authorities typically determine a threshold or cutoff score, below which households are deemed poor and are eligible for the program. Examples include the Mexico Progresa program (Buddelmeyer and Skoufias 2004) and Colombia's system for selecting beneficiaries of social spending, the so-called SISBEN (Barrera-Osorio, Linden, and Usquiola 2007).

Pension programs are another example of a type of program that targets units based on an eligibility index, albeit one of a different kind. Age constitutes a continuous index, and the retirement age constitutes the cutoff that determines eligibility. In other words, only people above a certain age are eligible to receive the pension. A third example of a continuous eligibility index would be test scores. Many countries award scholarships or prizes to the top performers on a standardized test, whose results are ranked from the lowest to the highest performer. If the number of scholarships is limited, then only students who score above a certain threshold score (such as the top 15 percent of students) will be eligible for the scholarship.

Key Concept:
Regression discontinuity design (RDD) is adequate for programs that use a continuous index to rank potential participants and that have a cutoff point along the index that determines whether or not potential participants receive the program.

The regression discontinuity design (RDD) is an impact evaluation method that can be used for programs that have a continuous eligibility index with a clearly defined cutoff score to determine who is eligible and who is not. To apply a regression discontinuity design, two main conditions are needed:

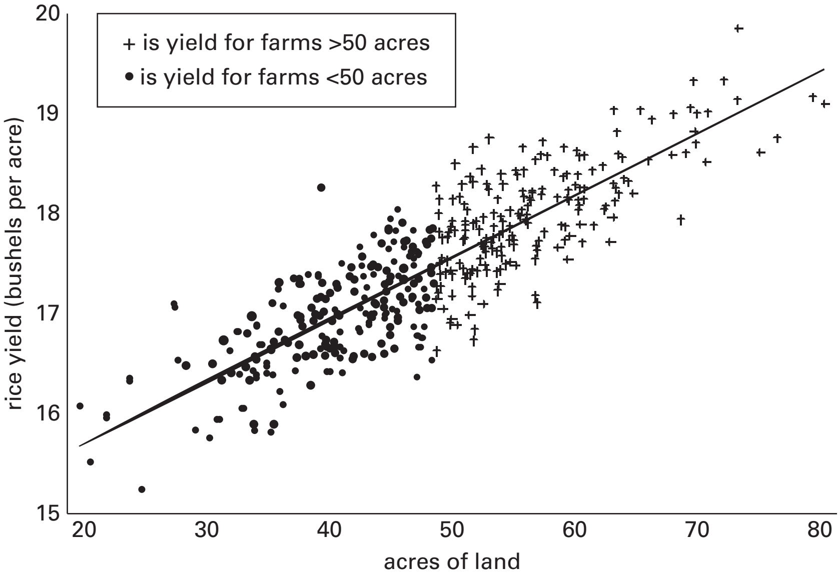
1. A continuous eligibility index, in other words, a continuous measure on which the population of interest can be ranked, such as a poverty index, a test score, or age.
2. A clearly defined cutoff score, that is, a point on the index above or below which the population is classified as eligible for the program. For example, households with a poverty index score less than 50 out of 100 might be classified as poor, individuals age 67 and older might be classified as pensioners, and students with a test score of 90 or more out of 100 might be eligible for a scholarship. The cutoff scores in these examples are 50, 67, and 90, respectively.

Case 1: Subsidies for Fertilizer in Rice Production

Consider an agriculture program that subsidizes rice farmers' purchase of fertilizer with the objective of improving total yields. The program targets small and medium-size farms, which it classifies as farms with fewer than 50 acres of total land. Before the program starts, we might expect the relationship between farm size and total rice production to be as shown in figure 5.1, in that smaller farms have lower total outputs than larger farms. The eligibility score in this case is the number of acres of the farm, and the cutoff is 50 acres. Under program eligibility rules, farms below the 50-acre cutoff are eligible to receive fertilizer subsidies, and farms with 50 or more acres are not. In this case, we might expect to see a number of farms with 48, 49, or even 49.9 acres that participate in the program. Another group of farms with 50, 50.1, and 50.2 acres will not participate in the program because they fell just to the wrong side of the cutoff. The group of farms with 49.9 acres is likely to be very similar to the group of farms with 50.1 acres in all respects, except that one group received the fertilizer subsidy and the other group did not. As we move further away from the eligibility cutoff, eligible and ineligible units will become more different by construction, but we have a measure of how different they are based on the eligibility criteria and therefore we can control for those differences.

Once the program rolls out and subsidizes the cost of fertilizer for small and medium farms, the program evaluators could use a regression discon-

Figure 5.1 Rice Yield



Source: Authors.

tinuity method to evaluate its impact. The regression discontinuity measures the difference in postintervention outcomes, such as total rice yields, between the units near the eligibility cutoff, which in our example is a farm size of 50 acres. The farms that were just too large to enroll in the program constitute the comparison group and generate an estimate of the counterfactual outcome for those farms in the treatment group that were just small enough to enroll. Given that these two groups of farms were very similar at baseline and are exposed to the same set of external factors over time (such as weather, price shocks, local and national agricultural policies, and so on), the only plausible reason for different outcomes in the postintervention period must be the program itself.

The regression discontinuity method allows us to successfully estimate the impact of a program without excluding any eligible population. However, note that the estimated impact is only valid in the neighborhood around the eligibility cutoff score. In our example, we have a valid estimate of the impact of the fertilizer subsidy program for the larger of the medium-size farms, that is, those with just under 50 acres of land. The impact evaluation will not necessarily be able to directly identify the impact of the program on small farms, say, those with 1 or 2 acres of land, where the effects of a fertilizer subsidy may differ in important ways from the effects observed on medium-size farms with 48 or 49 acres.

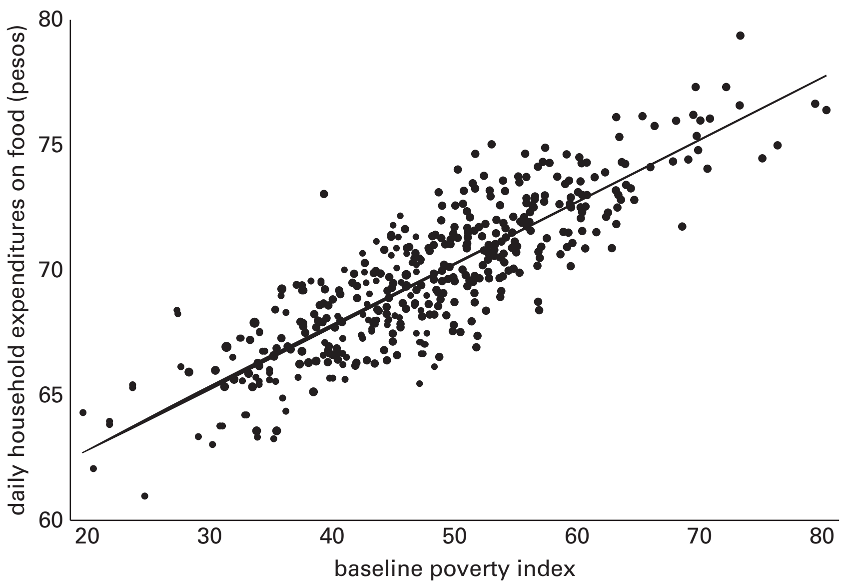
No comparison group exists for the small farms, since all of them are eligible to enroll in the program. The only valid comparison is for the farms near the cutoff score of 50.

Case 2: Cash Transfers

Assume that we are trying to evaluate the impact of a cash transfer program on the daily food expenditures of poor households. Also assume that we can use a poverty index,[1] which takes observations of a household's assets and summarizes them into a score between 0 and 100 that is used to rank households from the poorest to the richest. At the baseline, you would expect the poorer households to spend less on food, on average, than the richer ones. Figure 5.2 presents a possible relationship between the poverty index and daily household expenditures (the outcome) on food.

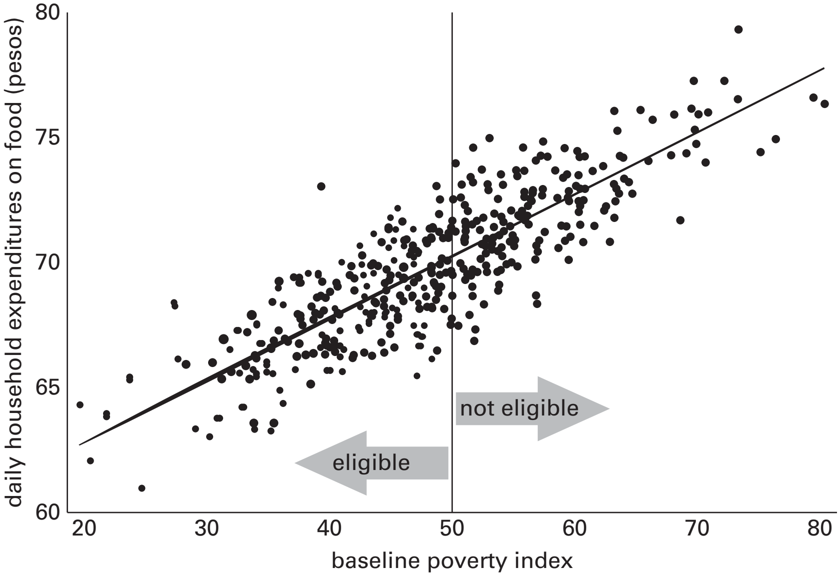
Now assume that the program targets only poor households, which are determined to be those with a score below 50. In other words, the poverty index can be used to determine eligibility: treatment will be offered only to households with a score of 50 or less. Households with a score above 50 are

Figure 5.2 Household Expenditures in Relation to Poverty (Preintervention)



Source: Authors.

Figure 5.3 A Discontinuity in Eligibility for the Cash Transfer Program

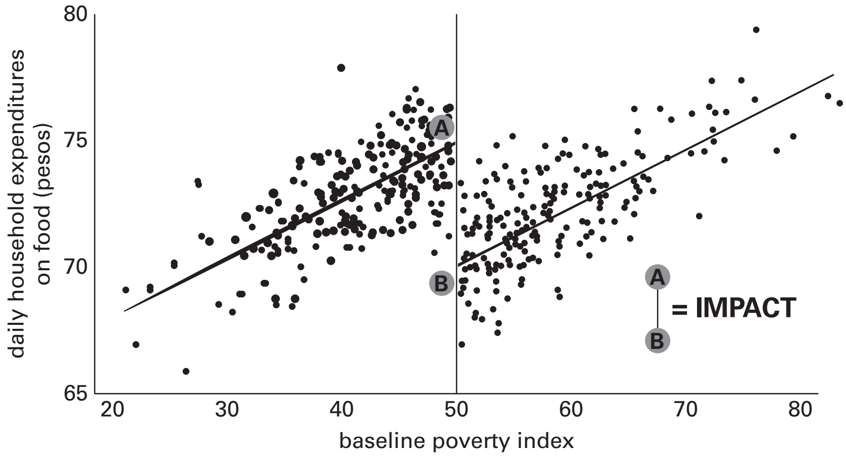


Source: Authors.

ineligible. In this example, the continuous eligibility index is simply the poverty index, and the cutoff score is 50. The continuous relationship between the eligibility index and the outcome variable (daily food expenditures) is illustrated in figure 5.3. Households just below the cutoff score are eligible for the program, while those just above the cutoff score are ineligible, even though the two types of households are very similar.

The RDD strategy exploits the discontinuity around the cutoff score to estimate the counterfactual. Intuitively, eligible households with scores just below the cutoff (50 and just below) will be very similar to households with a score just above the cutoff (for example, those scoring 51). On the continuous poverty index, the program has decided on one particular point (50) at which there is a sudden change, or discontinuity, in eligibility for the program. Since the households just above the cutoff score of 50 are similar to the ones that are just below it, except that they do not receive the cash transfers, the households just above can be used as a comparison group for the households just below. In other words, households ineligible for the program but close enough to the cutoff will be used as a comparison group to estimate the counterfactual (what would have happened to the group of eligible households in the absence of the program).

Figure 5.4 Household Expenditures in Relation to Poverty (Postintervention)



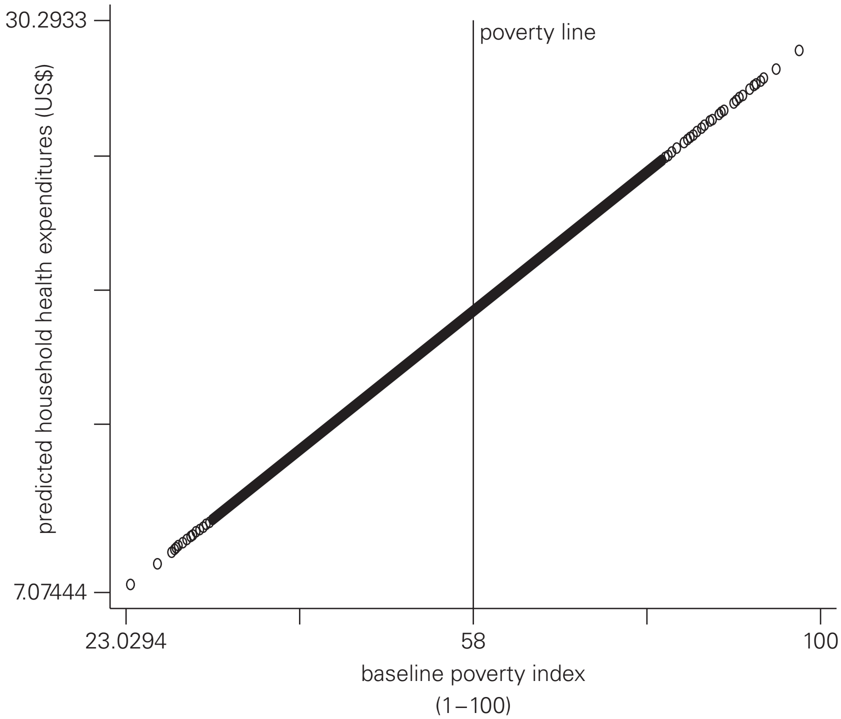
Source: Authors.

Figure 5.4 presents a possible postintervention situation conveying the intuition behind the RDD identification strategy. Average outcomes for (eligible) households with baseline poverty scores below the cutoff score are now higher than average outcomes for (ineligible) households with baseline scores just above the cutoff. Given the continuous relationship between scores on the poverty index and daily expenditures on food before the program, the only plausible explanation for the discontinuity that we observe postintervention must be the existence of the cash transfer program. In other words, since households in the vicinity (right and left) of the cutoff score had similar baseline characteristics, the difference in average food expenditures between the two groups is a valid estimate of the program's impact.

Using the Regression Discontinuity Design Method to Evaluate the Health Insurance Subsidy Program

Let us apply RDD to our health insurance subsidy program (HISP). After doing some more investigation into the design of the HISP, you find that in practice the authorities targeted the program to low-income households using the national poverty line. The poverty line is based on a poverty index that assigns each household in the country a score between 20 and 100 based on its assets, housing conditions, and sociodemographic struc-

Figure 5.5 Poverty Index and Health Expenditures at the Health Insurance Subsidy Program Baseline

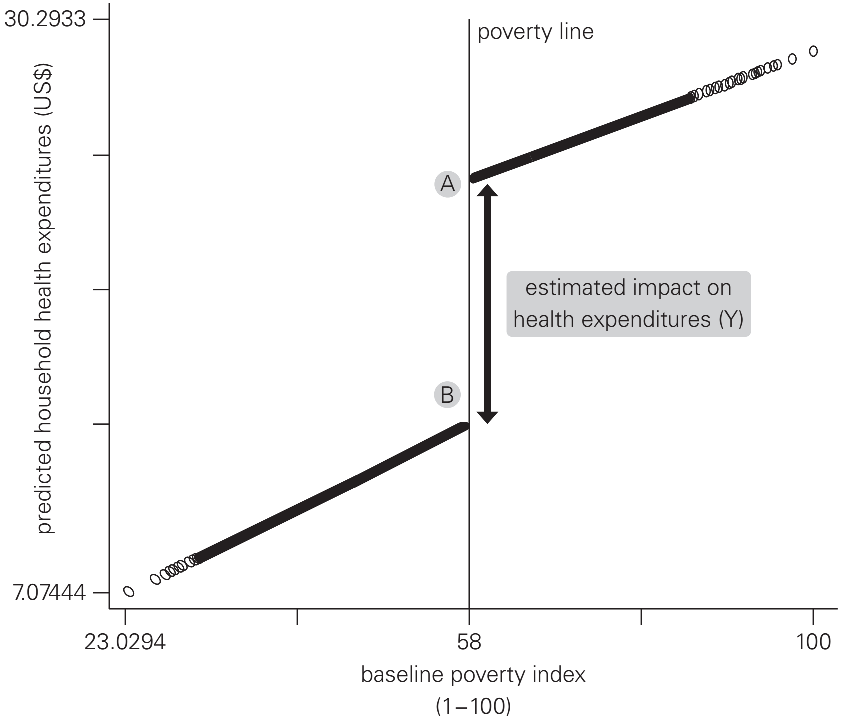


Source: Authors.

ture. The poverty line has been officially set at 58. This means that all households with a score of less than 58 are classified as poor, and all households with a score of more than 58 are considered to be nonpoor. Even in the treatment villages, only poor households were eligible to enroll in the HISP; nonetheless, your sample includes data on both poor and nonpoor households in the treatment villages.

Using the households in your sample of treatment villages, a colleague helps you run a multivariate regression and plot the relationship between the poverty index and predicted household health expenditures before HISP started (figure 5.5). The figure shows clearly that as a household's score on the poverty index rises, the regression predicts a higher level of health expenditures, reflecting the fact that wealthier households tended to have higher expenditures on, and consumption of, drugs and primary health services. Note that the relationship between the poverty index and health expenditures is continuous, that is, there is no evidence of a change in the relationship around the poverty line.

Figure 5.6 Poverty Index and Health Expenditures – Health Insurance Subsidy Program Two Years Later



Source: Authors.

Two years after the start of the pilot, you observe that only households with a score below 58 (that is, to the left of the poverty line) have been allowed to enroll in the HISP. Using follow-up data, you again plot the relationship between the scores on the poverty index and predicted health expenditures and find the relation illustrated in figure 5.6. This time, the relationship between the poverty index and the predicted health expendi-

Table 5.1 Case 5—HISP Impact Using Regression Discontinuity Design (Regression Analysis)

	Multivariate linear regression
Estimated impact on household health expenditures	−9.05** (0.43)

Source: Authors.

Note: Standard errors are in parentheses.

** Significant at the 1 percent level.

tures is no longer continuous—there is a clear break, or "discontinuity," at the poverty line.

The discontinuity reflects a decrease in health expenditures for those households eligible to receive the program. Given that households on both sides of the cutoff score of 58 are very similar, the only plausible explanation for the different level of health expenditures is that one group of households was eligible to enroll in the program and the other was not. You estimate this difference through a regression with the findings shown in table 5.1.

QUESTION 5

A. Is the result shown in table 5.1 valid for all eligible households?
B. Compared with the impact estimated with randomized assignment, what does this result say about those households with a poverty index of just under 58?
C. Based on this result from case 5, should the HISP be scaled up nationally?

The RDD Method at Work

Regression discontinuity design has been used in various contexts. Lemieux and Milligan (2005) analyzed the effects of social assistance on labor supply in Quebec. Martinez (2004) studied the effect of old age

Box 5.1: Social Assistance and Labor Supply in Canada

One of the classic studies using the RDD method took advantage of a sharp discontinuity in a social assistance program in Quebec, Canada, to understand the effects of the program on labor market outcomes. The welfare program, funded through the Canadian Assistance Plan, provides help to the unemployed. For many years, the program offered significantly lower payments to individuals under the age of 30 with no children, compared to individuals older than 30—$185 a month versus $507.

To rigorously evaluate this program, Lemieux and Milligan (2005) limited the sample to men without children and without a high school diploma and gathered data from the Canadian Census and the Labor Force Survey. To justify using the RDD approach, they showed that men close to the discontinuity (between the ages of 25 and 39) are very similar on observable characteristics.

Comparing men on both sides of the eligibility threshold, the authors found that access to greater social assistance benefits actually reduced employment by about 4.5 percent for men in this age range without children.

Source: Lemieux and Milligan 2005.

pensions on consumption in Bolivia. Filmer and Schady (2009) assessed the impact of a program that provided scholarships to poor students to encourage school enrollment and increase test scores in Cambodia. Buddelmeyer and Skoufias (2004) examined the performance of regression discontinuity relative to the randomized experiment in the case of Progresa and found that the impacts estimated using the two methods are similar for a large majority of the outcomes analyzed. A few of these examples are described in detail in boxes 5.1, 5.2, and 5.3.

Box 5.2: School Fees and Enrollment Rates in Colombia

In Colombia, Barrera-Osorio, Linden, and Urquiola (2007) used regression discontinuity design to evaluate the impact of a school fee reduction program (Gratuitad) on school enrollment rates in the city of Bogota. That program is targeted based on an index called the SISBEN, which is a continuous poverty index whose value is determined by household characteristics, such as location, the building materials of the home, the services that are available there, demographics, health, education, income, and the occupations of household members. The government established two cutoff scores along the SISBEN index: children of households with scores below cutoff score no. 1 are eligible for free education from grades 1 to 11; children of households with scores between cutoff scores no. 1 and no. 2 are eligible for a 50 percent subsidy on fees for grades 10 and 11; and children from households with scores above cutoff score no. 2 are not eligible for free education or subsidies.

The authors used regression discontinuity design for four reasons. First, household characteristics such as income or the education level of the household head are continuous along the SISBEN score at the baseline; in other words, there are no "jumps" in characteristics along the SISBEN score. Second, households on both sides of the cutoff scores have similar characteristics, suggesting that the design had produced credible comparison groups. Third, a large sample of households was available. Finally, the government kept the formula used to calculate the SISBEN index secret, so that households would not be able to manipulate their scores.

Using the RDD method, the researchers found that the program had a significant positive impact on school enrollment rates. Specifically, enrollment was three percentage points higher for primary school students from households below cutoff score no. 1 and 6 percent higher for high school students from households between cutoff scores no. 1 and no. 2. This study provides evidence on the benefits of reducing the direct costs of schooling, particularly for at-risk students. However, its authors also call for further research on price elasticities to better inform the design of subsidy programs such as this one.

Source: Barrera-Osorio, Linden, and Urquiola 2007.

Box 5.3: Social Safety Nets Based on a Poverty Index in Jamaica

The RDD method was also used to evaluate the impact of a social safety net initiative in Jamaica. In 2001, the government of Jamaica initiated the Programme of Advancement through Health and Education (PATH) to increase investments in human capital and improve the targeting of welfare benefits to the poor. The program provided health and education grants to children in eligible poor households, conditional on school attendance and regular health care visits. The average monthly benefit for each child was about $6.50 in addition to government waiver of certain health and education fees.

Because eligibility for the program was determined by a scoring formula, Levy and Ohls (2007) were able to compare households just below the eligibility threshold to households just above (between 2 and 15 points from the cutoff). The researchers justify using the RDD method with baseline data showing that the treatment and comparison households had similar levels of poverty, measured by proxy means scores, and similar levels of motivation, in that all of the households in the sample had applied to the program. The researchers also used the program eligibility score in the regression analysis to help control for any differences between the two groups.

Levy and Ohls (2007) found that the PATH program increased school attendance for children ages 6 to 17 by an average of 0.5 days per month, which is significant given an already fairly high attendance rate of 85 percent. Also, health care visits by children ages 0 to 6 increased by approximately 38 percent. While the researchers were unable to find any longer-term impacts on school achievement or health care status, they concluded that the magnitude of the impacts they did find was broadly consistent with conditional cash transfer programs implemented in other countries. A final interesting aspect of this evaluation is that it gathered both quantitative and qualitative data, using information systems, interviews, focus groups, and household surveys.

Source: Levy and Ohls 2007.

Limitations and Interpretation of the Regression Discontinuity Design Method

Regression discontinuity design estimates *local* average impacts around the eligibility cutoff at the point where treatment and comparison units are most similar. As we get closer to the cutoff, the units that are to the left and right of it will look more similar. In fact, when we get extremely close to the cutoff score, the units on the left and right of the line will be so similar that our comparison will be as good as if we had chosen the treatment and comparison groups using randomized assignment of the treatment.

Because the RDD method estimates the impact of the program around the cutoff score, or *locally*, the estimate cannot necessarily be generalized to units whose scores are further away from the cutoff score, this is, where eligible and ineligible individuals may not be as similar. The fact that the RDD method will not be able to compute an average treatment effect for all program participants can be seen as both a strength and a limitation of the method, depending on the evaluation question of interest. If the evaluation primarily seeks to answer the question, Should the program exist or not?, then the average treatment effect for the entire eligible population may be the most relevant parameter, and clearly the RDD will fall short of being perfect. However, if the policy question of interest is, Should the program be cut or expanded at the margin?, then the RDD produces precisely the local estimate of interest to inform this important policy decision.

The fact that the RDD method produces local average treatment effects also raises challenges in terms of the statistical power of the analysis. Since effects are estimated only around the cutoff score, fewer observations can be used than in other methods that would include all units. Relatively large evaluation samples are required to obtain sufficient statistical power when applying RDD. In practice, we determine a bandwidth around the cutoff score that will be included in the estimation by considering the balance in observed characteristics of the population above and below the cutoff. We can then do the estimation again using different bandwidths to check whether the estimates are sensitive to the chosen bandwidth. As a general rule, the wider the bandwidth, the greater the statistical power of the analysis, since more observations are included. However, moving further from the cutoff may also require additional functional form assumptions to obtain a credible estimate of impact.

An additional caveat when using the RDD method is that the specification may be sensitive to the functional form used in modeling the relationship between the eligibility score and the outcome of interest. In the example of the cash transfer program, we assumed that the baseline relation between the poverty index of households and their daily expenditures on food was simple and linear. In reality, the relation between the eligibility index and the outcome of interest (Y) at the baseline could be much more complex and could involve nonlinear relationships and interactions between variables. If we do not account for these complex relationships in the estimation, they might be mistaken for a discontinuity in the postintervention outcomes. In practice, we can estimate program impact using various functional forms (linear, quadratic, cubic, etc.) to assess whether, in fact, the impact estimates are sensitive to functional form.

Even with these limitations, regression discontinuity design yields unbiased estimates of the impact in the vicinity of the eligibility cutoff. The regression discontinuity strategy takes advantage of the program assignment rules, using continuous eligibility indexes, which are already common in many social programs. When index-based targeting rules are applied, it is not necessary to exclude a group of eligible households or individuals from receiving the treatment for the sake of the evaluation because regression discontinuity design can be used instead.

Note

1. This is sometimes called a "proxy-means test" because it takes the household's assets as a proxy or estimator for its means or purchasing power.

References

Barrera-Osorio, Felipe, Leigh Linden, and Miguel Urquiola. 2007. "The Effects of User Fee Reductions on Enrollment: Evidence from a Quasi-Experiment." Columbia University and World Bank, Washington, DC.

Buddelmeyer, Hielke, and Emmanuel Skoufias. 2004. "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA." World Bank Policy Research Working Paper 3386, IZA Discussion Paper 827, World Bank, Washington, DC.

Filmer, Deon, and Norbert Schady. 2009. "School Enrollment, Selection and Test Scores." World Bank Policy Research Working Paper 4998, World Bank, Washington, DC.

Lemieux, Thomas, and Kevin Milligan. 2005. "Incentive Effects of Social Assistance: A Regression Discontinuity Approach." NBER Working Paper 10541, National Bureau of Economic Research, Cambridge, MA.

Levy, Dan, and Jim Ohls. 2007. "Evaluation of Jamaica's PATH Program: Final Report." Mathematica Policy Research, Inc., Ref. 8966-090, Washington, DC.

Martinez, S. 2004. "Pensions, Poverty and Household Investments in Bolivia." University of California, Berkeley, CA.