## REGRESSION DISCONTINUITY DESIGNS

In regression discontinuity designs (RDDs), assignment to an intervention is based on whether a unit (that is, an individual, a firm, a state, and so on) scores above or below a known cutoff point on some measure. If the scores around the cutoff were not subject to manipulation (see details below), then researchers can estimate the effects of an intervention by comparing outcomes for units with scores just above the known cutoff (who did not receive the intervention) to units with scores just below it (who did receive the intervention).[1]

The typical approach to estimating intervention effects in RDDs is to fit separate regressions for the intervention and comparison units using data that lies within a specified range around the cutoff value and then to assess the difference of these regressions' predicted outcome values at the cutoff point. When all units below the cutoff receive the intervention and all units above the cutoff do not, then the difference in the regression predictions at the cutoff (the *discontinuity*) is the estimate of the effect of the intervention.

There are several key considerations involved in estimating treatment effects using an RDD design:

- The measure used to assign units to the intervention condition is typically referred to as the *forcing, assignment,* or *running* variable or score.
- The range of scores on the forcing variable above and below the cutoff defining the sample that researchers used to estimate impacts is known as the *bandwidth.*
- The model of the relationship between the forcing variable and outcome (that is, linear, quadratic, cubic, and so on) is known as the *functional form.*
- Assuming the forcing variable cannot be manipulated, then units within the defined bandwidth around the cutoff variable should be comparable, and an RDD should produce good (statistically consistent) estimates of the causal effects of an intervention.

An important distinguishing feature of RDDs is that they provide causal inferences of intervention effects that are more credible than those of other types of non-experimental designs reviewed by CLEAR (in particular, regression designs) (Lee and Lemieux, 2009). This is because the forcing variable cutoff value provides a known rule for assigning units to treatment and, with units just above and below cutoff value, units that are similar to each other in all ways other than their

[1] For simplicity, we assume that scoring below the threshold assigns study units to the intervention and that scoring above assigns them to the comparison. However, the reverse is also a valid assignment rule: a score above the threshold assigns units to the intervention, and a score below it assigns to the comparison.

exposure to the treatment. For units very close to the cutoff value, the similarity should mimic what would occur if the units had been randomly assigned (Imbens and Lemieux, 2008).

The selection of treated and control units can use a local or global regression approach. In a local regression, regressions are constrained to using only data within a set bandwidth (Cattaneo et al., 2023). In a global regression approach, all data points are used to estimate impacts. Because observations away from the cut point are more dissimilar than observations near the cut point, global regressions typically include additional polynomial terms to ensure a "best fit" to the data.

CLEAR reviews two types of RDDs: Sharp (SRDDs) and Fuzzy (FRDDs).[2] In an SRDD, everyone in the intervention group, but no one in the comparison group, receives the intervention. In an FRDD, there is some fuzziness in the treatment assignment mechanism. That is, some units who score below the cutoff may not receive the intervention or some units who score above the cutoff do receive the intervention (or both). Moreover, while there is still a discontinuity in the probability of intervention receipt at the cutoff score in an FRDD, the probability of receiving the intervention does not change from zero to one around the cutoff. In this case, the forcing variable is not the sole determinant of receipt of the intervention.[3]

A study should be reviewed under RDD guidelines if it meets the following criteria:

- A continuous scoring rule is used as the basis to assign the intervention to study units. For an SRDD, units with scores below a pre-set cutoff value are assigned to the intervention group, and units with scores above the cutoff are assigned to the comparison group, or vice versa. For an FRDD, there is a discontinuous change in the probability of receiving the intervention at the cutoff.
- The forcing variable must be continuous, near-continuous, or non-continuous but ordered. A continuous or near-continuous forcing variable will often have multiple values so that sample units either have their own unique score or share a score with other units but there are many shared distinct scores. When the forcing variable is continuous, there are likely to be multiple unique values above and below the cutoff. If the forcing variable is not continuous, there are likely to be fewer values, which means that multiple sample units will share a smaller number of unique values. In most cases, CLEAR expects forcing variables to be continuous or nearly so, and for there to be at least four unique values

[2] CLEAR's review standards for both SRDDs and FRDDs are adaptations of the RDD review standards from the What Works Clearinghouse (2022) Standards Handbook, version 5.0.

[3] This is similar to a randomized controlled trial with non-compliance.

above and below the cutoff.[4] However, if a forcing variable is not continuous or exhibits very few unique values above and below the cutoff (for example, one value above and one value below), reviewers should consult CLEAR leadership for guidance in conducting the review.[5]

- There must be no factor confounded with the forcing variable close to the cut-off determining assignment to the intervention. The cutoff value for the forcing variable must not be used to assign units to interventions other than the one under study. For example, if income is used as the eligibility criterion for both the intervention and other means-tested programs, then this could introduce a confound into the design. Additionally, categorical variables such as gender and race are not appropriate forcing variables because they are associated with many potential confounding factors (they are also not continuous or ordered). The absence of confounds is necessary to ensure that the study can isolate the causal effects of the tested intervention from the effects of other interventions; therefore, the forcing variable should be used only for assignment to a distinct intervention.

If the study does not satisfy all the above criteria, it should not be evaluated as an RDD (even if it claims to use an RDD) and instead should be assessed based on the most appropriate CLEAR non-experimental design standards. If reviewers are unsure what the appropriate non-experimental design should be, they should check with the topic area principal investigator or senior leadership.

Under the CLEAR Causal Evidence Guidelines, version 2.2, RDDs, along with RCTs and interrupted-time-series designs, can receive high, moderate, or low causal evidence ratings. As summarized in Table 1 and detailed below, a study that uses an RDD:

- Can receive high evidence rating if it completely satisfies Criteria RDD.1 through RDD.4 for an SRDD or Criteria RDD.1 through RDD.5 for an FRDD.[6]
- Can receive a moderate rating if it partially satisfies RDD.1 through RDD.4 (for SRDDs) and RDD.5 (for FRDDs).

---

[4] If a forcing variable has a small number of unique values, the relationship between the outcome and that variable cannot be estimated. CLEAR uses four values above and four values below the cutoff as it allows for a reasonable estimate.

[5] When a forcing variable is discrete so that very few unique values are above and below the cutoff, researchers may base their analysis on the local randomization framework, which is closely analogous to inference using an RCT. The local randomization approach can be used when the bandwidth around the cutoff point is narrow, and units above and below are considered as good as randomly assigned to study conditions. CLEAR anticipates that very few RDD studies will use a discrete forcing variable. If reviewers encounter this situation, they should consult with the CLEAR team about the most appropriate way to proceed following the RDD guidelines presented here.

[6] In the labor literature, it is uncommon to satisfy (partially or otherwise) Criterion RDD.2 on attrition. Similarly, it is uncommon to completely satisfy Criteria RDD.3 and RDD.4. It is therefore unlikely that an RDD study reviewed by CLEAR would satisfy all the specified guidelines necessary to receive a high causal-evidence rating.

- Will receive a low rating if it does not meet the criteria for a high or moderate rating.

**Table 1. RDD study ratings**

| Criterion | High | Moderate |
|---|---|---|
| RDD.1: Integrity of the forcing variable | Completely satisfy | Partially satisfy |
| RDD.2: Attrition and baseline equivalence | Completely satisfy | Partially satisfy |
| RDD.3: Continuity of relationship between outcome and the forcing variable | Completely satisfy | Partially satisfy |
| RDD.4: Functional form and bandwidth | Completely satisfy | Partially satisfy |
| RDD.5: Fuzzy regression discontinuity design (if applicable) | Completely satisfy | Partially satisfy |

In instances where a study provides impact estimates for both SRDDs and FRDDs, the impacts from the SRDD will be prioritized in the review.

## 1. CRITERIA FOR ALL RDDS

### Criterion RDD.1: Does the study establish the integrity of the forcing variable?

A key condition for an RDD to be valid is that there was no systematic manipulation of the forcing variable used to assign units to study conditions. *Manipulation* in this case means that forcing variable scores were not changed from their original values to influence assignment to the intervention. It should be established that there was no manipulation on the part of either the scorers (those who assign or determine the value of the forcing variable) or the study units (those whose eligibility to receive the intervention depends on the value of the forcing variable) to control whether they fall just above or just below the cutoff.

The integrity of the forcing variable should be established institutionally, statistically, and graphically.

- **Criterion RDD.1A**. **Description of institutional integrity.** The institutional integrity of the forcing variable must be established by an adequate description of the variable used, the cutoff value, who selected the cutoff value, whether the cut-off was known to individuals taking the test or undergoing the assessment, and who scored the test or assessment that was used as the variable. The description should indicate that systematic manipulation was unlikely because scorers had little opportunity or incentive to change the actual scores in order to influence who had access to the intervention. A clear opportunity or incentive to manipulate scores would invalidate this criterion. Situations in which manipulation is more or less likely include:

- A scorer who knew the cutoff scores for assigning people to the intervention condition and was able to manipulate scores to assign certain individuals to it would raise the risk that manipulation could undermine the design. For a workplace intervention, for example, manipulation is more likely if assignment is based on assessments of employees by supervisors who can influence which employees receive the intervention.
- This criterion is less likely to be violated if the processes for scoring the forcing variable and assigning individuals to the intervention condition are independent of each other. For example, manipulation is less likely if assignment decisions are based on existing forcing variables that were collected and scored prior to the implementation of the intervention and for unrelated reasons.

Similarly, those assigned to the intervention or comparison group should not be able to precisely control their own scores at the cutoff. For example, while some people score higher or lower on an assessment, they should not be able to manipulate their scores such that they can control whether their scores are just below or just above the cutoff value and therefore determine whether or not they would receive the intervention. Any purposeful manipulation of the forcing variable potentially compromises the RDD, because the groups just above and below the cutoff value may no longer be comparable. To ensure that manipulation does not undermine an RDD, the description of the forcing variable should be clearly written and show no opportunity or incentive for manipulation. Further, the following two sub-criteria should also indicate that manipulation was either unlikely or not pervasive.

- **Criterion RDD.1B. Statistical tests of the forcing variable.** The statistical integrity of the forcing variable must be demonstrated through established statistical tests (see McCrary, 2008) to establish the smoothness of the density of the forcing variable right around the cutoff. One way to assess evidence of manipulation is to check for a mass of points (bunching) or any asymmetry near the cutoff value of the forcing variable. Such a mass or asymmetry could indicate manipulation of the forcing variable in which certain units either just miss or just meet the intervention criterion. Statistical tests should provide evidence against the hypothesis that there is a mass of points near the cutoff value of the forcing variable. In this context, the McCrary test is often used, in which smoothness around the cutoff value is the null hypothesis. A study can demonstrate continuity in the density of the forcing variable around the cutoff using statistical tests other than a McCrary test, as long as the tests are based on established statistical methods and reference a peer-reviewed citation or an appropriate textbook.[7] For example, a study can demonstrate continuity using a test that shows balance on predetermined covariates (e.g., demographic characteristics) for the sample just above and below the cutoff. For this criterion to be satisfied, the McCrary or other statistical test must fail to reject the null hypothesis of continuity in the density of the forcing variable at the 5 percent level.

- **Criterion RDD.1C. Graphical analysis of the forcing variable.** The study should use a graphical analysis to establish the smoothness of the density of the forcing variable right around the cutoff (Cattaneo et al., 2020). A study can demonstrate the smoothness of the density function through the use of graphical tools such as histograms, density plots, and displays that apply polynomial estimators to histograms, among other tools. Visually, the graphs should show the absence of clustering around the cutoff value of the forcing variable. Reviewers should look for compelling evidence of discontinuity at the cutoff, which is visibly larger than discontinuities observed at other points along the forcing variable distribution.

A study *completely* satisfies Criterion RDD.1 if it meets Criterion RDD.1A through Criterion RDD.1C. A study *partially* satisfies Criterion RDD.1 if it meets any two of the three sub-criteria. Table 2 summarizes the requirements for completely and partially satisfying RDD.1.

**Table 2. Determining whether RDD.1 is satisfied**

| Conditions | Completely satisfy | Partially satisfy |
|---|---|---|
| Criterion RDD.1A. Institutional integrity | Must satisfy | Must satisfy any two criteria among RDD.1A, RDD.1B, or RDD.1C |
| Criterion RDD.1B. Statistical integrity | Must satisfy | |
| Criterion RDD.1C. Graphical integrity | Must satisfy | |

### Criterion RDD.2: Does the study demonstrate acceptable levels of overall and differential attrition around the cutoff value or establish the equivalence of the groups being compared at the cutoff value of the forcing variable?

In order to completely satisfy the criterion, the study must have acceptable levels of overall and differential attrition. The requirements presented here are similar to testing for overall and differential attrition (linked to treatment status) in the context of RCTs. The way attrition rates are calculated determines whether an RDD study satisfies this standard completely or not. An SRDD and FRDD study can only receive a high evidence rating if it completely satisfies the attrition criterion along with all other applicable criteria.

[7] Alternative tests may include local polynomial density estimators proposed by Cattaneo, Jansson, and Ma (2020) or g-order statistics proposed by Bugni and Canay (2021).

- **Criterion RDD.2A. Demonstrate attrition is low using the most accurate adjustment for the forcing variable.** The study must use at least one of the following approaches to show that the reported combination of overall and differential attrition is low using the threshold levels summarized in Table A.1 in the appendix. In addition, for both approaches, the value of the forcing variable must be known for all units.
  - The study can estimate the predicted mean attrition rate at the cutoff using data from below it or data from above it. Both numbers must be estimated using a statistical model that controls for the forcing variable using the same approach that was used to estimate the impact on the outcome. That is, the difference in attrition between the intervention and comparison groups must be estimated using either (1) the same bandwidth and/or functional form as was used to estimate the impact on the outcome, or (2) a bandwidth and/or functional form selected based on the same algorithm used for selecting the bandwidth and/or functional form to estimate the impact on the outcome. The overall attrition rate is defined as the average of the predicted mean attrition rate on either side of the cutoff. The differential attrition rate is defined as the difference in the predicted mean attrition rates on either side of the cutoff.
  - The study can report overall and differential attrition for the sample within the bandwidth used for the impact analysis. Using this method, the authors do not need to adjust for the forcing variable.
- **Criterion RDD.2B. Demonstrate attrition is low using a less accurate adjustment for the forcing variable.** The study can report overall and differential attrition for the *entire research sample* with or without adjusting for the forcing variable.

Regardless of the approach used, authors should be consistent when calculating overall and differential attrition. For example, if the overall attrition rate was calculated without an adjustment for the forcing variable, then the differential attrition rate should also be unadjusted (and vice versa). If attrition information is either not presented in the study or is incomplete, CLEAR will conduct an author query to obtain this information. If the overall and differential levels of attrition for a study cannot be determined, then CLEAR will consider the study to have high attrition.

- **Criterion RDD.2C. Baseline equivalence on key covariates (as identified in the review protocol).** The study should either (1) establish equivalence at the cutoff value of the forcing variable for the sample within the specified bandwidth used to estimate impacts, or (2) control for key covariates in the main impact analysis. CLEAR allows studies to satisfy this requirement in two ways. First, authors can calculate baseline effect sizes for required covariates and demonstrate that differences in effect size between the intervention and comparison groups are less than 0.05 standard deviations (SD). The second approach, which is more common in labor evaluation studies, is to include the covariates as control variables in the main impact model. Including covariates in the main impact models is not a substitute for showing that baseline effect sizes are small. Including covariates in the models allows for a wider bandwidth and controls for any differences in covariates (through the model estimated locally around the cutoff), obtaining a more precise estimate (Calonico et al., 2019).

- To satisfy RDD.2C, CLEAR adopts a similar approach. To demonstrate baseline equivalence for an RDD, a study could calculate an impact on the required covariate(s) at the cutoff value of the forcing variable. If a study uses this approach, it must either (1) use exactly the same bandwidth and/or functional form as was used to estimate the impact on the outcome variable, or (2) use the same algorithm for selecting the bandwidth and/or functional form as was used to estimate the outcome. With either approach, if a study demonstrates that the impact of assignment to the intervention group on the covariate(s) is less than 0.05 SD, then the criterion is completely satisfied. For covariates that are discrete, authors would need to calculate means of the covariates within the same bandwidths and test for differences in means. Alternatively, a study may also satisfy RDD.2C by including the required covariate(s) in the main impact model used to estimate impacts.

A study completely satisfies Criterion RDD.2 if it meets Criterion RDD.2A. A study partially satisfies Criterion RDD.2 if it meets RDD.2B or RDD.2C. Table 3 summarizes the requirements for completely and partially satisfying RDD.2.

**Table 3. Determining whether RDD.2 is satisfied**

| Conditions | Completely satisfy | Partially satisfy |
|---|---|---|
| Criterion RDD.2A. The reported combination of overall and differential attrition rates is low using an approach among those that have the potential to adjust for the forcing variable most accurately. | Must satisfy | Does not need to satisfy |
| Criterion RDD.2B. The reported combination of overall and differential attrition rates is low when calculated using an approach among those that may not provide as accurate an adjustment for the forcing variable. | Does not need to satisfy | Must satisfy 2B or 2C |
| Criterion RDD.2C. The study demonstrates baseline equivalence on key covariates. | | |

### Criterion RDD.3: Does the study establish the continuity of the relationship between the outcome and the forcing variable?

The study must show that in the absence of the intervention, there would be a smooth relationship between the outcome and the forcing variable at the cutoff score. This continuity requirement is necessary if RDD studies are to produce statistically consistent impact estimates. Conversely, the presence of discontinuities would suggest that that observed impacts at the cutoff might not be due to the intervention. The continuity requirement cannot be directly tested; instead, CLEAR allows the use of indirect approaches to establish that it is either completely or partially satisfied. The sub-criteria listed below are used to assess Criterion RDD.3.

- **Criterion RDD.3A. Graphical analysis of discontinuities.** There must be no evidence, using graphical analyses, of a discontinuity in the relationship between the outcome and the forcing variable other than at the cutoff value (unless a satisfactory explanation is provided). A study may provide, using either raw data or data aggregated within bins, a visual demonstration that discontinuities between the outcome and forcing variable are not evident anywhere other than at the cutoff value. An example of an acceptable graphical analysis could involve a scatter plot of the outcome and a forcing variable within the specified bandwidth used to estimate impacts; further, a graphical analysis could also show the absence of discontinuities for the full sample if the bandwidth is not used. The graph must not show a discontinuity larger than twice the standard error of the impact estimated at the cutoff value, unless that discontinuity is satisfactorily explained. Acceptable explanations of observed discontinuities may include structural features of the forcing variable and/or features of another intervention that were determined using cutoff values different from those of the same forcing variable. In these situations, the burden of proof is on the authors to provide a convincing explanation in the text of the study. If an explanation is lacking or unconvincing, CLEAR will not conduct an author query to obtain this information.
- **Criterion RDD.3B**. **Statistical analysis of discontinuities.** Statistical tests must show no evidence of a discontinuity in the relationship between the outcome and the forcing variable other than at the cutoff value (unless a satisfactory explanation is provided). The tests must use the same algorithm for selecting the bandwidth and/or functional form as used to estimate the impact on the outcome and should be conducted for values of the forcing variable below and above the cutoff, though the values can be either within or outside the bandwidth. A study may test for impacts at values of the forcing variable where there should be no impacts (a placebo test), such as the medians of points above or below the cutoff value. Based on the results of the statistical tests, the study should demonstrate that using other values of the forcing variable does not lead to statistically

significant impact estimates substantially more often than expected. A study may also use other statistical tests such as donut hole falsification tests to satisfy this requirement (Cattaneo et al., 2019).

A study completely satisfies Criterion RDD.3 if it satisfies sub-criteria RDD.3A and RDD.3B. A study partially satisfies Criterion RDD.3 if it satisfies either Criterion RDD.3A or RDD.3B. Table 4 summarizes the requirements.

**Table 4. Determining whether RDD.3 is satisfied**

| Conditions | Completely satisfy | Partially satisfy |
|---|---|---|
| RDD.3A. Graphical analyses | Must satisfy | Must satisfy 3A or 3B |
| RDD.3B. Statistical tests | Must satisfy | |

### Criterion RDD.4: Does the study justify the functional form and bandwidth used to estimate impacts and define the analysis sample?

The statistical model used to estimate impacts in an RDD is critical. Among the most important considerations are the specification of the functional form of the regression of the outcome variable on the forcing variable and the selection of the bandwidth around the cutoff value of the forcing variable; the units within the selected bandwidth constitute the analysis sample that is used to estimate impacts. CLEAR uses six sub-criteria to determine whether a study satisfies Criterion RDD.4:

- **Criterion RDD.4A**. **Statistical model controls for the forcing variable.** This criterion indicates that the impact analysis must account for the forcing variable in the model specification. In order to produce an unbiased estimate of the average treatment effect, the statistical model must always control for the forcing variable.
- **Criterion RDD.4B**. **Choice of bandwidth.** The impacts should be estimated using an appropriate nonparametric approach within a justified bandwidth. Because RDDs estimate average treatment effects within a specified bandwidth around the cutoff value of the forcing variable, the statistical procedure used to generate this estimate should be based on an appropriate nonparametric method, such as local linear regression. Moreover, the bandwidth should also be selected using a systematic procedure that is supported in the methodological literature, such as cross-validation or those outlined in Imbens and Kalyanaraman (2012).
- **Criterion RDD.4C**. **Model selection criteria.** If the study does not use a local regression or related nonparametric approach or uses such an approach but not within a justified bandwidth, then it may estimate impacts using methods based on minimizing the mean

squared error (MSE), bias-corrected methods for bandwidth selection, or a "best fit" regression. If a study uses a best fit regression approach, the functional form of the relationship between the outcome and the forcing variable must be shown to be a better fit to the data than at least two other functional forms, using any established measure of goodness of fit (such as the Akaike Information Criterion [AIC] or adjusted R-squared).

- **Criterion RDD.4D. Robustness analysis.** The study should show that the impact findings are robust to varying bandwidth or functional form choices. If Criterion RDD.4B applies, the authors should show that results are robust (signs and statistical significance of impact estimates are the same) for at least two justified bandwidths: the MSE-optimal choice and the (coverage error) CER-optimal choice. If Criterion RDD.4C applies, then the authors should show that the results are robust (signs and statistical significance of impact estimates are the same) for one bandwidth selected using either MSE methods or goodness-of-fit measures used to select the functional form (such as using the AIC, the regression R-squared, or another fit measure that is referenced in the methodological literature).
- **Criterion RDD.4E. Graphical analysis for functional form justification.** The study must graphically display the relationship between the outcome and the forcing variable, including a scatter plot and a fitted curve. If the study uses a particular functional form (linear, quadratic, or otherwise) for the outcome-forcing variable relationship, then the study must show graphically that this functional form fits the scatter plot reasonably well within the selected bandwidth. The study can demonstrate this relationship using either the raw data or data that have been aggregated into bins or intervals.
- **Criterion RDD.4F**. **Accounting for differences in model at cutoff.** The relationship between the forcing variable and the outcome must not be constrained to be the same on both sides of the cutoff. This criterion is meant to guard against overly restrictive specifications between the forcing variable and outcome.

A study completely satisfies RDD.4 if it meets criteria RDD.4A, RDD.4B, RDD.4D, RDD.4E, and RDD 4F. A study partially satisfies RDD.4 if it meets criteria RDD.4A, either RDD.4B or RDD.4C, and RDD.4E. Table 5 summarizes the requirements for completely and partially satisfying RDD.4.

**Table 5. Determining whether RDD.4 is satisfied**

<table>
<tr><th>Conditions</th><th>Completely satisfy</th><th>Partially satisfy</th></tr>
<tr><td>RDD.4A. Statistical model</td><td>Must satisfy</td><td>Must satisfy</td></tr>
<tr><td>RDD.4B. Choice of bandwidth</td><td>Must satisfy</td><td rowspan="2">Must satisfy 4B or 4C</td></tr>
<tr><td>RDD.4C. Model selection</td><td>Does not need to satisfy</td></tr>
<tr><td>RDD.4D: Robustness checks</td><td>Must satisfy</td><td>Does not need to satisfy</td></tr>
<tr><td>RDD.4E: Graphical analysis</td><td>Must satisfy</td><td>Must satisfy</td></tr>
<tr><td>RDD.4F: Accounting for differences in model at cutoff</td><td>Must satisfy</td><td>Does not need to satisfy</td></tr>
</table>

## 2. CRITERIA FOR FUZZY RDDS

In an SRDD, the treatment is clearly based on whether the score crosses a cutoff value. For example, all intervention group members receive intervention services if they score below the threshold for determining assignment to the intervention; correspondingly, no comparison group members receive services if they score above it. In an FRDD, some units with scores below the cutoff might not receive the intervention and/or some units with scores above the cutoff might receive the intervention, but there is still a substantial discontinuity in the *probability* of receiving services at the cutoff.

An FRDD is analogous to a treatment-on-treated effect analysis in RCTs with non-compliance in which the impact of taking up the intervention is estimated. In an FRDD analysis, the impact of service receipt is calculated as a ratio in which the numerator is the RDD impact on the outcome of interest and the denominator is the RDD impact on the probability of receiving the intervention. These analyses are typically modeled using a two-stage least-squares (2SLS) estimator common in estimation of effects using instrumental variables. In FRDDs, being above or below the cutoff is used as an instrument for treatment rather than as indicating treatment. In cases where compliance is imperfect but the RDD impact is estimated without accounting for the probability of intervention receipt (that is, using a reduced-form impact model in which an intervention participation indicator is not included), then no additional criteria have to be considered. This type of model is analogous to an intent-to-treat analysis in RCTs, and impacts should be interpreted accordingly. However, if the analysis accounts for the probability of receiving the intervention, then there are additional criteria.

### Criterion RDD.5: Does the study satisfy the requirements for an FRDD?

The following eight sub-criteria are relevant for FRDDs:

- **Criterion RDD.5A. Binary participation indicator.** The participation indicator must be a binary (yes/no) variable that represents receipt of at least some portion of the intervention.
- **Criterion RDD.5B. Single participation indicator.** The estimation model must have exactly one participation indicator.
- **Criterion RDD.5C. Binary study group indicator.** The indicator for whether the forcing variable is above or below the cutoff must be binary (yes/no) for the groups to which units were assigned.
- **Criterion RDD.5D. Covariates for two-stage model.** The model that estimates the impact on participation in the intervention must include the same covariates—including the forcing variable—as the model that estimates the impact on outcomes. In the typical situation in which 2SLS estimation is used, this requirement means that the same covariates must be used in the first- and second-stage equations.
- **Criterion RDD.5E. Exclusion restriction.** There should be no clear violations of the exclusion restriction. This requirement is analogous to satisfying the exclusion restriction for instrumental variable estimators. Specifically, this means that the indicator for intervention participation (the instrument) should affect the outcome only through the intervention itself and not through other channels. For example, the definition of intervention take-up should be consistent across both the intervention and comparison groups.
- **Criterion RDD.5F. Relevance of forcing variable.** The study must provide evidence that the forcing variable is a strong predictor of participation in the intervention.[8] For example, in a regression of program participation on a treatment indicator and other covariates, the coefficient on the treatment indicator must report a minimum F-statistic of 16 or a minimum t-statistic of 4 (Stock and Yogo, 2005).
- **Criterion RDD.5G. Bandwidths for two-stage model.** The FRDD impacts should be estimated through use of an appropriate nonparametric approach within a justified bandwidth. It is acceptable to use separate bandwidths for the numerator and denominator of the impact estimator, if both are selected using the same justified approach (for example, the Inverse Kinematics algorithm applied separately to the numerator and denominator). It is acceptable to use the bandwidth selected for the numerator if that bandwidth is smaller than (or equal to) a justified bandwidth selected for the denominator.

[8] This requirement is similar to Criterion IV.1 for instrumental variable estimators (sufficient strength of the instrument in a first-stage equation).

- **Criterion RDD.5H. Bandwidth for main impact model.** If Criterion RDD.5G is not met, the study can still partially satisfy Criterion RDD.5 if the FRDD impact is estimated using a bandwidth that is justified only for the numerator, or if the denominator is estimated using a "best fit" functional form (using any measure of goodness of fit such as the AIC or adjusted R-squared).

A study completely satisfies Criterion RDD.5 if it meets Criteria RDD.5A through RDD.5G. A study partially satisfies Criterion RDD.5 if it meets Criteria RDD.5A through RDD.5F, and RDD.5H. Table 6 summarizes the requirements for completely and partially satisfying RDD.5.

**Table 6. Determining whether Criterion RDD.5 is satisfied**

| Conditions | Completely satisfy | Partially satisfy |
|---|---|---|
| RDD.5A. Binary participation indicator | Must satisfy | Must satisfy |
| RDD.5B. Single participation indicator | Must satisfy | Must satisfy |
| RDD.5C. Binary study group indicator | Must satisfy | Must satisfy |
| RDD.5D. Covariates for two-stage model | Must satisfy | Must satisfy |
| RDD.5E. Exclusion restriction | Must satisfy | Must satisfy |
| RDD.5F. Relevance of forcing variable | Must satisfy | Must satisfy |
| RDD.5G. Bandwidths for two-stage model | Must satisfy | Does not need to satisfy |
| RDD.5H. Bandwidth for main impact model | Does not need to satisfy | Must satisfy |

## 3. APPLYING RDD CRITERIA TO STUDIES THAT INCLUDE AGGREGATE OR POOLED IMPACTS

Some RDD studies may report pooled or aggregate impacts for some combinations of forcing variables, cutoffs, and samples.

- **Pooled impacts** are based on data from each combination of forcing variable, cutoff, and sample that are standardized and grouped into a single dataset for which a single impact is calculated. For example, consider a study of transitional job programs that was conducted at seven sites. The authors could create a pooled impact by standardizing and combining the data from the seven sites to estimate a single intervention impact.
- **Aggregated impacts** are a weighted average[9] of impacts calculated separately for every combination of forcing variable, cutoff, and sample. For example, a study of transitional job programs was conducted at seven sites. The authors could aggregate impacts by estimating the impacts separately for each site and then combining the seven impact estimates using a weighted average.

[9] CLEAR does not require a specific approach to weighting, and authors could choose to use unit weighting. The selection of the weighting approach by study authors will not affect a study's evidence rating.

The study's overall research rating will be the highest rated impact—including pooled and aggregate impacts—presented in the study. Study authors may improve the rating of a pooled or aggregate impact by excluding combinations of forcing variables, cutoffs, and samples that do not meet the CLEAR RDD Guidelines. However, the exclusion reasons must be exogenous to intervention participation. The authors are responsible for showing that the exclusions were made for exogenous reasons.

If the study reports aggregate or pooled impacts, additional considerations are needed when applying the RDD Guidelines:

- **Criterion RDD.1: Does the study establish the integrity of the forcing variable?**
    - Each unique combination of forcing variable, cutoff, and sample that contributes to the pooled or aggregate impact must meet the institutional (1A) and statistical (1B) <u>or</u> graphical (1C) integrity of the forcing variable; otherwise, the study does not meet this criterion. However, this requirement would be met if the study only used data from those that met the criterion to calculate the intervention impact.
    - For pooled impacts, a statistical test of the pooled data can meet criterion RDD.1B and a graph of the pooled data can meet criterion RDD.1C.

- **Criterion RDD.2: Does the study demonstrate acceptable levels of overall and differential attrition around the cutoff value or establish the equivalence of the groups being compared at the cutoff value of the forcing variable?**
    - For pooled impacts, if the authors calculate overall and differential attrition for the pooled sample, Criterion RDD.2A, Criterion RDD.2B, and Criterion RDD.2C can be applied as is.
    - For aggregate impacts, if the authors calculate weighted averages of the overall and differential attrition rates for each unique combination of forcing variable, cutoff, and sample that contribute to the aggregate impact, Criterion RDD.2A and Criterion RDD.2B can be applied to the weighted overall and differential attrition. For Criterion RDD.2C, baseline equivalence can be established by applying the same aggregation method to the impacts on baseline covariates as is used to aggregate impacts on outcomes. A sample excluded from the pooled or aggregate impact due to endogeneity (i.e., the sample is potentially affected by the intervention) cannot be excluded from the attrition calculation.

- **Criterion RDD.3: Does the study establish the continuity of the relationship between the outcome and the forcing variable?**
    - For pooled impacts, Criterion RDD.3A and Criterion RDD.3B can be applied as is.

- For aggregate impacts, the requirements for Criterion RDD.3A and Criterion RDD.3B must be applied cumulatively across all combinations of forcing variables, cutoffs, and samples. For Criterion RDD.3A, the graphical displays must show no evidence of a discontinuity larger than twice the standard error of the impact at any noncutoff value within the bandwidth of any forcing variable for any sample. If impacts from nonoverlapping samples are being aggregated, the author may exclude from the aggregate impact any impacts from samples that do not satisfy this criterion. The tests must use the same algorithm for selecting the bandwidth and/or functional form as used to estimate the impact on the outcome and should be conducted for values of the forcing variable below and above the cutoff, though the values can be either within or outside the bandwidth. If impacts from nonoverlapping samples are being aggregated, it is acceptable to exclude from the aggregate impact any impacts from samples that do not satisfy this criterion.

- **Criterion RDD.4: Does the study justify the functional form and bandwidth used to estimate impacts and define the analysis sample?**
  - For pooled impacts, all criteria can be applied as is.
  - For aggregate impacts, criteria RDD.4A, RDD.4B, RDD.4C, RDD.4E, and RDD.4F must be applied to every impact included in the aggregate. Any impacts excluded from the aggregate because they do not satisfy one of the criteria will be treated as attrition. The aggregate impact will receive the lowest rating from among all of the impacts. Criterion RDD.4D does not need to be applied to every impact included in the aggregate, it is sufficient to demonstrate robustness of the aggregate impact only.

- **Criterion RDD.5: Does the study satisfy the requirements for an FRDD?**
  - For pooled impacts, all criteria can be applied as is.
  - For aggregate impacts, each criterion must be applied to every impact included in the aggregate. Any impacts excluded from the aggregate will be treated as attrition, with two exceptions. Impacts may be excluded if they do not meet Criterion RDD.5E (there is clear evidence of the violation of the exclusion restriction) or Criterion RDD.5F (the forcing variable is not a strong predictor of participation in the intervention). The aggregate impact will receive the lowest rating from among all of the impacts.

**Appendix Table A1. Thresholds of acceptable combinations of overall and differential attrition (percentages)**

| | Differential attrition | | | Differential attrition | |
|---|---|---|---|---|---|
| Overall attrition | Conservative boundary | Liberal boundary | Overall attrition | Conservative boundary | Liberal boundary |
| 0 | 5.7 | 10.0 | 34 | 3.5 | 7.4 |
| 1 | 5.8 | 10.1 | 35 | 3.3 | 7.2 |
| 2 | 5.9 | 10.2 | 36 | 3.2 | 7.0 |
| 3 | 5.9 | 10.3 | 37 | 3.1 | 6.7 |
| 4 | 6.0 | 10.4 | 38 | 2.9 | 6.5 |
| 5 | 6.1 | 10.5 | 39 | 2.8 | 6.3 |
| 6 | 6.2 | 10.7 | 40 | 2.6 | 6.0 |
| 7 | 6.3 | 10.8 | 41 | 2.5 | 5.8 |
| 8 | 6.3 | 10.9 | 42 | 2.3 | 5.6 |
| 9 | 6.3 | 10.9 | 43 | 2.1 | 5.3 |
| 10 | 6.3 | 10.9 | 44 | 2.0 | 5.1 |
| 11 | 6.2 | 10.9 | 45 | 1.8 | 4.9 |
| 12 | 6.2 | 10.9 | 46 | 1.6 | 4.6 |
| 13 | 6.1 | 10.8 | 47 | 1.5 | 4.4 |
| 14 | 6.0 | 10.8 | 48 | 1.3 | 4.2 |
| 15 | 5.9 | 10.7 | 49 | 1.2 | 3.9 |
| 16 | 5.9 | 10.6 | 50 | 1.0 | 3.7 |
| 17 | 5.8 | 10.5 | 51 | 0.9 | 3.5 |
| 18 | 5.7 | 10.3 | 52 | 0.7 | 3.2 |
| 19 | 5.5 | 10.2 | 53 | 0.6 | 3.0 |
| 20 | 5.4 | 10.0 | 54 | 0.4 | 2.8 |
| 21 | 5.3 | 9.9 | 55 | 0.3 | 2.6 |
| 22 | 5.2 | 9.7 | 56 | 0.2 | 2.3 |
| 23 | 5.1 | 9.5 | 57 | 0.0 | 2.1 |
| 24 | 4.9 | 9.4 | 58 | - | 1.9 |
| 25 | 4.8 | 9.2 | 59 | - | 1.6 |
| 26 | 4.7 | 9.0 | 60 | - | 1.4 |
| 27 | 4.5 | 8.8 | 61 | - | 1.1 |
| 28 | 4.4 | 8.6 | 62 | - | 0.9 |
| 29 | 4.3 | 8.4 | 63 | - | 0.7 |
| 30 | 4.1 | 8.2 | 64 | - | 0.5 |
| 31 | 4.0 | 8.0 | 65 | - | 0.3 |
| 32 | 3.8 | 7.8 | 66 | - | 0.0 |
| 33 | 3.6 | 7.6 | 67 | - | - |

Note: There are conservative and liberal standards for acceptable levels of attrition. The review protocols specify the attrition standard used for each topic area. When the attrition seems to be more endogenous (rather than exogenous) to the intervention—for example, disadvantaged youth choosing whether to participate in a residential career training program—the conservative standards are applied. When the attrition seems more exogenous to the intervention—for example, employers cutting back the number of slots in a training program because of reduced funding—the liberal standards are applied.

## 4. References

Bugni, F. A., & Canay, I. A. (2021). Testing continuity of a density via g-order statistics in the regression discontinuity design. *Journal of Econometrics*, *221*(1), 138-159.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, *101*(3), 442-451.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations, Cambridge Elements: Quantitative and Computational Methods for Social Science*. Cambridge University Press.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association, 115*, 1449–1455.

Cattaneo, M. D., Keele, L., & Titiunik, R. (2023). A guide to regression discontinuity designs in medical applications. *Statistics in Medicine, 42*(24), 4484-4513.

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies, 79*(3), 933–959.

Imbens, G., & Lemieux, T. (Eds.). (2008). The regression discontinuity design—Theory and applications [Special issue]. *Journal of Econometrics*, *142*(2).

Lee, D. S., & Lemieux, T. (2009). Regression discontinuity designs in economics (NBER Working Paper No. 14723). National Bureau of Economic Research.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics,* 1*24*(2), 698-714*.*

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In J. Stock & D. W. K. Andrews (Eds.), *Identification and inference for econometric models: Essays in Honor of Thomas J. Rothenberg* (pp. 80–108). Cambridge University Press.

What Works Clearinghouse. (2022). *What Works Clearinghouse Standards Handbook, Version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.