WWC Standards for Reviewing Fuzzy Regression Discontinuity Designs

Christina LiCalsi, Ph.D.
Principal Researcher
American Institutes for Research

Emily Tanner-Smith, Ph.D.
Associate Dean for Research
University of Oregon

Joshua Polanin, Ph.D.
Principal Researcher
American Institutes for Research

Learning Goals for the Webinar

During this webinar, you will learn:

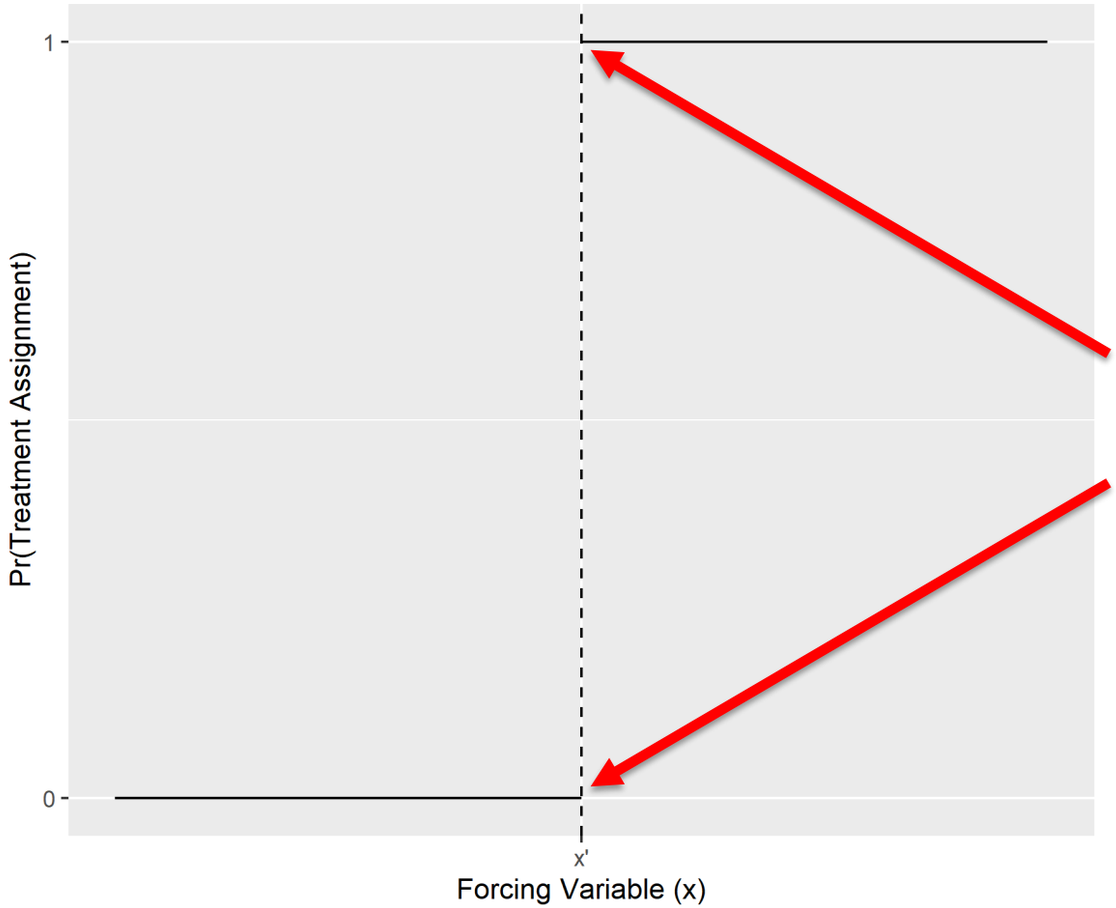
- How the instrumental variable approach is used to estimate intervention impacts in fuzzy regression discontinuity designs (RDDs)
- How the What Works Clearinghouse (WWC) reviews evidence from fuzzy RDDs
- How to apply the WWC's evidence standards to fuzzy RDDs

OVERVIEW OF FUZZY RDDS

Sharp vs. Fuzzy RDDs

- RDDs allocate participants to conditions based on how they score relative to a cutoff value on a continuous forcing variable
  - Sharp RDDs: All participants receive their assigned intervention or comparison condition.
    - Increase in probability of receiving intervention at the cutoff = 1
  - Fuzzy RDDs: Not all participants receive their assigned condition (due to no shows or crossovers).
    - But sharp increase in probability of receiving intervention at the cutoff of less than 1

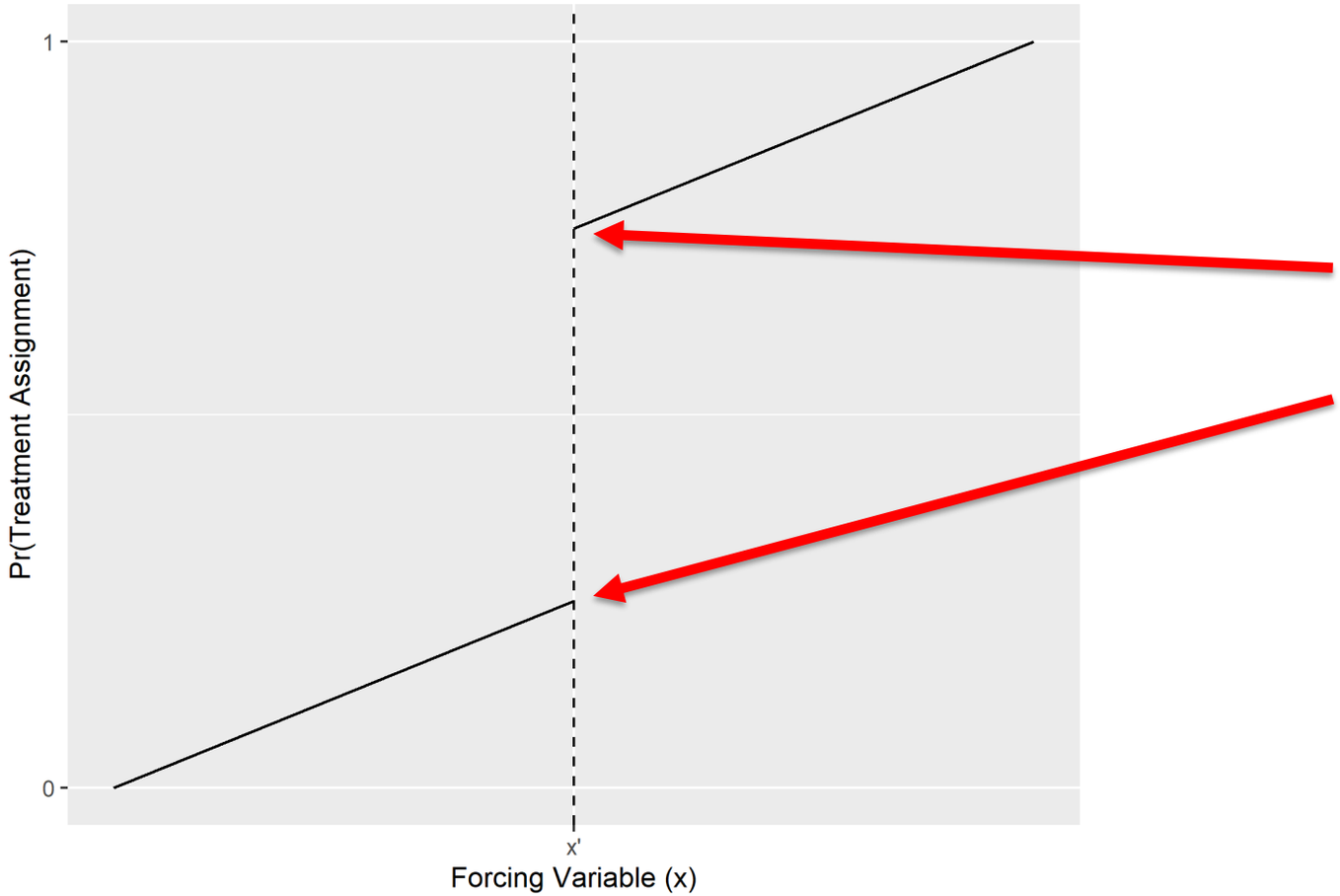
Sharp RDD: Visual



Probability of receiving intervention increases from 0 to 1 at threshold value (*X*)

Only *X* determines intervention receipt

Fuzzy RDD: Visual

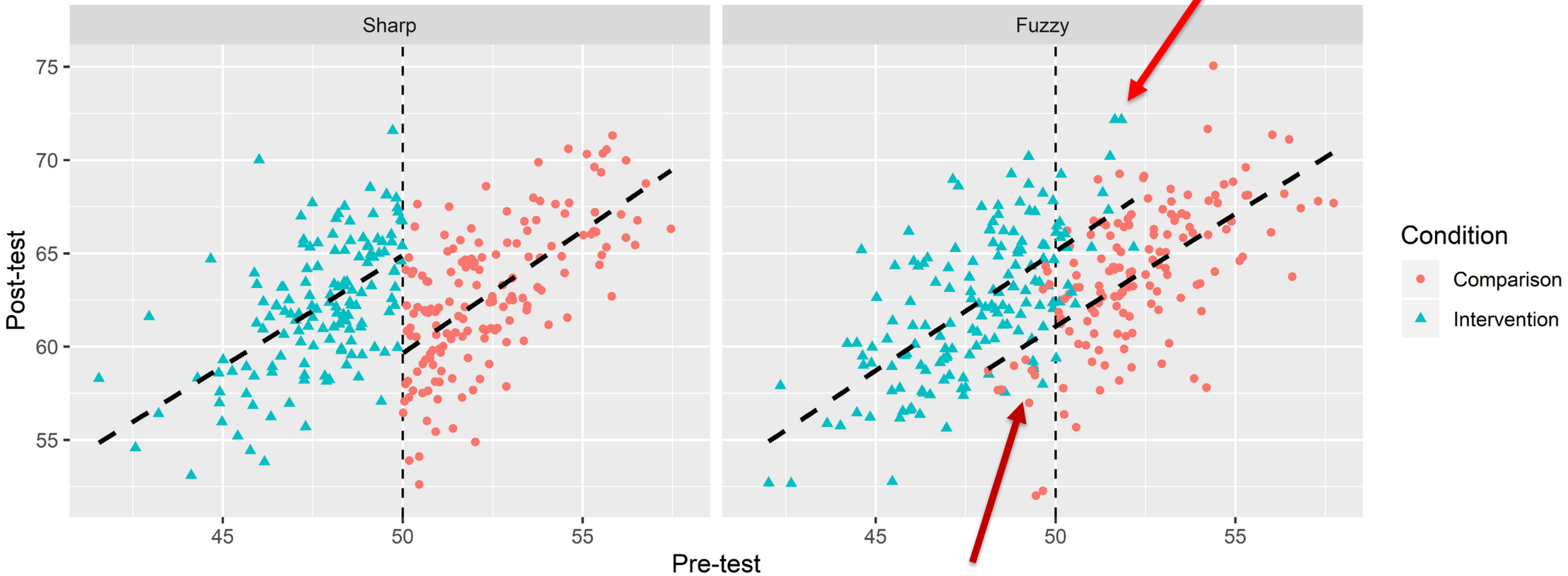


Probability of receiving intervention increases by less than 1 at threshold value (*X*); it may be greater than 0 probability on the no-intervention side, less than 1 on the intervention side, or both.

X and something else determines intervention receipt

Sharp vs. Fuzzy RDDs

Noncomplier (crossover): Not assigned to intervention but participated in the intervention anyway



Noncomplier (no-show): Assigned to intervention but did not participate in the intervention

Fuzzy RDD and Complier Average Causal Effect (CACE)

- Due to the noncompliance in fuzzy RDD, to determine an effect we calculate the average **effect of the intervention on participants at the cutoff value who received the intervention because they were assigned to it** (compliers).

Intervention Group		Comparison Group
T = 1	Compliers	T = 0
T = 1	Always-Takers	T = 1 (Crossovers)
T = 0 (No Shows)	Never-Takers	T = 0

Note: Complier Average Causal Effect (CACE) is often referred to as the Local Average Treatment Effect (LATE)

CACE and the Exclusion Restriction

- The only channel through which assignment to conditions can influence outcomes is by affecting take-up of the intervention (that is, assignment to conditions does not influence take-up status).
- Outcomes of always-takers and outcomes of never-takers do not differ between conditions.

Instrumental Variable Approach

- To calculate the CACE, employ an instrumental variable approach, which is often referred to as two-stage least squares (2SLS).
  - An instrumental variable is a variable that is associated with an outcome only through its association with intervention receipt.
  - In fuzzy RDD, the side of the cutoff is the instrument (impacts outcome only through its impact on probability of intervention receipt).

Instrumental Variable (2SLS) Approach

- The CACE is calculated as the ratio of two discontinuities.
  - Discontinuity 1 = Impact on outcome at cutoff
    - Outcome intervention side – Outcome comparison side
  - Discontinuity 2 = Impact on intervention participation at cutoff
    - Percent intervention on intervention side – Percent intervention on comparison side

$$\text{Impact of intervention receipt} = \frac{\text{Impact on outcome at cutoff}}{\text{Impact on intervention participation at cutoff}}$$

Knowledge Check

- Since 2002, all grade 3 students in Florida are required to meet the Level 2 benchmark or higher on the statewide reading test in order to be promoted to fourth grade.
- "Good cause exemptions" allow students to be promoted despite failing to score at the Level 2 benchmark or above.
- Researchers have used this natural experiment to examine the effect on later outcomes of being retained in grade 3.

- Question 1: Is this a sharp RDD or a fuzzy RDD?

Knowledge Check

- Question 1: Is this a sharp RDD or a fuzzy RDD?

- Answer: This is a fuzzy RDD. There are exemptions to being retained for students who fail to score at the Level 2 benchmark. Therefore, the probability of intervention receipt below the cutoff is not 1. Also, there is nothing stipulating that students who score at Level 2 or above cannot be retained. Therefore, the probability of intervention receipt above the cutoff is greater than 0.

Knowledge Check

- Question 2: Students who score just below the cutoff have an average reading test score of 300 one year later, while students just above the cutoff have an average reading test score of 288. If 80 percent of students who score just below the cutoff are retained and 20 percent of students above the cutoff are retained, then what is the CACE of being retained on reading test scores one year later?
  - Reminder: CACE = (Impact on outcome at cutoff/Impact on intervention participation at the cutoff)

- ❑ 12
- ❑ 15
- ❑ 20
- ❑ 25

Knowledge Check

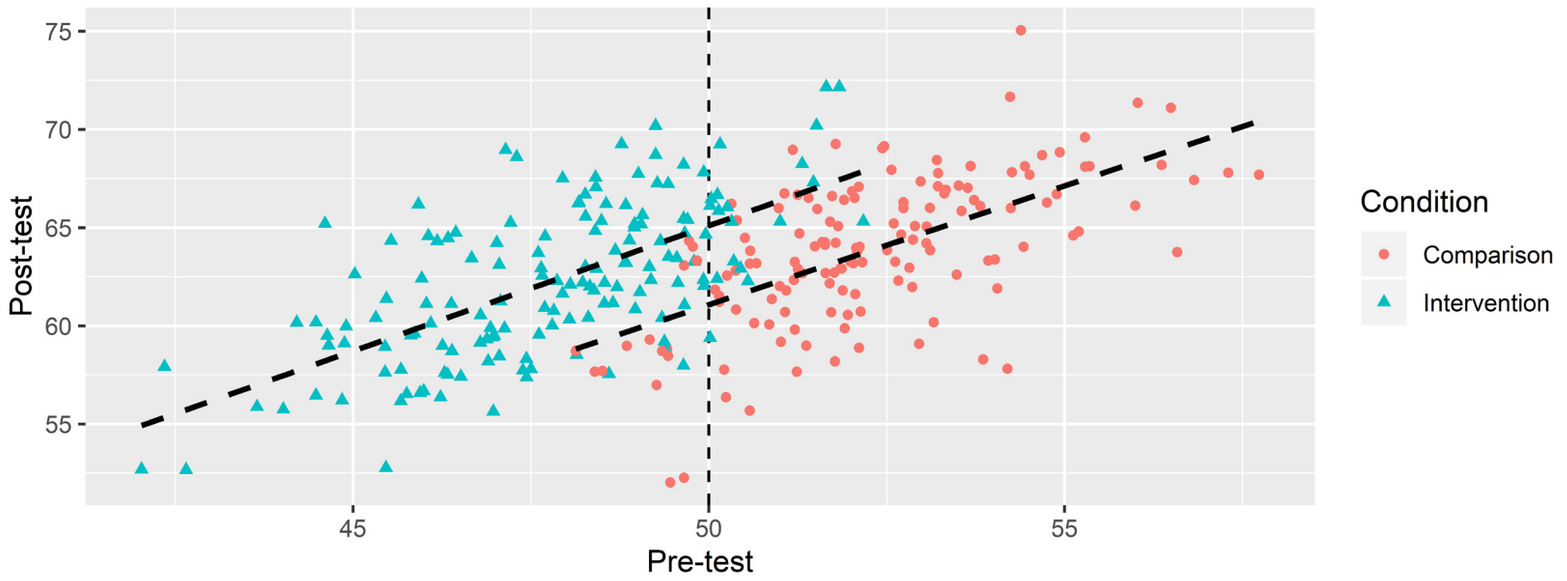
- Question 2: Students who score just below the cutoff have an average reading test score of 300 one year later, while students just above the cutoff have an average reading test score of 288. If 80 percent of students who score just below the cutoff are retained and 20 percent of students above the cutoff are retained, then what is the CACE of being retained on reading test scores one year later?

- Answer:
  - 20 points; (300 – 288)/(0.8 – 0.2)
  - First, we must calculate the impact on the outcome at the cutoff (discontinuity 1). This is the difference in later test scores between students just below (intervention side) and just above (comparison side), or 300 – 288, which equals 12.
  - Then, we must calculate the impact on retention (intervention participation) of scoring just below the cutoff (discontinuity 2). This is the difference in the percentage of students retained at the cutoff, or 0.8 – 0.2, which equals 0.6.
  - Finally, we must divide 12 by 0.6, which equals 20.
  - Thus, the average effect of being retained on reading test scores one year later for students who scored at the threshold for retention, and who were retained *because* they scored below the threshold, is 20 points.

WWC STANDARDS FOR REVIEWING FUZZY RDDS

Types of RDDs Eligible for WWC Review

- RDD studies are eligible for WWC review if they meet four criteria:
  1. The intervention assignments are based on a numerical forcing variable.
  2. The forcing variable is ordinal with at least four unique values each above or below the cutoff.
  3. No confounding factors in the study are perfectly aligned with either condition.
  4. The forcing variable used to calculate impacts is the actual forcing variable used for assignment to conditions.



Knowledge Check

- North Carolina successfully competed for federal Race to the Top Funds to turn around the lowest 5 percent of the state's schools through the Turning Around the Lowest Achieving Schools (TALAS) program. Assignment to TALAS was based on a school's 2010 composite score, which is the percentage of reading, mathematics, science, and end-of-course tests passed out of all such tests taken in a given school. The bottom 5 percent of schools in each type were to be placed in the TALAS program (additional schools were placed in the program based on low graduation rates).
- In all, 89 of the 1,772 North Carolina public elementary and middle schools were eligible for TALAS in 2010. The bottom 5 percent of composite scores cutoff was not used for any other non-TALAS programs or supports. All schools below the cutoff participated, as did two schools above the cutoff.
- A study used an RDD to examine the effect of the TALAS program participation on students' later mathematics and reading test scores. The study authors had access to the school composite scores used to determine eligibility for the TALAS program.

- Question: Is the study eligible for WWC review?

Knowledge Check

- Question: Is the study eligible for WWC review?

- Answer:
  - Yes
  - Criteria 1: Intervention assignments are based on a numerical forcing variable.
    - Yes. Intervention is based on a percentile rank.
  - Criteria 2: The forcing variable is ordinal with at least four unique values each above or below the cutoff.
    - Yes. There are 89 schools that fall below the cutoff. If percentiles are calculated to the first decimal place, there are 50 unique values below the cutoff and 950 above.
  - Criteria 3: No confounding factors in the study are perfectly aligned with either condition.
    - Yes. No other programs or policies are allocated based on the percentile criteria and perfectly aligned with the 5 percent cutoff.
  - Criteria 4: The forcing variable used to calculate impacts is the actual forcing variable used for assignment to conditions.
    - Yes. Schools were ranked by composite score and that composite score is the forcing variable that determined TALAS participation assignment and the one being used in the study.

Possible WWC Ratings for RDDs

MEETS WWC STANDARDS WITHOUT RESERVATIONS

MEETS WWC STANDARDS WITH RESERVATIONS

Standard	To be rated *Meets WWC RDD Standards <u>Without</u> Reservations*, studies must:	To be rated *Meets WWC RDD Standards <u>With</u> Reservations*, studies must:
1: Integrity of the forcing variable	Completely satisfy	Partially satisfy
2: Sample attrition	Completely satisfy	Must partially satisfy at least one of these two standards
3: Continuity	Completely satisfy	
4: Bandwidth/functional form	Completely satisfy	Partially satisfy
5: Fuzzy RDD	Completely satisfy (or waived)	Partially satisfy

The RDD Standard 5 is waived for studies that use a sharp RDD and is waived for studies that use a fuzzy RDD but calculate all impact estimates from a reduced form model (that is, outcome modeled as a function of forcing variable and indicator for being above or below the cutoff, but the participation indicator is excluded from the model).

Eight Criteria for Reviewing Fuzzy RDDs (RDD Standard 5)

Criterion	To completely satisfy the fuzzy RDD standard, the study:	To partially satisfy the fuzzy RDD standard, the study:
A: The participation indicator must be a binary indicator.	Must satisfy this criterion.	Must satisfy this criterion.
B: The estimation model must have exactly one participation indicator.	Must satisfy this criterion.	Must satisfy this criterion.
C: The indicator for being above or below the cutoff must be a binary indicator for the groups.	Must satisfy this criterion.	Must satisfy this criterion.
D: The same covariates must be included in (1) the analysis that estimates the impact on participation and (2) the analysis that estimates the impact on outcomes.	Must satisfy this criterion.	Must satisfy this criterion.
E: No clear violations of the exclusion restriction.	Must satisfy this criterion.	Must satisfy this criterion.
F: Evidence that the forcing variable is a strong predictor of participation in the intervention.	Must satisfy this criterion.	Must satisfy this criterion.
G: Local regression or related nonparametric approach with a justified bandwidth.	Must satisfy this criterion.	Does not need to satisfy this criterion.
H: Local regression or related nonparametric approach with a bandwidth that is only justified for the numerator, or the denominator is estimated using a best-fit functional form.	Does not need to satisfy this criterion.	Must satisfy this criterion.

Criteria A and B for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion A: The participation indicator must be a binary indicator for taking up at least a portion of the intervention.
  - The WWC does not currently synthesize evidence about intervention dosage; the WWC does not have standards for evaluating fuzzy RDD studies that use a continuous indicator for intervention dosage.

- Criterion B: The estimation model must have exactly one participation indicator.
  - The WWC does not currently have standards for evaluating fuzzy RDD studies that use more than one participation indicator.

Criteria C and D for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion C: The indicator for being above or below the cutoff must be a binary indicator for the groups to which participants are assigned.
  - The WWC does not currently have standards for evaluating fuzzy RDD studies that use nonbinary measures to indicate group assignment.

- Criterion D: The same covariates (including the forcing variable) must be included in (1) the analysis that estimates the impact on participation and (2) the analysis that estimates the impact on outcomes.

Criterion E for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion E: No clear violations of the exclusion restriction.
  - The only channel through which assignment to conditions can influence outcomes is by affecting take-up of the intervention (that is, assignment to conditions does not influence take-up status).
  - Outcomes of always-takers and outcomes of never-takers do not differ between conditions.
  - Definition of take-up must be the same across assigned conditions.

- Example violation: Defining participation inconsistently between the assigned intervention and comparison groups.
- Example violation: Assignment to the intervention group changes the behavior of participants even if they do not take up the intervention.

Criterion F for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion F: Study must provide evidence that the forcing variable is a strong predictor of participation in the intervention.
  - In a regression of program participation on intervention indicator and other covariates, the coefficient of the intervention indicator must report a minimum *F*-statistic of 16 or a minimum *t*-statistic of 4.

Criterion G for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion G: Study must use local regression or related nonparametric approach in which impacts are estimated within a justified bandwidth, using one of the following approaches:
  - Justified bandwidth selection procedure for fuzzy RDD impact estimate (ratio)
  - Separate justified bandwidths for numerator and denominator
  - Any bandwidth for numerator only, if it is ≤ justified bandwidth for denominator

What is considered a justified bandwidth selection procedure?

A bandwidth selected using a systematic procedure that is described in an article published in a peer-reviewed journal that describes the procedure and demonstrates its effectiveness (for example, cross-validation, plug-in/IK procedure, robust CCT procedure).

Criterion H for Reviewing Fuzzy RDDs (RDD Standard 5)

- Criterion H: The study must estimate fuzzy RDD impact using one of the following approaches:
  - Justified bandwidth for numerator only
  - Denominator estimated using a "best fit" functional form (that is, a form that is a better fit to the data than at least two other functional forms)

Eight Criteria for Reviewing Fuzzy RDDs (RDD Standard 5)

Criterion	To completely satisfy the FRDD standard, the study:	To partially satisfy the FRDD standard, the study:
A: The participation indicator must be a binary indicator.	Must satisfy this criterion	Must satisfy this criterion
B: The estimation model must have exactly one participation indicator.	Must satisfy this criterion	Must satisfy this criterion
C: The indicator for being above or below the cutoff must be a binary indicator for the groups.	Must satisfy this criterion	Must satisfy this criterion
D: The same covariates must be included in (1) the analysis that estimates the impact on participation and (2) the analysis that estimates the impact on outcomes.	Must satisfy this criterion	Must satisfy this criterion
E: No clear violations of the exclusion restriction.	Must satisfy this criterion	Must satisfy this criterion
F: Evidence that the forcing variable is a strong predictor of participation in the intervention.	Must satisfy this criterion	Must satisfy this criterion
G: Local regression or related nonparametric approach with a justified bandwidth.	Must satisfy this criterion	Does not need to satisfy this criterion
H: Local regression or related nonparametric approach with a bandwidth that is only justified for the numerator, or the denominator is estimated using a best-fit functional form.	Does not need to satisfy this criterion	Must satisfy this criterion

Eight Criteria for Reviewing Fuzzy RDDs (RDD Standard 5)

Criterion	To completely satisfy the FRDD standard, the study:	To partially satisfy the FRDD standard, the study:
A: The participation indicator must be a binary indicator.	Must satisfy this criterion	Must satisfy this criterion
B: The estimation model must have exactly one participation indicator.	Must satisfy this criterion	Must satisfy this criterion
C: The indicator for being above or below the cutoff must be a binary indicator for the groups.	Must satisfy this criterion	Must satisfy this criterion
D: The same covariates must be [illegible] mates the impact on participation an[illegible] impact on outcomes.	Mus[illegible]	Must satisfy this criterion
E: No clear violations of the exclusion restriction.	Must satisfy this criterion	Must satisfy this criterion
F: Evidence that the forcing variable is a strong predictor of participation in the intervention.	Must satisfy this criterion	Must satisfy this criterion
G: Local regression or related nonparametric approach with a justified bandwidth.	Must satisfy this criterion	Does not need to satisfy this criterion
H: Local regression or related nonparametric approach with a bandwidth that is only justified for the numerator, or the denominator is estimated using a best-fit functional form.	Does not need to satisfy this criterion	Must satisfy this criterion

Highest possible rating

MEETS WWC STANDARDS WITHOUT RESERVATIONS

MEETS WWC STANDARDS WITH RESERVATIONS

APPLICATION OF THE WWC REVIEW CRITERIA FOR FUZZY RDDs

Example

- A U.S. state requires schools to retain third-grade students who do not perform at a basic proficiency level on the state reading exam (pass/fail). However, some students who pass the reading exam may be eligible for an exemption and thus are not retained. Furthermore, teachers may also elect to retain students who pass the reading exam but are deemed in need of retention.
- Researchers used a fuzzy RDD to estimate the complier average causal effect of third-grade retention on student's subsequent academic performance on standardized reading achievement tests. Students are assigned to the intervention condition (retained) and comparison condition (promoted) based on a continuous, numerical forcing variable (state reading exam) that ranges from 0 to 100 with the cutoff value defined as 50. The researchers have access to state administrative records containing the actual state reading exam scores that were used to assign students to condition. The researchers provide no indication that this cutoff value was used to assign students to any other interventions or services.

✓ Eligible for WWC review as an RDD

Example

- The authors estimated complier average causal effects using instrumental variables estimation in a 2SLS model as follows:

First stage: $$R = \gamma_1 C + \gamma_2 F + \gamma_3 (C \text{ x } F) + \Gamma Z + \varepsilon$$

Second stage: $$Y = \delta_1 R + \delta_2 F + \delta_3 (C \text{ x } F) + \Lambda Z + \eta$$

- *R* indicates whether a student was retained in grade 3
- *C* is a dummy indicator for whether student fell below the cutoff value on the forcing variable
- *F* is the continuous measures of the forcing variable centered around the cutoff
- *Z* is a vector of student demographic characteristics (age, gender, race, grade 2 reading achievement)
- *Y* indicates reading achievement in grade 4
- ε and η are error terms

Example

- The authors estimated complier average causal effects using instrumental variables estimation in a 2SLS model as follows:

First stage: $$R = \gamma_1 C + \gamma_2 F + \gamma_3 (C \text{ x } F) + \Gamma Z + \varepsilon$$

Second stage: $$Y = \delta_1 R + \delta_2 F + \delta_3 (C \text{ x } F) + \Lambda Z + \eta$$

- *R* indicates whether a student was retained in grade 3
- *C* is a dummy indicator for whether student fell below the cutoff value on the forcing variable
- *F* is the continuous measures of the forcing variable centered around the cutoff
- *Z* is a vector of student demographic characteristics (age, gender, race, grade 2 reading achievement)
- *Y* indicates reading achievement in grade 4
- ε and η are error terms

Criterion A: Binary participation indicator for taking up some portion of intervention? ✓

Criterion B: Does the model include only one participation indicator? ✓

Criterion C: The participation indicator is binary? ✓

Example

- The authors estimated complier average causal effects using instrumental variables estimation in a 2SLS model as follows:

First stage: $R = \gamma_1 C + \gamma_2 F + \gamma_3 (C \text{ x } F) + \Gamma Z + \varepsilon$

Second stage: $Y = \delta_1 R + \delta_2 F + \delta_3 (C \text{ x } F) + \Lambda Z + \eta$

- *R* indicates whether a student was retained in grade 3
- *C* is a dummy indicator for whether student fell below the cutoff value on the forcing variable
- *F* is the continuous measures of the forcing variable centered around the cutoff
- *Z* is a vector of student demographic characteristics (age, gender, race, grade 2 reading achievement)
- *Y* indicates reading achievement in grade 4
- ε and η are error terms

Criterion D: ✓
Same covariates used in estimates of impact on participation and impact on outcomes?

Example

- The authors estimated complier average causal effects using instrumental variables estimation in a 2SLS model as follows:

First stage: $R = \gamma_1 C + \gamma_2 F + \gamma_3 (C \text{ x } F) + \Gamma Z + \varepsilon$

Second stage: $Y = \delta_1 R + \delta_2 F + \delta_3 (C \text{ x } F) + \Lambda Z + \eta$

Criterion E:
No clear violations of the exclusion restriction? ✗

- The authors explicitly stated that grade retention (vs. promotion) was defined consistently for the intervention and comparison groups. However, the authors stated that many parents of retained students reported seeking supplementary reading tutoring and instruction after being notified of the students' failing grade on the grade 3 reading exam.

Example

- The authors estimated complier average causal effects using instrumental variables estimation in a 2SLS model as follows:

First stage: $R = \gamma_1 C + \gamma_2 F + \gamma_3 (C \text{ x } F) + \Gamma Z + \varepsilon$

Second stage: $Y = \delta_1 R + \delta_2 F + \delta_3 (C \text{ x } F) + \Lambda Z + \eta$

- The authors estimated the two-stage least squares model using a bandwidth of 16 (that is, 8 test score points on either side of the cutoff value of 50). This bandwidth was selected using the Imbens and Kalyanaraman (2012) optimal bandwidth algorithm, which identified an optimal bandwidth of 16 for the impact on participation and an optimal bandwidth of 24 for the impact on outcomes. The authors reported that the *t*-statistic for the instrument was 3.87.

Criterion F:
Evidence that the forcing variable is a strong predictor of participation? ✗

Criterion G:
Fuzzy RDD impacts estimated within a justified bandwidth? ✓

Criterion H:
Justified bandwidth for numerator only or denominator estimated using a "best fit" functional form? n/a

Example: Ratings on Fuzzy RDD Standard 5

Criterion	To completely satisfy the FRDD standard, the study:	To partially satisfy the FRDD standard, the study:
A: The participation indicator must be a binary indicator.	Must satisfy this criterion.	Must satisfy this criterion.
B: The estimation model must have exactly one participation indicator.	Must satisfy this criterion.	Must satisfy this criterion.
C: The indicator for being above or below the cutoff must be a binary indicator for the groups.	Must satisfy this criterion.	Must satisfy this criterion.
D: The same covariates must be included in (1) the analysis that estimates the impact on participation and (2) the analysis that estimates the impact on outcomes.	Must satisfy this criterion.	Must satisfy this criterion.
E: No clear violations of the exclusion restriction.	Must satisfy this criterion.	Must satisfy this criterion.
F: Evidence that the forcing variable is a strong predictor of participation in the intervention.	Must satisfy this criterion.	Must satisfy this criterion.
G: Local regression or related nonparametric approach with a justified bandwidth.	Must satisfy this criterion.	Does not need to satisfy this criterion.
H: Local regression or related nonparametric approach with a bandwidth that is only justified for the numerator, or the denominator is estimated using a best-fit functional form.	Does not need to satisfy this criterion.	Must satisfy this criterion.

Highest possible rating:
Does Not Meet WWC RDD Standards

Questions?

Contact

Christina LiCalsi, Principal Researcher
American Institutes for Research
clicalsi@air.org

Emily Tanner-Smith, Associate Dean for Research
University of Oregon
etanners@uoregon.edu

Joshua Polanin, Principal Researcher
American Institutes for Research
jpolanin@air.org

WWC Help Desk
https://ies.ed.gov/ncee/wwc/help

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) for the What Works Clearinghouse under contract 919900180019, administered by American Institutes for Research. The content of this product does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

References and Resources

Heissel & Ladd. (2016). *School Turnaround in North Carolina: A regression Discontinuity Analysis. Economics of Education Review*. Center for Analysis of Longitudinal Data in Education Research. Accessible at https://caldercenter.org/sites/default/files/WP%20156.pdf.

Imbens & Kalyanaraman. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, 79 (3), 933-959.

What Works Clearinghouse. (n.d.). *WWC Procedures Handbook*, *Version 4.0.* Washington, DC: Author. Retrieved September 26, 2019, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

What Works Clearinghouse. (2018). *WWC Reporting Guide for Study Authors: Regression Discontinuity Design Studies*. Washington, DC: Author. Retrieved September 26, 2019, from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_rdd_guide_022218.pdf

What Works Clearinghouse. (n.d.). *WWC Standards Handbook*, *Version 4.0.* Washington, DC: Author. Retrieved September 26, 2019, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.