Regression Discontinuity

RMDA II — Spring 2026

Sebastian Tello-Trillo

University of Virginia

Lecture 7

Overview

Part 1

Intuition

- Why thresholds matter
- Motivating example
- Selection bias

Part 2

Application

- RD vocabulary
- Sharp vs. Fuzzy RD
- Implementation
- Optimal bandwidth

Part 3

Assumptions

- Continuity
- Local randomization
- What must not jump

Part 4

Providing Evidence

- Graphical evidence
- Covariate balance
- Density test
- Placebo tests

+ Other Topics

Internal/external validity, checklists

Part 1: Intuition

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly

Policies are discontinuous

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65

Policies are discontinuous

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65

Nature Doesn’t Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65
- A test score of 70 passes; 69 fails

Nature Doesn’t Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65
- A test score of 70 passes; 69 fails
- Blood alcohol $\geq$ 0.08% triggers a DUI charge

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65
- A test score of 70 passes; 69 fails
- Blood alcohol $\geq$ 0.08% triggers a DUI charge
- Income $<$ \$$X$ qualifies for a cash transfer

Nature Doesn’t Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65
- A test score of 70 passes; 69 fails
- Blood alcohol $\geq$ 0.08% triggers a DUI charge
- Income $<$ \$$X$ qualifies for a cash transfer

Nature Doesn't Make Jumps. But Policies Do.

Nature is continuous

- Height, weight, age vary smoothly
- A person 64 years, 11 months old is nearly identical to one who just turned 65
- No natural reason outcomes should jump at arbitrary thresholds

Policies are discontinuous

- Medicare eligibility at **exactly** age 65
- A test score of 70 passes; 69 fails
- Blood alcohol $\geq 0.08\%$ triggers a DUI charge
- Income $< \$X$ qualifies for a cash transfer

Key Insight

If we observe a **jump in outcomes** exactly at a policy threshold, and nothing else changes there, the jump must be caused by the policy.

The Core Insight: Rules Create Quasi-Random Assignment

The thought experiment:

- Suppose admission requires a test score ≥ 70

The Core Insight: Rules Create Quasi-Random Assignment

The thought experiment:

- Suppose admission requires a test score ≥ 70
- A student scoring **69** is nearly identical to one scoring **71** in ability, background, and unobservables

The Core Insight: Rules Create Quasi-Random Assignment

The thought experiment:

- Suppose admission requires a test score ≥ 70
- A student scoring **69** is nearly identical to one scoring **71** in ability, background, and unobservables
- But one was admitted; the other was not

The Core Insight: Rules Create Quasi-Random Assignment

The thought experiment:

- Suppose admission requires a test score ≥ 70
- A student scoring **69** is nearly identical to one scoring **71** in ability, background, and unobservables
- But one was admitted; the other was not

The Core Insight: Rules Create Quasi-Random Assignment

The thought experiment:

- Suppose admission requires a test score ≥ 70
- A student scoring **69** is nearly identical to one scoring **71** in ability, background, and unobservables
- But one was admitted; the other was not

Key Insight

Near the cutoff, assignment is "as good as random"

The running variable is continuous, so there is no reason for units to sort perfectly around the cutoff.

Selective Schools: The Selection Bias Problem

Question: Does attending a selective school cause better outcomes?

Selective Schools: The Selection Bias Problem

Question: Does attending a selective school cause better outcomes?

Question

Before we talk about the RD solution, try to think: how would you set up a regression to answer this question? What would the regression look like? What is the fundamental problem with that regression?

Selective Schools: The Identification Problem

Why Would This Be Biased?

Students who attend selective schools are already different—higher ability, more motivated, better resources. Comparing them to students in non-selective schools conflates the school effect with the selection effect.

Selective Schools: The Identification Problem

Why Would This Be Biased?

Students who attend selective schools are already different—higher ability, more motivated, better resources. Comparing them to students in non-selective schools conflates the school effect with the selection effect.

What would we need? A group of students who are identical to selective-school students in every way except they happened not to attend. That sounds impossible… unless we look near a cutoff.

The RD Solution: Barely In vs. Barely Out

The insight:
Among students scoring just above or just below the cutoff:

- Same academic preparation

The RD Solution: Barely In vs. Barely Out

The insight:
Among students scoring just above or just below the cutoff:

- Same academic preparation
- Similar family background

The RD Solution: Barely In vs. Barely Out

The insight:
Among students scoring just above or just below the cutoff:

- Same academic preparation
- Similar family background
- Only difference: admission to the school

The RD Solution: Barely In vs. Barely Out

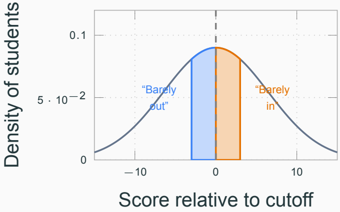
The insight:
Among students scoring just above or just below the cutoff:

- Same academic preparation
- Similar family background
- Only difference: admission to the school

The RD Solution: Barely In vs. Barely Out

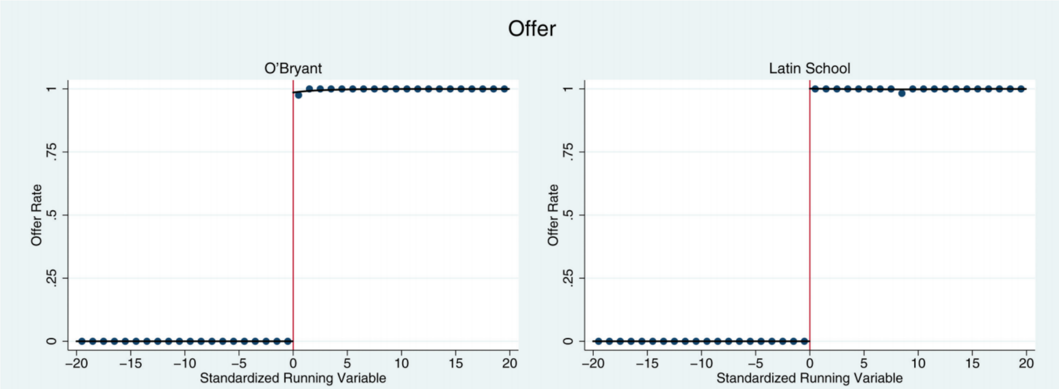
The insight:
Among students scoring just above or just below the cutoff:

- Same academic preparation
- Similar family background
- Only difference: admission to the school

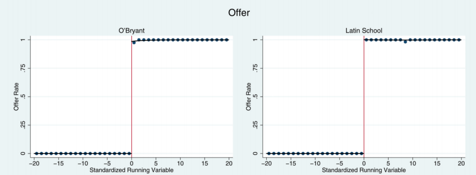


Key Insight

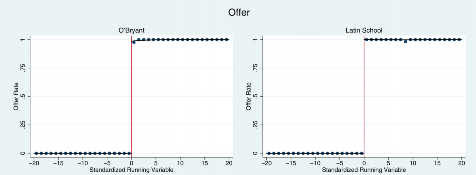
Comparing "barely in" vs. "barely out" gives us a clean causal estimate. The admission cutoff acts like a **local randomization**.


Offer
O'Bryant
Latin School
Offer Rate
Standardized Running Variable
1
.75
.5
.25
0
−20
−15
−10
−5
0
5
10
15
20

Does the Cutoff Generate a Jump in Offers?



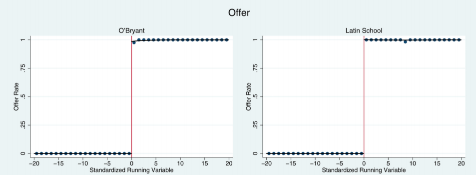
Does the Cutoff Generate a Jump in Offers?



What we observe:

- The probability of receiving an offer jumps sharply at the test score cutoff

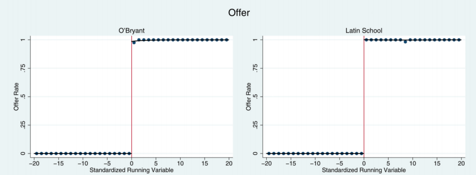
Does the Cutoff Generate a Jump in Offers?



What we observe:

- The probability of receiving an offer jumps sharply at the test score cutoff
- Below cutoff: near-zero offer rate

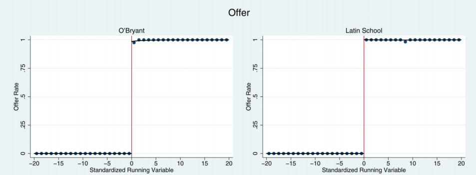
Does the Cutoff Generate a Jump in Offers?



What we observe:

- The probability of receiving an offer jumps sharply at the test score cutoff
- Below cutoff: near-zero offer rate
- Above cutoff: close to 100% offer rate

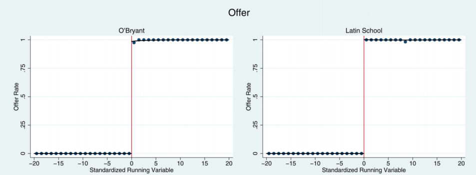
Does the Cutoff Generate a Jump in Offers?



What we observe:

- The probability of receiving an offer jumps sharply at the test score cutoff
- Below cutoff: near-zero offer rate
- Above cutoff: close to 100% offer rate

Does the Cutoff Generate a Jump in Offers?



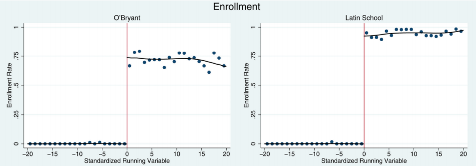
What we observe:

- The probability of receiving an offer jumps sharply at the test score cutoff
- Below cutoff: near-zero offer rate
- Above cutoff: close to 100% offer rate

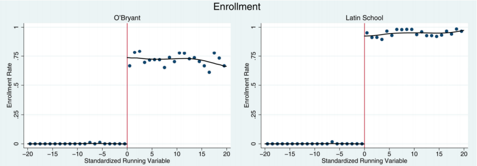
Key Insight

No offer jump $\Rightarrow$ no RD. Always verify that the running variable actually predicts treatment at the cutoff.

Does the Offer Translate to Enrollment?



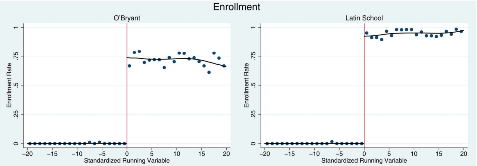
Does the Offer Translate to Enrollment?



Enrollment also jumps:

- Enrollment is the **actual treatment** (attending the selective school)

Does the Offer Translate to Enrollment?

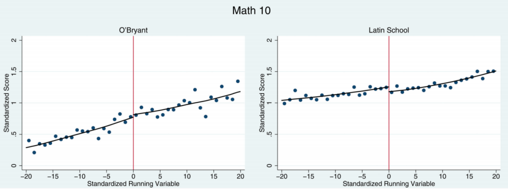


Enrollment also jumps:

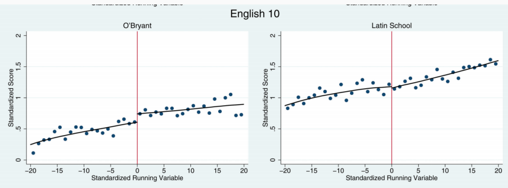
- Enrollment is the **actual treatment** (attending the selective school)
- The jump in enrollment is smaller than in offers — not all students accept

Do Outcomes Jump? Math and English Scores

Math Scores after 2 Years

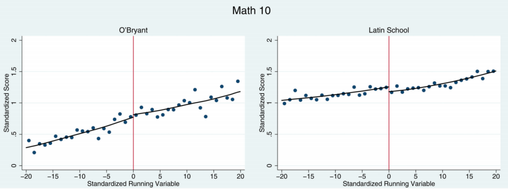


English Scores after 2 Years

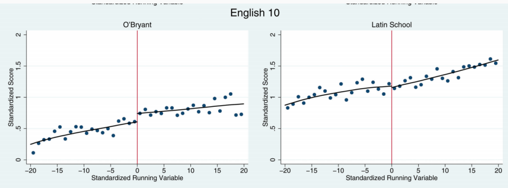


Do Outcomes Jump? Math and English Scores

Math Scores after 2 Years



English Scores after 2 Years



Key Insight

If the outcome **jumps at the same cutoff** as treatment, and continuity holds, the jump is the causal effect. The graph is the primary evidence.

Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)

Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)

Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)

Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)

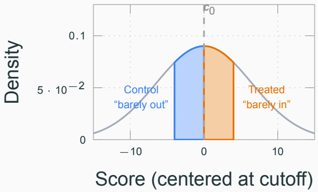
Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)

Scoring **70** vs. **69** is largely determined by luck near the cutoff.

- These tiny fluctuations are unrelated to unobservable ability



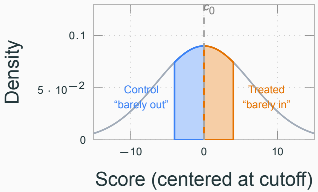
Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)

Scoring **70** vs. **69** is largely determined by luck near the cutoff.

- These tiny fluctuations are unrelated to unobservable ability
- Near the cutoff: treatment assignment $\approx$ random



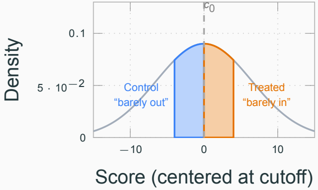
Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)

Scoring **70** vs. **69** is largely determined by luck near the cutoff.

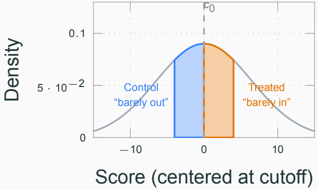
- These tiny fluctuations are unrelated to unobservable ability
- Near the cutoff: treatment assignment ≈ random



Why “Barely In vs. Barely Out” Works

Think about what determines a test score:

- Preparation and ability (systematic)
- How well you slept the night before (luck)
- Whether you guessed right on an ambiguous question (luck)



Scoring **70** vs. **69** is largely determined by luck near the cutoff.

- These tiny fluctuations are unrelated to unobservable ability
- Near the cutoff: treatment assignment ≈ random

Key Insight

Near the cutoff, the running variable acts like a **local lottery**. This is the magic of RD.

Exercise 1: Is This a Valid RD Design?

Question

For each scenario, decide: **(i)** Is there a valid RD design? **(ii)** If yes, name the running variable and cutoff. **(iii)** What could go wrong?

(a) Students scoring ≥ 70 on a placement test are placed in honors classes.
(b) A lottery randomly assigns 500 families to receive a cash transfer.
(c) Firms with > 50 employees must provide health insurance.
(d) Counties with a poverty rate below 20% do not receive a federal education grant.

Exercise 1: Answers

Answer

(a) **Valid RD.** Running variable: test score; cutoff: 70. Risk: students retake the test to cross 70 (manipulation).

(b) **Not an RD** — a lottery is an RCT. There is no continuous score or threshold.

(c) **Valid RD.** Running variable: number of employees; cutoff: 50. Risk: firms may strategically keep headcount at 49 (bunching/manipulation).

(d) **Valid RD.** Running variable: county poverty rate; cutoff: 20%. Risk: other federal programs may also use a 20% cutoff (compound treatment).

Part 2: Application

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility

- **Cutoff** c_0

  The value of X at which treatment status switches

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility

- **Cutoff** c_0

  The value of X at which treatment status switches

- **Treatment** D

  The intervention triggered by crossing the threshold

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility
- **Cutoff** c_0

  The value of X at which treatment status switches
- **Treatment** D

  The intervention triggered by crossing the threshold
- **Outcome** Y

  The variable we want to measure the causal effect on

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility

- **Cutoff** c_0

  The value of X at which treatment status switches

- **Treatment** D

  The intervention triggered by crossing the threshold

- **Outcome** Y

  The variable we want to measure the causal effect on

The Language of RD: Core Variables

Four Key Concepts

- **Running variable** X

  A continuous, pre-determined score that determines eligibility
- **Cutoff** c_0

  The value of X at which treatment status switches
- **Treatment** D

  The intervention triggered by crossing the threshold
- **Outcome** Y

  The variable we want to measure the causal effect on

Key Insight

The running variable must be **continuous** and pre-determined.

The Language of RD: Formal Notation

Treatment Indicator (Sharp RD)

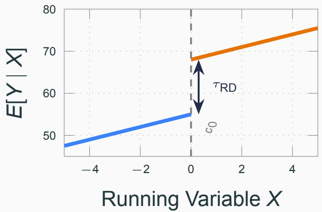
$$D_i = \mathbf{1}(X_i \geq c_0)$$

Equals 1 if the running variable crosses the cutoff, 0 otherwise.

The RD Estimand

$$\tau_{\mathrm{RD}} = \lim_{x \downarrow c_0} \mathbb{E}[Y|X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y|X{=}x]$$

The **jump** in the conditional expectation of Y exactly at the cutoff.



Examples of Running Variables and Cutoffs

Study	**Running Variable**	**Cutoff**	**Treatment**
School admissions	Test score	Passing grade	Admission
Medicare	Age (months)	65 years	Insurance
ESL program	WIDA score	District threshold	ESL placement
Vietnam draft	Lottery number	#195	Military service
DUI penalties	Blood alcohol	0.08%	Criminal charge
Head Start	County poverty rate	Federal cutoff	Program funds

Exercise 2: Identify the RD Components

Question

A scholarship program awards a \$10,000 grant to students who score $\geq$ 1200 on the SAT. You want to study whether receiving the scholarship increases four-year college graduation rates.

(a) What is the **running variable**?
(b) What is the **cutoff**?
(c) What is the **treatment**?
(d) What is the **outcome**?
(e) Is this a **sharp** or **fuzzy** RD? Why?

Exercise 2: Answers

Answer

(a) **Running variable:** SAT score (continuous, pre-determined).

(b) **Cutoff:** 1200 points.

(c) **Treatment:** Receiving the $10,000 scholarship.

(d) **Outcome:** Four-year college graduation rate (or a binary indicator for graduating).

(e) **Fuzzy RD** — not everyone who scores $\geq$ 1200 necessarily applies for or accepts the scholarship. The probability of treatment jumps at 1200, but likely not from 0 to 1. Use $\mathbf{1}(\text{SAT} \geq 1200)$ as an instrument for scholarship receipt.

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0
- No one above the cutoff goes untreated

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0
- No one above the cutoff goes untreated
- No one below the cutoff gets treated

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0
- No one above the cutoff goes untreated
- No one below the cutoff gets treated
- Estimate directly via OLS

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

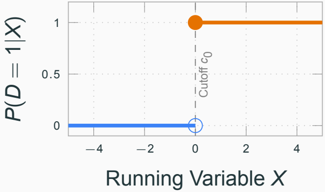
Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0
- No one above the cutoff goes untreated
- No one below the cutoff gets treated
- Estimate directly via OLS

Sharp RD: Perfect Compliance

Definition: Treatment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$



Key properties:

- $P(D = 1|X)$ jumps from 0 to 1 at c_0
- No one above the cutoff goes untreated
- No one below the cutoff gets treated
- Estimate directly via OLS

```
gen rv_c = rv - cutoff
gen D = (rv >= cutoff)

reg y D rv_c, robust
```

Fuzzy RD: Imperfect Compliance

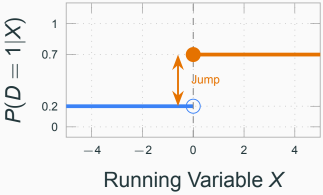
Definition: The **probability** of treatment jumps at c_0, but not from 0 to 1:

$$\lim_{x \downarrow c_0} P(D{=}1|X{=}x) \neq \lim_{x \uparrow c_0} P(D{=}1|X{=}x)$$

Fuzzy RD: Imperfect Compliance

Definition: The **probability** of treatment jumps at c_0, but not from 0 to 1:

$$\lim_{x \downarrow c_0} P(D=1|X=x) \neq \lim_{x \uparrow c_0} P(D=1|X=x)$$

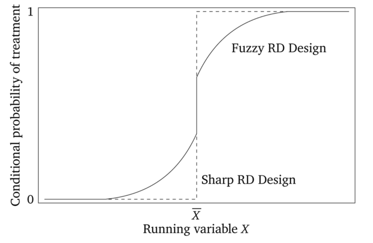


Analogy to IV: $\mathbf{1}(X \geq c_0)$ is the instrument.

Fuzzy RD estimate via 2SLS:

$$\hat{\tau}_{\text{Fuzzy}} = \frac{\text{Jump in } E[Y|X] \text{ at } c_0}{\text{Jump in } E[D|X] \text{ at } c_0}$$

Sharp vs. Fuzzy: Visual Comparison



Sharp: $P(D = 1|X)$ jumps $0 \to 1$. Estimate by OLS with indicator $D = \mathbf{1}(X \geq c_0)$.

Fuzzy: $P(D = 1|X)$ jumps by less than 1. Use $\mathbf{1}(X \geq c_0)$ as instrument in 2SLS. Identifies LATE.

The Research Question Determines the Design

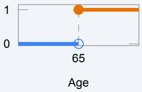
The same threshold can support a sharp or a fuzzy RD—it depends on what you define as "treatment," which depends on your **research question**.

The Research Question Determines the Design

The same threshold can support a sharp or a fuzzy RD—it depends on what you define as "treatment," which depends on your **research question**.

RQ 1: Does Medicare eligibility affect health?

Sharp RD. At age 65 everyone becomes eligible by law. $P(\text{eligible} = 1|X)$ jumps from 0 to 1.

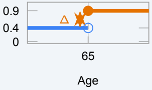


Treatment D_i: being eligible

Estimator: OLS

RQ 2: Does having Medicare affect health?

Fuzzy RD. Not everyone who turns 65 enrolls; some under-65s have private coverage. The cutoff shifts insurance probability.



Treatment D_i: actually insured

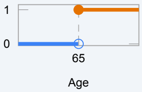
Estimator: 2SLS (IV)

The Research Question Determines the Design

The same threshold can support a sharp or a fuzzy RD—it depends on what you define as "treatment," which depends on your **research question**.

RQ 1: Does Medicare eligibility affect health?

Sharp RD. At age 65 everyone becomes eligible by law. $P(\text{eligible} = 1|X)$ jumps from 0 to 1.

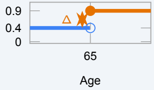


Treatment D_i: being eligible

Estimator: OLS

RQ 2: Does having Medicare affect health?

Fuzzy RD. Not everyone who turns 65 enrolls; some under-65s have private coverage. The cutoff shifts insurance probability.



Treatment D_i: actually insured

Estimator: 2SLS (IV)

Key Insight

Deciding sharp vs. fuzzy is about **what treatment you want to study**. Define your research question first; the design follows.

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score
- **Cutoff:** District-determined threshold

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score
- **Cutoff:** District-determined threshold
- **Treatment:** ESL program enrollment

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score
- **Cutoff:** District-determined threshold
- **Treatment:** ESL program enrollment

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score
- **Cutoff:** District-determined threshold
- **Treatment:** ESL program enrollment

Why fuzzy? Teachers and parents have some discretion. Not everyone below the cutoff gets placed; not everyone above escapes placement.

ESL Placement: A Real Fuzzy RD

Setting: WIDA language proficiency test determines ESL/newcomer placement.

- **Running variable:** WIDA composite score
- **Cutoff:** District-determined threshold
- **Treatment:** ESL program enrollment

Why fuzzy? Teachers and parents have some discretion. Not everyone below the cutoff gets placed; not everyone above escapes placement.

```
. reg yl T  x_c

      Source |       SS           df       MS      Number of obs   =       972
-------------+----------------------------------   F(2, 969)       =   3767.84
       Model |  30511.6419         2  15255.821    Prob > F        =    0.0000
    Residual |  3923.44067       969  4.04895838   R-squared       =    0.8861
-------------+----------------------------------   Adj R-squared   =    0.8858
       Total |  34435.0826       971  35.4635248   Root MSE        =    2.0122

------------------------------------------------------------------------------
          yl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           T |   4.276015   .2239165    19.10   0.000     3.836598    4.715432
         x_c |   2.172231   .0569372    38.15   0.000     2.060497    2.283966
       _cons |   44.45318   .1309122   339.56   0.000     44.19628    44.71008
------------------------------------------------------------------------------
```

Real-World Application

Estimating the causal effect of ESL programs on academic achievement—without comparing very different students.

Exercise 3: Sharp or Fuzzy?

Question

For each setting, a research question is given. Classify the design as **Sharp RD**, **Fuzzy RD**, or **Not an RD**. Explain why.

(a) *Research question:* Does **facing a DUI charge** reduce future drunk-driving incidents? Setting: a DUI charge is automatically filed whenever BAC is measured at $\geq 0.08\%$.

(b) *Research question:* Does **attending this selective university** raise long-run earnings? Setting: the university admits everyone scoring ≥ 1350 on the SAT, but some admitted students choose to enroll elsewhere.

(c) *Research question:* Does **participating in job training** increase employment? Setting: the program is available to households earning below $20,000/year, but a social worker has discretion to waive or grant the requirement.

Exercise 3: Answers

Answer

(a) **Sharp RD.** Treatment is facing a DUI charge. The charge is mechanically triggered by BAC $\geq 0.08\%$: every driver above the threshold is charged. $P(D = 1|X)$ jumps $0 \to 1$. Running variable: BAC; Cutoff: 0.08.

(b) **Fuzzy RD.** Treatment is enrolling at this university. The offer is sharp (everyone ≥ 1350 receives it), but enrollment is fuzzy—some admitted students choose elsewhere. Use the offer as an instrument for enrollment. Running variable: SAT score; Cutoff: 1350.

(c) **Fuzzy RD.** Treatment is actual program participation. Social-worker discretion means crossing the $20,000 threshold shifts but does not fully determine participation. Use the income cutoff as an instrument. Running variable: household income; Cutoff: $20,000.

Part 2 — continued

How to Apply It

Regression equations, Stata implementation,

non-linearities, and optimal bandwidth

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator
- $(X_i - c_0)$ — running variable, **centered at cutoff**

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator
- $(X_i - c_0)$ — running variable, **centered at cutoff**
- δ — the **RD estimate**: the jump in Y at c_0

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator
- $(X_i - c_0)$ — running variable, **centered at cutoff**
- δ — the **RD estimate**: the jump in Y at c_0
- β_1 — the slope of the running variable

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator
- $(X_i - c_0)$ — running variable, **centered at cutoff**
- δ — the **RD estimate**: the jump in Y at c_0
- β_1 — the slope of the running variable

The Basic RD Regression: The Equation

Estimating equation:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \varepsilon_i$$

What each term does:

- $D_i = \mathbf{1}(X_i \geq c_0)$ — treatment indicator
- $(X_i - c_0)$ — running variable, **centered at cutoff**
- δ — the **RD estimate**: the jump in Y at c_0
- β_1 — the slope of the running variable

Key Insight

δ captures the **discontinuous jump** at the cutoff.

β_1 captures the continuous relationship between X and Y on each side.

The Basic RD Regression: In Stata

Step-by-step implementation

```
* Step 1: Center the running variable at the cutoff
gen rv_c = rv - cutoff

* Step 2: Create treatment indicator
gen D = (rv >= cutoff)

* Step 3: Run the RD regression
reg y D rv_c, robust

* Step 4: Read the treatment effect
display _b[D]
```

The Basic RD Regression: In Stata

Step-by-step implementation

```
* Step 1: Center the running variable at the cutoff
gen rv_c = rv - cutoff

* Step 2: Create treatment indicator
gen D = (rv >= cutoff)

* Step 3: Run the RD regression
reg y D rv_c, robust

* Step 4: Read the treatment effect
display _b[D]
```

Warning

Always **center** the running variable. This makes δ directly interpretable as the jump at c_0. Without centering, β_0 will not equal the outcome at the cutoff.

Allowing Different Slopes on Each Side

Basic model: Same slope β_1 on both sides.

More flexible specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \beta_2 D_i (X_i - c_0) + \varepsilon_i$$

- β_1 = slope **below** cutoff (control side)

Allowing Different Slopes on Each Side

Basic model: Same slope β_1 on both sides.

More flexible specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1(X_i - c_0) + \beta_2 D_i(X_i - c_0) + \varepsilon_i$$

- β_1 = slope **below** cutoff (control side)
- $\beta_1 + \beta_2$ = slope **above** cutoff (treated side)

Allowing Different Slopes on Each Side

Basic model: Same slope β_1 on both sides.

More flexible specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \beta_2 D_i (X_i - c_0) + \varepsilon_i$$

- β_1 = slope **below** cutoff (control side)
- $\beta_1 + \beta_2$ = slope **above** cutoff (treated side)
- δ still captures the jump at c_0

Allowing Different Slopes on Each Side

Basic model: Same slope β_1 on both sides.

More flexible specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \beta_2 D_i (X_i - c_0) + \varepsilon_i$$

- β_1 = slope **below** cutoff (control side)
- $\beta_1 + \beta_2$ = slope **above** cutoff (treated side)
- δ still captures the jump at c_0

Allowing Different Slopes on Each Side

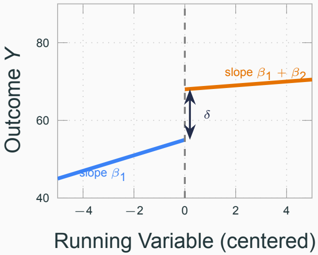
Basic model: Same slope β_1 on both sides.

More flexible specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1(X_i - c_0) + \beta_2 D_i(X_i - c_0) + \varepsilon_i$$

- β_1 = slope **below** cutoff (control side)
- $\beta_1 + \beta_2$ = slope **above** cutoff (treated side)
- δ still captures the jump at c_0

```
gen Dx = D * rv_c
reg y D rv_c Dx, robust
```



Adding Controls to an RD

Extended specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \mathbf{X}_i' \gamma + \varepsilon_i$$

where $\mathbf{X}_i$ are pre-determined covariates.

Why controls help:

- Identification comes from the cutoff, not from controls

Adding Controls to an RD

Extended specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \mathbf{X}_i' \gamma + \varepsilon_i$$

where $\mathbf{X}_i$ are pre-determined covariates.

Why controls help:

- Identification comes from the cutoff, not from controls
- Controls reduce residual variance $\Rightarrow$ smaller SEs

Adding Controls to an RD

Extended specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \mathbf{X}_i' \gamma + \varepsilon_i$$

where $\mathbf{X}_i$ are pre-determined covariates.

Why controls help:

- Identification comes from the cutoff, not from controls
- Controls reduce residual variance $\Rightarrow$ smaller SEs
- Do **not** change the causal interpretation if continuity holds

Adding Controls to an RD

Extended specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \mathbf{X}_i' \gamma + \varepsilon_i$$

where $\mathbf{X}_i$ are pre-determined covariates.

Why controls help:

- Identification comes from the cutoff, not from controls
- Controls reduce residual variance $\Rightarrow$ smaller SEs
- Do **not** change the causal interpretation if continuity holds

Adding Controls to an RD

Extended specification:

$$Y_i = \beta_0 + \delta D_i + \beta_1 (X_i - c_0) + \mathbf{X}_i' \gamma + \varepsilon_i$$

where $\mathbf{X}_i$ are pre-determined covariates.

Why controls help:

- Identification comes from the cutoff, not from controls
- Controls reduce residual variance $\Rightarrow$ smaller SEs
- Do **not** change the causal interpretation if continuity holds

Warning

Do **not** include controls that could be affected by the treatment ("bad controls").

Only include **pre-determined** characteristics: baseline scores, demographics, family income before the policy.

Exercise 5: Interpret This RD Output

Question

You run an RD regression studying the effect of a tutoring program (cutoff: score = 50) on final exam performance (0–100).

Variable	Coef.	SE	p-value
D (treatment)	4.8	1.2	0.000
rv_c	0.22	0.05	0.000
Constant	51.3	0.9	0.000
$N = 1{,}240$	Bandwidth: $\pm$10 points		Robust SEs

(a) How much does exam performance jump at the cutoff?

(b) Interpret the coefficient on rv_c.

(c) What does the constant represent?

(d) Should you add controls? If so, what kind?

Exercise 5: Answers

Answer

(a) **4.8 points.** Students just above the cutoff (who receive tutoring) score 4.8 points higher on the final exam than students just below.

(b) **0.22:** For each additional point in the entrance score (within the bandwidth), the final exam score increases by 0.22 points. This is the continuous slope of the running variable.

(c) **51.3:** The predicted final exam score for a student with $rv\_c = 0$ and $D = 0$ — i.e., a student right at the cutoff who did not receive the tutoring program.

(d) **Yes, pre-determined controls** (e.g., gender, prior GPA, school fixed effects). They reduce residual variance and tighten standard errors. Do not control for any outcome measured after tutoring began.

The Non-Linearity Problem

Problem: If the true relationship is curved, a **linear model** can create a **phantom jump** at the cutoff.

Example DGP (no true effect):

$$Y = 1000 - 100X + X^2 + \varepsilon$$

The Non-Linearity Problem

Problem: If the true relationship is curved, a **linear model** can create a **phantom jump** at the cutoff.

Example DGP (no true effect):

$$Y = 1000 - 100X + X^2 + \varepsilon$$

Treatment effect $= 0$ by construction — yet a linear fit sees the curve as a jump!

The Non-Linearity Problem

Problem: If the true relationship is curved, a **linear model** can create a **phantom jump** at the cutoff.

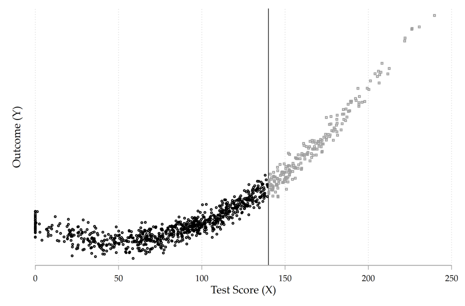
Example DGP (no true effect):

$$Y = 1000 - 100X + X^2 + \varepsilon$$

Treatment effect $= 0$ by construction — yet a linear fit sees the curve as a jump!

Warning

Always inspect the scatter plot before running any regression. A curved relationship is visible to the eye but will fool a linear model.



The Fix: Polynomial Controls

Solution: Add higher-order terms in $(X - c_0)$:

$$Y_i = \beta_0 + \delta D_i + \sum_{p=1}^{P} \beta_p (X_i - c_0)^p + \varepsilon_i$$

Allow **different polynomials on each side:**

$$+ \sum_{p=1}^{P} \gamma_p D_i \cdot (X_i - c_0)^p$$

The Fix: Polynomial Controls

Solution: Add higher-order terms in $(X - c_0)$:

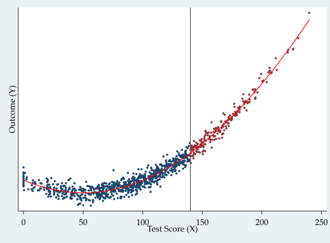
$$Y_i = \beta_0 + \delta D_i + \sum_{p=1}^{P} \beta_p (X_i - c_0)^p + \varepsilon_i$$

Allow **different polynomials on each side:**

$$+ \sum_{p=1}^{P} \gamma_p D_i \cdot (X_i - c_0)^p$$

In Stata:

```
gen rv2 = rv_c^2
gen Drv2 = D*rv2
reg y D rv_c rv2 Dx Drv2, robust
```



Quadratic fit correctly captures the curve. The “jump” from the linear model disappears.

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)
- Too **many** terms: overfit near the boundary, inflated standard errors

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)
- Too **many** terms: overfit near the boundary, inflated standard errors

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)
- Too **many** terms: overfit near the boundary, inflated standard errors

Practical recommendations:

- Start with **linear**, show robustness to quadratic

```
. reg y D x x2 x3

      Source |       SS           df       MS      Number of obs   =     1,000
-------------+----------------------------------   F(4, 995)       =   7798.40
       Model |  3.2207e+10         4  8.0518e+09   Prob > F        =    0.0000
    Residual |  1.0273e+09       995  1032494.42   R-squared       =    0.9691
-------------+----------------------------------   Adj R-squared   =    0.9690
       Total |  3.3235e+10       999  33267799.9   Root MSE        =    1016.1

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   44.96491   147.8125     0.30   0.761    -245.0952     335.025
           x |  -106.7774   5.371053   -19.88   0.000    -117.3173   -96.23752
          x2 |   1.083471   .0586708    18.47   0.000     .9683381    1.198604
          x3 |  -.0002625   .0001779    -1.48   0.141    -.0006117    .0000867
       _cons |   10081.61   139.5867    72.22   0.000     9807.687    10355.52
------------------------------------------------------------------------------
```

Cubic vs. local linear. Both are valid robustness checks.

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)
- Too **many** terms: overfit near the boundary, inflated standard errors

Practical recommendations:

- Start with **linear**, show robustness to quadratic
- Prefer **local linear regression** (narrow bandwidth) over global high-degree polynomials

```
. reg y D x x2 x3

      Source |       SS           df       MS      Number of obs   =     1,000
-------------+----------------------------------   F(4, 995)       =   7798.40
       Model |  3.2207e+10         4  8.0518e+09   Prob > F        =    0.0000
    Residual |  1.0273e+09       995  1032494.42   R-squared       =    0.9691
-------------+----------------------------------   Adj R-squared   =    0.9690
       Total |  3.3235e+10       999  33267799.9   Root MSE        =    1016.1

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   44.96491   147.8125     0.30   0.761    -245.0952     335.025
           x |  -106.7774   5.371053   -19.88   0.000    -117.3173   -96.23752
          x2 |   1.083471   .0586708    18.47   0.000     .9683381    1.198604
          x3 |  -.0002625   .0001779    -1.48   0.141    -.0006117    .0000867
       _cons |   10081.61   139.5867    72.22   0.000     9807.687    10355.52
------------------------------------------------------------------------------
```

Cubic vs. local linear. Both are valid robustness checks.

How Many Polynomial Terms?

Balancing act:

- Too **few** terms: risk fitting a curve as a jump (non-linearity bias)
- Too **many** terms: overfit near the boundary, inflated standard errors

Practical recommendations:

- Start with **linear**, show robustness to quadratic
- Prefer **local linear regression** (narrow bandwidth) over global high-degree polynomials
- Use `rdrobust` for optimal bandwidth selection (next section!)

```
. reg y D x x2 x3

      Source |       SS           df       MS      Number of obs   =     1,000
-------------+----------------------------------   F(4, 995)       =   7798.40
       Model |  3.2207e+10         4  8.0518e+09   Prob > F        =    0.0000
    Residual |  1.0273e+09       995  1032494.42   R-squared       =    0.9691
-------------+----------------------------------   Adj R-squared   =    0.9690
       Total |  3.3235e+10       999  33267799.9   Root MSE        =    1016.1

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   44.96491   147.8125     0.30   0.761    -245.0952     335.025
           x |  -106.7774   5.371053   -19.88   0.000    -117.3173   -96.23752
          x2 |   1.083471   .0586708    18.47   0.000     .9683381    1.198604
          x3 |  -.0002625   .0001779    -1.48   0.141    -.0006117    .0000867
       _cons |   10081.61   139.5867    72.22   0.000     9807.687    10355.52
------------------------------------------------------------------------------
```

Cubic vs. local linear. Both are valid robustness checks.

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp RD OLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Treatment indicator: $D_i = \mathbf{1}[X_i \geq c_0]$

Fuzzy RD 2SLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Eligibility instrument: $Z_i = \mathbf{1}[X_i \geq c_0]$

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp RD OLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Treatment indicator: $D_i = \mathbf{1}[X_i \geq c_0]$
2. Single OLS regression:

$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$

Fuzzy RD 2SLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Eligibility instrument: $Z_i = \mathbf{1}[X_i \geq c_0]$
2. **First stage** — compliance jump:

$$D_i = \alpha_0 + \pi Z_i + \alpha_1 \widetilde{X}_i + u_i$$

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp RD OLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Treatment indicator: $D_i = \mathbf{1}[X_i \geq c_0]$
2. Single OLS regression:
$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$
3. $\hat{\delta}$ = jump at c_0 = ATE for cutoff units

Fuzzy RD 2SLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Eligibility instrument: $Z_i = \mathbf{1}[X_i \geq c_0]$
2. **First stage** — compliance jump:
$$D_i = \alpha_0 + \pi Z_i + \alpha_1 \widetilde{X}_i + u_i$$
3. **Second stage** — instrument D_i with Z_i:
$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp RD (OLS)

1. Center: $\widetilde{X}_i = X_i - c_0$
   Treatment indicator: $D_i = \mathbf{1}[X_i \geq c_0]$
2. Single OLS regression:
$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$
3. $\hat{\delta}$ = jump at c_0 = ATE for cutoff units
4. **Stata:**

```
gen rv_c = rv - c
gen D = (rv >= c)
reg y D rv_c, robust
```

Fuzzy RD (2SLS)

1. Center: $\widetilde{X}_i = X_i - c_0$
   Eligibility instrument: $Z_i = \mathbf{1}[X_i \geq c_0]$
2. **First stage** — compliance jump:
$$D_i = \alpha_0 + \pi Z_i + \alpha_1 \widetilde{X}_i + u_i$$
3. **Second stage** — instrument D_i with Z_i:
$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$
4. $\hat{\delta}$ = LATE (compliers at c_0)

   Stata:

```
gen rv_c = rv - c; gen Z = (rv >= c)
ivregress 2sls y rv_c (D = Z), robust
```

Sharp vs. Fuzzy RD: How the Process Differs

Both designs exploit the same cutoff—but the **estimation strategy** differs because the treatment variable changes.

Sharp RD — OLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Treatment indicator: $D_i = \mathbf{1}[X_i \geq c_0]$
2. Single OLS regression:

$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$

3. $\hat{\delta}$ = jump at c_0 = ATE for cutoff units
4. **Stata:**

```
gen rv_c = rv - c
gen D = (rv >= c)
reg y D rv_c, robust
```

Fuzzy RD — 2SLS

1. Center: $\widetilde{X}_i = X_i - c_0$
   Eligibility instrument: $Z_i = \mathbf{1}[X_i \geq c_0]$
2. **First stage** — compliance jump:

$$D_i = \alpha_0 + \pi Z_i + \alpha_1 \widetilde{X}_i + u_i$$

3. **Second stage** — instrument D_i with Z_i:

$$Y_i = \beta_0 + \delta D_i + \beta_1 \widetilde{X}_i + \varepsilon_i$$

4. $\hat{\delta}$ = LATE (compliers at c_0)

   Stata:

```
gen rv_c = rv - c; gen Z = (rv >= c)
ivregress 2sls y rv_c (D = Z), robust
```

Key Insight

Sharp RD = **OLS** with a jump dummy. Fuzzy RD = **2SLS** using the threshold as an instrument.

Part 2 — continued

Choosing a Bandwidth

Bias–variance tradeoff, MSE-optimal selection,

and the CCT approach

Bandwidth Shapes the Jump Estimate

If the true relationship is curved, a wider bandwidth forces a linear fit onto a curve—the extrapolated value at c_0 is wrong, biasing $\hat{\delta}$.

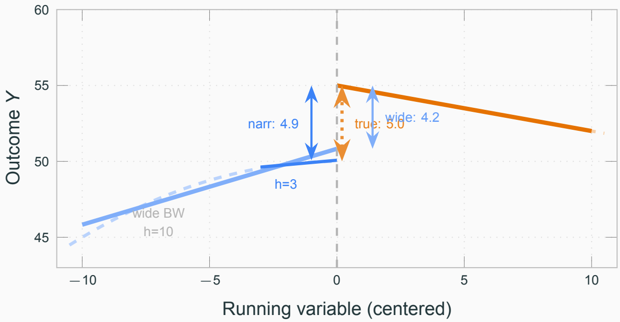
Bandwidth Shapes the Jump Estimate

If the true relationship is curved, a wider bandwidth forces a linear fit onto a curve—the extrapolated value at c_0 is wrong, biasing $\hat{\delta}$.

Numerical example.

True effect = 5 pts. Left side concave.

h	$\hat{\delta}$	Bias
3	4.9	≈ 0
6	4.7	0.3
10	4.2	0.8



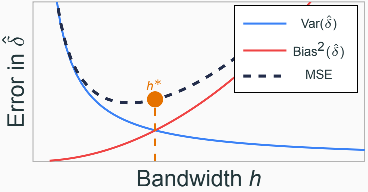
Dashed: true relationship. Solid: OLS fits. Wide-BW left line overshoots, shrinking $\hat{\delta}$.

Why Bandwidth Matters

Bandwidth h defines how far from the cutoff to include observations when fitting each side.

The Fundamental Trade-off

- **Narrow** h: near the cutoff, linearity is credible (low bias). But few observations $\Rightarrow \hat{\delta}$ **is noisy** (high variance)

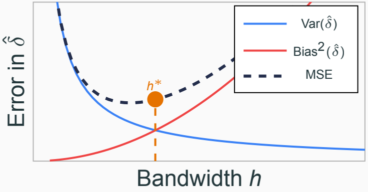


Why Bandwidth Matters

Bandwidth h defines how far from the cutoff to include observations when fitting each side.

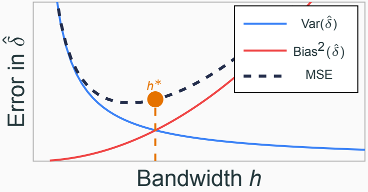
The Fundamental Trade-off

- **Narrow** h: near the cutoff, linearity is credible (low bias). But few observations $\Rightarrow \hat{\delta}$ **is noisy** (high variance)
- **Wide** h: more observations $\Rightarrow \hat{\delta}$ **is more precise** (lower variance). But units far from c_0 may violate local linearity $\Rightarrow$ biased $\hat{\delta}$



Why Bandwidth Matters

Bandwidth h defines how far from the cutoff to include observations when fitting each side.

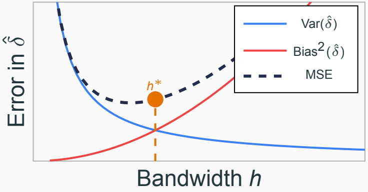


The Fundamental Trade-off

- **Narrow** h: near the cutoff, linearity is credible (low bias). But few observations $\Rightarrow$ $\hat{\delta}$ **is noisy** (high variance)
- **Wide** h: more observations $\Rightarrow$ $\hat{\delta}$ **is more precise** (lower variance). But units far from c_0 may violate local linearity $\Rightarrow$ biased $\hat{\delta}$

Why Bandwidth Matters

Bandwidth h defines how far from the cutoff to include observations when fitting each side.



The Fundamental Trade-off

- **Narrow** h: near the cutoff, linearity is credible (low bias). But few observations $\Rightarrow$ $\hat{\delta}$ **is noisy** (high variance)
- **Wide** h: more observations $\Rightarrow$ $\hat{\delta}$ **is more precise** (lower variance). But units far from c_0 may violate local linearity $\Rightarrow$ biased $\hat{\delta}$

Key Insight

Choosing h is a **bias–variance tradeoff for** $\hat{\delta}$ specifically. Neither extreme is right.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*
- Also provides **robust bias-corrected** (RBC) standard errors

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*
- Also provides **robust bias-corrected** (RBC) standard errors
- Different from conventional SEs — RBC SEs are wider and more honest

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*
- Also provides **robust bias-corrected** (RBC) standard errors
- Different from conventional SEs — RBC SEs are wider and more honest
- Implemented in Stata via `rdrobust`

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*
- Also provides **robust bias-corrected** (RBC) standard errors
- Different from conventional SEs — RBC SEs are wider and more honest
- Implemented in Stata via `rdrobust`

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

MSE-Optimal Bandwidth: The CCT Approach

Calonico, Cattaneo, and Titiunik (2014) propose selecting h to minimize **Mean Squared Error**:

$$h^* = \arg\min_h \underbrace{\text{Bias}^2(h)}_{\text{more with wide } h} + \underbrace{\text{Variance}(h)}_{\text{more with narrow } h}$$

Key features of the CCT approach:

- Uses a **data-driven** formula for h^*
- Also provides **robust bias-corrected** (RBC) standard errors
- Different from conventional SEs — RBC SEs are wider and more honest
- Implemented in Stata via `rdrobust`

Intuition

The formula estimates:

- How fast the running variable–outcome relationship curves (drives bias)
- The conditional variance of the outcome (drives variance)

It then solves for the h that makes these two forces equal at the margin.

Real-World Application

Two bandwidths: CCT also suggests reporting results for h^* and $2h^*$ to show robustness.

Implementing Optimal Bandwidth in Stata

Install and use `rdrobust`

```
* Install the package (once)
capture ssc install rdrobust

* Data-driven bandwidth selection
rdbwselect y rv, c(cutoff)

* Run RD with optimal BW + robust bias-corrected SEs
rdrobust y rv, c(cutoff)

* Plot: data-driven bins on each side
rdplot y rv, c(cutoff)

* Report robustness: use 0.5*h* and 2*h*
rdrobust y rv, c(cutoff) h(0.5)
rdrobust y rv, c(cutoff) h(2)
```

Implementing Optimal Bandwidth in Stata

Install and use `rdrobust`

```
* Install the package (once)
capture ssc install rdrobust

* Data-driven bandwidth selection
rdbwselect y rv, c(cutoff)

* Run RD with optimal BW + robust bias-corrected SEs
rdrobust y rv, c(cutoff)

* Plot: data-driven bins on each side
rdplot y rv, c(cutoff)

* Report robustness: use 0.5*h* and 2*h*
rdrobust y rv, c(cutoff) h(0.5)
rdrobust y rv, c(cutoff) h(2)
```

Key Insight

`rdrobust` is the modern standard. Always report the **robust bias-corrected (RBC)** estimate and check that results are stable across a range of bandwidths.

Exercise 6: The Bandwidth Tradeoff

Question

You study the effect of financial aid (triggered at income $<$ \$30,000) on college enrollment. The CCT optimal bandwidth is $h^* = \$2{,}100$.

(a) What is the advantage of using a narrow bandwidth of \$500? What is the risk?

(b) What is the advantage of using a wide bandwidth of \$10,000? What is the risk?

(c) The `rdrobust` output shows $h^* = \$2{,}100$. What does this tell you?

(d) Your estimate changes from 8 pp to 3 pp when you widen from $h^* = \$2{,}100$ to $h = \$6{,}000$. Is this reassuring or concerning?

Exercise 6: Answers

Answer

(a) **Narrow ($500):** Units are very comparable across the threshold — strong internal validity. Risk: very few observations, large standard errors, noisy estimates.

(b) **Wide ($10,000):** More observations, more statistical power. Risk: families at $20,000 and $29,000 income may differ in many ways; local linearity assumption becomes strained, introducing bias.

(c) $h^* = \$2{,}100$ is the MSE-minimizing bandwidth. It balances the trade-off between bias and variance. Use it as the main specification, and report sensitivity to $0.5h^*$ and $2h^*$.

(d) **Concerning.** A large change in the estimate when widening the bandwidth suggests the relationship is not locally linear over the wider range. The functional form assumption may be violated. Report robustness checks and potentially use a more flexible polynomial.

Part 3: Assumptions

The ONE Key Assumption

Formal statement:

$$\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] = \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]$$

and symmetrically for $Y(1)$ above the cutoff.

Plain language:

The ONE Key Assumption

Formal statement:

$$\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] = \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]$$

and symmetrically for $Y(1)$ above the cutoff.

Plain language:

Without the policy, outcomes would have evolved smoothly through the threshold. Any observed jump in Y at c_0 is therefore caused by the treatment.

The ONE Key Assumption

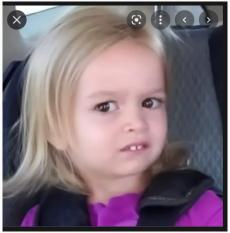
Formal statement:

$$\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] = \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]$$

and symmetrically for $Y(1)$ above the cutoff.

Plain language:

Without the policy, outcomes would have evolved smoothly through the threshold. Any observed jump in Y at c_0 is therefore caused by the treatment.



$\mathbb{E}[Y(0) \mid X]$ is continuous; the jump at c_0 is the treatment effect.

Key Insight

Continuity is **not directly testable**, but we can gather indirect evidence supporting it.

▸ Appendix: Assumptions in Potential Outcomes (Sharp & Fuzzy)

Implication 1: Valid Counterfactuals

What Continuity Gives Us

Units **just below** the cutoff are valid counterfactuals for units **just above**.

They are identical in expectation in all observed and unobserved characteristics. The only thing that differs is which side of c_0 they landed on.

Implication 1: Valid Counterfactuals

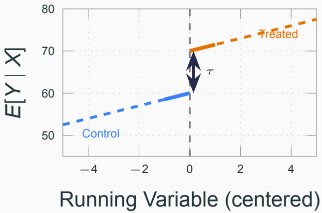
What Continuity Gives Us

Units **just below** the cutoff are valid counterfactuals for units **just above**.

They are identical in expectation in all observed and unobserved characteristics. The only thing that differs is which side of c_0 they landed on.

Key Insight

This is the same logic as a local experiment. If potential outcomes are continuous, then near the cutoff we have something close to random assignment.



Implication 2: Nothing Else Should Jump

The "No Other Jumps" Requirement

No other variable affecting Y should jump at c_0.

If a confounder also jumps at the threshold, we cannot distinguish its effect from the treatment effect.

Implication 2: Nothing Else Should Jump

The "No Other Jumps" Requirement

No other variable affecting Y should jump at c_0.

If a confounder also jumps at the threshold, we cannot distinguish its effect from the treatment effect.

Warning

Assumption violations occur when:

- Units **manipulate** the running variable
- **Other policies** switch at the same threshold
- Units **select** into the running variable on either side

Implication 2: Nothing Else Should Jump

The "No Other Jumps" Requirement

No other variable affecting Y should jump at c_0.

If a confounder also jumps at the threshold, we cannot distinguish its effect from the treatment effect.

Warning

Assumption violations occur when:

- Units **manipulate** the running variable
- **Other policies** switch at the same threshold
- Units **select** into the running variable on either side

Real-World Application

Medicare example:
Both Medicare and Social Security eligibility change around age 65. If we only study Medicare, we may attribute the combined effect of both programs to Medicare alone.

This is called **compound treatment**: two treatments changing at the same threshold.

Exercise 4: Does the Continuity Assumption Hold?

Question

For each scenario, assess whether the continuity assumption is plausible. Identify any specific threat.

(a) You study the effect of **Head Start** (federal preschool) using a county poverty rate cutoff. Counties with poverty rate $\geq 40\%$ received funding first.

(b) You study the effect of **traffic cameras** on future accidents. Cameras were installed when a road's accident rate exceeded a threshold.

Exercise 4: Answers

Answer

(a) **Plausible, but check for compound treatment.** Counties near the poverty cutoff may differ along many dimensions (tax base, local governance, other federal programs also triggered by poverty thresholds). Run RD with covariates as outcomes to check for jumps.

(b) **Likely violated — regression to the mean.** Roads that just crossed the accident-rate threshold were probably experiencing an unusually bad period. Their accident rate would have fallen back toward the mean even without cameras. The "treatment effect" partially reflects reversion to trend, not the camera. This is a **mean-reversion bias** common when thresholds are based on noisy outcomes.

Part 4: Providing Evidence for Assumptions

Validity Tests 1 & 2: Graphical Evidence and Covariate Balance

Test 1: Graphical Evidence

Plot outcome vs. running variable. A **visible jump** at c_0 is the primary evidence.

- `binscatter y rv, rd(c)`
- `twoway (scatter y rv) (lfit y rv if D==0 || D==1)`

Validity Tests 1 & 2: Graphical Evidence and Covariate Balance

Test 1: Graphical Evidence

Plot outcome vs. running variable. A **visible jump** at c_0 is the primary evidence.

- `binscatter y rv, rd(c)`
- `twoway (scatter y rv) (lfit y rv if D==0 || D==1)`

Key Insight

In RD, **the graph is the regression**. If the visual is not convincing, no regression will save the paper.

Validity Tests 1 & 2: Graphical Evidence and Covariate Balance

Test 1: Graphical Evidence

Plot outcome vs. running variable. A **visible jump** at c_0 is the primary evidence.

- `binscatter y rv, rd(c)`
- `twoway (scatter y rv) (lfit y rv if D==0 || D==1)`

Key Insight

In RD, **the graph is the regression**. If the visual is not convincing, no regression will save the paper.

Test 2: Covariate Balance

Run the same RD with pre-determined covariates as the outcome. They should **not jump**.

- `reg age D rv_c, robust`
- `reg female D rv_c, robust`
- `reg baseline_score D rv_c, robust`

If a covariate jumps at the cutoff, continuity may be violated.

Validity Tests 3 & 4: No Other Jumps and Density Test

Test 3: No Other Policy Jumps

Does any **other policy** change at the same threshold?

If so, we cannot attribute the effect to our treatment alone. Search for **compound treatments** carefully.

Validity Tests 3 & 4: No Other Jumps and Density Test

Test 3: No Other Policy Jumps

Does any **other policy** change at the same threshold?

If so, we cannot attribute the effect to our treatment alone. Search for **compound treatments** carefully.

Test 4: McCrary Density Test

Do units **manipulate** the running variable to cross the cutoff?

A **spike** in the density just above c_0 is a red flag.

- `histogram rv, xline(c)`
- `rddensity rv, c(c) plot`

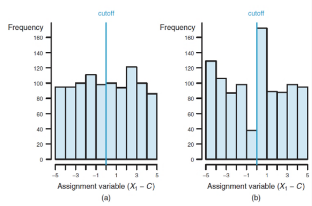
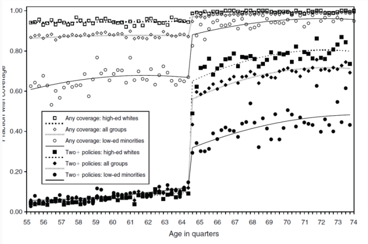


FIGURE 11.10: Histograms of Assignment Variable for RD Analysis

McCrary (2008) density test: look for a spike just above the cutoff. A smooth density supports the no-manipulation assumption.

Other Considerations

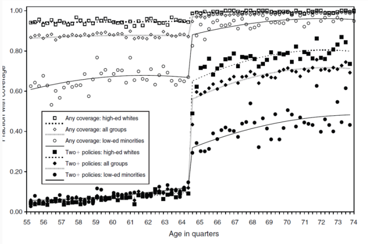
Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65

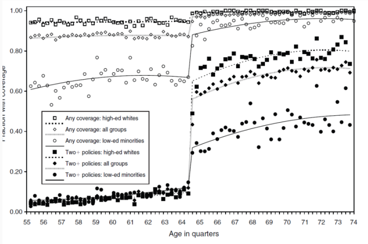
Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65
- Medicare provides publicly subsidized health insurance

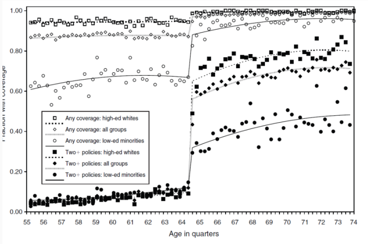
Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65
- Medicare provides publicly subsidized health insurance
- Before 65, most workers are covered through employers

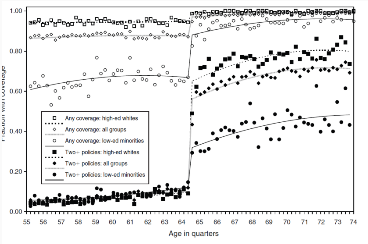
Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65
- Medicare provides publicly subsidized health insurance
- Before 65, most workers are covered through employers
- At 65, a sharp increase in insurance coverage occurs

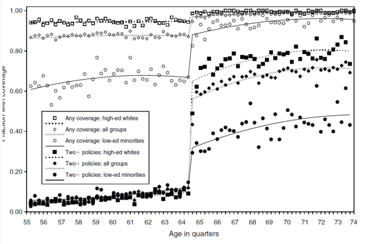
Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65
- Medicare provides publicly subsidized health insurance
- Before 65, most workers are covered through employers
- At 65, a sharp increase in insurance coverage occurs

Medicare Eligibility at Age 65



Background:

- Most Americans become eligible for **Medicare** at exactly age 65
- Medicare provides publicly subsidized health insurance
- Before 65, most workers are covered through employers
- At 65, a sharp increase in insurance coverage occurs

Key paper: Card, Dobkin, and Maestas (2008). *The Impact of Nearly Universal Insurance Coverage on Health Care Utilization.*

Medicare at 65: Question

Question

Identify the key elements of an RD design for the Medicare study:

(a) What is the **running variable**?
(b) What is the **treatment**?
(c) What is the **cutoff**?
(d) Is this a **sharp** or **fuzzy** RD? Explain.
(e) What is the key **identifying assumption**? Is it plausible?
(f) What **other things** change at age 65 that could confound the estimate?

Medicare at 65: Answers

Answer

(a) **Running variable:** Age (in months), centered at 65 years.

(b) **Treatment:** Medicare insurance coverage.

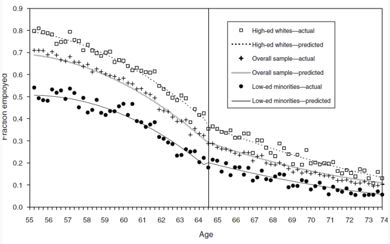
(c) **Cutoff:** Exactly age 65 (= 780 months of age).

(d) **Fuzzy RD:** Not all newly eligible individuals immediately enroll. Insurance coverage jumps but not from 0 to 100%.

(e) **Identifying assumption:** Health outcomes would have evolved smoothly through age 65 absent Medicare. Plausible since age is a continuous biological variable.

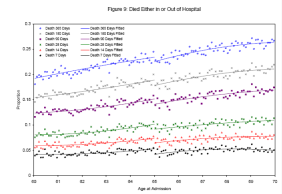
(f) **Confounders at 65:** Social Security eligibility, retirement incentives, Medicare Part B, behavioral changes from retirement.

Medicare at 65: What Does the Data Show?

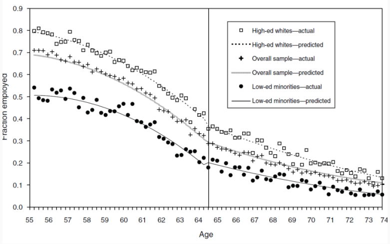


Key findings:

- Insurance coverage **jumps sharply** at 65

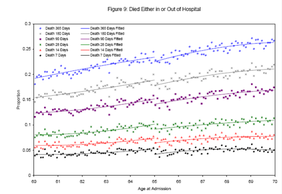


Medicare at 65: What Does the Data Show?

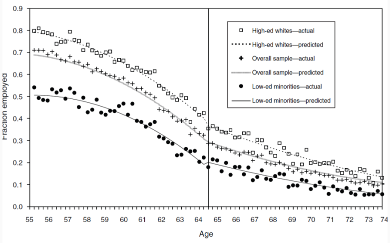


Key findings:

- Insurance coverage **jumps sharply** at 65
- Doctor visits and hospitalizations increase

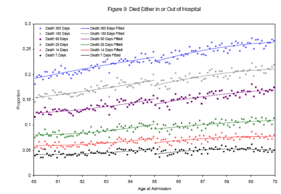


Medicare at 65: What Does the Data Show?



Key findings:

- Insurance coverage **jumps sharply** at 65
- Doctor visits and hospitalizations increase
- Mortality **decreases** slightly at 65



Internal Validity: What RD Gets Right

RD has very high internal validity:

Near the cutoff, units on both sides are **essentially identical** in all observed <u>and unobserved</u> characteristics.

This means:

- No omitted variable bias

Internal Validity: What RD Gets Right

RD has very high internal validity:

Near the cutoff, units on both sides are **essentially identical** in all observed and unobserved characteristics.

This means:

- No omitted variable bias
- No selection bias (near the cutoff)

Internal Validity: What RD Gets Right

RD has very high internal validity:

Near the cutoff, units on both sides are **essentially identical** in all observed and unobserved characteristics.

This means:

- No omitted variable bias
- No selection bias (near the cutoff)
- The estimate is **as credible as an RCT**—but only locally

Internal Validity: What RD Gets Right

RD has very high internal validity:

Near the cutoff, units on both sides are **essentially identical** in all observed and unobserved characteristics.

This means:

- No omitted variable bias
- No selection bias (near the cutoff)
- The estimate is **as credible as an RCT**—but only locally

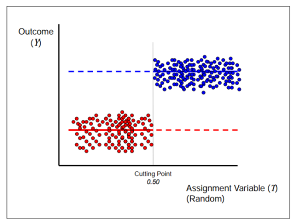
Internal Validity: What RD Gets Right

RD has very high internal validity:

Near the cutoff, units on both sides are **essentially identical** in all observed and unobserved characteristics.

This means:

- No omitted variable bias
- No selection bias (near the cutoff)
- The estimate is **as credible as an RCT**—but only locally



Angrist (1990): draft lottery as a natural experiment with a sharp discontinuity at the draft number cutoff.

Real-World Application

The Vietnam draft lottery: birth dates were randomly assigned draft numbers, creating a sharp threshold at lottery number 195. This provides clean causal identification.

External Validity: The Limitation of RD

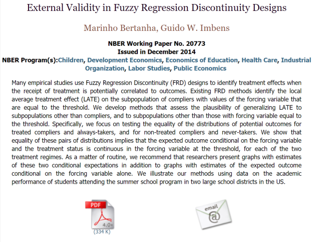
What RD estimates:

$$\tau_{RD} = \mathbb{E}[Y^1 - Y^0 \mid X = c_0]$$

A **Local Average Treatment Effect (LATE)**—the average effect only for units at the cutoff.

Key limitation: Cannot extrapolate to units far from the cutoff.

- Effect for students scoring 90 may differ from those scoring 70 (near the cutoff)


External Validity in Fuzzy Regression Discontinuity Designs

Marinho Bertanha, Guido W. Imbens

NBER Working Paper No. 20773
Issued in December 2014
NBER Program(s): Children, Development Economics, Economics of Education, Health Care, Industrial Organization, Labor Studies, Public Economics

Many empirical studies use Fuzzy Regression Discontinuity (FRD) designs to identify treatment effects when the receipt of treatment is potentially correlated to outcomes. Existing FRD methods identify the local average treatment effect (LATE) on the subpopulation of compliers with values of the forcing variable that are equal to the threshold. We develop methods that assess the plausibility of generalizing LATE to subpopulations other than compliers, and to subpopulations other than those with forcing variable equal to the threshold. Specifically, we focus on testing the equality of the distributions of potential outcomes for treated compliers and always-takers, and for non-treated compliers and never-takers. We show that equality of these pairs of distributions implies that the expected outcome conditional on the forcing variable and the treatment status is continuous in the forcing variable at the threshold, for each of the two treatment regimes. As a matter of routine, we recommend that researchers present graphs with estimates of these two conditional expectations in addition to graphs with estimates of the expected outcome conditional on the forcing variable alone. We illustrate our methods using data on the academic performance of students attending the summer school program in two large school districts in the US.

We estimate the effect at the cutoff. Extending to the full distribution requires strong assumptions.

External Validity: The Limitation of RD

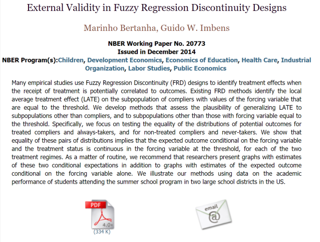
What RD estimates:

$$\tau_{\text{RD}} = \mathbb{E}[Y^1 - Y^0 \mid X = c_0]$$

A **Local Average Treatment Effect (LATE)**—the average effect only for units at the cutoff.

Key limitation: Cannot extrapolate to units far from the cutoff.

- Effect for students scoring 90 may differ from those scoring 70 (near the cutoff)
- External validity requires that "cutoff units" are representative of the broader population



External Validity in Fuzzy Regression Discontinuity Designs

Marinho Bertanha, Guido W. Imbens

NBER Working Paper No. 20773
Issued in December 2014
NBER Program(s): Children, Development Economics, Economics of Education, Health Care, Industrial Organization, Labor Studies, Public Economics

Many empirical studies use Fuzzy Regression Discontinuity (FRD) designs to identify treatment effects when the receipt of treatment is potentially correlated to outcomes. Existing FRD methods identify the local average treatment effect (LATE) on the subpopulation of compliers with values of the forcing variable that are equal to the threshold. We develop methods that assess the plausibility of generalizing LATE to subpopulations other than compliers, and to subpopulations other than those with forcing variable equal to the threshold. Specifically, we focus on testing the equality of the distributions of potential outcomes for treated compliers and always-takers, and for non-treated compliers and never-takers. We show that equality of these pairs of distributions implies that the expected outcome conditional on the forcing variable and the treatment status is continuous in the forcing variable at the threshold, for each of the two treatment regimes. As a matter of routine, we recommend that researchers present graphs with estimates of these two conditional expectations in addition to graphs with estimates of the expected outcome conditional on the forcing variable alone. We illustrate our methods using data on the academic performance of students attending the summer school program in two large school districts in the US.

We estimate the effect at the cutoff. Extending to the full distribution requires strong assumptions.

Extrapolating Beyond the Cutoff

ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework

Joshua Angrist, Ivan Fernandez-Val

NBER Working Paper No. 16566
Issued in December 2010
NBER Program(s):Children, Economics of Education, Health Economics, Labor Studies, Public Economics

This paper develops a covariate-based approach to the external validity of instrumental variables (IV) estimates. Assuming that differences in observed complier characteristics are what make IV estimates differ from one another and from parameters like the effect of treatment on the treated, we show how to construct estimates for new subpopulations from a given set of covariate-specific LATEs. We also develop a reweighting procedure that uses the traditional overidentification test statistic to define a population for which a given pair of IV estimates has external validity. These ideas are illustrated through a comparison of twins and sex-composition IV estimates of the effects childbearing on labor supply.

(284 K)

The linear fit from the RD may not represent the relationship far from the cutoff.

Wanna Get Away? RD Identification Away from the Cutoff

Joshua Angrist, Miikka Rokkanen

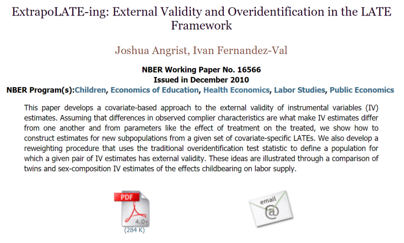
NBER Working Paper No. 18662
Issued in December 2012
NBER Program(s):Children, Economics of Education, Health Economics, Labor Studies, Public Economics

In the canonical regression discontinuity (RD) design for applicants who face an award or admissions cutoff, causal effects are nonparametrically identified for those near the cutoff. The impact of treatment on inframarginal applicants is also of interest, but identification of such effects requires stronger assumptions than are required for identification at the cutoff. This paper discusses RD identification away from the cutoff. Our identification strategy exploits the availability of dependent variable predictors other than the running variable. Conditional on these predictors, the running variable is assumed to be ignorable. This identification strategy is illustrated with data on applicants to Boston exam schools. Functional-form-based extrapolation generates unsatisfying results in this context, either noisy or not very robust. By contrast, identification based on RD-specific conditional independence assumptions produces reasonably precise and surprisingly robust estimates of the effects of exam school attendance on inframarginal applicants. These estimates suggest that the causal effects of exam school attendance for 9th grade applicants with running variable values well away from admissions cutoffs differ little from those for applicants with values that put them on the margin of acceptance. An extension to fuzzy designs is shown to identify causal effects for compliers away from the cutoff.

(1693 K)

A well-designed RCT estimates effects for the full experimental sample. RD is limited to the neighborhood of the threshold.

Extrapolating Beyond the Cutoff



ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework

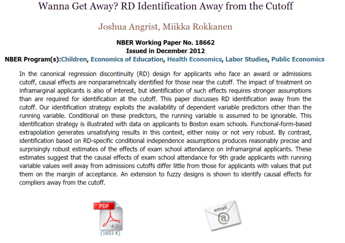
Joshua Angrist, Ivan Fernandez-Val

NBER Working Paper No. 16566
Issued in December 2010
NBER Program(s): Children, Economics of Education, Health Economics, Labor Studies, Public Economics

This paper develops a covariate-based approach to the external validity of instrumental variables (IV) estimates. Assuming that differences in observed complier characteristics are what make IV estimates differ from one another and from parameters like the effect of treatment on the treated, we show how to construct estimates for new subpopulations from a given set of covariate-specific LATEs. We also develop a reweighting procedure that uses the traditional overidentification test statistic to define a population for which a given pair of IV estimates has external validity. These ideas are illustrated through a comparison of twins and sex-composition IV estimates of the effects childbearing on labor supply.

(284 K)

The linear fit from the RD may not represent the relationship far from the cutoff.



Wanna Get Away? RD Identification Away from the Cutoff

Joshua Angrist, Miikka Rokkanen

NBER Working Paper No. 18662
Issued in December 2012
NBER Program(s): Children, Economics of Education, Health Economics, Labor Studies, Public Economics

In the canonical regression discontinuity (RD) design for applicants who face an award or admissions cutoff, causal effects are nonparametrically identified for those near the cutoff. The impact of treatment on inframarginal applicants is also of interest, but identification of such effects requires stronger assumptions than are required for identification at the cutoff. This paper discusses RD identification away from the cutoff. Our identification strategy exploits the availability of dependent variable predictors other than the running variable. Conditional on these predictors, the running variable is assumed to be ignorable. This identification strategy is illustrated with data on applicants to Boston exam schools. Functional-form-based extrapolation generates unsatisfying results in this context, either noisy or not very robust. By contrast, identification based on RD-specific conditional independence assumptions produces reasonably precise and surprisingly robust estimates of the effects of exam school attendance on inframarginal applicants. These estimates suggest that the causal effects of exam school attendance for 9th grade applicants with running variable values well away from admissions cutoffs differ little from those for applicants with values that put them on the margin of acceptance. An extension to fuzzy designs is shown to identify causal effects for compliers away from the cutoff.

(1693 K)

A well-designed RCT estimates effects for the full experimental sample. RD is limited to the neighborhood of the threshold.

Key Insight

RD has **high internal validity** and **limited external validity**. Its strength is transparency and credibility; its limitation is scope.

The RD Checklist: Primary Evidence

What you must show:

✓ Treatment **jumps** at the cutoff (first stage)

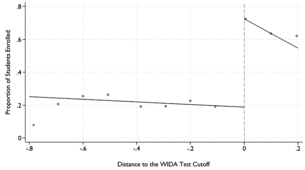


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)

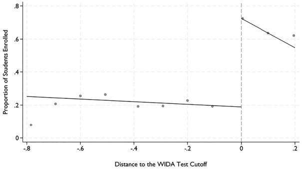


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)
- ✓ Jump is clearly **visible in the graph**

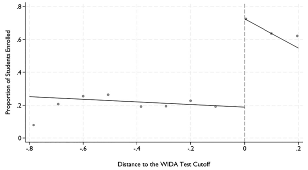


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)
- ✓ Jump is clearly **visible in the graph**
- ✓ Robust to different bandwidths

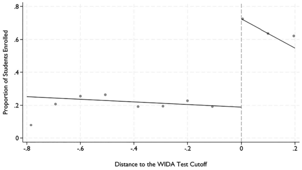


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)
- ✓ Jump is clearly **visible in the graph**
- ✓ Robust to different bandwidths
- ✓ Robust to different polynomial specifications

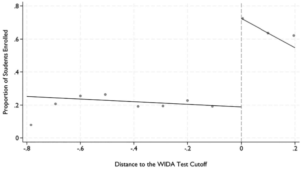


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)
- ✓ Jump is clearly **visible in the graph**
- ✓ Robust to different bandwidths
- ✓ Robust to different polynomial specifications

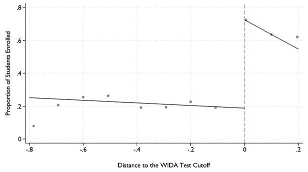


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

The RD Checklist: Primary Evidence

What you must show:

- ✓ Treatment **jumps** at the cutoff (first stage)
- ✓ Outcomes **jump** at the cutoff (reduced form)
- ✓ Jump is clearly **visible in the graph**
- ✓ Robust to different bandwidths
- ✓ Robust to different polynomial specifications

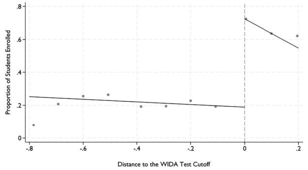


FIGURE 1 Probability of newcomer program enrollment around the eligibility cutoff.

Notes: Figure 1 shows the proportion of students who are reportedly enrolled in the newcomer program as a function of the distance of students' English Language Proficiency screening score from the eligibility cutoff. The figure also shows fitted regression lines from the baseline specification using the optimal bandwidth length. Scoring just below the cutoff for program eligibility increases the likelihood of program participation by 53.7 percentage points, significant at the 1% level. The sample is limited to students who were initially identified as English Learners and who took the WIDA assessment in grades 3 through 7.

Key Insight

The most convincing RD papers require almost no explanation. The **graph speaks for itself**.

The RD Checklist: Supporting Evidence

What you should also show:

✓ Pre-determined covariates do **not jump** (balance test)

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)
- ✓ No spurious jumps at **placebo cutoffs**

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)
- ✓ No spurious jumps at **placebo cutoffs**
- ✓ No other policy changes at the same threshold

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)
- ✓ No spurious jumps at **placebo cutoffs**
- ✓ No other policy changes at the same threshold

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)
- ✓ No spurious jumps at **placebo cutoffs**
- ✓ No other policy changes at the same threshold

Warning

A single failed check doesn't kill the paper, but each failure requires explanation. Multiple failures are fatal.

The RD Checklist: Supporting Evidence

What you should also show:

- ✓ Pre-determined covariates do **not jump** (balance test)
- ✓ Density of running variable is **smooth** (no manipulation)
- ✓ No spurious jumps at **placebo cutoffs**
- ✓ No other policy changes at the same threshold

Warning

A single failed check doesn't kill the paper, but each failure requires explanation. Multiple failures are fatal.

Real-World Application

Readers and referees will look for all of these. Build the checklist into your workflow from the start.

Final Practice Question

Question

A cash transfer program provides $500/month to households with income below $25,000/year. You want to evaluate its effect on children's educational outcomes.

(a) What is the running variable? The cutoff? The treatment?

(b) What identification assumption is needed?

(c) What would concern you about manipulation? How to test?

(d) Why might the RD estimate have limited external validity?

(e) Sketch what the ideal graph would look like.

Final Practice: Answers

Answer

(a) **Running variable:** Household income. **Cutoff:** $25,000/year. **Treatment:** $500/month cash transfer.

(b) **Continuity:** Educational outcomes would vary smoothly with income through the $25,000 threshold in the absence of the program.

(c) **Manipulation:** Households may under-report income to fall below $25,000. Test: McCrary density test—look for a spike just below $25,000 in the income distribution.

(d) **External validity:** Estimate captures the effect for families with income very close to $25,000. Effects may differ substantially for very low- or middle-income households.

(e) **Ideal graph:** Binscatter of educational outcome vs. income—smooth on both sides with a visible upward jump just below $25,000 (at the cutoff, from the treated side).

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment
- Estimates a **local** treatment effect at the threshold

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment
- Estimates a **local** treatment effect at the threshold
- Requires only the **continuity assumption**: potential outcomes vary smoothly through c_0

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment
- Estimates a **local** treatment effect at the threshold
- Requires only the **continuity assumption**: potential outcomes vary smoothly through c_0
- Has **high internal validity** but **limited external validity**

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment
- Estimates a **local** treatment effect at the threshold
- Requires only the **continuity assumption**: potential outcomes vary smoothly through c_0
- Has **high internal validity** but **limited external validity**

Key Takeaways: What RD Does

- RD exploits **arbitrary administrative thresholds** to create quasi-random assignment near the cutoff
- Compares units **just below** and **just above** the cutoff — they are near-identical except for treatment
- Estimates a **local** treatment effect at the threshold
- Requires only the **continuity assumption**: potential outcomes vary smoothly through c_0
- Has **high internal validity** but **limited external validity**

Key Insight

RD is one of the most **credible and transparent** quasi-experimental methods. Its strength: graphical, intuitive, and hard to fake.

Key Takeaways: Implementation Checklist

In Stata, step by step:

✓ Center running variable: `gen rv_c = rv - c`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`
- ✓ Interact slopes: `gen Dx=D*rv_c; reg y D rv_c Dx, robust`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`
- ✓ Interact slopes: `gen Dx=D*rv_c; reg y D rv_c Dx, robust`
- ✓ Optimal bandwidth: `rdrobust y rv, c(c)`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`
- ✓ Interact slopes: `gen Dx=D*rv_c; reg y D rv_c Dx, robust`
- ✓ Optimal bandwidth: `rdrobust y rv, c(c)`
- ✓ Plot the discontinuity: `rdplot y rv, c(c)`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`
- ✓ Interact slopes: `gen Dx=D*rv_c; reg y D rv_c Dx, robust`
- ✓ Optimal bandwidth: `rdrobust y rv, c(c)`
- ✓ Plot the discontinuity: `rdplot y rv, c(c)`
- ✓ Density test: `rddensity rv, c(c) plot`

Key Takeaways: Implementation Checklist

In Stata, step by step:

- ✓ Center running variable: `gen rv_c = rv - c`
- ✓ Create threshold indicator: `gen D = (rv >= c)`
- ✓ Basic RD regression: `reg y D rv_c, robust`
- ✓ Interact slopes: `gen Dx=D*rv_c; reg y D rv_c Dx, robust`
- ✓ Optimal bandwidth: `rdrobust y rv, c(c)`
- ✓ Plot the discontinuity: `rdplot y rv, c(c)`
- ✓ Density test: `rddensity rv, c(c) plot`
- ✓ Covariate balance: run RD with pre-treatment variables as outcome

Regression Discontinuity

The threshold separates everything.

The comparison reveals causality.

Questions? See the walkthrough do-file.

lecture7_walkthrough.do

Appendix A: Sharp RD — Potential Outcomes Derivation

Setup: $Y_i = Y_i(0) + D_i[Y_i(1) - Y_i(0)], \quad D_i = \mathbf{1}(X_i \geq c_0)$

The RD estimand reads the **observed jump** at c_0. Substituting $D = 1$ above and $D = 0$ below:

$$\tau_{\text{RD}} = \lim_{x \downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y \mid X{=}x] = \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(1) \mid X{=}x]}_{\text{above cutoff}} - \underbrace{\lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{below cutoff}}$$

Appendix A: Sharp RD — Potential Outcomes Derivation

Setup: $Y_i = Y_i(0) + D_i[Y_i(1) - Y_i(0)], \quad D_i = \mathbf{1}(X_i \geq c_0)$

The RD estimand reads the **observed jump** at c_0. Substituting $D = 1$ above and $D = 0$ below:

$$\tau_{\text{RD}} = \lim_{x \downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y \mid X{=}x] = \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(1) \mid X{=}x]}_{\text{above cutoff}} - \underbrace{\lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{below cutoff}}$$

Add and subtract $\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x]$ to decompose into effect + bias:

$$\tau_{\text{RD}} = \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(1) - Y(0) \mid X{=}x]}_{\text{True Effect}=\mathbb{E}[Y(1)-Y(0) \mid X=c_0]} + \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{Bias}=\text{jump in counterfactual } Y(0) \text{ at } c_0}$$

Appendix A: Sharp RD — Potential Outcomes Derivation

Setup: $Y_i = Y_i(0) + D_i[Y_i(1) - Y_i(0)], \quad D_i = \mathbf{1}(X_i \geq c_0)$

The RD estimand reads the **observed jump** at c_0. Substituting $D = 1$ above and $D = 0$ below:

$$\tau_{\text{RD}} = \lim_{x \downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y \mid X{=}x] = \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(1) \mid X{=}x]}_{\text{above cutoff}} - \underbrace{\lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{below cutoff}}$$

Add and subtract $\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x]$ to decompose into effect + bias:

$$\tau_{\text{RD}} = \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(1) - Y(0) \mid X{=}x]}_{\text{True Effect}=\mathbb{E}[Y(1)-Y(0)|X=c_0]} + \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{Bias}=\text{jump in counterfactual } Y(0) \text{ at } c_0}$$

Key Insight

Continuity $\Rightarrow \mathbb{E}[Y(0) \mid X]$ continuous at $c_0 \Rightarrow$ **Bias = 0**

$\therefore \quad \tau_{\text{RD}} = \mathbb{E}[Y(1) - Y(0) \mid X = c_0]$

Appendix B: Fuzzy RD — Potential Outcomes Derivation

Setup: D_i is not deterministic. Define the **compliance jump** at c_0:

$$\Delta_D = \lim_{x \downarrow c_0} \mathbb{E}[D \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[D \mid X{=}x]$$

The **Wald (Fuzzy RD) estimator** scales the outcome jump by compliance:

$$\tau_{\text{FRD}} = \frac{\Delta Y}{\Delta_D}, \qquad \Delta Y = \lim_{x \downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y \mid X{=}x]$$

Appendix B: Fuzzy RD — Potential Outcomes Derivation

Setup: D_i is not deterministic. Define the **compliance jump** at c_0:

$$\Delta_D = \lim_{x \downarrow c_0} \mathbb{E}[D \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[D \mid X{=}x]$$

The **Wald (Fuzzy RD) estimator** scales the outcome jump by compliance:

$$\tau_{\text{FRD}} = \frac{\Delta Y}{\Delta_D}, \qquad \Delta Y = \lim_{x \downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y \mid X{=}x]$$

Decompose the numerator ΔY into effect and bias:

$$\Delta Y = \underbrace{\mathbb{E}[Y(1){-}Y(0) \mid X{=}c_0, \text{compliers}] \cdot \Delta_D}_{\text{True Effect} \times \text{Compliance}} + \underbrace{\lim_{x \downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] - \lim_{x \uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{Bias}}$$

Dividing by Δ_D:

$$\tau_{\text{FRD}} = \underbrace{\mathbb{E}[Y(1){-}Y(0) \mid X{=}c_0, \text{compliers}]}_{\text{True causal effect (LATE)}} + \underbrace{\frac{\text{Bias}}{\Delta_D}}_{\to 0}$$

Appendix B: Fuzzy RD — Potential Outcomes Derivation

Setup: D_i is not deterministic. Define the **compliance jump** at c_0:

$$\Delta_D = \lim_{x\downarrow c_0} \mathbb{E}[D \mid X{=}x] - \lim_{x\uparrow c_0} \mathbb{E}[D \mid X{=}x]$$

The **Wald (Fuzzy RD) estimator** scales the outcome jump by compliance:

$$\tau_{\text{FRD}} = \frac{\Delta Y}{\Delta_D}, \qquad \Delta Y = \lim_{x\downarrow c_0} \mathbb{E}[Y \mid X{=}x] - \lim_{x\uparrow c_0} \mathbb{E}[Y \mid X{=}x]$$

Decompose the numerator ΔY into effect and bias:

$$\Delta Y = \underbrace{\mathbb{E}[Y(1){-}Y(0) \mid X{=}c_0, \text{compliers}] \cdot \Delta_D}_{\text{True Effect} \times \text{Compliance}} + \underbrace{\lim_{x\downarrow c_0} \mathbb{E}[Y(0) \mid X{=}x] - \lim_{x\uparrow c_0} \mathbb{E}[Y(0) \mid X{=}x]}_{\text{Bias}}$$

Dividing by Δ_D:

$$\tau_{\text{FRD}} = \underbrace{\mathbb{E}[Y(1){-}Y(0) \mid X{=}c_0, \text{compliers}]}_{\text{True causal effect (LATE)}} + \underbrace{\frac{\text{Bias}}{\Delta_D}}_{\to 0}$$

Key Insight

Continuity $\Rightarrow$ Bias = 0

$\tau_{\text{FRD}} = \mathbb{E}[Y(1) - Y(0) \mid X{=}c_0, \text{compliers}]$

— the LATE for cutoff compliers.