UNIVERSITY
OF OSLO

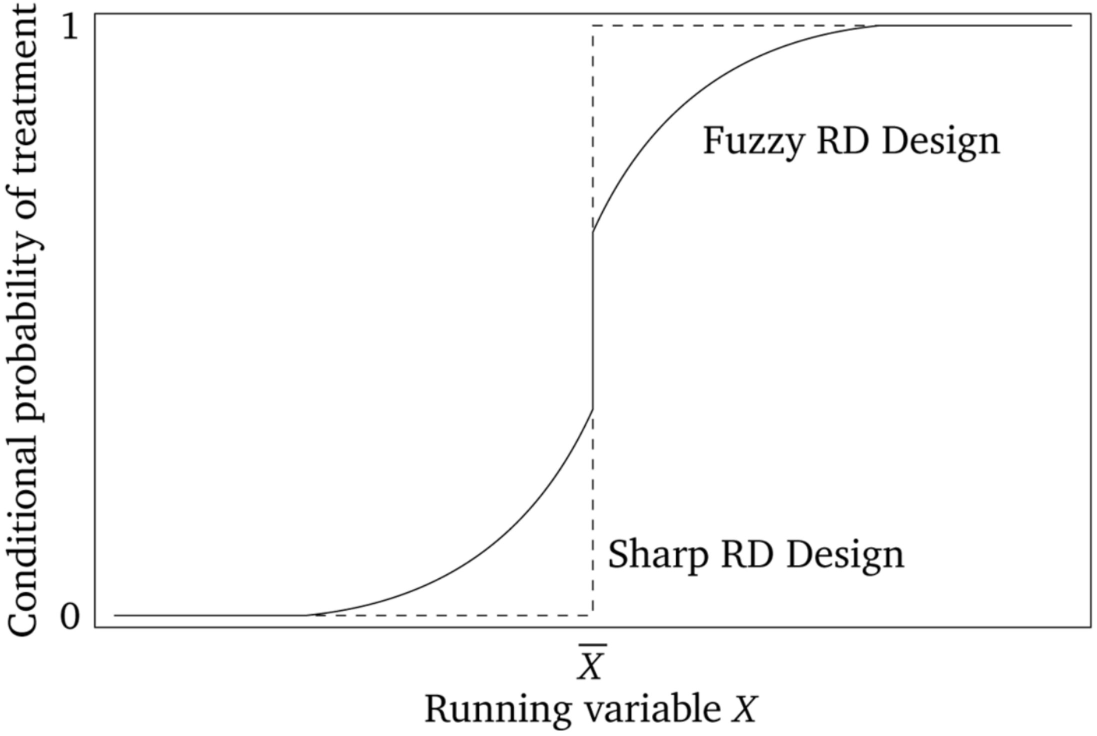


Figure 6.4. from Cunningham (2021)

REGRESSION DISCONTINUITY I

José Manuel Arencibia Alemán & Isa Steinmann

Session 8

UiO : Centre for Educational Measurement
Det utdanningsvitenskapelige fakultet

February 10, 2022

5	Regression Models – Time: 31 January 2022, 12:15 – 14:00h – Main Instructor: Isa Steinmann – Required Reading: Angrist & Pischke (2015), Chapter 2
6	Further Control Strategies – Time: 03 February 2022, 12:15 – 14:00h – Main Instructor: Isa Steinmann – Required Reading: -
7	Instrumental Variable Approaches – Time: 07 February 2022, 12:15 – 14:00h – Main Instructor: José Manuel Arencibia Alemán – Required Reading: Angrist & Pischke (2015), Chapter 3
8	Regression Discontinuity Designs I – Time: 10 February 2022, 12:15 – 14:00h – Main Instructor: José Manuel Arencibia Alemán – Required Reading: Angrist & Pischke (2015), Chapter 4

Last session’s take-away messages

- IV allow us to draw causal inferences when treatment is endogenous
- Good instruments (relevant, independent and satisfying the exclusion restriction) are hard to find.
- External validity of LATEs should be well argued for, specially if instruments are “naturally occurring”.

Session's learning goal

- Understanding the role of rules in generating "as good as random" circumstances to infer causal effects
- Getting the intuition of RD Designs through analysis of DAGs and graphs, and being able to formalize simple specifications

Content

1. Introduction
   - ✓ Rules on chance
   - ✓ Four elements of RDD
2. Birthdays and Funerals
   - ✓ Sharp RD
   - ✓ RD Formalization
   - ✓ Parametric and Non-parametric RD
3. The Elite Illusion
   - ✓ Fuzzy RD
   - ✓ Fuzzy RD Is IV

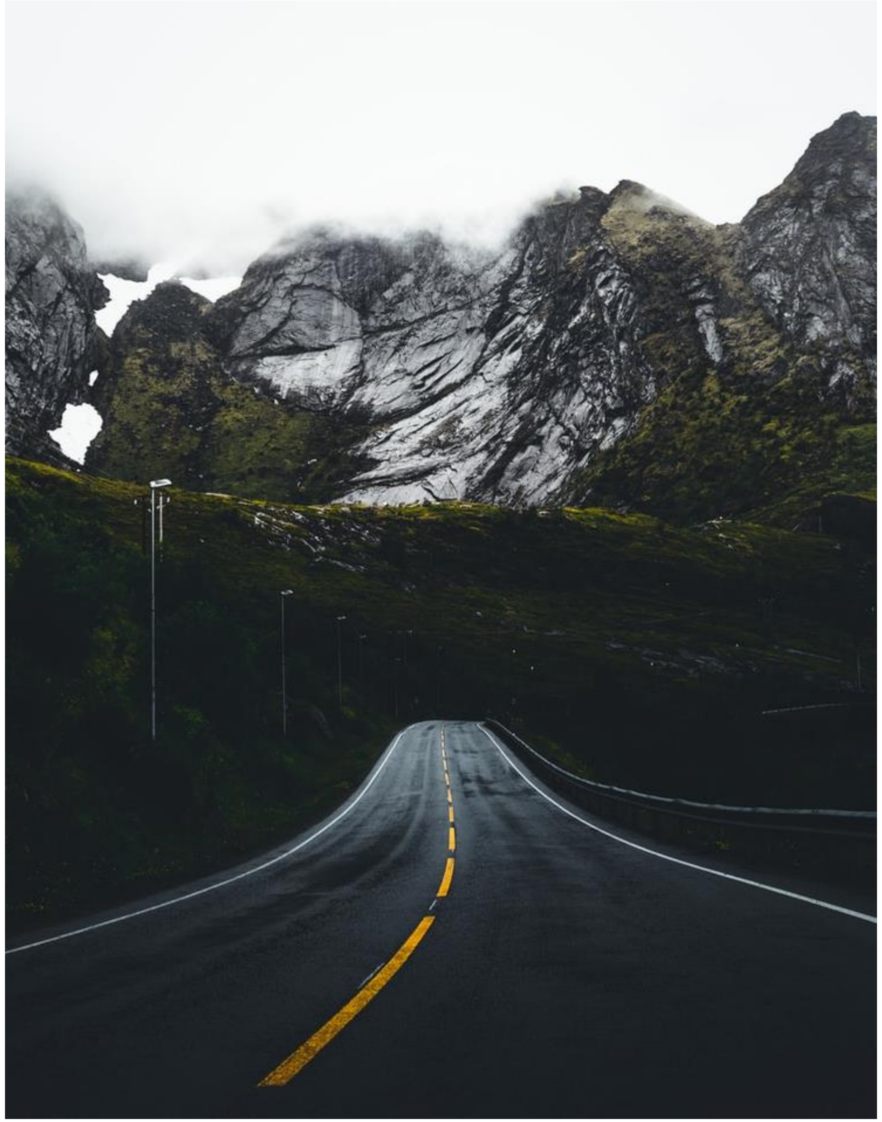


Photo by Taneli Lahtinen on Unsplash

Introduction

- One of the main challenges (if not “the main challenge”) of Social Sciences is to deal with the vast amount of variables potentially confounding (causal) relations of interest
- To study these (causal) relations, the gold standard is the RCT for its ability to isolate treatment effects from those of confounders
- Control strategies may eliminate the confounding effect of observable variables, and IV isolate causal paths exploiting naturally occurring or researcher induced partial randomization.
- In this session, we introduce another potentially powerful tool in our quest to find causal effects

Regression Discontinuity Designs:
Rules on chance

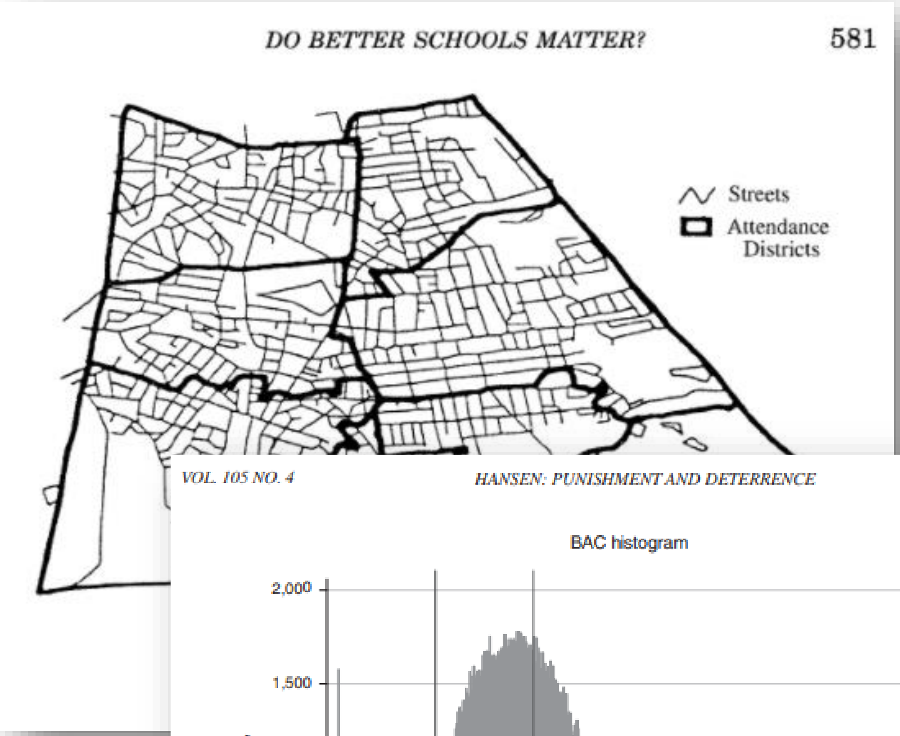
”Human behavior is constrained by rules. [e.g., on class size, age of retirement, recruitment] Although many of these rules seem arbitrary, we say: bring ‘em on!

For rules that constrain the role of chance in human affairs often generate interesting experiments.”

Angrist and Pischke (2015, p. 147)

Introduction

- The deterministic and, often times, arbitrary nature of rules provides opportunities to answer causal questions otherwise impossible to explore, e.g.,
  - How much more are parents willing to pay for better schools? (Black, 1999)
    - In district boundaries, the only difference in the value of a house might be determined by where a child is mandated to attend. *Finds that parents are willing to pay 2,5% more for a 5% increase in test scores.*
  - What is the effect of punishment on drinking and driving recidivism? (Hansen, 2015)
    - Different levels of alcohol in blood are associated with different punishment. *Finds that fines reduce recidivism in 2 percent points (-17%) and jail in 1 additional point (-9%).*



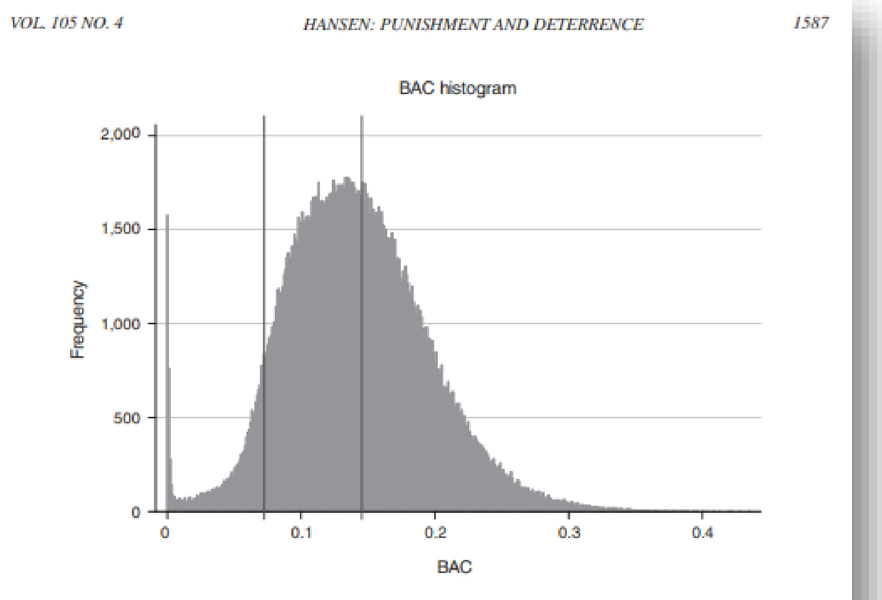


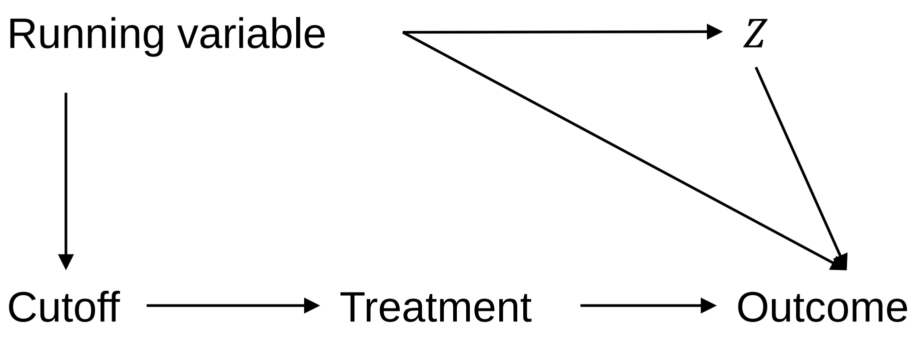
FIGURE 1. BAC DISTRIBUTION

Notes: Based on administrative records from the Washington State Impaired Driver Testing Program, 1999–2007. The histogram height on the vertical axis is based on frequency of observations, with BAC on the horizontal axis. The vertical black lines represent the two legal thresholds at 0.08 and 0.15. The bin width is 0.001, the original precision used on the breathalyzers.

Introduction

RDD: The Intuition

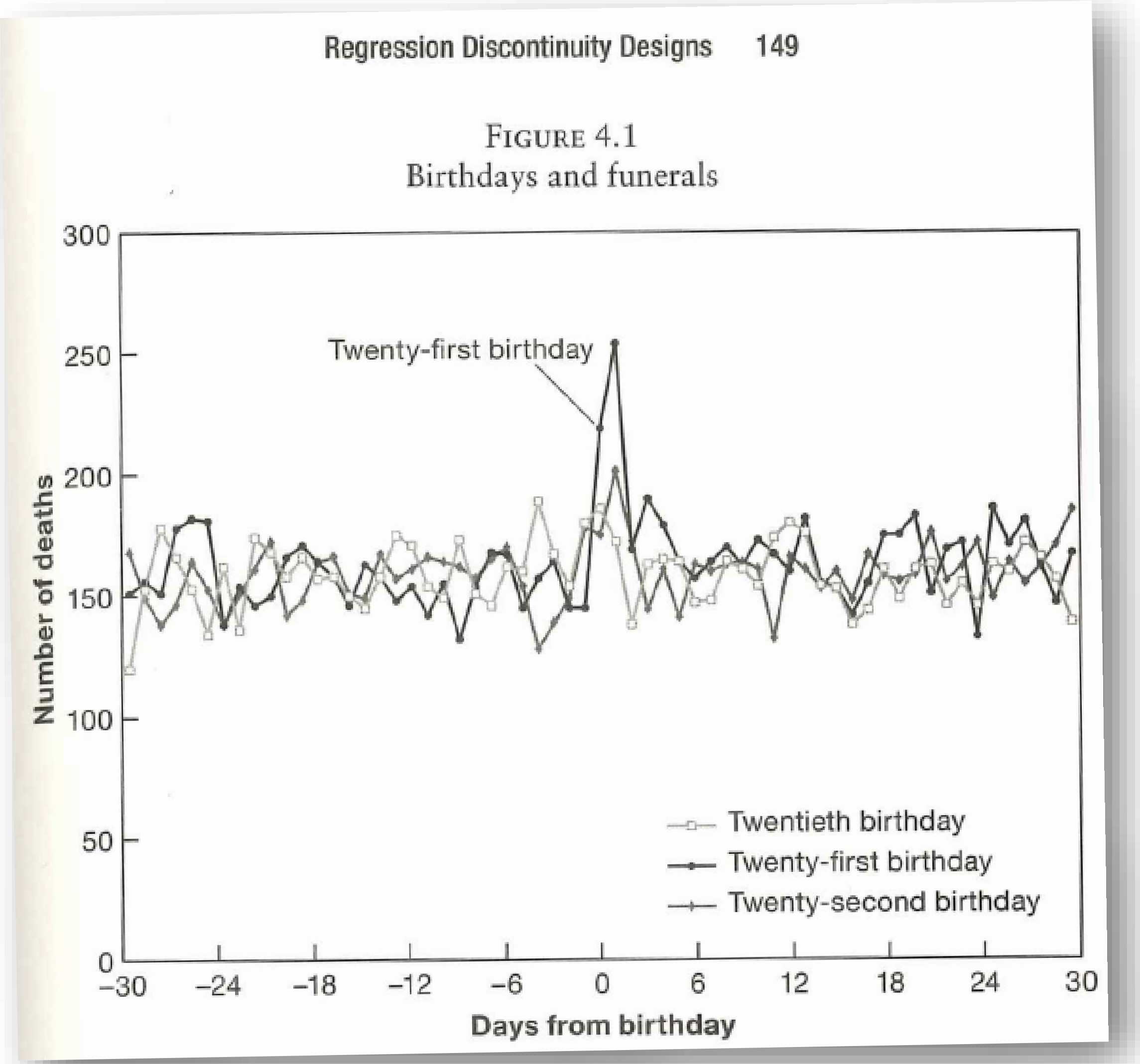
- Four elements characterize RDD
  - **Running variable**: a variable associated with the outcome of interest (and some endogenous variables, Z, potentially unobservable) that determines treatment
  - **Outcome**
  - **«Cutoff» score**: the value at which the running variable induces treatment
  - **Treatment**
- If the cutoff is exogenous so is treatment and, thus, the association between treatment and outcome is a causal effect.



Huntington (2022)

Birthdays and Funerals

What does figure 4.1 suggest?



Setting the Stage

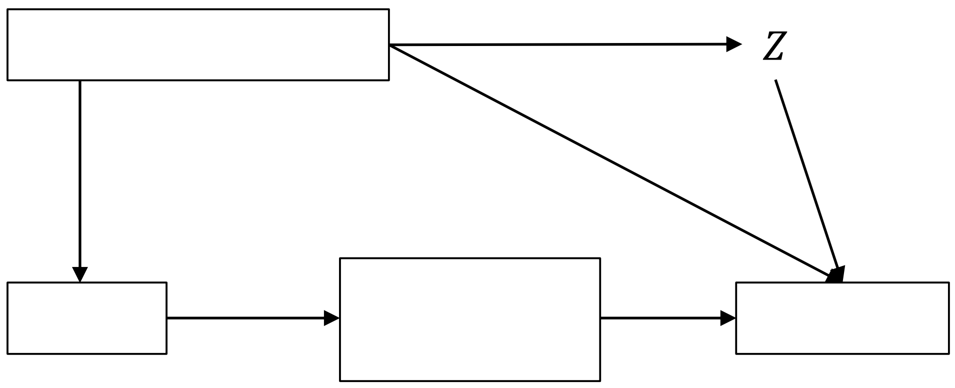
- Despite a Minimum Legal Drink Age (MLDA) of 21-years-old, underage drinking is common and not necessarily moderate
- A group of American Colleges presidents have lobbied to return to the Vietnam-era threshold of 18
  - **Hypothesis**: Lowering the legal drinking age, "would discourage binge drinking and promote a culture of mature alcohol consumption"
  - **Alternative hypothesis**: "age-21 reduces youth access to alcohol, thereby preventing some harm"

Angrist & Pischke (2015)

Birthdays and Funerals

Setting the Stage

- Which are the four elements of RDD to study the (natural) experiment just described and where do they go in this DAG?
  - **Running variable:**
  - **Outcome:**
  - **«Cutoff» score:**
  - **Treatment:**



Huntington (2022)

Birthdays and Funerals

Sharp RD

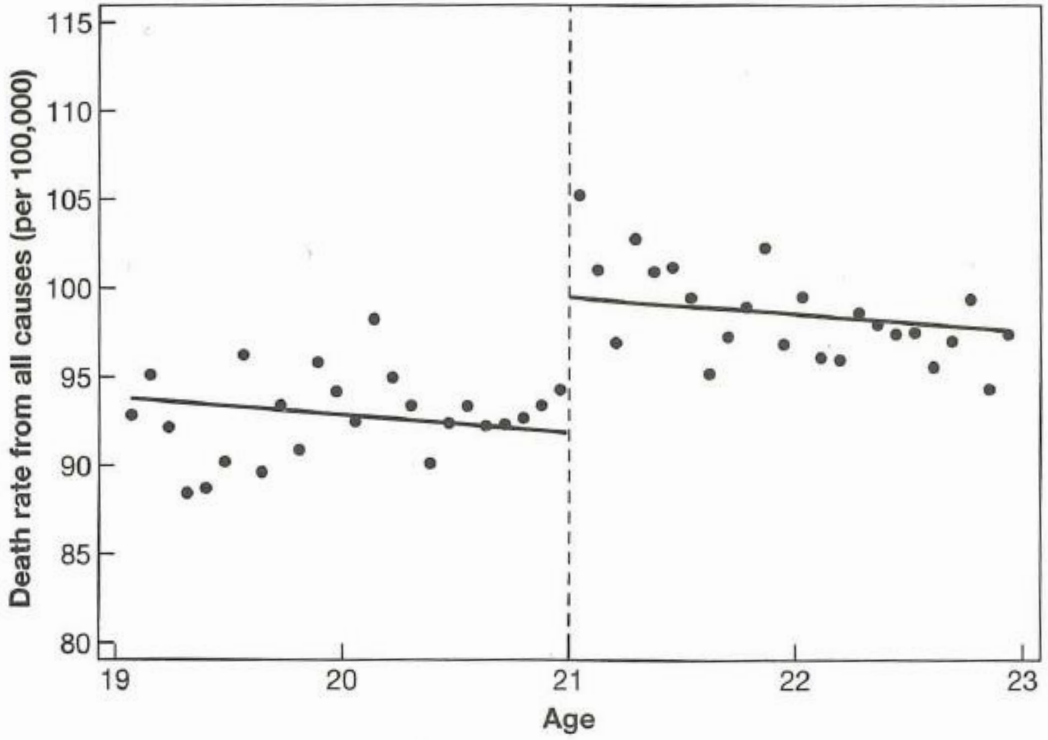
- What is the story told by the graph?
  a) Negative trend in death rates
  b) Significant shift at 21 in deaths per 100,000 $(7.66\ (1.51\ s.e.)$, p. 160)

"*The **causal question** addressed by Figure 4.2 is the effect of legal access to alcohol on death rates. The treatment variable in this case can be written*"

$$D_a = \begin{cases} 1\ if\ a \geq 21 \\ 0\ if\ a\ < 21 \end{cases}$$

150 Chapter 4

FIGURE 4.2
A sharp RD estimate of MLDA mortality effects



Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

Angrist & Pischke (2015)

Birthdays and Funerals

Sharp RD

- In all RDDs, treatment status, D_a, is a deterministic and discontinuous (at the cutoff value, a_0) function of the running variable
$$D_a = f(a)$$
- If the probability of treatment goes from 0 to 1 at the cutoff value, we face sharp RD. Otherwise, we are talking fuzzy RD.

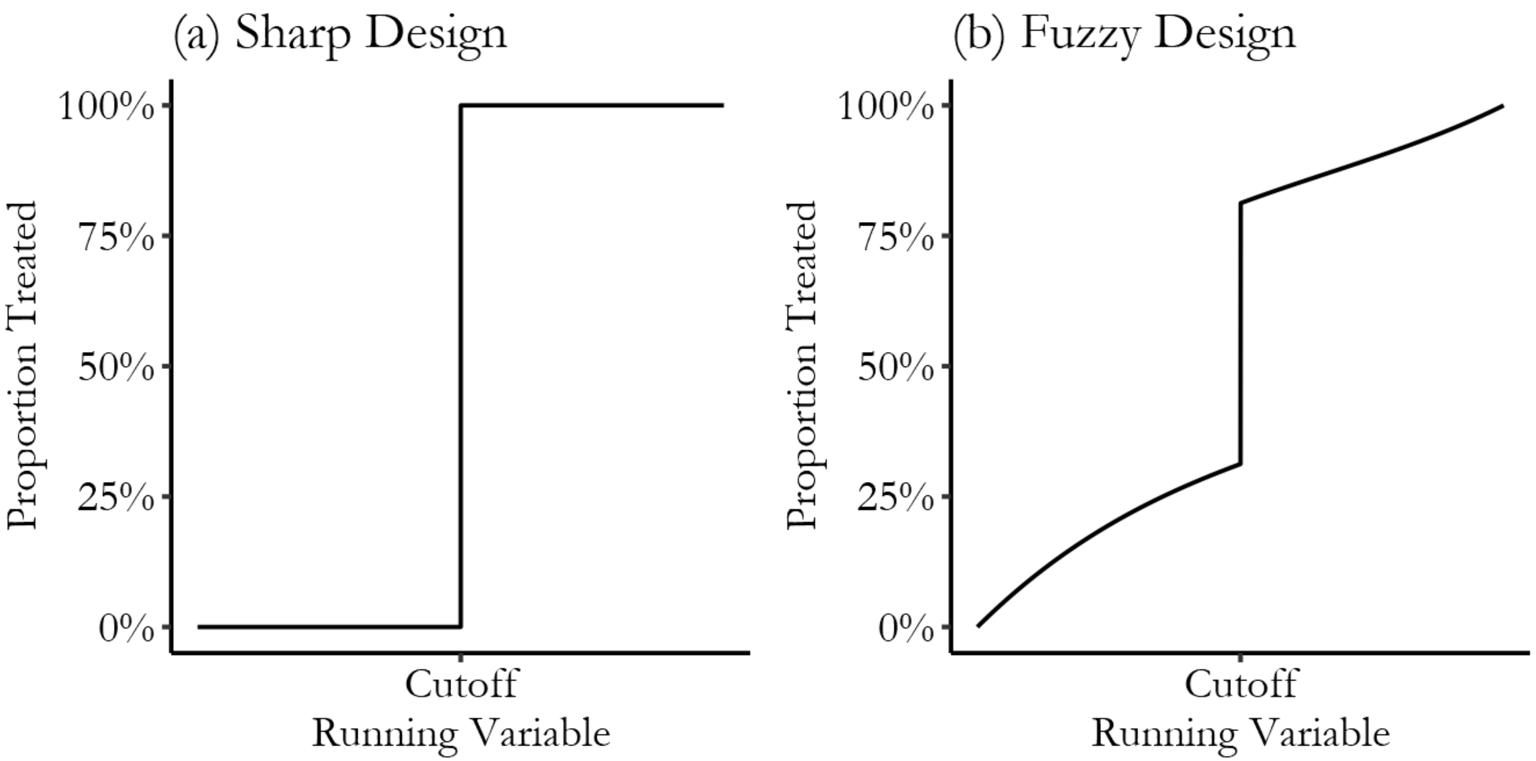


Figure 20.3 from Huntington (2022)

What kind of RD does generate the MLDV?

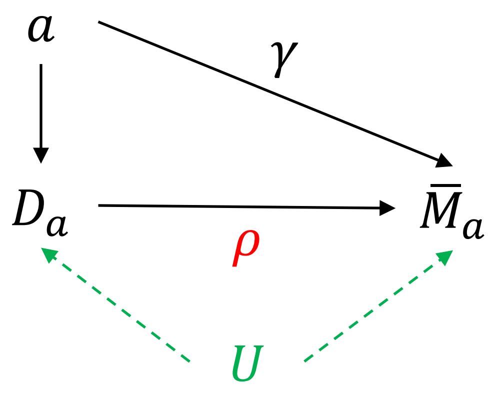
Birthdays and Funerals

RD Formalization: Linear trend

- In our example, let $\boldsymbol{\bar{M}_a}$ be the outcome variable (death rate in a month), $\boldsymbol{a}$ the running variable (age), and $\boldsymbol{D_a}$ the treatment (legal access to alcohol) determined by the cutoff value, $\boldsymbol{a_0}$ (MLDA). **Assuming linearity** (like in Figure 4.2 of Angrist and Pischke (2015)), the causal effect of treatment, $\boldsymbol{\rho}$, can be estimated from:

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$$

- Effectively, we have regular lineal regression controlling for the linear trend of the running variable



What about OVB?

- If OVB, the estimate of ρ would depend on the correlation of D_a with U.
- BUT, by definition, D_a is function solely of a. That is, if the effect of a is adequately represented by our function, there is no OVB

Birthdays and Funerals

RD Formalization: Non-linear trends

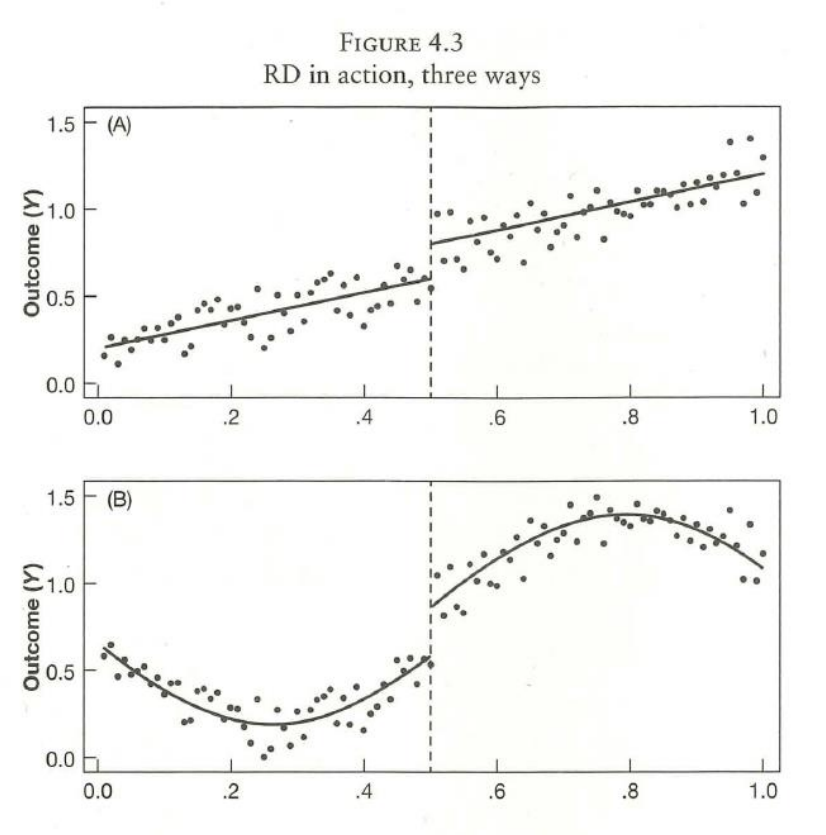
- RDDs core boils down to (using the notation from our example) measuring:

$$\lim_{a \to a_0^+} \widehat{\bar{M}_a} - \lim_{a \to a_0^-} \widehat{\bar{M}_a}$$

- For this measure to be precise, the relation between outcome and running variable has to be precise.

- For example, we can estimate the causal effect of treatment in panel B of figure 4.3 (Angrist et al., 2015, p. 154) with a regression of the form (?):

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1 a + \boldsymbol{\gamma_2 a^2} + e_a$$



Angrist & Pischke (2015)

Birthdays and Funerals

RD Formalization: Interactions between treatment and running variables

- We can also account for different coefficients to the left and right of the cutoff by simply adding an interaction term between the running variable and treatment (for easier interpretation of the interaction, it is useful to center the running variable around the cutoff):

$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta(a - a_0)D_a + e_a$$

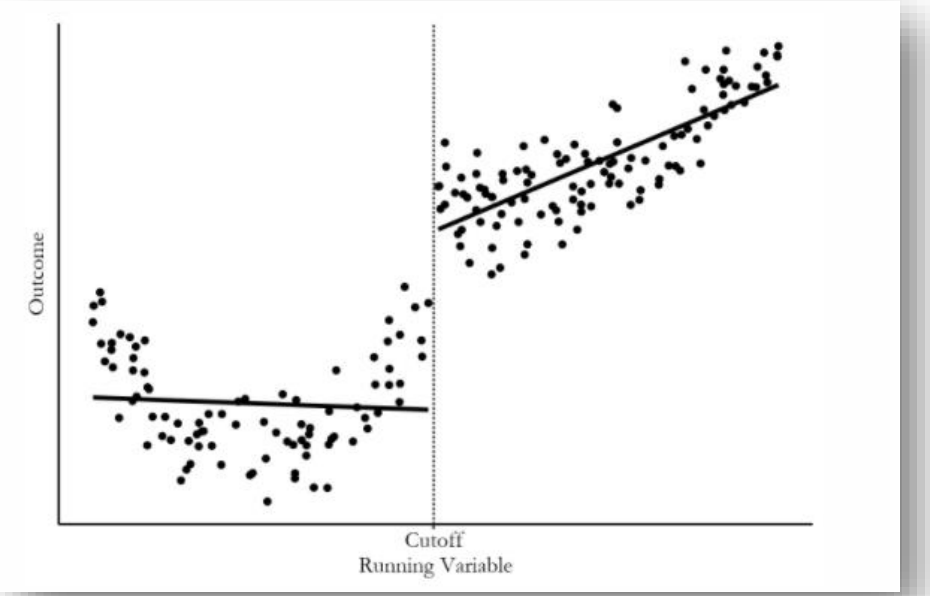


Figure 20.5 from Huntington (2022)

Angrist & Pischke (2015)

Birthdays and Funerals

RD Formalization: Interactions between treatment and running variables

- Given:

$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta(a - a_0)D_a + e_a$$

- *What is the causal effect of treatment, $\frac{\partial \bar{M}_a}{\partial D_a}$?*

$$\bar{M}_a{}'_{D_a} = \rho + \delta(a - a_0)$$

  – *Notice: the effect of treatment changes (increases if $\delta > 0$, decreases if $\delta < 0$) with a*

- *What is the average jump in death rates at the cutoff, $\left.\frac{\partial \bar{M}_a}{\partial D_a}\right|_{a=a_0}$?*

$$\left.\bar{M}_a{}'_{D_a}\right|_{a=a_0} = \rho$$

  – *Notice: the effect of treatment at the cutoff continues to be captured by ρ*

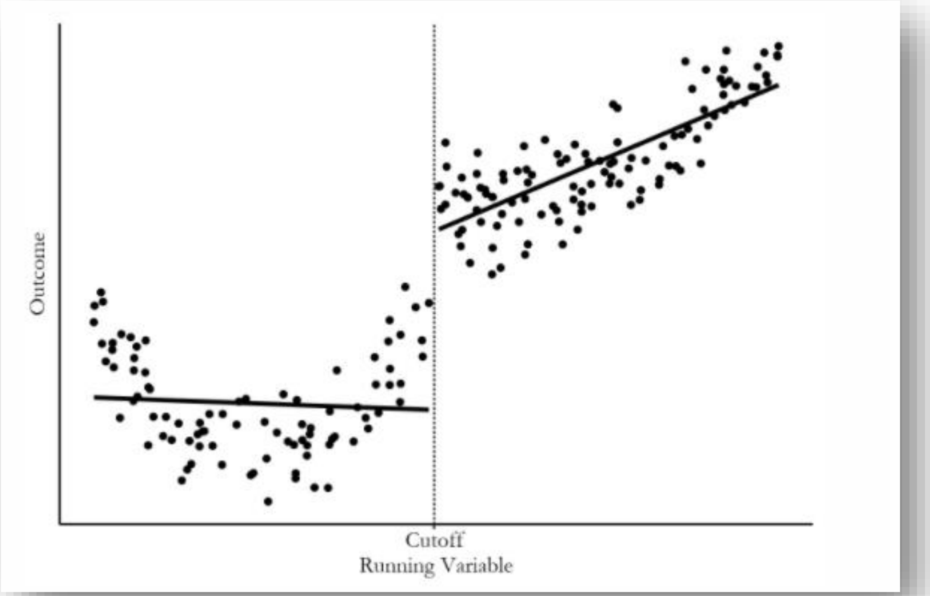


Figure 20.5 from Huntington (2022)

WARNING

- Interactions accommodate differential effects of treatment across the running variable, **BUT** as we get further away from the cutoff, treated and untreated might deviate in ways other than treatment.
  - **Angrist & Pischke (2015, p. 153):** «the validity of RD turns on our willingness to extrapolate across values of the running variable, at least for values in the ***neighborhood*** of the cutoff at which treatment switches on»
- Appropriate modelization is necessary to precisely estimate the treatment effect on the treated at the cutoff (ρ), but be aware of not extrapolating far beyond the cutoff without reason. May I add:
  - You will often **not** find reasons and still be tempted to do it. **Please, don't!**
  - This applies to baroque and the simplest of linear model specifications (although, again, ***limited*** extrapolation might be justified when the relationship between running variable and outcome is linear).

Birthdays and Funerals

RD Formalization: Non-linear trends & interaction terms

- Finally, when data and theory suggest **non-linear trends** between running variable **and** outcome, as well as **changes in the relation between running variable and outcome as result of the treatment**, this can be introduced in a model of the following form:

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2$$
$$+\delta_1(a - a_0)D_a + \delta_2(a - a_0)^2 D_a + e_a$$

- Which model is better?
  - “no general rules here, and no substitute for a thoughtful look at the data. We’re especially fortunate when results are not highly sensitive to the details of our model choices.“ (p. 157)

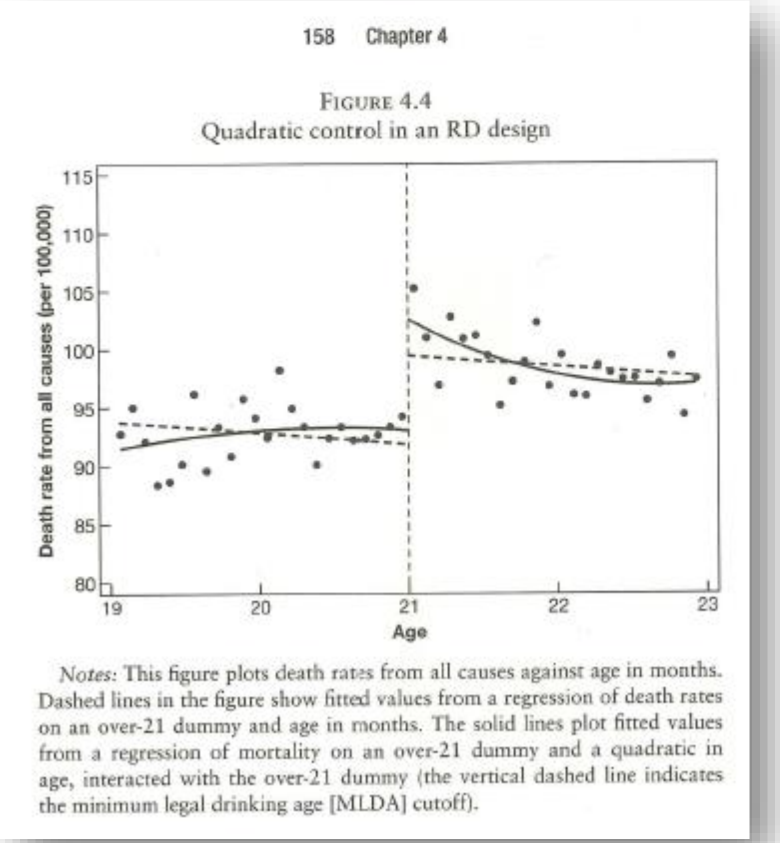

158 Chapter 4

FIGURE 4.4
Quadratic control in an RD design

Notes: This figure plots death rates from all causes against age in months. Dashed lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months. The solid lines plot fitted values from a regression of mortality on an over-21 dummy and a quadratic in age, interacted with the over-21 dummy (the vertical dashed line indicates the minimum legal drinking age [MLDA] cutoff).

Angrist & Pischke (2015)

Birthdays and Funerals

- *Is there a causal effect of legal access to alcohol on death rates?*

RD strategies: Regression vs. Parametric RD

- **Parametric RD**: "For a small set of points close to the boundary, nonlinear trends need not concern us at all. This suggest an approach that compares averages in a narrow window just to the left and just to the right of the cutoff." (p. 161)

$$\rho = E[\bar{M}_a | D_a = 1]\Big|_{a \to a_0} - E[\bar{M}_a | D_a = 0]\Big|_{a \to a_0}$$

- Problem:
  - If the window is too narrow, the sample size might be not enough to obtain reliable estimates... but if it is too large, potential selection bias will flourish.

160 Chapter 4

TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22 (1)	(2)	Ages 20–21 (3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

Angrist & Pischke (2015)

Birthdays and Funerals

RD strategies: Parametric RD vs. Non-Parametric RD

- Non-parametric RD deals with the “bias-variance” trade off of parametric models by regressing in a window around the cutoff called bandwidth, $\boldsymbol{b}$ (see prior slide)

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + e_a$$
$$in\ a\ sample\ such\ that\ a_0 - b \leq a \leq a_0 + b$$

- Linear function is chosen because the form of the function matters little when the estimation is near the cutoff (i.e., if $b \to 0$; $\bar{M}_a \equiv f(a)$ approaches linearity).

- *For the interested (recommended):*
  - *Bandwidth should be small enough to obviate concerns about polynomial choice, but large enough to avoid lack of precision in the estimates (Imbens and Kalyanaraman 2012)*
  - *Non-parametric models are typically weighted using kernel functions (Cunningham [2021])*

GROUP EXCERCISE

Get together and try to come up with possible RD designs (real or imaginary). What are the running variable, cutoff, treatment and outcomes in your RD design?

- Find inspiration in the examples we have seen until now:
  - MLDA on mortality rates (*age, MLDA, access to alcohol, death rates*)
  - Punishment on recidivism (*BAC, punishable levels of alcohol, fine/jail, recidivism*)
  - School attendance district boundaries on house prices (*location, district boundaries, house prices, demand on houses*)

The Elite Illusion

Setting the stage

- **Exam schools**: highly selective (on the basis of ISEE tests) public schools
- **RQ**: consequences of exam school treatment
  - ***Hypothesis***: exam schools have a positive effect on academic achievement driven by peer-effects
- **Problem**: High-achievers will be high-achievers
- Why does RCTs do not work here to answer the causal question?
  - If admission is randomized, peer-effects disappear and, thus, so does the potential causal effect

Fuzzy RD

- Applicants with scores close to admissions cutoff, whether to the right or to the left, might be considered as good as randomly assigned.
  - **Nuance**: at the cutoff, some students choose to go elsewhere while many of those rejected at one exam school end up at another (with a slightly lower cutoff slightly lower).

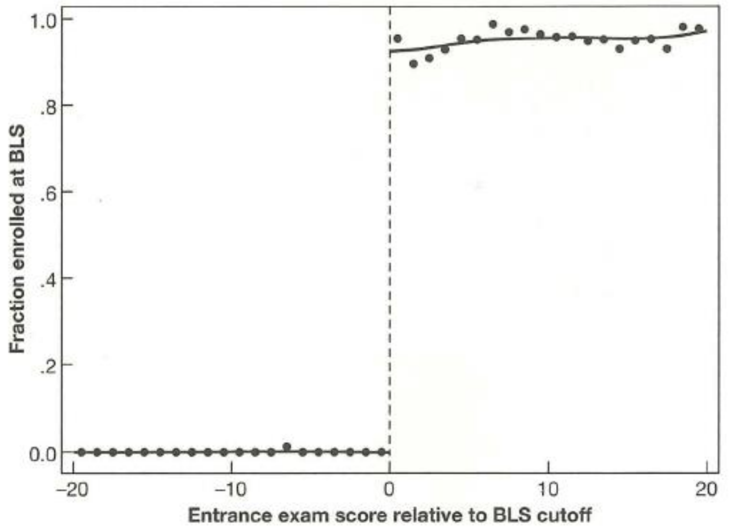
"When discontinuities change treatment probabilities or average characteristics (**treatment intensity**, for short), […] the resulting RD design is said to be ***fuzzy***." (pp. 165-166)

Angrist & Pischke (2015)

The Elite Illusion

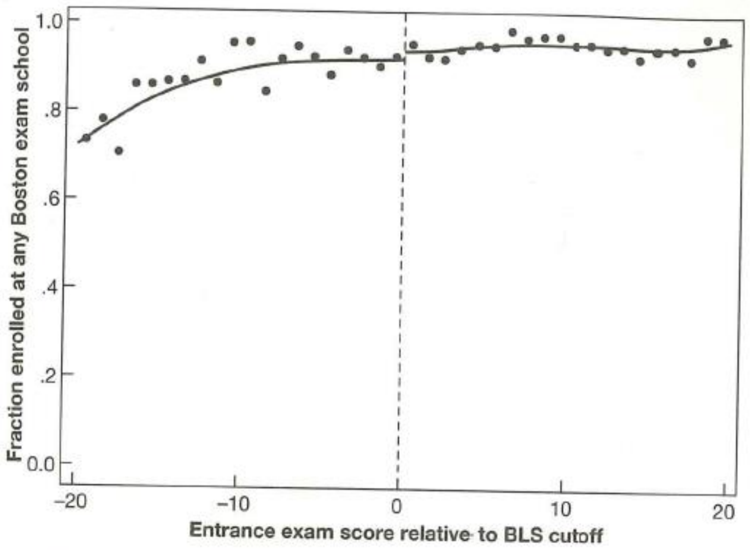
Fuzzy RD

FIGURE 4.6
Enrollment at BLS



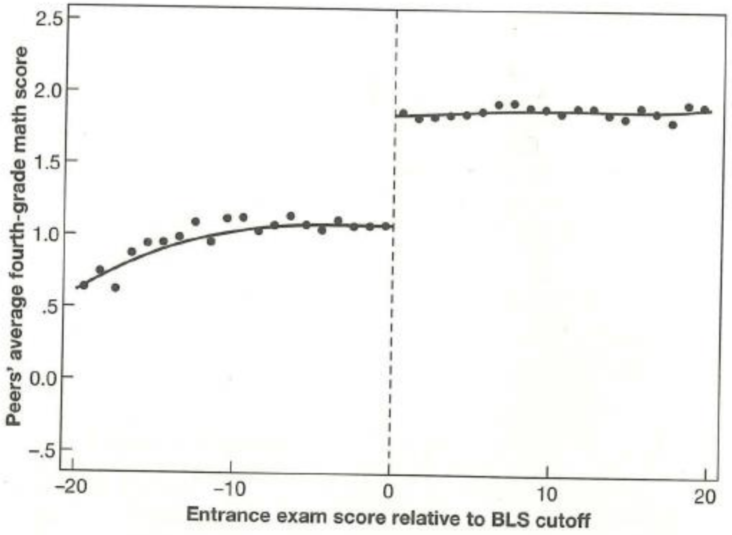
Notes: This figure plots enrollment rates at Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the BLS admissions cutoff. Solid lines show fitted values from a local linear regression estimated separately on either side of the cutoff (indicated by the vertical dashed line).

FIGURE 4.7
Enrollment at any Boston exam school



Notes: This figure plots enrollment rates at any Boston exam school, conditional on admissions test scores, for Boston Latin School (BLS) applicants scoring near the BLS admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

FIGURE 4.8
Peer quality around the BLS cutoff



Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Angrist & Pischke (2015)

The Elite Illusion

Fuzzy RD Is IV

- A naive approach would regress students' 7th-grade math scores, $\boldsymbol{Y_i}$, on the average 4th-grade math score of their classmates in 7th-grade, $\boldsymbol{\bar{X}_{(i)}}$, controlling on their 4th-grade math scores, $\boldsymbol{X_i}$. This analysis typically suggest an increase of $.25\sigma$ in outcomes for each σ increase in $\bar{X}_{(i)}$:

$$Y_i = \theta_0 + \theta_1 \bar{X}_{(i)} + \theta_2 X_i + u_i$$

What is the problem?

$$\begin{tikzcd}
\bar{X}_{(i)} \arrow[rr] & & Y_i \\
& U \arrow[ul, dashed] \arrow[ur, dashed] &
\end{tikzcd}$$

"While peer quality jumps at the cutoff, cross-cutoff comparisons of variables related to applicants' own abilities, motivation, and family background—the sources of selection bias we usually worry about—show no similar jumps. […] ***Peers change discontinuously at admissions cutoffs, but exams school applicants' own characteristics do not.****"* (p. 171) → **RD Design**

Angrist & Pischke (2015)

The Elite Illusion

Fuzzy RD Is IV

- **RD approach**: Use the enrollment score in BLS as cutoff on our running variable and, thus, treatment becomes "qualifying for BLS":

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}$$

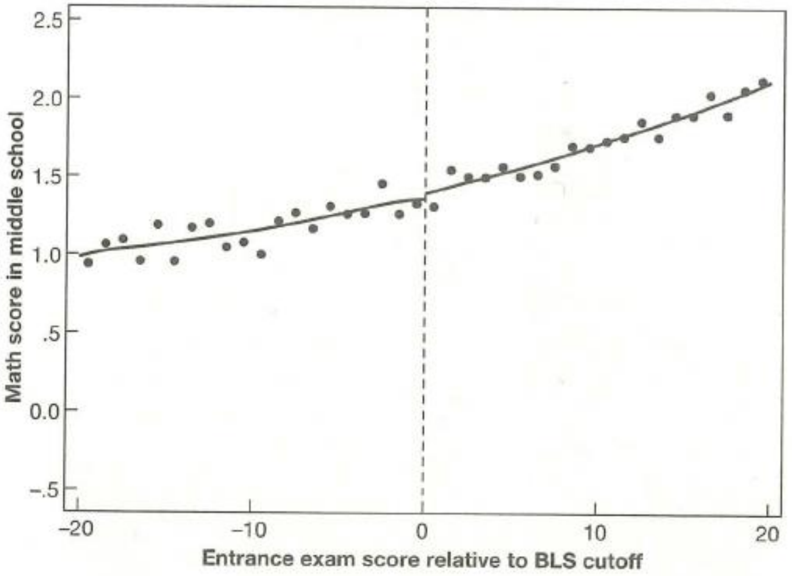
- How to interpret $\rho = -.02\ (.10)$?
  - ρ is the jump in outcomes at the discontinuity produced by the treatment.
  - BUT ALSO, is the reduced form of an **IV** in which peer-effects are instrumentalized by an «as good random in the neighborhood of the cutoff» (**independence**) assignment to different intensities of peers math ability. The first-stage and 2SLS forms are:

$$\bar{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i} \quad ; \quad Y_i = \alpha_2 + \lambda \hat{\bar{X}}_{(i)} + \beta_2 R_i + e_{2i}$$

- Since the first stage is close to zero and non-significant, the 2SLS estimate will be zero too

172 Chapter 4

FIGURE 4.9
Math scores around the BLS cutoff



Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Angrist & Pischke (2015)

The Elite Illusion

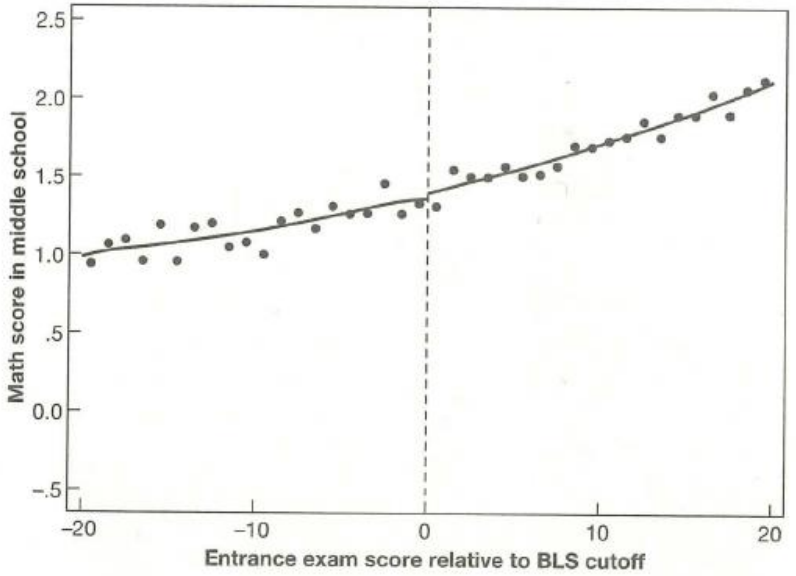
More interpretation

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}$$

- Implicit in our interpretation, is that effects of exam schools in outcomes are driven solely by characteristics of students which, around the cutoff, are as good as randomly assigned
- However, some characteristics of the exam schools changing at the cutoff might also affect our outcome variable. In that case, our specification would be affected by OVB.
- Which effect would this have in our estimate?
    - If at the cutoff characteristics of the exam school change (e.g., hired teachers have to have more experience), our estimate would be biased (too large, and thus, peer-achievement would be at most neutral, and possibly negative for achievement)

172 Chapter 4

FIGURE 4.9
Math scores around the BLS cutoff



Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Angrist & Pischke (2015)

Take away messages

- RDD take advantage of more or less arbitrary cutoffs in running variables to, under the assumption that near the cutoff individuals average characteristics are equal, infer causal effects of treatment
- Sharp RDDs allow causal inference when treatment switches at the cutoff.
- Fuzzy RDDs allow causal inference when treatment probability or intensity changes at the cutoff.
- Fuzzy RDDs are methodologically equivalent to IV

9	Regression Discontinuity Designs II – Time: 14 February 2022, 12:15 – 14:00h – Main Instructor: Isa Steinmann – Required Reading: Luyten (2006)
10	Differences-in-Differences Designs I – Time: 17 February 2022, 12:15 – 14:00h – Main Instructor: Isa Steinmann – Required Reading: Angrist & Pischke (2015), Chapter 5
11	Differences-in-Differences Designs II – Time: 21 February 2022, 12:15 – 14:00h – Main Instructor: Isa Steinmann – Required Reading: Strello, Strietholt, & Siepmann (2021)
12	Lessons Learned and Closing – Time: 23 February 2022, 12:15 – 14:00h – Main Instructor: José Manuel Arencibia Alemán – Required Reading: Rutkowski & Delandshere (2016)

Recommended Literature

- Cunningham, S. (2021). Causal inference : the mixtape. New Haven, Connecticut, Yale University Press. (Chapter 6)
- N. Huntington-Klein (2022). The effect : an introduction to research design and causality. Boca Raton, Chapman and Hall/CRC Press. (Chapter 20)
- Angrist, J. D. and J.-S. Pischke (2008). Mostly Harmless Econometrics : An Empiricist's Companion. Princeton, NJ, Princeton University Press. (Chapter 6, available online at through the University Library)

References

Main reference:

- Angrist, J. D. and J.-S. Pischke (2015). Mastering 'metrics : the path from cause to effect. Princeton, N.J, Princeton University Press.

Other sources:

- Cunningham, S. (2021). Causal inference : the mixtape. New Haven, Connecticut, Yale University Press.
- Huntington-Klein, N. and N. Huntington-Klein (2022). The effect : an introduction to research design and causality. Boca Raton, Chapman and Hall/CRC Press.

Example papers:

- Black, S. E. (1999). "Do Better Schools Matter? Parental Valuation of Elementary Education." The Quarterly Journal of Economics **114**(2): 577-599.
- Hansen, B. (2015). "Punishment and Deterrence: Evidence from Drunk Driving." American Economic Review **105**(4): 1581-1617.