PHS 2000
Econometric Methods 2022
Lecture 3
Regression Discontinuity Design

Maggie McConnell

Department of Global Health and Population

Harvard T. H. Chan School of Public Health

Plan of presentation

1. Explaining Regression Discontinuity Designs (RDD)
2. Key assumptions
3. Testing the assumptions (where possible)
4. Model choices in estimation (lots of choices)
5. Examples (the importance of pictures)
6. Some practical tips

Unobserved variables and observational biases

- Recap: omitted variable bias, endogeneity can be a big problem in estimating causal effects

- What can we do?
  - RCTs – but RCTs are costly, not always desirable and may not be the right design for questions around local, state or federal policy
  - IV estimation – but good instruments are not easy to find and IV only valid under very strict assumptions and face challenges with external validity

Regression Discontinuity Designs (RDD) -- intuition

- RDDs offer an alternative to estimate causal effects in observational data

- The idea is relatively simple: for many variables of interest, institutional rules or laws have generated *arbitrary* eligibility criteria, which means that the probability of receiving exposure to a given variable "jumps" at certain points in a continuous distribution

- These "jumps" or discontinuities provide us with quasi-experimental variation that we can explore to pin down causal effects of interest

- Econometrics by graphs

Example 1 for discontinuities: Scholarship

- In general, scholarships are strongly correlated with other variables of interest such as academic performance or parental background
  - Very hard to determine the impact of scholarships on later life success
- In many states, scholarship rules are very strict
  - For example, if GPA > x, a student gets the scholarship
- If we compare students just below the cutoff to students just above the cutoff, the only difference in their later life outcomes should be scholarship since both groups should be otherwise similar

Example 2A: Guidelines -- birthweight

- RDD designs can even work in settings where there is a sufficiently powerful *informal or non-binding* rule
  - This is especially true in medical care settings where non-binding guidelines common
- Example: Very low birth weight (birth weight > 1500) for infants
  - No sharp discontinuity in medical risk, more of a “convention”
- By comparing infants just below this threshold to patients just above it, we can infer how much “more intensive” care for infants affects outcomes
  - What outcomes would you be interested in?

Example 2B: Guidelines – pregnancy age

- Another example: Advanced maternal age (age>35) for pregnancy ("geriatric pregnancy")
  - No actual discontinuity in medical risk
  - Threshold used to decide who should get genetic testing in 1979
  - Now commonly used as rule of thumb / billing justification
- By comparing patients just below this threshold to patients just above it, we can infer how much "more intensive" care during pregnancy affects pregnancy outcomes
  - What outcomes would you be interested in?

Many other examples of RDD in the literature

- Question: What is the impact of a longer postpartum stay on hospital readmission?
- Running variable: time
- Cutoff: midnight

- Question: What is the impact of aging into Medicare on racial / ethnic differences in access to health care?
- Running variable: age
- Cutoff: 65

- Question: What is the impact of having a female mayor on mortality from COVID-19?
- Running variable: female vote margin
- Cutoff: 0%

RDD – how it works

- Treatment is discontinuous in a continuous “running variable”
  - i.e. GPA, birth weight, vote share, age, time

- We examine only individuals in a small window around the cutoff and compare treated and controls
  - Why only in a small window?

- If confounders are continuous in the running variable, they should be balanced on treated and control in the small window

Under what conditions can we estimate RDD?

Conditions

- A continuous eligibility index: a continuous measure on which the population of interest is ranked (i.e. test score, health indicator age).

- A clearly defined cutoff point: a point on the index above or below which the population is determined to be eligible for the program
  - Sometimes referred to as the “threshold”

What assumptions required for RDD to be valid?

Key assumption

The only thing that differs for groups above and below the threshold is the likelihood of program participation/eligibility

Implications of this assumption

- Individuals can't be able to manipulate the running variable in order to increase the chances of being included / excluded
- Individuals close to the cutoff point should be very similar, on average, in observed and unobserved characteristics

Stop for questions

RDD, RCTs, and IV estimation

- Under ideal conditions, RDDs are very similar conceptually to RCTs: if it is true that treatment assignment is close to “random” within the analyzed range, then RDDs become a quasi-experiment

- RDDs can also be seen as a special case of IV estimation
  - Under the RDD assumption, we know that there is an increase in the treatment at the cutoff, and we explore this exogenous variation in the treatment to identify the causal effect of interest

- **Key question:** How plausible is it that the cutoff affects the outcome only through treatment?
  - Recall, this is the similar to the exclusion assumption in an IV setting

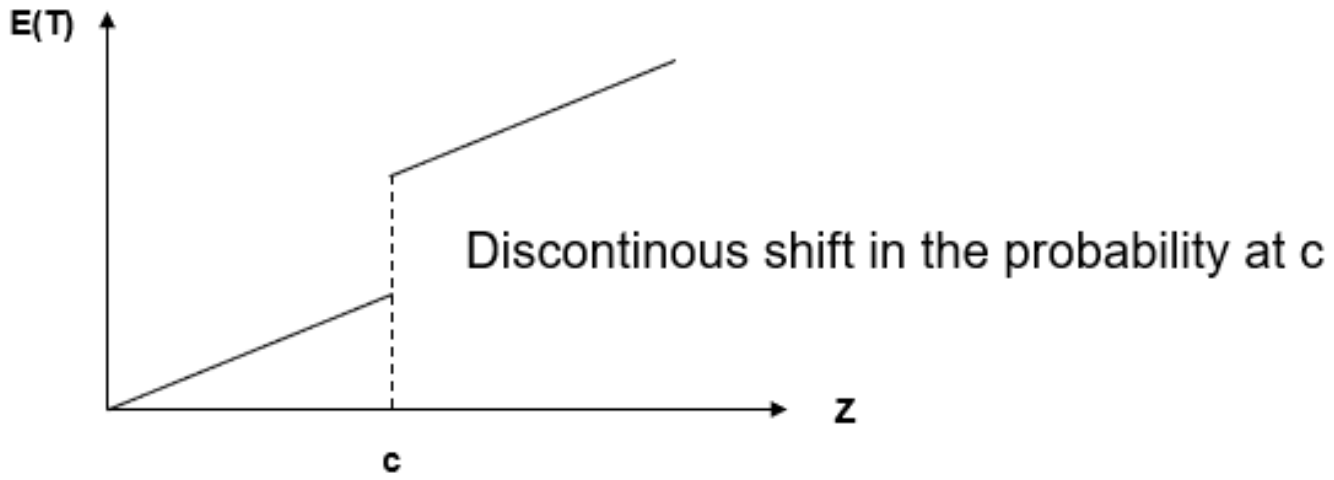
How do we feel about the identifying assumption in these cases

- How do we feel about the identifying assumption in these cases?
  - LBW designation for more intensive newborn care and follow-up
  - Advanced maternal age for more “intensive” medical treatment
  - Births around midnight for longer postpartum stay
  - Close election for representation by a Republican candidate

The critical threshold

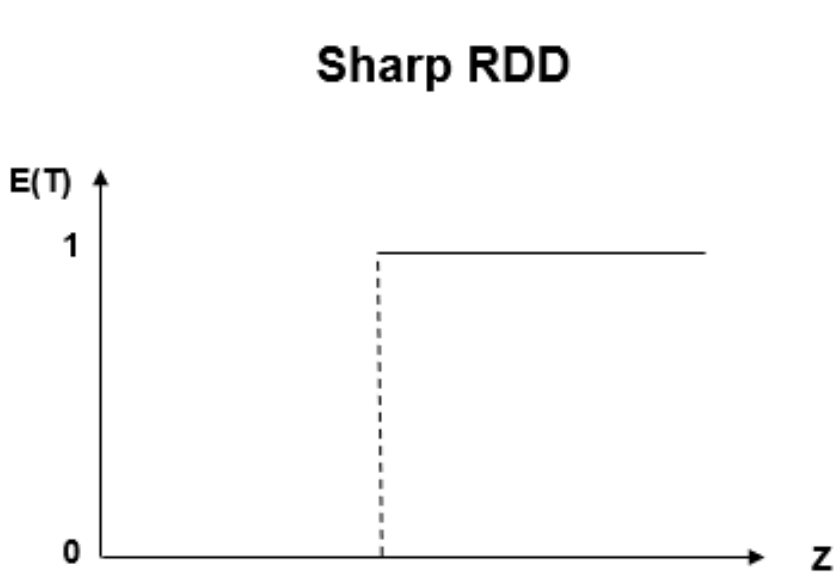
- Assume that there is a continuous variable z which determines eligibility for treatment discontinuously at some cutoff (c)

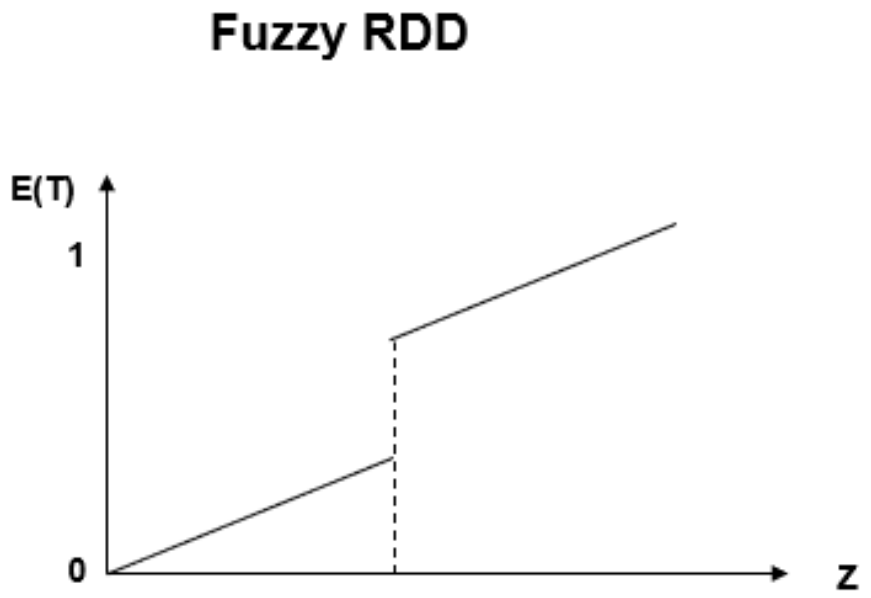
$$\lim_{Z \downarrow c} \Pr(T_i = 1) \neq \lim_{Z \uparrow C} \Pr(T_i = 1)$$



Sharp discontinuities vs. fuzzy discontinuities

- With sharp discontinuities, the probability of receiving treatment at the cutoff goes from 0 to 1
- With fuzzy discontinuities, the change in probability of receiving treatment at the cutoff is less than 1





Sharp or fuzzy

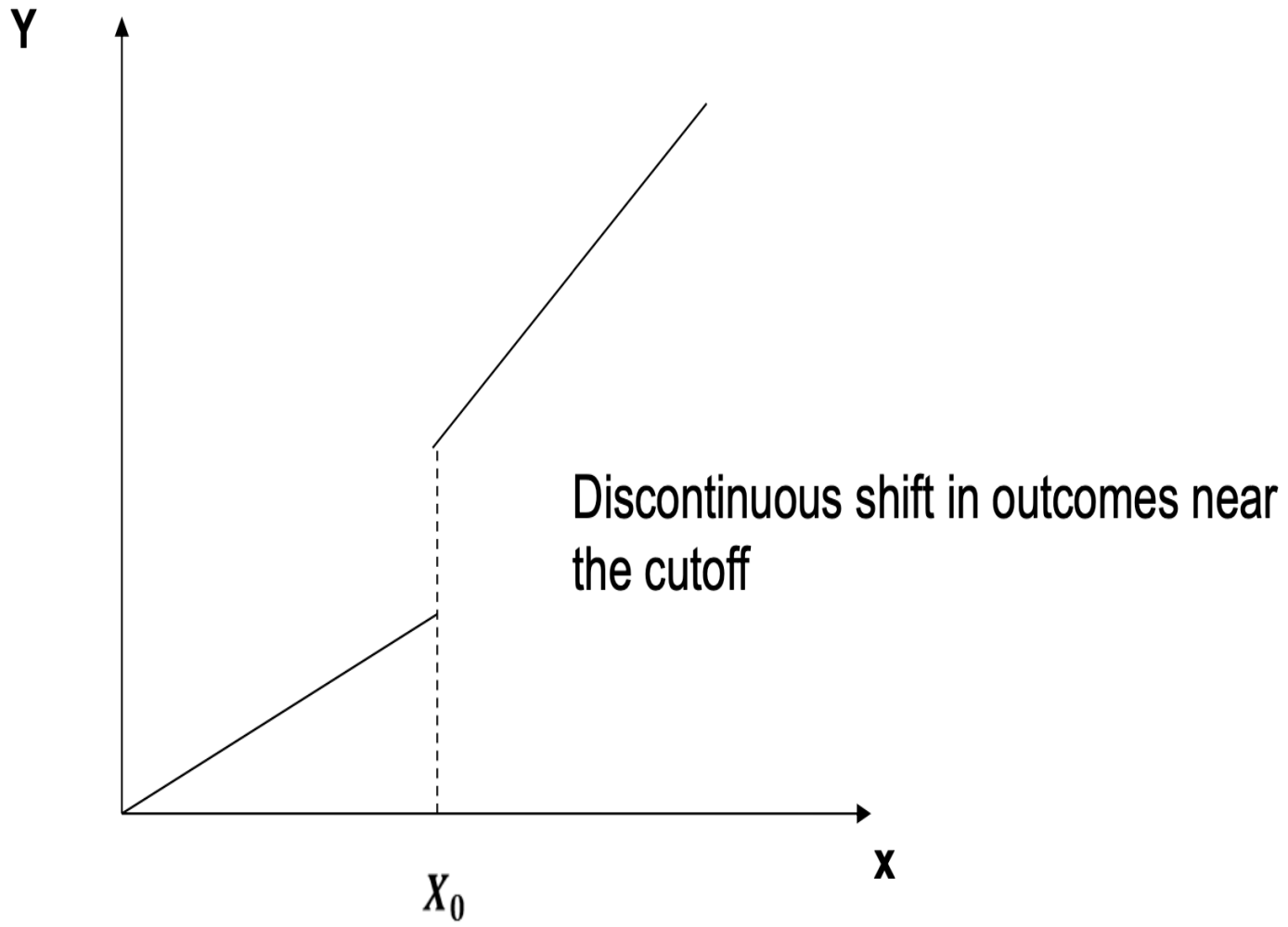
- Would these be sharp or fuzzy RDD?
  - LBW designation for more intensive newborn care and follow-up
  - Advanced maternal age for more “intensive” medical treatment
  - Births around midnight for longer postpartum stay
  - Close election for representation by a female political leader

Sharp or fuzzy

- Would these be sharp or fuzzy RDD?
  - *LBW designation for more intensive newborn care and follow-up*
  - *Advanced maternal age for more "intensive" medical treatment*
  - *Births around midnight for longer postpartum stay*
  - **Close election for representation by a female political leader**

Look for “discontinuities” in outcomes

- If we think we have met the conditions for valid RDDs we can look for “jumps” in outcomes
- We could observe shift in intercept **and** change in slope of relationship



Identifying the causal effect

- We want to know the causal effect of getting “more” of the treatment

- We should be able to get this by comparing outcomes just below and just above the threshold
  - Under the assumption that once we are sufficiently close to the cutoff, treatment assignment is “quasi-random”

- The difference in mean outcomes (called the Average Causal Effect or ACE) at the threshold is given by:

$$ACE = \lim_{Z \uparrow c} E[Y_i(1) | Z_i = c] - \lim_{Z \downarrow c} E[Y_i(0) | Z_i = c]$$

Interpreting sharp vs. fuzzy designs

- With sharp designs, the difference between the observed outcomes directly provide us with the treatment effect of interest

- With fuzzy designs, the probability of getting the treatment does not increase by 1 but by some value p where $0 < p < 1$
  - Therefore, to get the impact of receiving one "full unit" of treatment, we need to scale the estimated difference in the outcome at the cutoff by the change in probability of treatment receipt

Complier Average Causal Effect (CACE)

- To get the marginal impact of a unit increase, we can divide the average difference in the outcome (Y) by the average difference in the probability of treatment receipt ($\Pr(T = 1)$)

$$CACE_{RDD} = \frac{\lim_{Z \uparrow c} E[Y_i(1) | Z_i = c] - \lim_{Z \downarrow c} E[Y_i(0) | Z_i = c]}{\lim_{Z \uparrow c} \Pr[T_i = 1 | Z_i = c] - \lim_{Z \downarrow c} \Pr[T_i = 1 | Z_i = c]}$$

- This is called the Complier Average Causal Effect (CACE) because it estimates the causal impact of the treatment among those subjects whose treatment status changes around the threshold (i.e., the comply with official protocols)
- Same as adherence adjustment in RCT
  - Requires same assumptions as estimation of LATE with IV

Conceptual challenge 1: Plausibility of identifying assumption

- The critical concept underlying RDD is that probability of treatment assignment jumps at a given threshold

- This is plausible in many scenarios, but will be violated if:
  - Subjects can manipulate the assignment variable
  - Program administrators can manipulate the assignment variable

- Quasi-random assignment assumption is more likely to be plausible if subjects don't know their score / only have a noisy signal
  - We can check this assumption empirically by looking at density distributions of the running variable around the threshold – should be smooth

- Goodhart's law: any measure used for policy becomes distorted

Conceptual challenge 2: External validity

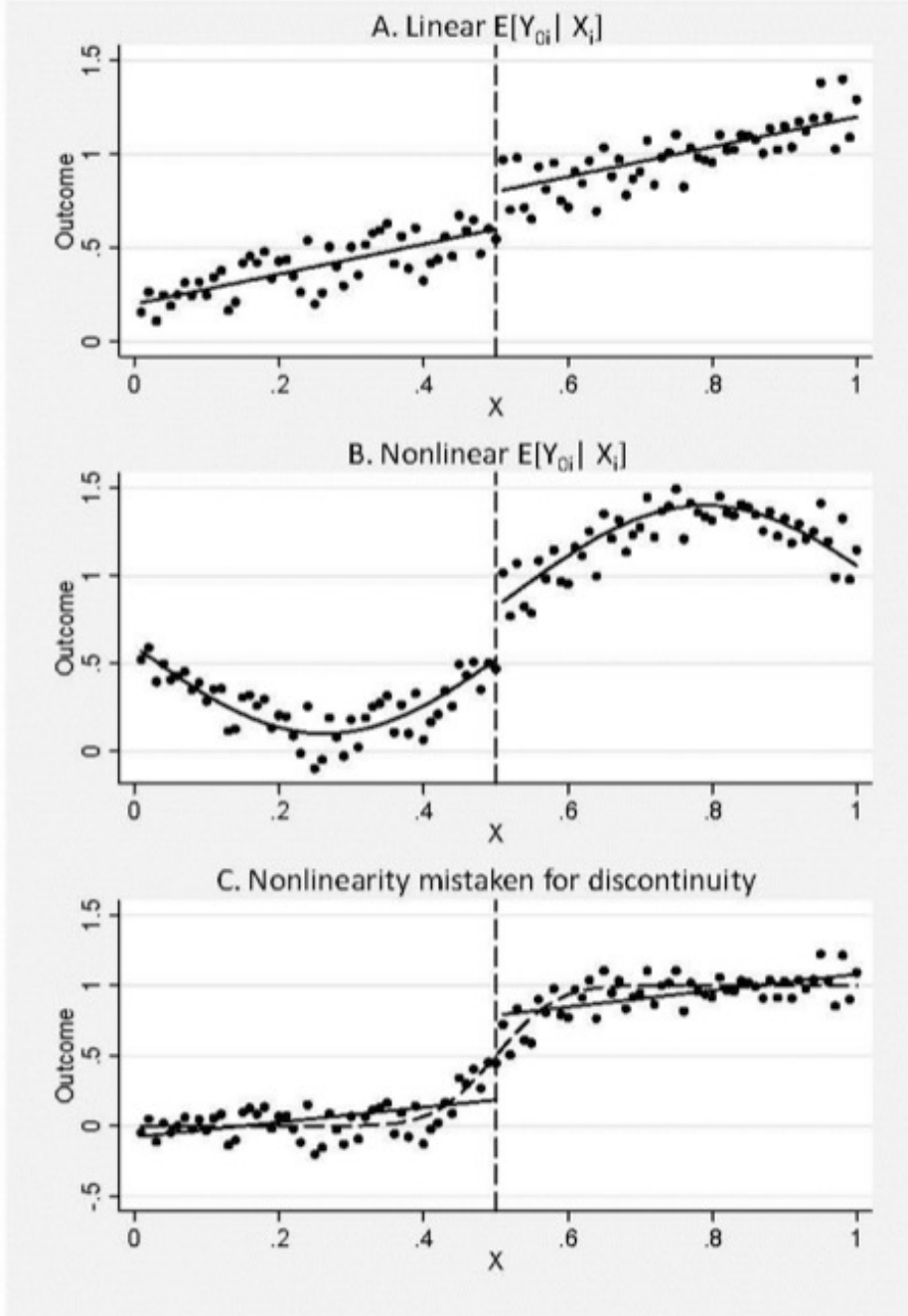
- RDDs focus on causal effects for very specific populations close to the threshold (e.g. well-performing students close to getting scholarships)

- The ACE estimator will give us the average causal impact among subjects of this “type”, which may not be representative of the ATE in the population

- The CACE is even more restricted: it will give us the average impact among subjects of this type who “comply” with policy, i.e., do not get the treatment below the threshold and get it above

Conceptual challenge 3: Bandwidth and counterfactual outcomes

- Estimate gives the average causal effect as we limit the sample to those very close to the threshold
- But we need to have a reasonable bandwidth to have a decently sized sample and reasonable standard errors
- As we widen the bandwidth the there is greater potential for confounding –consistency result only holds in the limit as bandwidth tends to zero
  - With a wider bandwidth some modeling of functional form might come into play
- Very important to have the correct functional form relationship between X and Y or else we could “find a discontinuity” where there isn’t one...


A. Linear E[Y0i| Xi]
Outcome
X
B. Nonlinear E[Y0i| Xi]
Outcome
X
C. Nonlinearity mistaken for discontinuity
Outcome
X

Avoiding mis-specification

- Best way to prevent fake discontinuities is to estimate RD in a small bandwidth around the cutoff
  - In this region, functional form assumptions matter less
- With a larger bandwidth you might use a more complex model
  - If confounders vary smoothly with the running variable (Z), we can model their effect as a continuous function of the running variable ($f(z)$)
  - This function can be different above and below the cutoff due to interaction with the treatment
  - We can approximate $f(z)$ by a polynomial or by a nonparametric curve
- Could also use a larger bandwidth but weight observations near the cutoff more heavily

Practical Tips: Assessing the Validity of RDD

1) Show characteristics (SES, age...) are balanced around cutoff.
   - One way to do this is to estimate the same main regression as RD but your "outcomes" are the covariates. Look for jumps near the cutoff.
   - Similar to balance tables in RCTs
2) Estimate causal effect using different bandwidths and check stability of estimates — instability suggests wrong functional form
3) Look for "bunching" in the running variable around cutoff
   - Should not be extra mass in distribution near cutoff
   - Can use McCrary 2008 density test to test this formally
4) Estimate "false cutoff"
   - Should not find significant effects at the wrong cutoff

Example: “Very Low Birth Weight”

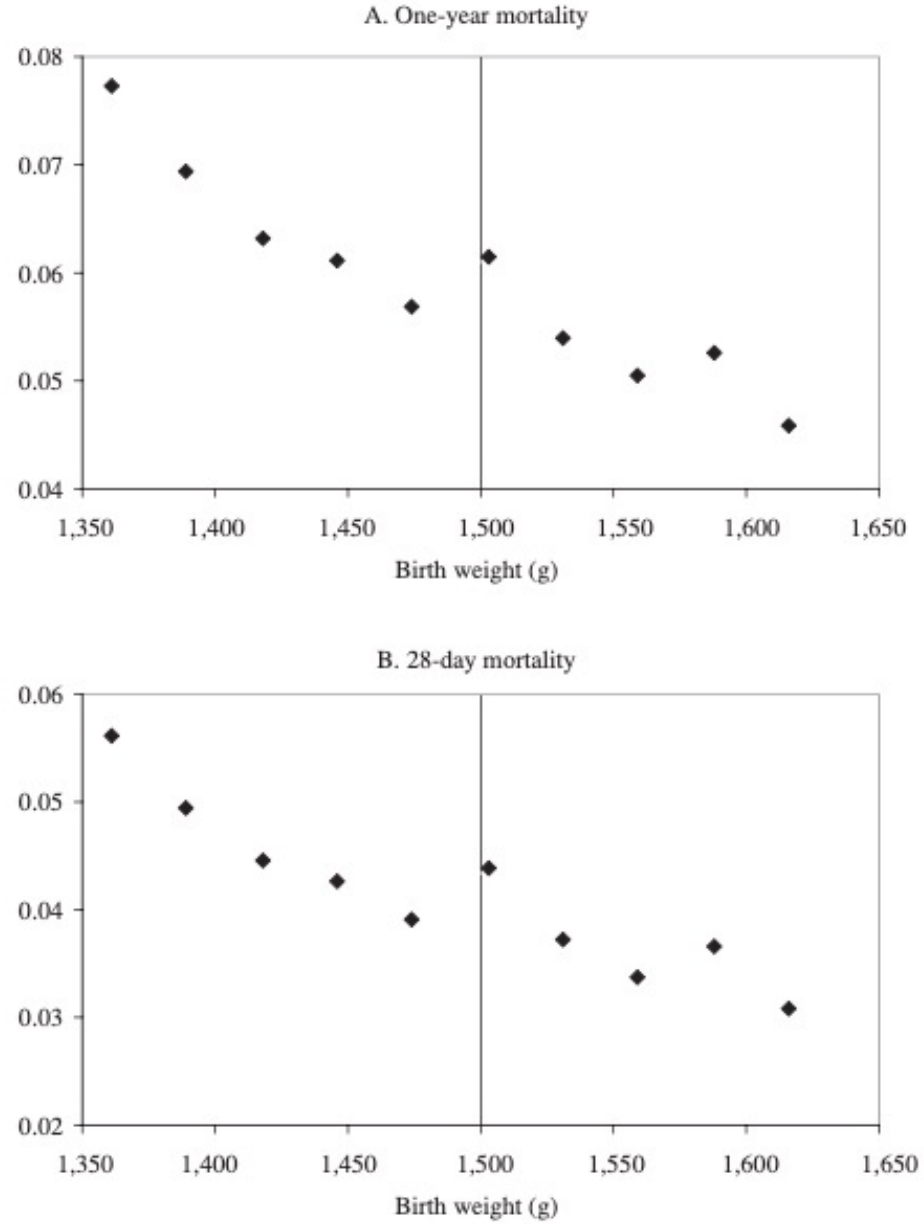


FIGURE II
One-Year and 28-Day Mortality around 1,500 g
NCHS birth cohort–linked birth/infant death files, 1983–1991 and 1995–2003, as described in the text. Points represent gram-equivalents of ounce intervals, with births grouped into one-ounce bins radiating from 1,500 g; the estimates are plotted at the median birth weight in each bin.

Understanding mechanisms

Who are the "compliers"?

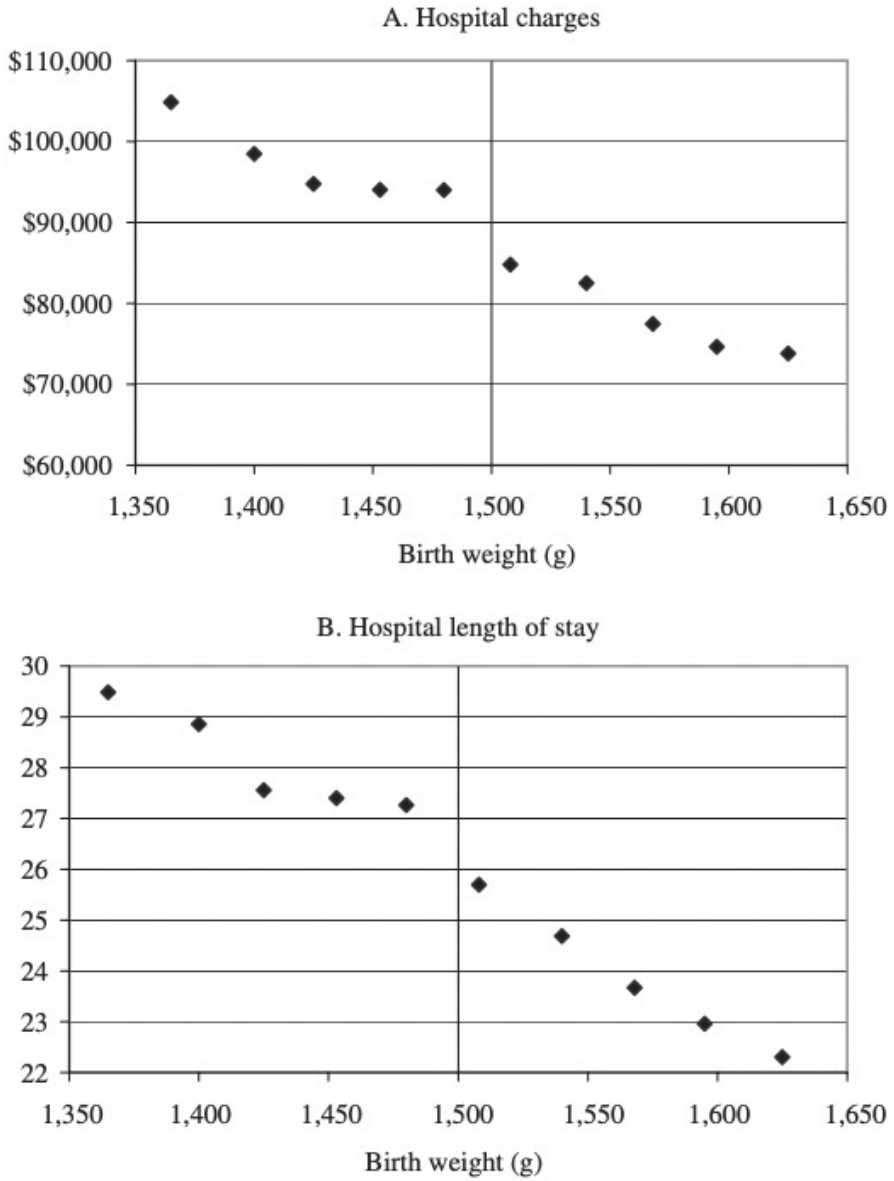


FIGURE III
Summary Treatment Measures around 1,500 g

Data are all births in the five-state sample (AZ, CA, MD, NY, and NJ), as described in the text. Charges are in 2006 dollars. Points represent gram-equivalents of ounce intervals, with births grouped into one-ounce bins radiating from 1,500 g; the estimates are plotted at the median birth weight in each bin.

Long-term follow-up

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

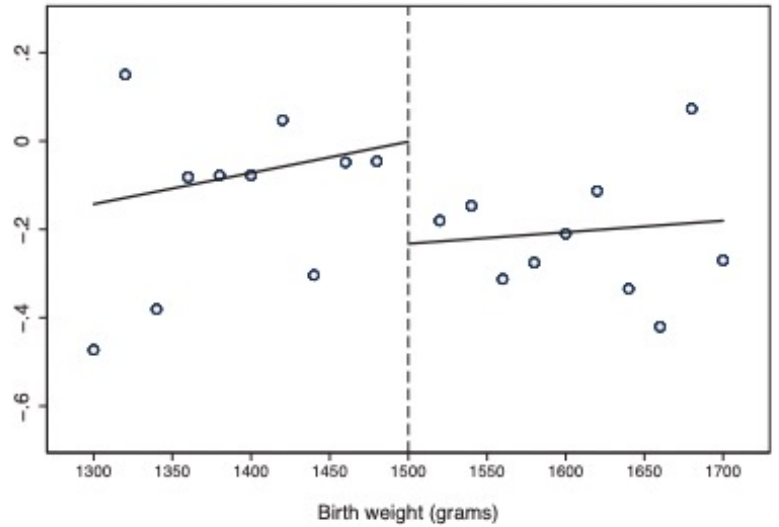
The returns to early-life interventions for very low birth weight children

Eric Chyn[a,*], Samantha Gold[b], Justine Hastings[c]

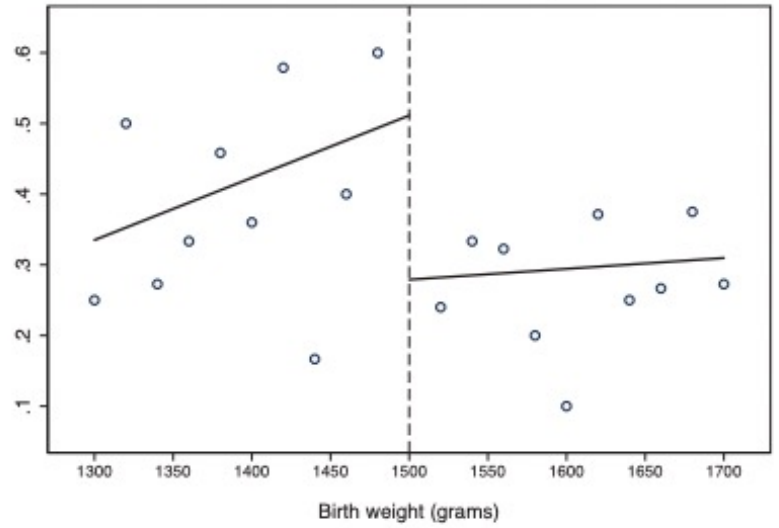
What would be the mechanism for these effects?

Figure 1: Impacts for Selected Education Outcomes

A. Test Scores (3-8)



B. 4-Year College Enrollment by 22



Notes: Each panel shows the relationship between birth weight and a selected education outcome. Dots represent means within 20 gram bins of the running variable. The dark lines are predictions from a linear model using the individual-level data.

Distribution of birth weight by grams

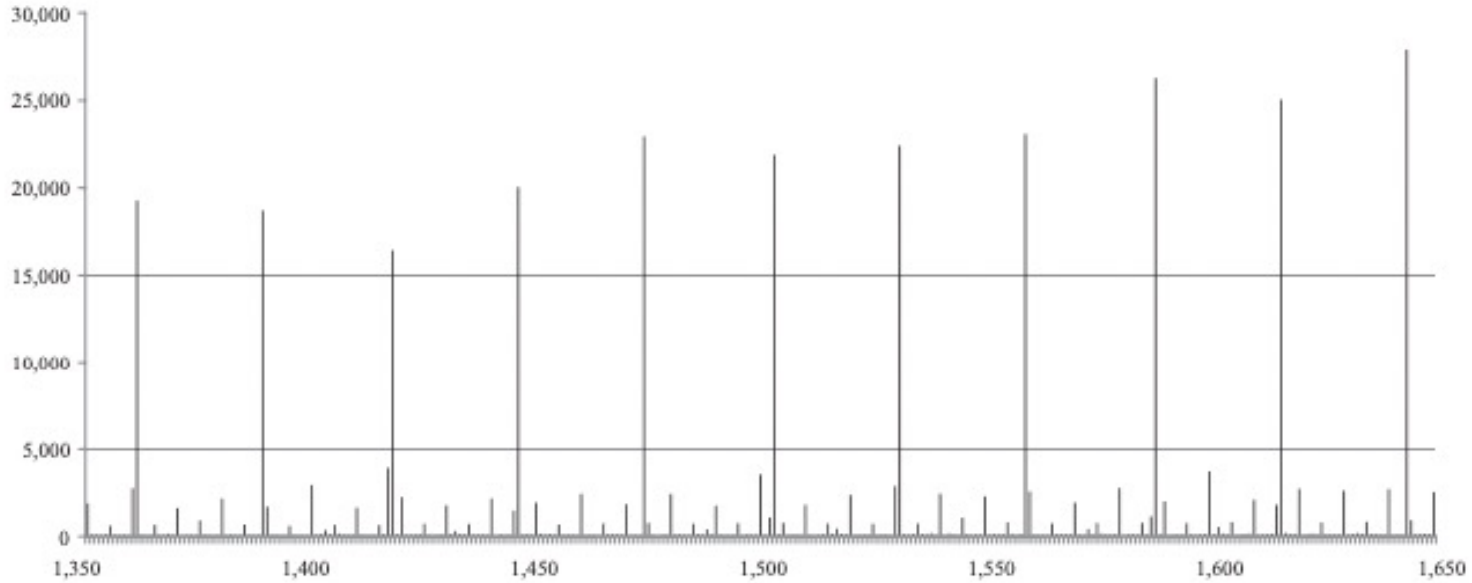
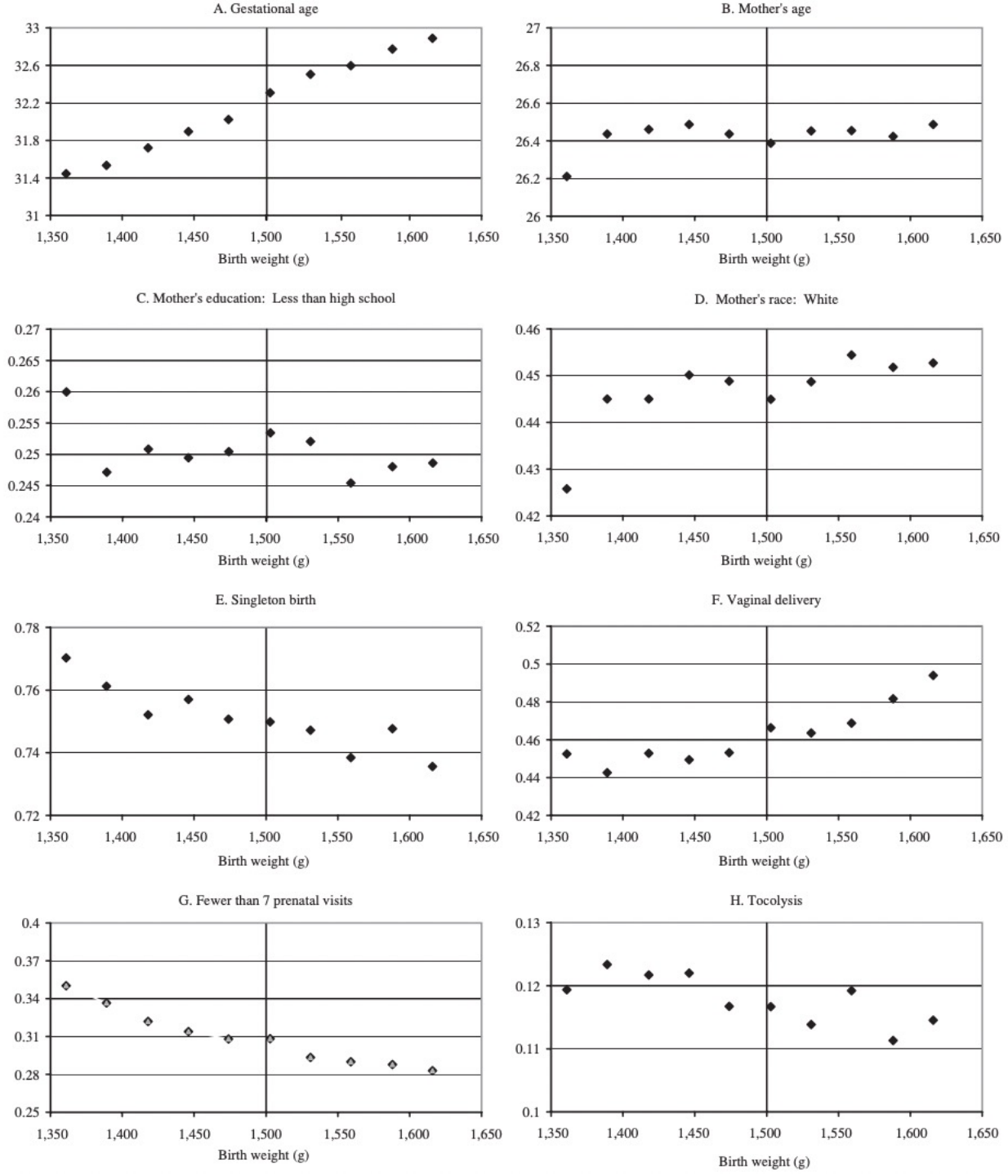


FIGURE I
Frequency of Births by Gram: Population of U.S. Births between 1,350 and 1,650 g
NCHS birth cohort linked birth/infant death files, 1983–1991 and 1995–2003, as described in the text.

Maternal characteristics



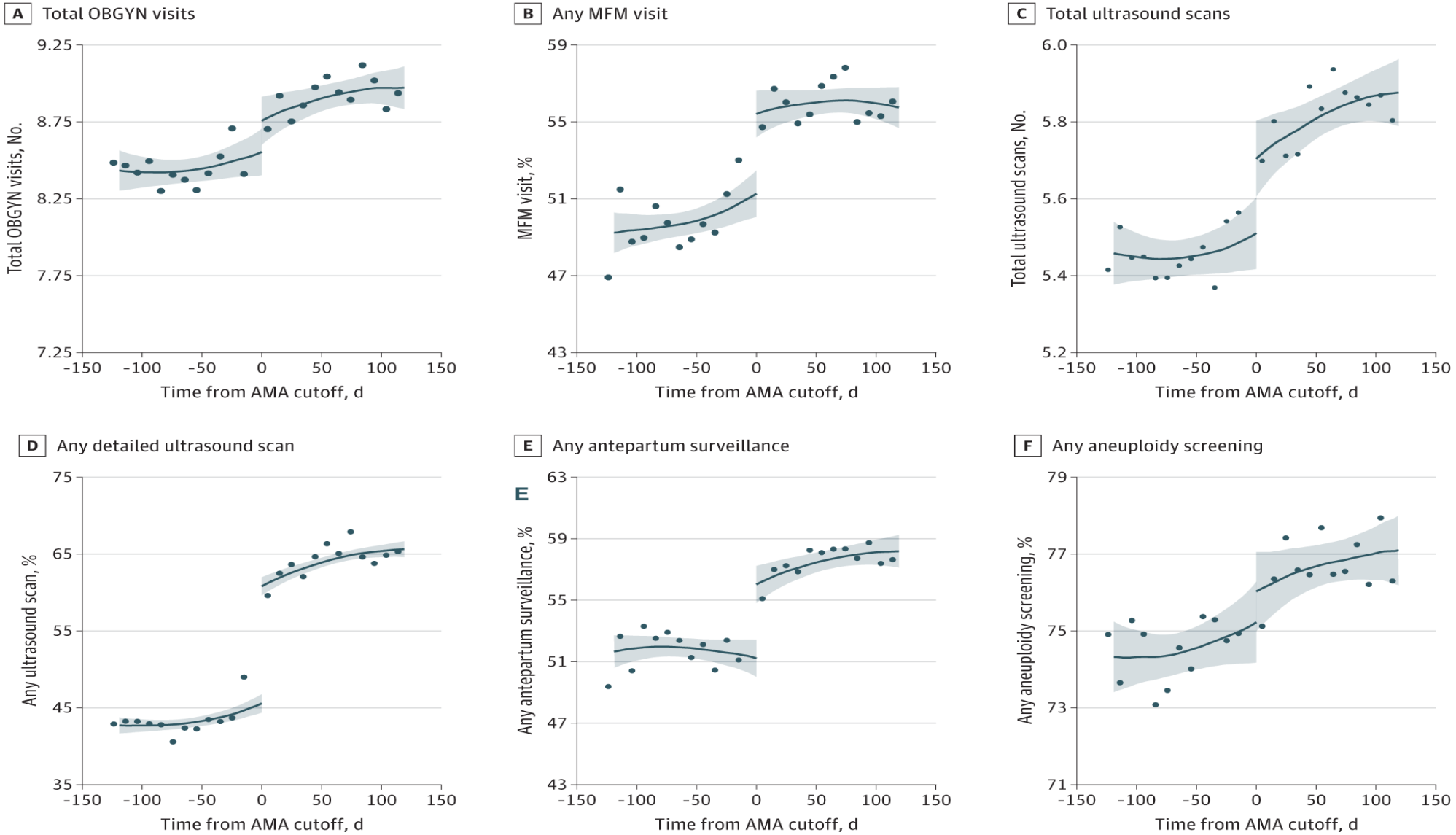
Another specification concern: heaping

SAVING BABIES? REVISITING THE EFFECT OF VERY LOW BIRTH WEIGHT CLASSIFICATION

Alan I. Barreca, Melanie Guldi, Jason M. Lindo and Glen R. Waddell

- Heaping: when a continuous variable because discrete rounding of numbers
  - Could be strategic or could be accidental
- Potential solution – donut RDD where observations in extremely close proximity to threshold are removed and RDD is re-estimated

Example A: “Advanced Maternal Age”

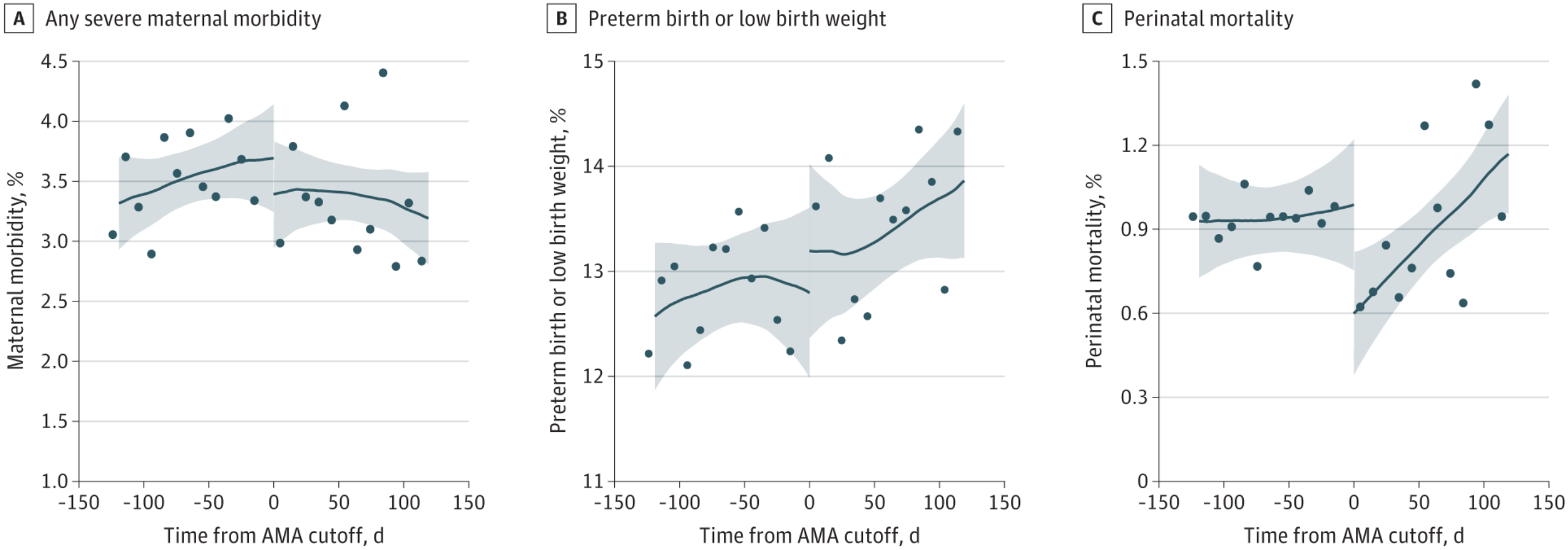


Geiger et al 2021

Estimate results for all births and for subgroup of ”low-risk births” – why?

Who are the “compliers” here?

Example A: “Advanced Maternal Age”

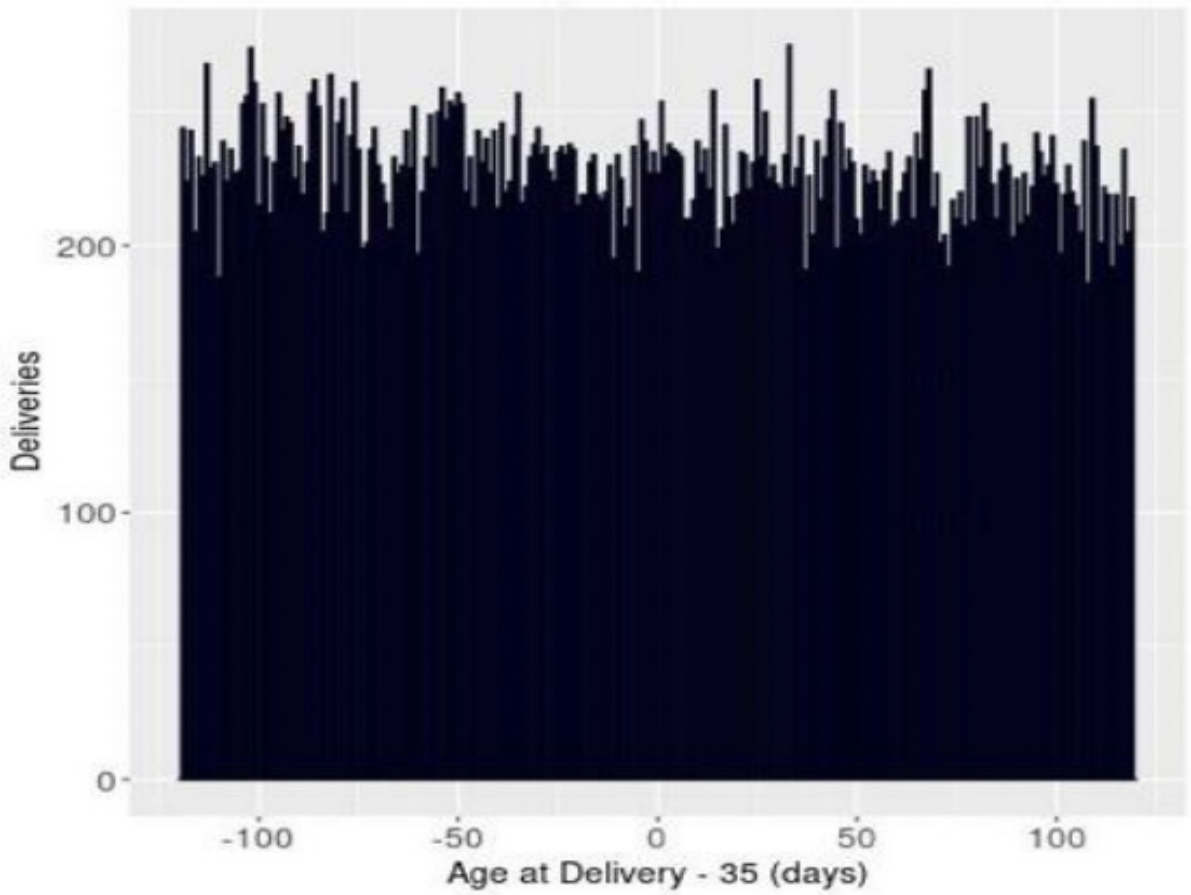


Sensitivity tests

- Check smoothness of the running variable
- Check for differences in maternal, infant, and zip code or county characteristics across the cutoff, why?
- Check for changes in the likelihood of a pregnancy ending in termination or miscarriage at the cutoff, why?
- Checked sensitivity to bandwidth selection (90, 120, 150 days)
- Checked that results were unique to the 35-year cutoff by running placebo tests using cutoffs above and below 35 years, why?

Example: “Advanced Maternal Age”

eFigure 4. Histogram of All Deliveries to Individuals Within 120 Days of the AMA Cutoff



Legend: Shown is the number of deliveries by the running variable, i.e., the number of days between the expected date of delivery and the individual's 35th birthday. Sample includes all individuals with an expected date of delivery within 120 days of the 35th birthday.

Original Investigation

December 3, 2021

Association of Prenatal Care Services, Maternal Morbidity, and Perinatal Mortality With the Advanced Maternal Age Cutoff of 35 Years

Caroline K. Geiger, PhD[1,2]; Mark A. Clapp, MD, MPH[3]; Jessica L. Cohen, PhD[4]

» Author Affiliations | Article Information

JAMA Health Forum. 2021;2(12):e214044. doi:10.1001/jamahealthforum.2021.4044

Maternal characteristics around threshold

eTable 6. Test for Changes in Sample Characteristics at Age 35 Cutoff

	Full Sample (N = 51,290)	
Outcomes	**Coefficient (95% CI)**	**P-value**
Infant characteristics		
Trisomy 21	0.001 (0.000, 0.003)	0.107
Maternal characteristics		
Any high-risk pregnancy diagnoses	0.003 (-0.017, 0.023)	0.771
Chronic hypertension	0.003 (-0.006, 0.012)	0.469
Pregestational diabetes	-0.006 (-0.015, 0.003)	0.167
Obesity	0.002 (-0.011, 0.014)	0.809
Multiple gestation	0.003 (-0.006, 0.012)	0.469
Zip code/county of residence characteristics		
Median household income	-490.109 (-1699.641, 719.422)	0.427
% White	-0.074 (-0.774, 0.625)	0.835
% Hispanic	-0.019 (-0.581, 0.542)	0.946
Urban RUCA	0.003 (-0.012, 0.019)	0.677
Any NICU	0.001 (-0.014, 0.017)	0.891
OBGYNs per 10,000 deliveries	1.192 (-1.444, 3.828)	0.376

Abbreviations: CI = confidence interval; RUCA = rural-urban commuting areas; NICU = neonatal intensive care unit; OBGYN = obstetrician-gynecologist

Notes:

Sample includes all individuals with an expected date of delivery within 120 days of her 35th birthday. Individuals with an expected date of delivery within 7 days of the 35th birthday were excluded. All regressions for maternal and infant characteristics control for zip-code characteristics (percent white, percent Hispanic, median household income, and whether the zip code is urban) and county-level characteristics (any hospital with neonatal intensive care unit and OBGYNs per 10,000 deliveries). All regressions for zip code/county of residence control for chronic hypertension, pregestational diabetes, obesity, and multiple gestation. All regressions include state of residence, year, and month of delivery fixed effects.

Age as a running variable – special issues

- Some policies are age dependent – old age pension, Medicare eligibility
- Age cannot be manipulated – but can people *report* false ages to get access
- Several policies may change at same age
  - Age 65: Medicare, some pensions, senior discounts.
- Change at 65 is not random and is foreseeable. We may see actions below 65 in anticipation of treatment at age 65
  - Consumption smoothing of pension income
  - Delaying health care utilization until Medicare access
- Is that likely to be a concern in the case of “advanced maternal age”?

Strength of RDDs

- RDD models have many potential applications
- The assumption that treatment assignment is discontinuous around a very specific threshold is plausible in many settings and can be tested
- The degree to which populations are the same to the “left” and “right” of the threshold can be tested empirically – test balance of observables
- RDD estimate the causal effect of treatments in practice, which may be very different from the controlled environments generated for RCTs

IV vs. RDD

- Both IV and RDD strategies explore random variations generated by specific programs or cutoffs

- While this variation could be randomly distributed across the entire population in theory (like the variation in exposure generated by an RCT), most IVs and RDD explore variations generated by government laws or other quasi-random rules

- If treatment impact is heterogeneous, the estimates are not necessarily the average treatment effects for the population.
  - We refer to RDD estimates as ACE or CACE, and to IV estimates as LATE.

- Big advantage of RDD – relies on continuity of missing covariates with z, not on no correlation between covariates and z.

Formal RDD Assumptions

- Requires takings limits as we narrow the window around the threshold.
- Otherwise fairly easy if we assume treatment effect is homogenous (ACE)
  - Continuity of outcome at threshold if untreated – no natural jump
  - Discontinuity in treatment at threshold
- If effect is heterogeneous we need more assumptions
  - Assuming treatment effect continuous in z and probability of treatment is independent of size of treatment effect we can estimate complier average causal effect (CACE)
- See additional slides for details

RDD summary

- RDD offers a conceptually very elegant approach to identify causal effects in observational data

- Like IV, RDD often identifies treatment effects for particular subpopulations, those near the cutoff, which may not be universally applicable (i.e., external validity may be limited)

- Theory only holds as bandwidth and window gets very small; however, under such a scenario, the sample size goes to zero as well. Therefore, some modeling is required for finite bandwidths

Advanced topics in RDD

- The bandwidth choice is flexible
  - Try different bandwidths and plot estimates (and standard errors) as a function of bandwidth
  - Optimal bandwidth selection on some criteria such as confidence intervals is available. This may prevent cherry picking results
- Degree of polynomial in running variable around bandwidth is flexible
  - With very small bandwidth, linear form is common; with wider bandwidths, higher order polynomials or non-parametric function should be estimated as well
- Method of calculating standard errors is flexible
  - Different formulae have been developed – at least 4.
- Lots of technical issues, but graphs are key. We should be able to see the results in pictures to be convincing!

Additional Slides

Recommended readings on approach

1. Bor, J., Moscoe, E., Mutevedzi, P., Newell, M. L., & Baernighausen, T. (2014). Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology, 25*(5), 729-737

2. Lee DS, Lemieux T. Regression discontinuity designs in economics. *Journal of Economic Literature* 2010;48:281–355

3. Robin Jacob, Pei Zhu, Marie-Andrée Somers, Howard Bloom. A Practical Guide to Regression Discontinuity [link]

Recommended readings for application of RDD

1. Almond, Douglas, Doyle, Joseph J., Kowalski, Amanda E., & Williams, Heidi. (2010). “Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns”. *The Quarterly Journal of Economics, 125*(2), 591-634. doi: 10.1162/qjec.2010.125.2.591

2. Michael Anderson & Carlos Dobkin & Tal Gross, 2012. "The Effect of Health Insurance Coverage on the Use of Medical Services," American Economic Journal: Economic Policy, American Economic Association, vol. 4(1), pages 1-27, February

Recommended readings for application of RDD

3. Anderson, Michael, & Magruder, Jeremy. (2012). Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database*. *The Economic Journal, 122*(563), 957-989. doi: 10.1111/j.1468-0297.2012.02512.x

4. Lee, David S., Moretti, Enrico, & Butler, Matthew J. (2004). Do Voters Affect or Elect Policies? Evidence from the U. S. House. *The Quarterly Journal of Economics, 119*(3), 807-859. doi: 10.1162/0033553041502153

RDD Formal Assumptions

- Outcome y, treatment x (either 0 or 1) , running variable z

$$y_i = \alpha_i + x_i \beta_i$$

$$y_i(z_i, w_i, x_i) = \alpha_i(z_i, w_i) + x_i(z_i, w_i)\beta_i(z_i, w_i)$$

- Outcome a function of running variable, treatment, and covariates
- Treatment can depend on running variable and covariates
- General formulation allows for heterogeneous treatment effect

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and estimation of treatment effects with a regression-discontinuity design." Econometrica 69.1 (2001): 201-209.

RDD Formal Assumptions

ASSUMPTIONS:

(i) RD $x^+ = \lim_{z \to z_0^+} E[x_i \,|\, z_i = z], x^- = \lim_{z \to z_0^-} E[x_i \,|\, z_i = z]$ exist $x^+ \neq x^-$

(ii) Assumption A1 $E[\alpha_i \,|\, z_i = z]$ is continuous at $z = z_0$

Homogeneity and Heterogeneity

- Theorem: If we have RD and A1 and homogeneity $\beta_i = \beta$ a constant then

$$\beta = \frac{y^+ - y^-}{x^+ - x^-} \quad where \quad y^+ = \lim_{z \to z_0^+} E[y_i | z_i = z], y^- = \lim_{z \to z_0^-} E[y_i | z_i = z]$$

- Assumption A2 $E[\beta_i | z_i = z]$ is continuous at $z = z_0$

- Theorem: If we have RD, A1, A2 and x_i is independent of β_i given z_i then

$$E[\beta_i | z_i = z_0] = \frac{y^+ - y^-}{x^+ - x^-}$$

Exogeneity and independence assumptions not needed in Fuzzy Regression Discontinuity

- We do not need additional exogeneity and independence assumptions
- The running variable z can affect the outcome directly but it most do so in in continuous way by assumption A1. A jump in y at the regression discontinuity must be the result of an effect through x rather than a direct effect.
- There maybe omitted variables correlated with both the running variable z and the outcome y (non –independence). However by A1 they lead to only continuous changes in the outcome as z changes. They cannot explain a jump in y at the discontinuity.

Independence of treatment with respect to treatment effect

- Probability of treatment not correlated with expected treatment effect
- Violated if those with larger treatment effects more likely to be treated – selection into treatment
- Eg scholarship as a function of test scores and going to university
  - Those with scholarship more likely to go at threshold
  - We see earnings effects at threshold
  - Those that go and take up the scholarship may select on have large returns to education due to higher earnings

However probability of treatment may be correlated with expected gain

- Assumption A3

(i) $(\beta_i, x_i(z))$ is jointly independent of z near z_0

(ii) $There\ exists\ \varepsilon > 0, such\ that\ x_i(z_0 + e) \geq x_i(z_0 - e), \forall e \leq \varepsilon$

- (i) Means treatment effect may be correlated with treatment but effect of treatment for each individual does not vary with z near the discontinuity
  - Example multiple subgroups each with different constant treatment effects beta and different probabilities of being compliers
- (ii) Monotonicity – implies no defiers in the neighborhood of the discontinuity

Local Average Treatment Effect

- Suppose assumptions RD, A1, and A3 hold then

$$\lim_{e \to 0^+} E[\beta_i \,|\, x(z_0 + e) - x(z_0 - e) = 1] = \frac{y^+ - y^-}{x^+ - x^-}$$

- Under these assumptions the effect of treatment in a fuzzy RDD is the average treatment effect on compliers at the cutoff.
- Example treatment effect on women different than men (both constant) then the LATE depends on the ratio of women to men among compliers.

Multiple running variables

- Two approaches (assume sharp RDD)
- Approach 1: Define single running variable by distance (in some metric) from the threshold border with positive distance above and negative distance below threshold. Do all analysis in terms of this single running variable
- If cutoffs are of the form treat if $z1>z1^*$ or $z2>z2^*$ we can define
  - $z = \max\{z1-z1^*, z2-z2^*\}$ with a threshold at $z^*=0$
- Approach 2: then we can do two separate RDD
  - Estimate effect at $z1^*$ for $z2<z2^*$
  - Estimate effect at $z2^*$ for $z1<z1^*$
- Each approach estimates different ATE
  - Estimate effect at $z1^*$ for $z2<z2^*$ - ATE crossing boundary $z1^*$ given $z2<z2^*$
  - Estimate effect at $z2^*$ for $z1<z1^*$ - ATE crossing boundary $z2^*$ given $z1<z1^*$
  - Single running variable gives an average of these two effects weighted by the proportion of times we cross the threshold on each boundary