An Introduction to Regression Discontinuity Design

Laura Wherry
Assistant Professor
Division of GIM & HSR

RCMAR/CHIME Methodological Seminar

November 20, 2017

Introduction to Regression Discontinuity Design (RDD)

- A relatively *new* nonexperimental research approach (emerged in late 1990s in economics)
- Strong research design that approximates random assignment, potentially more credible than other quasi-experimental approaches
- Yet to be fully taken advantage of in the health care research setting
  - See BMJ article by Venkataramani, Bor and Jena (2016)

What Is RDD?

- Estimates treatment effect in nonexperimental setting when treatment is determined by whether an observed "assignment" variable exceeds a known cutoff point
  - e.g. Eligibility for Medicare begins at age 65
- Compares individuals with values just above and below the cutoff point to estimate a treatment effect

Example 1: Medicare at Age 65

Card, Dobkin, Maestas QJE 2009

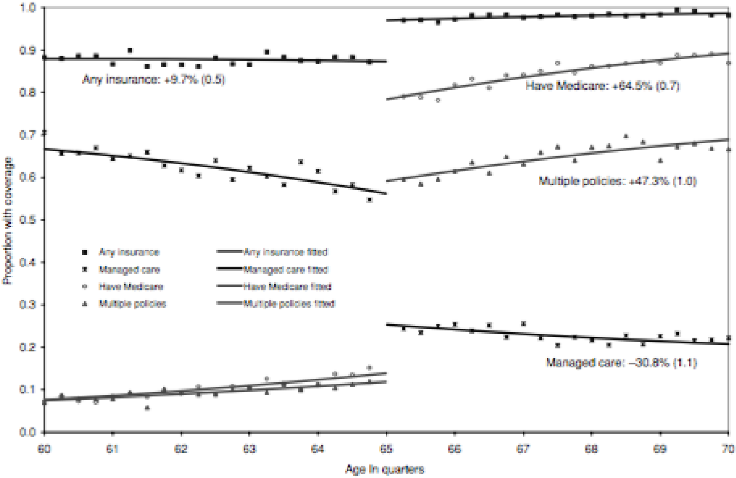


FIGURE I

Changes in Health Insurance at Age 65 in National Health Interview Survey

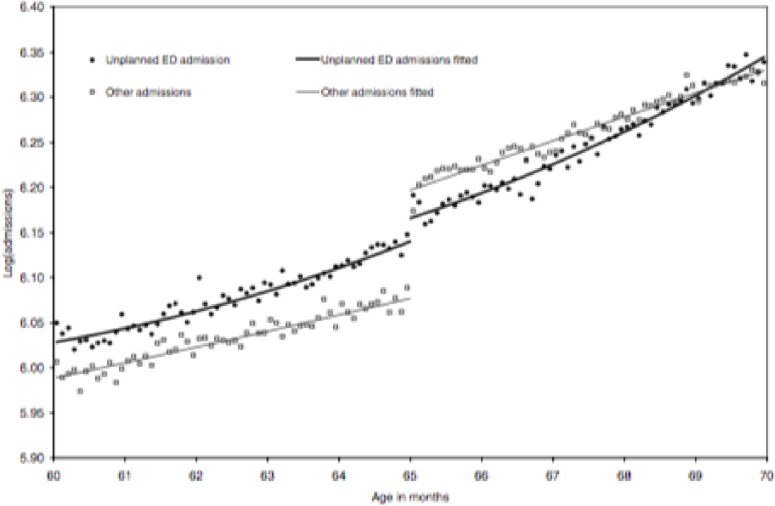


FIGURE II

Number of Admissions by Route into Hospital, California, 1992–2002

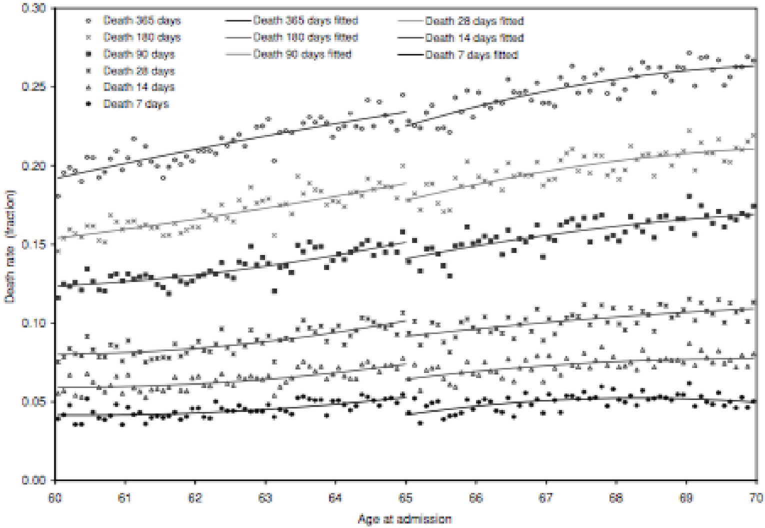
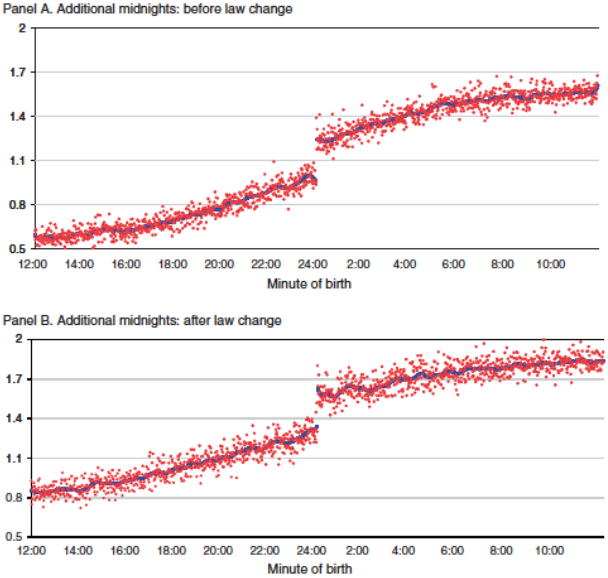


FIGURE VI
Patient Mortality Rates over Different Follow-Up Intervals

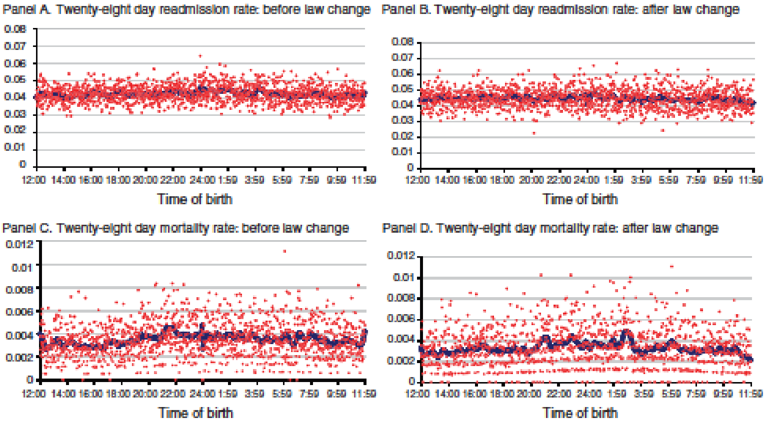
Example 2: Length of Hospital Stay After Birth

Almond and Doyle AEJ:EP 2011



Example 2: Length of Hospital Stay After Birth

Almond and Doyle AEJ:EP 2011



Crucial RD Assumption

- Assumes that individuals on either side of cutoff are otherwise similar in absence of treatment
- What will invalidate the RD design?
  - If individuals can manipuate their value of the "assignment variable" to gain treatment
  - If individuals are not similar on either side of cutoff
- How to check this?
  - Make sure that no heaping at cutoff
  - Make sure observable baseline covariates do not change at the cutoff

Advantages and Disadvantages

- Advantages:
  - Relies only on the assumption that variation in treatment is as good as random in a neighborhood around the cutoff
  - Assumption can be tested by looking at density and distribution of observed covariates around cutoff
- Disadvantages:
  - Tells you information on treatment effect that may not necessarily generalize to broader population (i.e. external validity)

Three Steps to Implement

1. Graph the data for visual inspection
2. Estimate the treatment effect using regression methods
3. Run checks on assumptions underlying research design

Steps Differ Slightly for Two Types of RDD

1. Sharp RD: Probability of receiving treatment jumps from 0 to 1 at cutoff
   - Treatment effect estimated by comparing outcomes for individuals right above and below cutoff
2. Fuzzy RD: Probability of receiving treatment increases at cutoff but depends on other factors
   - Comparing outcomes for individuals right above and below cutoff gives effect of assignment to treatment by the cutoff rule
   - Instrumental variables methods are used to estimate effect among compliers

Step 1: Graph the Data

- Major advantage of the RDD approach is its transparency
- Graph the average value of the outcome variable (Y) for different bins of the assignment variable (X)
  - Pick bins of large enough size to show smooth picture, but small enough to allow to see jump at cutoff
- (Optional) overlay a flexible regression model to "smooth" the graph
- FRD only: Also graph the probability of treatment

Step 1: Graph the Data

- Major advantage of the RDD approach is its transparency
- Graph the average value of the outcome variable (Y) for different bins of the assignment variable (X)
    - Pick bins of large enough size to show smooth picture, but small enough to allow to see jump at cutoff
- (Optional) overlay a flexible regression model to "smooth" the graph
- FRD only: Also graph the probability of treatment

Note: In Stata, rdplot command will pick the optimal bin size for you, download from https://sites.google.com/site/rdpackages/rdrobust

Example RD Graph

Lee and Lemieux 2010

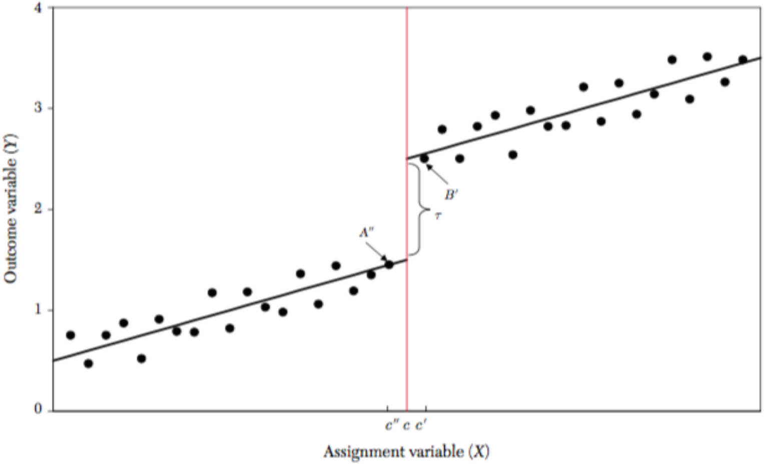


Figure 1. Simple Linear RD Setup

Step 2: Estimate the Treatment Effect

Two different types of regression methods

1. Parametric regression: polynomial regression methods require functional form assumption and use data points further away from cutoff in estimation
   - Ex. linear RD regression (next slide)
2. Nonparametric regression: local linear regression methods do not require functional form assumption and put more weight on observations closest to the cutoff
   - Implement using rdrobust in Stata, download from https://sites.google.com/site/rdpackages/rdrobust

FRD only: estimate the treatment effect using Two-Stage Least Squares

Basic Linear RD Regression

Estimate the following linear regression

$$Y = \alpha + D\tau + X\beta + \epsilon \tag{1}$$

where Y is the outcome variable
X is the assignment variable
$D = 1$ if $X \geq$ cutoff value c
$D = 0$ otherwise

Basic Linear RD Regression

Estimate the following linear regression

$$Y = \alpha + D\tau + X\beta + \epsilon \tag{1}$$

where Y is the outcome variable
X is the assignment variable
$D = 1$ if $X \geq$ cutoff value c
$D = 0$ otherwise

In Stata: (1) generate D=1*(xvar>=c)
(2) regress yvar D xvar, robust
Coefficient estimate for D will give you the treatment effect

Step 3: Run Checks on Research Design

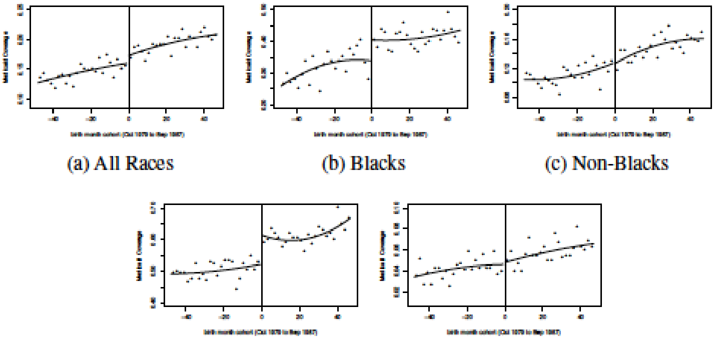
- Check specification choice: examine sensitivity of regression estimates to functional form assumption (ex. linear model) and bandwidth of observations around cutoff
- Make sure good comparison: check for jumps in value of baseline covariates at the cutoff point
- Check for sign of manipulation: look for discontinuities in density of assignment variable
- Placebo test: look at whether outcome is discontinuous at other values of assignment variable

Example: Wherry and Meyer *Journal of Human Resources* 2016

- Medicaid expansion for children implemented in 1991 applied to kids born after September 30, 1983 only
- Kids (or parents) born right before and after this date were not able to manipulate their birthdates
- Kids born right before and after this date are otherwise similar
- Therefore, this rule created variation in childhood exposure to Medicaid that was as good as random for kids born right before and after September 30, 1983

RD Graph of Treatment

Figure 1: Medicaid Coverage in Childhood, Ages 8 to 13, NHIS



(a) All Races

(b) Blacks

(c) Non-Blacks

(d) Households below poverty level

(e) Households above poverty level

RD Graph of Outcome: Later Life Mortality (Ages 15-18)

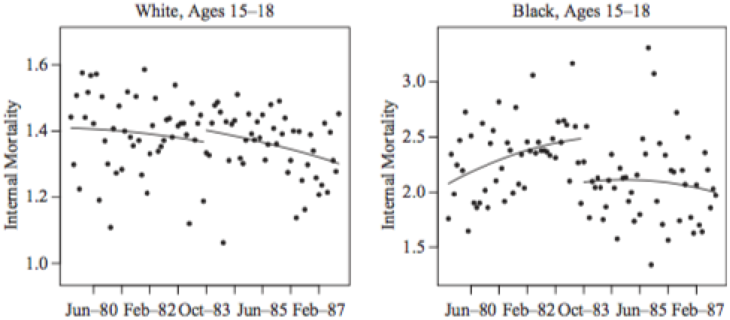


Figure 3
Child Mortality from Internal Causes by Child Race

RD Regression Estimates

Table 3
Change in Annual Internal-Cause Mortality Rate for Children Born After September 30, 1983, by Race and Age Group

	Black Children			White Children		
	Four-Year Window	Three-Year Window	Two-Year Window	Four-Year Window	Three-Year Window	Two-Year Window
Ages 15–18						
Linear	–0.443***	–0.465***	–0.460**	0.028	–0.000	–0.009
	(0.126)	(0.151)	(0.195)	(0.048)	(0.052)	(0.068)
Linear spline	–0.447***	–0.465***	–0.453**	0.025	–0.000	–0.006
	(0.124)	(0.148)	(0.195)	(0.045)	(0.050)	(0.065)
Quadratic	–0.448***	–0.471***	–0.467**	0.022	–0.001	–0.010
	(0.125)	(0.148)	(0.194)	(0.046)	(0.051)	(0.066)
Quadratic spline	–0.456**	–0.349	0.053	0.008	0.011	0.077
	(0.197)	(0.258)	(0.320)	(0.071)	(0.072)	(0.085)
Baseline mean	2.322	2.387	2.440	1.393	1.378	1.384
N	192	144	96	192	144	96

- Checked for robustness to other functional forms and window sizes
- Checked to make sure no heaping (i.e. manipulation) at cutoff
- Checked for discontinuities in baseline covariates: birth weight and gestational age, mother's characteristics (age, marital status, educational attainment)
- Checked for jumps at non-discontinuity points

Great Resources

- Venkataramani AS, Bor J, Jena AB. 2016. Regression Discontinuity Designs in Healthcare Research. *BMJ* 352: i1216.
- Lee DS, Lemieux T. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2): 281-355.
- Imbens GW, Lemieux T. 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics* 142: 615-635.

Questions?

Thank you!

lwherry@mednet.ucla.edu