Regression-Discontinuity Design

Day 2 of
Quasi-Experimental Workshop
Agenda

• Design Overview
• Estimation approaches
• Addressing threats to validity
• Example 1: Shadish, Galindo, Wong, V., Steiner, & Cook (2011)
• Additional issues
  – Improving statistical power
  – Improving generalization of treatment effects across multiple cutoffs, samples, and assignment variables
Design Overview

Design
Examples
Visual depiction
Rationales
Regression Discontinuity

Resource allocation can be by a merit score, need (or risk) score, first come..., date of birth

How prevalent are allocation mechanisms like this in a given society?

Some first examples in health are ...

RD is the design for such circumstances

What is it; and why does it have this name?

Need to learn the language of an assignment variable, cutoff and outcome
Real Examples of RD

Reading First by MDRC
Ludwig & Miller - Head Start on High School graduation
RDD Visual Depiction

Comparison
RDD Visual Depiction
Two Rationales for RDD

1. Selection process is completely known and can be modeled through a regression line of the assignment and outcome variables
   - Untreated portion of the AV serves as a counterfactual
2. It is like an experiment around the cutoff
   - Benefit: Functional form need not be identified
Two Rationales for RDD

1. Selection process is completely known and can be modeled through a regression line of the assignment and outcome variables
   - Untreated portion of the AV serves as a counterfactual

2. It is like an experiment around the cutoff
   - Benefit: Functional form need not be identified away from the cutoff
Two Rationales for RDD

1. Selection process is completely known and can be modeled through a regression line of the assignment and outcome variables
   – Untreated portion of the AV serves as a counterfactual

2. It is like an experiment around the cutoff
   – Benefit: Functional form need not be identified away from the cutoff

• Empirically validated by 6 within-study comparisons (Aiken et al., 1998; Buddelmeyer & Skoufias, 2005; Black, Galdo & Smith, 2007; Berk et al., in press; Greene, 2010; Shadish et al., 2011), even though estimands differ between RE and RD and power too
Required Assumptions for RD design

1. Prob. of treatment receipt must be discontinuous at cutoff. More of those receiving treatment should be on treatment side of cutoff than the other side. If all are, then “sharp” RD. Otherwise, “fuzzy” RD

2. No discontinuity in potential outcomes in the cutoff (the “continuity restriction”). That is, no alternative interpretation should also show a discontinuity at the cutoff. If so, it would serve as a causal confound.
1. Overrides to the cutoff: Sharp, Fuzzy, and “Non-RD” Designs
Sharp RD Designs

• Requires that the probability of treatment changes discontinuously from 1 to 0 at the cutoff such that \( \lim_{z \downarrow z_c} E[D_i | z_i = z_c] - \lim_{z \uparrow z_c} E[D_i | z_i = z_c] = 1. \)

• In practice, this means there are no cross-overs and no no-shows in the study.

• Parametric, non-parametric, and semi-parametric approaches may be applied for estimating treatment effects.

• Treatment effects are identified at the cutoff.
Fuzzy RD Designs

• Requires a discontinuity in the probability of treatment at the cutoff such that
\[
\lim_{z \downarrow z_c} E[D_i | z_i = z_c] \neq \lim_{z \uparrow z_c} E[D_i | z_i = z_c], \text{ but not a 1 to 0 change.}
\]

• In practice, this means some individuals “crossed-over” treatments or did not “show-up” for treatment.

• Hahn, Todd, and van der Klaauw (2001) show that when fuzziness occurs, the local average treatment effect (LATE) may be inferred for the subset of individuals who are induced into treatment at the cutoff.
Estimating Treatment Effects in Fuzzy RDDs

- LATE is equivalent to the difference in mean outcomes for the treatment and comparison groups (ITT estimate) divided by the difference in treatment receipt rates for both groups within a close neighborhood around the cutoff.

\[
\lim_{z \downarrow z_c} E[Y_i | z_i = z_c] - \lim_{z \uparrow z_c} E[Y_i | z_i - z_c] \\
\lim_{z \downarrow z_c} E[D_i | z_i = z_c] - \lim_{z \uparrow z_c} E[D_i | z_i - z_c]
\]

- In a regression framework, this is estimated through a two-stage least squared IV approach.
Requirements for Estimating Treatment Effects in Fuzzy RDDs

• Similar to how non-compliance is handled in randomized experiments

• Requires the following:
  1. That the instrument (treatment assignment) is correlated with treatment receipt; and
  2. That the instrument is not correlated with errors in the outcome model.
“Non”-RD designs

• No discontinuity in the probability of treatment at the cutoff.
• Occurs when there are so many overrides to the cutoff that the assignment mechanism is meaningless.
• Treatment effects cannot be estimated at the cutoff.
Implementation Threats to RD Design

1. Overrides to the cutoff
   a. Sharp design
   b. Fuzzy design
   c. Extreme “fuzziness” with no discontinuity in probability of treat at cutoff

2. Manipulation of assignment scores by units
2. Manipulation of the assignment variable

- Occurs when participants manipulate assignment scores to receive or avoid treatment.
- Different from “overrides to the cutoff” because researcher does not know what scores – and treatment assignment – participants should have received.
- No definitive test for knowing when it occurs, but graphical analysis can help detect when it does.
Example: AYP Data from Texas

Histogram

Kernel Density Plot

Drop in density of observations before cutoff

Jump in density of observations at the cutoff
Diagnostic test

• Can perform “McCrary Test” (2008)
  – Basically a statistical test for assessing whether there is a discontinuity in the density of observations at the cutoff.
  – Test is “reassuring” for the analyst, but not sufficient for proving the validity of the design.
  – Must combine thoughtful consideration of the assignment process with observation of distribution of data around the cutoff.
What to do?

• This is a big problem, because there is no work (we are aware of) for addressing manipulation of assignment process when it occurs.

• Can – and should -- examine qualitatively how manipulation of assignment scores occurred and control for those cases
Analytic Threats to RD Design: Parametric Analyses

1. Misspecification of the response function
Analytic Threats to RD Design: Misspecification of functional form in Parametric Analyses

• Recall, in RDD, we measure the size of the effect as the size of the discontinuity in regression lines at the cutoff:
The size of the discontinuity at the cutoff is the size of the effect.
Nonlinearities in Functional Form

• Anything that affects the size of that discontinuity other than treatment is a threat.
• In the example, we assumed the relationship between assignment and outcome was linear—regressions are straight lines.
• But functional form can be nonlinear due to:
  – Nonlinear Relationships between the assignment variable and the outcome
  – Interactions between the assignment variable and treatment.
Functional Form

• Put more technically, effects are unbiased only if the functional form of the relationship between the assignment variable and the outcome variable is correctly modeled

• Consider first the example of a nonlinear relationship between the assignment variable and the outcome variable:
Here we see a discontinuity between the regression lines at the cutoff, which would lead us to conclude that the treatment worked. But this conclusion would be wrong because we modeled these data with a linear model when the underlying relationship was nonlinear.
If we super-impose a nonlinear regression line\textsuperscript{1} onto the data, a line that seems to match the curves in the data pretty well, we see no discontinuity at the cutoff anymore, and correctly conclude that the treatment had no effect.

\textsuperscript{1} In this case, a cubic function ($X^3$)
Functional Form: Interactions

• Sometimes the treatment works better for some people than for others
  – For example, it is common to find that more advantaged children (higher SES, higher pretest achievement scores) benefit more from treatment than do less advantaged children.

• If this interaction (between the assignment variable and treatment) is not modeled correctly, a false discontinuity will appear:
Here we see a discontinuity that suggests a treatment effect. However, these data are again modeled incorrectly, with a linear model that contains no interaction terms, producing an artifactual discontinuity at the cutoff…
If we superimpose the regression lines that would have been obtained had an interaction term been included, we would find no discontinuity at the cutoff…
The interpretation of this example is important to understand. The title of the graph says "false treatment main effect". However, the treatment did have an *interaction effect*: Treatment helped children with higher scores on the assignment variable more than children with lower scores on the assignment variable...
Here we see an example where the treatment had both a main effect and an interaction effect, correctly modeled.

Regression Discontinuity: Treatment Effect and Interaction

main

Posttest Scores

Assignment Variable Scores
How to Detect Nonlinearities

• Visual Inspection of relationship between assignment and outcome prior to treatment (e.g., if archival data is used).

• Visual Inspection of the Graph

• Computer Programs (e.g., Cook and Weisberg)

• Analysis: Overfitting the model (more on this later).
Adding Nonlinear Terms to the Model

- Include nonlinear functions of the assignment variable in the equation, for example:

\[ Y_i = \hat{\beta}_0 + \hat{\beta}_1 d + \hat{\beta}_2 (z_i - z_c) + \hat{\beta}_3 (z_i - z_c)^2 + e_i \]

Where \(d\) is the treatment indicator and \(z\) is the assignment variable.

- There are many such nonlinear functions, so selecting the correct one is crucial.
Adding Interaction Terms to the Model

• One can also add interactions between treatment assignment \((d)\) and the assignment variable \((z)\), for example:

\[
Y_i = \hat{\beta}_0 + \hat{\beta}_1 d_i + \hat{\beta}_2 (z_i - z_c) + \hat{\beta}_3 x_i (z_i - z_c) + e_i
\]
Adding Nonlinear and Interaction Terms to the Model

- And one can add both nonlinear and interaction terms to the model:

\[ Y_i = \hat{\beta}_0 + \hat{\beta}_1 d_i + \hat{\beta}_2 (z_i - z_c) + \hat{\beta}_3 (z_i - z_c)^2 \\
+ \hat{\beta}_4 d_i (z_i - z_c) + \hat{\beta}_5 d_i (z_i - z_c)^2 + e_i \]

- As you can imagine, the model can get quite large.
- Though it may seem complex, all this can easily be done in SPSS, and all the relevant terms can easily be defined with simple SPSS Compute statements.
How Do You Know Which Terms to Add?

• If you did a pretest run of the design (before treatment was begun), use that data to model the baseline functional form.
• Visually inspect the data for clues about what form seems likely.
• Use available programs for curve fitting in order to explore possible functional forms (e.g., Cook & Weisberg, 1994)
Adding Terms, Continued

• When in doubt, start by overfitting the model:
  – Add more polynomial and interaction terms than you think are needed, and then eliminate the nonsignificant ones (from higher order to lower order).
  – When in doubt, keep terms in the equation; such overfitting will yield unbiased estimates, but will reduce power the more nonsignificant terms are kept.
Adding Terms, Continued

• Sensitivity Analyses: Do substantive conclusions vary with different assumptions about which terms to add?
• If sample size is large, split the sample in half randomly:
  – Develop the model on one half
  – Cross-validate the model on the other half.
Example of Parametric Plot used again later

RD data from Shadish, Galindo, Wong, V., Steiner, & Cook (2011)
But Parametric Analyses being Replaced by Non-Parametric and Semi-Parametric ones
Non-parametric Approach

\[
\lim_{x \downarrow c} E[Y_i \mid z_i = z_c] - \lim_{x \uparrow c} E[Y_i \mid z_i = z_c]
\]

• Where \( z_i \) is the assignment score, \( z_c \) is the cutoff, and \( Y_i \) are the outcomes.
• Treatment effects estimated by examining the size of the discontinuity in the conditional expectation of the outcome given the assignment variable for observations \textit{at the cutoff}.
• The limits are often estimated by non-parametric local linear kernel regression.
• Allows analyst to relax assumptions about functional form away from the cutoff.
Local Linear

kernel = normal
bandwidth = 2 | span = FALSE

(see animated file here)
Non-parametric plot
Local linear kernel regression in STATA

RD data from Shadish, Galindo, Wong, V., Steiner, & Cook (2011)
3. Semi-Parametric Approach

\[ Y_i = \tau D_i + f(Z_i) + X_i^t \tilde{a} + \varepsilon_i \]

- Where \( Y_i, D_i, \) and \( X_i \) are the same as before, and \( f(Z_i) \) in the generalized additive model is the smooth function of the assignment variable for the control and treatment groups.
- The smooth function is typically modeled using a series of splines.
- Treatment effect (\( \tau \)) may be estimated via thin plate regression splines using the \textit{mcgv}-package in R (R Development Core Team, 2009; Wood, 2006).
- Advantage is that it allows for flexibility in modeling the response function while controlling for baseline covariates in the model.
Semi-Parametric Plot

RD data from Shadish, Galindo, Wong, V., Steiner, & Cook (2011)
Problem of Incorrect Bandwidth Selection
Bias versus efficiency tradeoff
What to do?

• Check to see whether treatment effects are robust to alternative specifications of bandwidth
• State of art is to use Imbens & Kalyanaraman (2010) (IK) optimal bandwidth module that takes account of boundary issues
  – IK programs for bandwidth selection for Stata and Matlab: http://www.economics.harvard.edu/faculty/imbens/software_imbens
  - IK Program for bandwidth selection in R http://web.me.com/devin.caughey/Site/Code.html
RDD Example 1
Shadish, Galindo, Wong, V., Steiner, & Cook (2011)
Vocabulary intervention

• Example of how we probed all required assumptions in our analysis of an RDD
• Assignment was based on a pretest score on a vocabulary exam.
• Students who scored below the cutoff (20) received online vocabulary training, those who scored above the cutoff received a math training.
• Sample was a pool of students at UC Merced who agreed to participate in an online study.
RD Design assumption 1

- Probability of treatment discontinuous at cutoff
**RD Design assumption 2**

- **Continuity in potential outcomes at the cutoff**
  - No discontinuity at cutoff in 20+ covariates examined
  - Not a conclusive test, but reassuring

**Female**
- Non-parametric est: (1.83): .01
- TR est: .69 / CI for TR: .54 - .85
- CO est: .7 / CI for CO: .53 - .87

**Age**
- Non-parametric est (2.55): .04
- TR est: 18.79 / CI for TR: 18.05 - 19.53
- CO est: 18.83 / CI for CO: 18.49 - 19.17
Implementation threat 1

• Overrides to the cutoff (misallocation of treatment)
  – Examined whether participants entered appropriate assigned treatment conditions by looking at assignment scores, the cutoff, and treatment receipt status.
  – No evidence of individual changing treatment status based on what they were assigned.
  – 0% cross-over; 0% no-shows
Implementation assumption 2

- No evidence of manipulation of the assignment variable
Analytic assumption 1

- Correct specification of functional form for parametric analyses

Linear function
Analytic assumption 1 continued

• Correct specification of functional form for parametric analyses

Cubic function
What to do?

• Conduct graphical analyses
• Examine residuals
• Check to see whether treatment effects are robust to alternative specifications of functional form
• Check to see whether treatment effects are robust to alternative pseudo-cutoffs
• Estimate treatment effects using non-parametric and semi-parametric approaches
<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>RDD mean (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Linear regression</td>
<td>5.05 (0.57)***</td>
</tr>
<tr>
<td>(2) Quadratic regression*</td>
<td>5.91 (0.77)***</td>
</tr>
<tr>
<td>(3) Cubic regression</td>
<td>6.04 (1.07)***</td>
</tr>
<tr>
<td>(4) Generalized additive Model (GAM)</td>
<td>5.76 (0.72)***</td>
</tr>
<tr>
<td>(5) Local linear regression</td>
<td>6.28 (1.12)***</td>
</tr>
</tbody>
</table>

*The quadratic function was selected as the optimal RDD model.
Analytic assumption 2

- Correct specification of bandwidth for non-parametric estimates
  - Used Imbens & Kalyanaraman (2010) algorithm for choosing optimal bandwidth
  - Estimated treatment effects using bandwidths that were ½ and double the size of the IK optimal bandwidth to check robustness of estimates.

**Optimal BW:** Treatment effect for Non-parametric (4): 6.48
TR est: 13.21 / CI for TR: 10.94 - 15.48
CO est: 6.73 / CI for CO: 5.92 - 7.55

½ Optimal BW: Treatment effect for Non-parametric (2): 5.1
TR est: 12 / CI for TR: 10.75 - 13.25
CO est: 6.9 / CI for CO: 5.93 - 7.87

Double optimal BW: Treatment effect for Non-parametric (8): 5.89
TR est: 12.27 / CI for TR: 10.86 - 13.69
CO est: 6.39 / CI for CO: 5.74 - 7.04
RDD Example 2
State Pre-Kindergarten Example

• In this example, we were particularly concerned about two issues: (1) functional form assumptions and (2) overrides to the cutoff:

• Study Design
  – Pre-K available by birth date cutoff in 38 states, here scaled as 0 (zero)
  – 5 chosen for study and summed here
  – How does pre-K affect PPVT (vocabulary) and print awareness (pre-reading)
1. Mis-specification of Functional Form  
(Pre-K Example) 

Best case scenario – regression line is linear and parallel  
(NJ Math)
Sometimes, form is less clear

Example: Oklahoma Print Awareness

Mis-specification of Functional Form
(Pre-K Example)
Correct specification of functional form assumption
What to do about functional form (1)

• Graphical approaches
Graphical analysis (New Jersey PPVT)

Lowess

Local Linear Regression (LLN)

Regression
What to do (2)

• Parametric approaches
  – Alternate specifications and samples
    • Include interactions and higher order terms
      – Linear, quadratic, & cubic models
      – Look for statistical significance for higher order terms
      – When functional form is ambiguous, overfit the model (Sween 1971; Trochim 1980)
    • Truncate sample to observations closer to cutoff
  – Bias versus efficiency tradeoff
What to do (3)

• Non-parametric approaches
  – What are non-parametric approaches?
    • Eliminates functional form assumptions
    • Performs a series of regressions within an interval, weighing observations closer to the boundary
    • Use local linear regression because it performs better at the boundaries
  – What depends on selecting correct bandwidth?
    • Key tradeoff in NP estimates: bias vs precision
  – How do you select appropriate bandwidth?
    – Ocular/sensitivity tests
    – Imbens & Kalyanaraman (2010)
    – Cross-validation methods
      » “Leave-one-out” method
What to do (4)

- State-of-art is imperfect

- So we test for robustness and present multiple estimates

<table>
<thead>
<tr>
<th></th>
<th>Parametric estimates</th>
<th>Non-Parametric estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear (1)</td>
<td>Quadratic (2)</td>
</tr>
<tr>
<td>New Jersey PPVT</td>
<td>5.71* (1.44)</td>
<td>5.37* (2.02)</td>
</tr>
</tbody>
</table>
What to do (4)

- Another example
- Where function appears to be cubic

<table>
<thead>
<tr>
<th></th>
<th>Parametric estimates</th>
<th>Non-Parametric estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Quadratic</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>21.01*</td>
<td>15.55*</td>
</tr>
<tr>
<td>Print Awareness</td>
<td>(3.52)</td>
<td>(4.84)</td>
</tr>
</tbody>
</table>
What to do (5)

- Move to a tie breaker experiment, with both RD and random assignment around cutoff
Tie-breaking experiment design

- Cutoff control group: outpatient rehabilitation
- Cutoff interval where patients get randomized to inpatient or outpatient rehabilitation
- Cutoff treatment group: inpatient rehabilitation
- Treatment effect

Outcome Measure vs. Baseline Assignment Measure
Example
Cocaine Project

• Included 500 patients addicted to cocaine
• Purpose was to show whether intensive in-patient rehabilitation showed better improvement for patients as opposed to outpatient care
• Assignment variable was a composite score based on the following: 1) employment and legal status; 2) family relationship and recovery; 3) alcohol and drug history; 4) psychological status..
• Patients with scores above 60 assigned to inpatient rehab; patients with scores less than 40 assigned outpatient rehab; those with scores 40-60 in randomized trial.
• Used RD-RE design b/c randomized experiment was consider neither ethical nor feasible.
What to do (6)

• Curtail at the tails and use cases to sample as densely as possible around the cutoff
What to do (7)

• Estimation through design
  – Cohort Comparison: becomes an almost independent estimate of the counterfactual functional form of the missing treatment group

  – Other design RDD features: pretests, nonequivalent dependent variable
Summary concerning mis-specification of the response function

• Do Graphical analyses
• In parametric models, do alternative specifications
• Also do non-parametric estimates
• Present multiple estimates to check robustness
• If possible, move to tie-breaker experiment around the cutoff
• If possible, sample densely at the cutoff
• If possible, use pretest or other comparison RD functions
2. Dealing with Overrides to the cutoff

- Observe misallocation in the data

<table>
<thead>
<tr>
<th>States</th>
<th>Fuzzy cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michigan</td>
<td>2%</td>
</tr>
<tr>
<td>New Jersey</td>
<td>4%</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>4%</td>
</tr>
<tr>
<td>South Carolina</td>
<td>1%</td>
</tr>
<tr>
<td>West Virginia</td>
<td>8%</td>
</tr>
</tbody>
</table>
What to do about fuzzy allocation?

• RDD as an instrumental variable
  – Provides causal estimates if instrument is correlated with treatment and not with errors in the outcome.
  – Proof: Hahn, Todd, and van der Klaauw, 2001
  – In this case, we implemented in a regression (two stage least squares) framework as opposed to the non-parametric wald estimator
2. Overrides to the cutoff

- Multiple estimations

<table>
<thead>
<tr>
<th></th>
<th>% Fuzzy</th>
<th>Full sample</th>
<th>Restricted sample</th>
<th>Instrumental Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(8)</td>
<td>(9)</td>
<td>(11)</td>
</tr>
<tr>
<td>Michigan PPVT</td>
<td>2%</td>
<td>-2.20</td>
<td>-1.91</td>
<td>-2.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.64)</td>
<td>(4.15)</td>
<td>(4.53)</td>
</tr>
<tr>
<td>West Virginia PPVT</td>
<td>8%</td>
<td>2.75</td>
<td>3.66</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.23)</td>
<td>(2.39)</td>
<td>(1.94)</td>
</tr>
<tr>
<td>Michigan Print Awareness</td>
<td>2%</td>
<td>25.21*</td>
<td>24.98*</td>
<td>22.14*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.48)</td>
<td>(3.58)</td>
<td>(3.11)</td>
</tr>
<tr>
<td>West Virginia Print Awareness</td>
<td>8%</td>
<td>22.25*</td>
<td>24.49*</td>
<td>20.15*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.59)</td>
<td>(3.50)</td>
<td>(2.98)</td>
</tr>
</tbody>
</table>
Summary of Analytic Plan for Addressing Mis-specification of the Functional Form and Overrides to the Cutoff in the Pre-K data
### Specification of Functional Form

**Parametric approaches (1)**

<table>
<thead>
<tr>
<th></th>
<th>Functional form</th>
<th>Parametric models in analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Quad</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>PPVT</td>
<td>Quad</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.49)</td>
</tr>
<tr>
<td>Student covariates</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
## Specification of Functional Form

**Parametric approaches (2)**

<table>
<thead>
<tr>
<th>PPVT</th>
<th>Functional form</th>
<th>Linear</th>
<th>Quad</th>
<th>Cubic</th>
<th>6 mnth s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Student covariates</td>
<td>Quad</td>
<td>0.33</td>
<td>-1.91</td>
<td>-3.18</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(4.15)</td>
<td>(5.62)</td>
<td></td>
<td>(6.10)</td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
## Specification of Functional Form

Non-parametric approaches

<table>
<thead>
<tr>
<th>Functional form</th>
<th>Parametric models used in analysis</th>
<th>Non-parametric estimates by bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT</td>
<td>Linear 0.33, Quad -1.91, Cubic -3.18, 6 mnths -3.74</td>
<td>50 BW -4.99, 75 BW -1.66</td>
</tr>
<tr>
<td></td>
<td>Students Yes, Yes, Yes, Yes</td>
<td>No, No</td>
</tr>
<tr>
<td></td>
<td>Fuzzy cases No, No, No</td>
<td>No, No</td>
</tr>
</tbody>
</table>
## Overrides to cutoff (1)

### Alternative samples

<table>
<thead>
<tr>
<th>Function form</th>
<th>Linear</th>
<th>Quad</th>
<th>Cubic</th>
<th>6 mnths</th>
<th>50 BW</th>
<th>75 BW</th>
<th>OLS estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) PPVT</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>-1.91</td>
<td>-3.18</td>
<td>-3.74</td>
<td>-4.99</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(4.15)</td>
<td>(5.62)</td>
<td>(6.10)</td>
<td>(4.40)</td>
<td>(3.91)</td>
<td></td>
</tr>
<tr>
<td>Student covariates</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OLS estimates</th>
<th>Full (ITT)</th>
<th>Resticted (TOT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2.20</td>
<td>-1.91</td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(4.15)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OLS estimates</th>
<th>Full (ITT)</th>
<th>Resticted (TOT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Note: The table above shows the parametric models used in analysis, non-parametric estimates by bandwidth, and OLS estimates for different conditions. The table includes linear, quadratic, and cubic models, along with 50 and 75 bandwidth estimates. The OLS estimates are provided for Full (ITT) and Restricted (TOT) conditions.
### Overrides to cutoff (2)
**Alternative samples (with and w/o covariates)**

<table>
<thead>
<tr>
<th>Function form</th>
<th>Linear</th>
<th>Quad</th>
<th>Cubic</th>
<th>6 mths</th>
<th>50 BW</th>
<th>75 BW</th>
<th>OLS estimates</th>
<th>IV estimates with and without covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full sample (ITT)</td>
<td>Resticted sample (TOT)</td>
</tr>
<tr>
<td>PPVT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
</tr>
<tr>
<td>Quad</td>
<td>0.33</td>
<td>-1.91</td>
<td>-3.18</td>
<td>-3.74</td>
<td>-4.99</td>
<td>-1.66</td>
<td>-2.20</td>
<td>-1.91</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(4.15)</td>
<td>(5.62)</td>
<td>(6.10)</td>
<td>(4.40)</td>
<td>(3.91)</td>
<td>(3.64)</td>
<td>(4.15)</td>
</tr>
<tr>
<td>Student covariates</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
## All Outcomes

<table>
<thead>
<tr>
<th>Functiona l form</th>
<th>Parametric models used in analysis</th>
<th>Non-parametric estimates by bandwidth</th>
<th>OLS estimates with and without covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Quad</td>
<td>Cubic</td>
</tr>
<tr>
<td>PPVT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quad</td>
<td>0.33</td>
<td>-1.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>Linear</td>
<td>2.03*</td>
<td>2.25*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student covariates</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td></td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* denotes significance.
## Final estimates

<table>
<thead>
<tr>
<th>Function form</th>
<th>Parametric models used in analysis</th>
<th>Non-parametric estimates by bandwidth</th>
<th>OLS estimates</th>
<th>IV estimates with and without covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Quad</td>
<td>Cubic</td>
<td>6 mths</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>PPVT</td>
<td>Quad</td>
<td>0.33</td>
<td>-1.91</td>
<td>-3.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.49)</td>
<td>(4.15)</td>
<td>(5.62)</td>
</tr>
<tr>
<td>Math</td>
<td>Linear</td>
<td>2.03*</td>
<td>2.25*</td>
<td>2.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.56)</td>
<td>(0.90)</td>
<td>(1.29)</td>
</tr>
<tr>
<td>Print awareness</td>
<td>Linear</td>
<td>24.98*</td>
<td>21.58*</td>
<td>21.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.58)</td>
<td>(5.68)</td>
<td>(7.77)</td>
</tr>
<tr>
<td>Student covariates</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fuzzy cases</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Potential Issues with using Age-Related, but not other, Cutoffs

Differential entrance and attrition time in TR and CO groups
- Treatment group consists of students who are in kindergarten in the fall
- Control group consists of students who are enrolled in pre-K in the fall
- Are there compositional differences between children who stay in pre-K until the fall of their kindergarten year (treatment group) and children who are just starting pre-K and so have not had a chance to attrit (control group)

• Instrumentation issues between treatment and control group
- Assessments such as the PPVT have different start rules for 4-year-olds than 5-year-olds. 12 points are automatically added to all 5 year old group. Measurement difference may be confounded with treatment that depends on age.
Addressing Potential Issues with using Age-Related Cutoffs

• Complications with calculating the ITT
  – Treatment group consists of students who are in kindergarten in the fall
  – Control group consists of students who are enrolled in pre-K in the fall
  – Lacks initial sample for both treatment and control groups
    • Consider various definitions of the initial sample: eligibility defined by “age + residency” or “age, residency, & interest”

• Instrumentation issues between treatment and control group
  – Assessments such as the PPVT have different start rules for 4-year-olds then they do for 5-year-olds. Measurement difference may be confounded with treatment.
    • Students near the cutoff should begin with the same start items on assessments.
Addressing Potential Issues with using Age-Related Cutoffs

• Differential entrance and attrition time in TR and CO groups
  – Because RD study lacks an initial defined sample, cannot assess attrition and selection effects
  – The issue then is: Are there compositional differences between children who make it to the fall of their kindergarten year (treatment group) and children who are just starting pre-K (control group)

• Particularly troubling concern because it could introduce serious bias in the design and make the approach invalid
  – Prompting some critics to ask, Is this an RD design at all?
<table>
<thead>
<tr>
<th>Fallible RD</th>
<th>Institutional Cohort Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children’s birthdates and state cutoffs are a mostly deterministic</td>
<td>Assume that children’s birthdates and state cutoffs are not a deterministic mechanism for</td>
</tr>
<tr>
<td>mechanism for treatment assignment for the sub-population of units near</td>
<td>treatment assignment, so we should regard this design as an institutional cohort design</td>
</tr>
<tr>
<td>the cutoff, but should probe for selection effects.</td>
<td>approach.</td>
</tr>
<tr>
<td>The RD focuses on the comparability of units at the cutoff.</td>
<td>An institutional cohort design focuses on the comparability of all 5-year-olds in kindergarten</td>
</tr>
<tr>
<td></td>
<td>versus all 4-year-old cohorts in pre-K</td>
</tr>
<tr>
<td>Leads the researcher to look for discontinuities in potential outcomes at</td>
<td>Leads the researcher to consider matching methods between cohorts to improve internal</td>
</tr>
<tr>
<td>the cutoff</td>
<td>validity (will address later) ...</td>
</tr>
</tbody>
</table>
Potential Issues with using Age-Related Cutoffs

• The foregoing concerns suggest null hypotheses we could examine:
  – Since treatment children had opportunities to attrite that control children did not have, there should be more “missing” children in the treatment group.
    • Examine the data to determine whether there is a discontinuity in the DENSITY of observations at the cutoff.
  – Attrition could result in compositional differences between groups
    • Examine the data to determine whether there are discontinuities in third variables CONDITIONAL MEANS at the cutoff
• Both of these concern possible violations to the RD assumption of no discontinuities in the potential outcomes (called Assumption 2)
• Begin with conditional means on 3rd variables
African-American

Home language Spanish

Girl

Free/Reduced lunch
Fallible RD?

- Looked for discontinuities in control variables and saw:
  - No significant differences in any of our control covariates at the cutoff (within states and across states)
  - But we had access to only a few demographic covariates (gender, race/ethnicity, free/reduced price lunch, child’s home language)
Fallible RD?

• Now we look for discontinuities in the density of observations (McCrary test)
  – The McCrary test can help us probe whether there was unobserved selection
McCrary Test for All States
McCrary Test for New Jersey
Fallible RD?

• When we probed the New Jersey data, we found that the discontinuity in the McCrary test was driven by cases with a September 1\textsuperscript{st} cutoff.
• Five years ago, children who were born on September 1\textsuperscript{st} were born on a Monday and children who were born on August 31\textsuperscript{st}, were born on Sunday.
• Birth record data shows that there are large “day” effects (i.e. more births on Fridays and Mondays, fewer births on weekends and holidays), so the discontinuity may be driven by day effects
• Not a problem if the “day effects” are not correlated with students’ achievement, but still under investigation.
Summarizing Checks

– Conditional Mean Differences? No observable compositional differences between treatment and control group at the cutoff, but examined only four variables

– (Note that if we were to move to an institutional cohort design approach, we would have few covariates for matching too)

– When states were combined, no significant differences in the density of observations at the cutoff for all cases

– However, some significant differences in density of cases when we looked within some states and evidence that this may be because of day and date effects on births (more births on Fridays, fewer births on weekends and holidays)
Take Home Message

• Like RE, the RD design must be implemented carefully – with careful consideration of all possible selection effects and violations to the assignment mechanism.

• Use state-of-the-art methods for probing the validity of the RD (such as checking whether there are discontinuities in third variables and in the density of observations).

• But even these implementation tests require careful consideration of the data (i.e. “day effects”).
Other Kinds of Threats with RD

1. Lower statistical power than the experiment
2. Treatment effects estimated from an RDD may have less generalization than treatment effects estimated from a randomized experiment
Inadequate statistical power

• Need adequate power for estimating regression line on both sides of the cutoff.
  – Randomized experiments more efficient than RDD by a factor of about 2.73 (Goldberger, 1972b)
Variance of Impact Estimator in RA

\[ \text{Var}_{ra}(\hat{\alpha}_0) = \frac{\sigma^2 (1-R_{ra}^2)}{np(1-p)(1-ns-co)^2} \]

- \( \sigma^2 \) is the variance of mean student outcomes across schools within the treatment or comparison group
- \( R_{ra}^2 \) is the square of the correlation between school outcomes and the assignment variable
- \( np(1-p) \) is the total variation in treatment status across schools
- \( ns \) is the no-show (via appeals) rate
- \( co \) is the cross-over (via participation) rate
Power Considerations Common to Randomized Experiments and RDD

1. Sample size -- increase N
2. Distribution of outcome variable - normal if possible
3. Misallocation rates - lower the better
4. R-square of school- and student-level covariates - as high as possible
5. Clustered designs and intra-class correlations -- as low as possible or as reduced as possible through the use of covariates explaining clustering
Variance of Impact Estimator in RDD

$$\text{Var}_{rd}(\hat{\alpha}_1) = \frac{\sigma^2 (1-R_{rd}^2)}{np(1-p)(1-ns-co)^2(1-R_{ts}^2)}$$

- All other terms are the same, except:
- $R_{ts}^2$ is the square of the correlation between treatment status and assignment variable – higher the value less power (and more co-linearity)
Power Considerations Unique to RDD

1. Bandwidth around cutoff - smaller the less power
2. Non-linear response function - higher order the less power
3. Distribution around cutoff - dense as possible
4. Location of cutoff in distribution – complex issue – not at center where co-linearity greatest but nor at extreme where few cases on one side of cutoff
Power Considerations Unique to RDD

1. Bandwidth around cutoff
2. Non-linear response function
3. Shape of the distribution around cutoff
4. Location of cutoff in distribution
5. Emerging issue: Increased variances when assignment variables are discrete (Lee & Card, 2008) - big problem if becomes state of art
Other Kinds of Threats

1. Lower statistical power than the experiment
2. Treatment effects estimated from an RDD may have less generalization than treatment effects estimated from a randomized experiment
Generalizing Beyond The Cut-off In RD

• Strict interpretation of RDD: treatment effects apply to members of a cut-off sub-population.
  – What is a strict cut-off sub-population?
    • Students who score exactly 110 on a math test.
    • Schools in which exactly 35% of students are below the poverty line.
    • People born on September 1\textsuperscript{st} in a given year.

• Many cut-off sub-populations are not very interesting.
Treatment effect is interpreted as the average treatment effect at the cutoff ($\tau_s$).

This is in contrast to what is usually estimated by the RE, which is the average treatment effect ($\tau_{ATE}$).

Without further assumptions about functional form, $\tau_s$ generalizes only to this cutoff point on this assignment variable.

Cutoff may or may not be of policy or practical relevance.

RD data from Shadish et al. (2011)
Observed vs Counterfactual Outcomes in a Hypothetical RD

Below The Cut-off (No Treatment)

Above The Cut-off (Treatment)

Observed Treated Outcomes

Counterfactual Untreated Outcomes

Counterfactual Treated Outcomes
Observed vs Counterfactual Outcomes in a Hypothetical RD

Below The Cut-off (No Treatment)  Above The Cut-off (Treatment)

-10  0  10  20  30

Outcome

Assignment Variable

-10 -5 0 5 10

Observed Treated Outcomes

Counterfactual Treated Outcomes

Observed Untreated Outcomes

Counterfactual Untreated Outcomes

Assignment Variable
Questions About Generalizing From RD

• In any RD study, ask:
  1. Is this cut-off sub-population relevant to policy or scientific questions?
  2. Does this assignment variable define other sub-populations that are more relevant to policy or science?
  3. What population actually is of interest? What randomized experiment would I like to run?
Two Methods For Generalization To Assignment Variable Populations

• Extrapolate using only functional form assumptions.

• Extrapolate using both functional form assumptions and a comparison group.
What Do We Mean By Comparison Groups?

- **RDD Group**: Units are assigned to treatment based on a regression discontinuity design.

- **Non-Equivalent Comparison Groups**: All units are untreated.
  
  - Example:
    - One school district tries out the new treatment and uses an RD assignment rule to allocate.
    - Other school districts don’t offer the treatment.
Cash and Counseling Demonstration Project

• **Treatment:** Instead of Medicaid procuring home care and community services for you, you receive a fixed budget to allocate as you see fit.

• **Control:** Status Quo. Medicaid agency procures services on your behalf.

• **Study population:** disabled Medicaid recipients who volunteered to participate. Three states: Arkansas, Florida, New Jersey.

• **Key Outcomes:**
  – Post-Treatment Medicaid Expenditures
  – Number of Unpaid Home Care Workers

• **Results:**
  – Treatment leads to higher Medicaid Expenditures
  – Treatment leads to a reductions in the use of unpaid home care workers
Our Purpose

• Use experimental data as a causal benchmark
• Transform the experimental data into RD and RD + non-equivalent comparison group. (We’ll show you how.)
  – Compare benchmarks with RD extrapolations
  – Compare benchmarks with RD + comparison group extrapolations
  – Compare RD vs RD + comparison group extrapolations.
• Examine both the bias and variance of the various extrapolations.
Creating A Regression Discontinuity

• Choose Arkansas to be the RDD group.
  – Choose an assignment variable and a cut-off value
  – Delete treatment group cases below the cut-off
  – Delete control group cases above the cut-off.

• Two assignment variables:
  – AGE with cut-off = 50
  – Pre-Test Medicaid Expenditures with cut-off = $7000

• Two Outcomes:
  – Number of Unpaid Workers
  – Post-Test Medicaid Expenditures

• Two Non-Equivalent Control Groups:
  – Control group from New Jersey
  – Control group from Florida
What Is The Target Parameter?

• Average untreated outcome above the cut-off score.
  – The corresponding experiment is random assignment among people above the cut-off.
  – Here, that means random assignment of people over age 50 for one assignment variable and random assignment of people with expenditures over $7000 last year.

• We also look at the functional forms that are extrapolated beyond the cut-off. This gives a sense of the RD estimates that would result if different cut-off values were selected.
Presenting Results

• Focus: Post-Treatment Medicaid Expenditures.
  – Graphs of extrapolated functions.
  – Estimates of mean above the cut-off parameter.
  – Bootstrap estimates of variance, average bias, and mean square error.
Extrapolated Functional Forms vs Experimental Benchmark

Number of Unpaid Workers As A Function Of Age

Number of Unpaid Workers

Age
Extrapolated Functional Forms vs Experimental Benchmark

Number of Unpaid Workers As A Function of Pre-Treatment Medicaid Expenditures

- Benchmark Untreated Outcomes Function
- Linear Extrapolation
- Non-Equivalent Comparison Group (Semi-Parametric Estimation)
What About The Above The Cut-Off Average?

- We just looked at functional forms.
- But not all points of the function matter equally in the overall average.
- Age distributions and pre-treatment spending distributions are not rectangular.
- We now look at density weighted averages.
Age and Expenditure Distributions

Density of Observations By Assignment Variables In Arkansas

Age

Pre-Treatment Expenditures
Root Mean Square Error In Experimental Benchmark vs Extrapolations

Number of Unpaid Workers as a Function of Pre-Treatment Expenditures

- Experimental
- Linear Extrapolation
- Partially Linear Model
- Polynomial Group

Square Root of Mean Square Error

0

0.1

0.2

0.3
Root Mean Square Error In Experimental Benchmark vs Extrapolations

- Experimental
- Linear Extrapolation
- Partially Linear Model
- Polynomial Group

Number of Unpaid Workers as a Function of Age

Square Root of Mean Square Error

0 0.2 0.4 0.6 0.8
Post-Treatment Medicaid Expenditures as a Function of Pre-Treatment Expenditures

Root Mean Square Error In Experimental Benchmark vs Extrapolations

Square Root of Mean Square Error
Root Mean Square Error In Experimental Benchmark vs Extrapolations

Post-Treatment Medicaid Expenditures as a Function of Age

- Experimental
- Linear Extrapolation
- Partially Linear Model
- Polynomial Group

Square Root of Mean Square Error
### Results For Number of Unpaid Workers With Expenditure Based Assignment

<table>
<thead>
<tr>
<th>Design</th>
<th>Benchmark</th>
<th>Bootstrap Bias</th>
<th>Bootstrap SD</th>
<th>Percent Bias</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>2.26</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Comparison Group Semi-Parametric</td>
<td>2.26</td>
<td>0.00</td>
<td>0.12</td>
<td>0.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Model</td>
<td>2.26</td>
<td>-0.03</td>
<td>0.12</td>
<td>-0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Comparison Group Parametric Model</td>
<td>2.26</td>
<td>-0.20</td>
<td>0.26</td>
<td>-0.09</td>
<td>0.32</td>
</tr>
<tr>
<td>No Comparison Group -- Linear Extrapolation</td>
<td>2.26</td>
<td>-0.20</td>
<td>0.26</td>
<td>-0.09</td>
<td>0.32</td>
</tr>
</tbody>
</table>
## Results For Number of Unpaid Workers With Age Based Assignment

<table>
<thead>
<tr>
<th>Design</th>
<th>Benchmark</th>
<th>Bootstrap Bias</th>
<th>Bootstrap SD</th>
<th>Percent Bias</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>2.20</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Comparison Group Semi-Parametric Model</td>
<td>2.20</td>
<td>-0.11</td>
<td>0.20</td>
<td>-0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Comparison Group Parametric Model</td>
<td>2.20</td>
<td>-0.17</td>
<td>0.19</td>
<td>-0.08</td>
<td>0.25</td>
</tr>
<tr>
<td>No Comparison Group -- Linear Extrapolation</td>
<td>2.20</td>
<td>-0.40</td>
<td>0.63</td>
<td>-0.18</td>
<td>0.75</td>
</tr>
</tbody>
</table>
# Results For Expenditures With Expenditure Based Assignment

<table>
<thead>
<tr>
<th>Design</th>
<th>Benchmark</th>
<th>Bootstrap Bias</th>
<th>Bootstrap SD</th>
<th>Percent Bias</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>9946.77</td>
<td>22.16</td>
<td>282.80</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Comparison Group Semi-Parametric Model</td>
<td>9946.77</td>
<td>184.49</td>
<td>452.52</td>
<td>0.02</td>
<td>488.68</td>
</tr>
<tr>
<td>Comparison Group Parametric Model</td>
<td>9946.77</td>
<td>378.30</td>
<td>401.02</td>
<td>0.04</td>
<td>551.30</td>
</tr>
<tr>
<td>No Comparison Group -- Linear Extrapolation</td>
<td>9946.77</td>
<td>-95.32</td>
<td>636.10</td>
<td>-0.01</td>
<td>643.20</td>
</tr>
</tbody>
</table>
## Results For Expenditures With Age Based Assignment

<table>
<thead>
<tr>
<th>Design</th>
<th>Benchmark</th>
<th>Bootstrap Bias</th>
<th>Bootstrap SD</th>
<th>Percent Bias</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>6800.21</td>
<td>2.17</td>
<td>182.50</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Comparison Group</td>
<td>6800.21</td>
<td>2194.96</td>
<td>1011.61</td>
<td>0.32</td>
<td>2416.86</td>
</tr>
<tr>
<td>Semi-Parametric Model</td>
<td>6800.21</td>
<td>2238.94</td>
<td>971.78</td>
<td>0.33</td>
<td>2440.74</td>
</tr>
<tr>
<td>Parametric Model</td>
<td>6800.21</td>
<td>2716.04</td>
<td>3538.79</td>
<td>0.40</td>
<td>4460.93</td>
</tr>
<tr>
<td>No Comparison Group -- Linear Extrapolation</td>
<td>6800.21</td>
<td>2194.96</td>
<td>1011.61</td>
<td>0.32</td>
<td>2416.86</td>
</tr>
</tbody>
</table>
Conclusions re Using a Comparison RD to add to Generalization

• Non-equivalent comparison group often reduces bias relative to pure functional form extrapolations.

• Huge gains in statistical precision.

• In terms of MSE the reductions in variance swamp the effects of bias.

• Non-equivalent comparison groups are a useful way of studying the consistency of causal effects beyond the cut-off.
Summary for RDD

• Well warranted in theory and compared to RCT
• One big threat is mis-specified functional form - overfit and use non-parametric
• Another big threat is manipulation - easier to describe than control for
• Another big threat is partial treatment - easy to deal with via instrumental variable approach
• Generalization is limited to cutoff, but a comparison RD can be used to extend generalization
• Research into additional complexities is fast and furious right now and state of the art developing
Implications for Practice

• Which approach to use?
  – Use univariate approach first for estimating $\tau_R$ and $\tau_M$ and to assess whether treatment effects are constant.
  – If treatment effects are constant, use frontier or centering approach for estimating $\tau_{MRD}$.
    • Use frontier approach only if functional form of the response surface is known
    • For the centering approach, may reduce heterogeneity in the outcome by using difference scores for the outcome
  – The IV approach is not recommended because we do not know when the potential outcomes are discontinuous and because of reduced efficiency
• Additional Topic not covered in Class:
• Multiple Assignment Scores
RD with Multiple Cutoffs Design (as commonly implemented)

- Design has a single assignment variable.
- Multiple cutoffs along the assignment variable.
- Cutoffs often vary by sample (i.e. across sites & time).
- Purpose is to estimate average treatment effects across an interval on the assignment variable ($\tau_{mc}$) may be estimated.
  - Black, Galdo, and Smith call it a “discontinuity frontier” but we won’t use that terminology here.
Sample Data with Multiple Cutoffs

<table>
<thead>
<tr>
<th>Observation</th>
<th>Site</th>
<th>Cutoff</th>
<th>Assignment score</th>
<th>Centered Assignment score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>45</td>
<td>43</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>45</td>
<td>49</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>45</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>45</td>
<td>70</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>50</td>
<td>41</td>
<td>-9</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>50</td>
<td>43</td>
<td>-7</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>50</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Analyzing RDs with Multiple Cutoffs

Standard approach: Convert design to a 2 dimensional RD
1. For each cutoff, center assignment variable at zero.
2. Pool observations into a single dataset.
3. Analyze as a two dimensional RD with a single assignment variable and a cutoff at zero.
4. Estimate $\tau_{mc}$ using pooled data.
Typical estimation strategy for RDs with Multiple Cutoffs

\[ Y_{ij} = \alpha + \tau D_{ij} + \beta_1 \overline{AV}_i + \beta_2 D_{ij} \overline{AV}_i + X_i^t \gamma + Z_j^t \delta + \mu_j + \varepsilon_{ij} \]

Where \( Y_{ij} \) is the outcome for participant \( i \) in cutoff group \( j \); \( D \) is whether unit \( i \) is in treatment or comparison in cutoff group \( j \); \( AV_i \) is assignment score for unit \( i \) (standardized and centered by cutoff group); \( X_i \) and \( Z_j \) are vectors of unit and cutoff-group level covariates (respectively); \( u_j \) and \( \varepsilon_{ij} \) are the unit and cutoff group level random error terms, assumed to be iid, and \( \tau \) is the treatment effect for unit \( i \) in cutoff group \( j \).

- Alternatively, cutoff group-based fixed effects might be included in the model.
- This model assumes that 1) each cutoff group has the same functional form; 2) constant treatment effects across cutoffs (though we could include treatment by cutoff interaction terms); and 3) standardizing and centering assignment scores by cutoff group weighs units appropriately for estimating desired treatment effects.
Examples

• Black, Galdo, and Smith (2007)
  – AV: Profiling score between 1 and 20
  – For each employment office each week, applicants at most risk for were assigned to services until the number of slots were nearly filled. Those with tie profiling scores were randomly assigned to treatment and control conditions. All remaining applicants assigned to comparison condition.
  – 32 employment offices, with 272 RDD cutoff groups (ranging in size from 4 to 241)

• Buddelmeyer and Skoufias (2007)
  – AV: Score on scale of material wealth
  – Scale was based on a discriminant score analysis conducted separately for different geographic regions in Mexico. Each region had its own cutoff score.
  – Seven regions, with each region having between 383 and 10,790 households

• Reading First (2007)
  – AV: Scale used by district to assign Reading First funds
  – Assignment variable and cutoff varied by site. Assignment variables included rankings based on student achievement, poverty, ratings on proposals, and other.
  – 18 sites, with each site having between 5 and 32 schools
Issue 1: To standardize or not standardize the assignment variable?
Issue 2: To pool or not pool across multiple cutoff points?

- The benefit of the design is that in theory, we can estimate the average treatment effect across an interval of cutoffs ($\tau_{mc}$).
  - If we include treatment x cutoff group interactions, we could also look at heterogeneous effects by cutoff.

- The concern, however, is that pooling across multiple cutoffs obscures sources of heterogeneity because cutoffs often vary by sites, time, cohorts, grades, etc.
Generalizing Treatment Effects Across Multiple Samples
Pooling is Already Common in Social Research

• Multi-site RCTs in medicine
• Examine each local RCT for obvious sources of bias
• Analyze an OLS model with site dummies and some interactions
• Conclude about average effect across sites and about some interactions.
...Also Common in

- Meta-Analysis
- Take effect sizes from studies of all kinds, including experiments or RD studies
- Test for heterogeneity
- If heterogeneous, explore methodological and substantive reasons for variation in effect sizes
- Method includes RE, RD, Q-E or whatever
What is Special about Pooling RD Estimates?
One Argument

• Very little -- just a different kind of causal estimate to pool
• Actually, it is a particularly unproblematic case since no observational studies and a strong presumption of bias-free estimates
• Hence pooling is only about precision of whatever average estimate you come up with; not really about bias at all.
What is Special about Pooling RD Estimates? A Different Argument

- Each site or study with RD has less power than RE and so SE’s larger
- RD has less transparent assumptions about bias than RE, and so
- Greater uncertainty about bias in estimates and in standard errors compared to RE
Estimating Treatment Effects across Samples
Option 1

1) Aggregated data approach in one big analysis:
   – Pool data from multiple states and cohorts into a single dataset and analyze in HLM
   – Re-center and standardize assignment variable for all sites and pool data to identify weighted average effect across the various discontinuity points.
   – Interaction terms to accommodate heterogeneous samples, treatments, and response functions.
   – Models aggregated response functions across all cohorts, grades and states.
Benefits and Tradeoffs of Option 1

• Benefits (relative to Option 2)
  – Increased sample size to detect effects
  – Most common current strategy for multi-site studies with causal estimates, including Reading First with its RD estimates

• Limitations
  – Less transparent assumptions, especially with interactions between treatment and third variables
  – Differences between disaggregated functional forms are purely a function of sampling error. What happens if they systematically differ?
Estimating Treatment Effects across Samples

**Option 2: Meta-analytic approach**

First, RD-based treatment effects are estimated for each grade within each cohort and each state.

– Second, these treatment effect are successively pooled. Grade-specific effects are averaged within each cohort and state. Then, resulting cohort-specific estimates are averaged within each state and, finally, treatment effects for each state are pooled across states into a single effect.
Benefits and Tradeoffs of Option 2

The main advantage is that at each pooling level the homogeneity of treatment effects can be assessed by comparing the pooled within-effect variance with the corresponding between-effect variance obtained from the estimated parameter covariance matrix of the three-level model. If effects can be considered as homogeneous, pooling is warranted, if not different treatment effects may be reported.

Limitations:

• To get the estimated covariance matrices one must assume that correlation structures of treatment effects are constant across grades, cohort, and states.
  – Or, one can use Hedges et al. robust standard errors method for dealing with correlation between estimates.

• Reduced efficiency relative to option 1
Generalizing Across Multiple Assignment Mechanisms
Distribution of Units in an RD with Two Assignment Variables

A visual depiction
Multivariate RDD with Two Assignment Variables
A visual depiction
Multivariate RDD with Two Assignment Variables
Treatment effects estimated
Recent Education Examples of RDDs with Multiple Assignment Variables

- College financial aid offer (van der Klaauw, 2002; Kane, 2003)
- Remedial education (Jacob & Lefgren, 2004a)
- Teacher professional development (Jacob & Lefgren, 2004b)
- High school exit exams (Martorell, 2005; Papay et al. 2010)
- No Child Left Behind (Gill et al., 2007)
The Average Treatment Effect along the Cutoff Frontier ($\tau_{\text{MRD}}$)

$\tau_{\text{MRD}}$ is the weighted average of conditional expectations given the single frontiers $F_R$ and $F_M$:

$$
\tau_{\text{MRD}} = E[G_i \mid (R_i, M_i) \in F] = w_R E[G_i \mid R_i \in F_R] + w_M E[G_i \mid M_i \in F_M] \\
= w_R \tau_R + w_M \tau_M
$$

Where $G_i$ is the average size of the discontinuity at the R and M cutoff frontiers, and $f(r,m)$ is the joint density function for assignment variables $R$ and $M$. 
Frontier-specific Effect ($\tau_R$)

$$\tau_R = E[G_i \mid R_i \in F_R] = \frac{\int \limits_{m \geq m_c} g(r, m) f(r = r_c, m) dm}{\int \limits_{m \geq m_c} f(r = r_c, m) dm}$$

Where $g(r, m)$ is the treatment function for the $R$ frontier along the $M$ assignment variable, and $f_r(r_i = r_c, m)$ is the conditional density function for the $F_R$.

• To get the conditional expectation $F_R$, we integrate the treatment function with the conditional density function along $F_R$.
• Note that no weights are needed because there is no pooling of treatment effects across $F_R$ and $F_M$.
• Average treatment effect for the $M$ frontier is calculated in a similar with corresponding treatment and density functions.
Treatment Weights for $\tau_{\text{MRD}}$

Weights $w_r$ and $w_m$ reflect the probabilities for observing a subject at the $R$- or $M$-frontiers

$$w_M = \frac{\int_{r \geq r_c} f(r, m = m_c) dr}{\int_{m \geq m_c} \int_{r \geq r_c} f(r = r_c, m) dm + \int_{r \geq r_c} f(r, m = m_c) dr}$$

$$w_R = \frac{\int_{m \geq m_c} \int_{r \geq r_c} f(r = r_c, m) dm}{\int_{m \geq m_c} \int_{r \geq r_c} f(r = r_c, m) dm + \int_{r \geq r_c} f(r, m = m_c) dr}$$

However, note that weights are sensitive to the scaling and distribution of the assignment variables.
Requirements for a valid Multivariate RDD

Similar to RD case with single assignment mechanism

1. A discontinuity in probability of treatment receipt across the frontier;
2. Continuity in the expectations of potential outcomes at $F_R$ and $F_M$. 
Estimating Treatment Effects
Four Proposed Approaches

1. Frontier approach
2. Centering approach
3. Univariate approach
4. IV Approach
Frontier Approach

Estimates the discontinuity along each frontier simultaneously, and applies appropriate weights to obtain the overall effect.

- First, estimate the treatment function, which is the average size of the discontinuity along the cutoff frontiers using parametric, semi-parametric, or non-parametric approaches.
- Second, estimate the joint density function by using a bivariate kernel density estimator or by estimating conditional density functions for $R$ and $M$ separately for observations that lie within a narrow bandwidth around the frontier.
- Third, numerically integrate the product of treatment and joint density functions at the cutoff frontiers to obtain conditional expectations across both frontiers.
- Third, apply appropriate treatment weights to each discontinuity frontier.

Estimates $\tau_{MRD}$, $\tau_M$, $\tau_R$
Centering Approach

Procedure allows researcher to address the “curse of dimensionality” issue by collapsing multiple assignment scores for unit \( i \) to a single assignment variable.

First, for each unit \( i \), center assignment variables \( r \) and \( m \) to their respective cutoffs, that is \( r_i - r_c \) and \( m_i - m_c \).

Second, choose the minimum centered value \( z_i = \min(r_i - r_c, m_i - m_c) \) is chosen as the unit’s sole assignment score.

Third, pool units and analyze as a standard RD design with \( z \) as the single assignment variable.

Estimates \( \tau_{MRD} \)
Univariate Approach

Addresses dimensionality problem by estimating treatment effects for each frontier separately.

First, exclude all observations with $r$ values less than its respective cutoff ($r_c$), and choosing a single assignment variable (say, $m$) and cutoff ($m_c$).

Second, estimate treatment effects by measuring size of discontinuity of the conditional outcomes at the cutoff for the designated assignment variable using parametric, semi-parametric, or non-parametric approaches

Estimates $\tau_R$ or $\tau_M$
IV Approach (1)

Rather than exclude observations assigned to treatment by alternative mechanisms, delegate these cases as “fuzzy” units.

First, designate a treatment assignment mechanism serves as the instrument for treatment receipt.

Second, estimate the local average treatment effect, which is the difference in conditional mean outcomes for treatment and comparison groups divided by the difference in treatment receipt rates for both groups within a neighborhood around the cutoff.
Monte Carlo Study

Wong, V., Steiner, and Cook (2010) examines the performance of the four approaches when the following factors are varied:

1. Complexity of the true response surface
2. Distribution and scale of the assignment variables
3. Methodological approach (frontier, centering, univariate, and IV) for analyzing MRDDs

• Simulations based on 500 replications with a sample size of 5,000 for each repetition.
Summary for RDD

• Well warranted in theory and compared to RCT
• One big threat is mis-specified functional form - overfit and use non-parametric
• Another big threat is manipulation - easier to describe than control for
• Another big threat is partial treatment - easy to deal with via instrumental variable approach
• Generalization is limited to cutoff, but a comparison RD can be used to extend generalization
• Research into additional complexities is fast and furious right now and state of the art developing
Implications for Practice

• Which approach to use?
  – Use univariate approach first for estimating $\tau_R$ and $\tau_M$ and to assess whether treatment effects are constant.
  – If treatment effects are constant, use frontier or centering approach for estimating $\tau_{MRD}$.
    • Use frontier approach only if functional form of the response surface is known
    • For the centering approach, may reduce heterogeneity in the outcome by using difference scores for the outcome
  – The IV approach is not recommended because we do not know when the potential outcomes are discontinuous and because of reduced efficiency