Regression Discontinuity Designs: Potential & Key Challenges

September 6, 2012

OPRE Methodological Advancement Meeting

Philip Gleason

MATHEMATICA Policy Research

RD History

- RD methods have been around for 50 years
  - Thistlewaite & Campbell (1960)
  - See Cook (2008) for history
- Widespread use in policy evaluation more recent
  - Federal evaluations: e.g., Reading First (Gamse et al. 2008)
  - Policy analyses: e.g., Head Start (Ludwig & Miller 2007)
- With RD methods still evolving…
  - Under what circumstances can RD be successfully implemented?
  - What are the key challenges to successful implementation?

Comparing RCT, RD, and Matching Designs

	RCT	RD	Matching (non-exp.)
Selection process fully known	Yes	Yes	No
Selection process controlled by researcher	Yes	No	No
Control/comparison group eligible for program	Yes	No	Usually
Statistical modeling critically important	No	Yes	Yes
Impact estimates can generalize to full participant population	Yes	No	Yes

How Does RD Work?

- **Possible when assignment to treatment (T) determined fully be one or more continuous variables**
  - **Rating variable (R), assignment variable, forcing variable**

- **If R > cutoff value → T**

  If R < cutoff value → C **(or vice versa)**

- **Key assumptions**
  - **Underlying relationship between R and outcome (Y) is continuous**
  - **R and cutoff value are determined independently**
  - **No other interventions based on same cutoff**

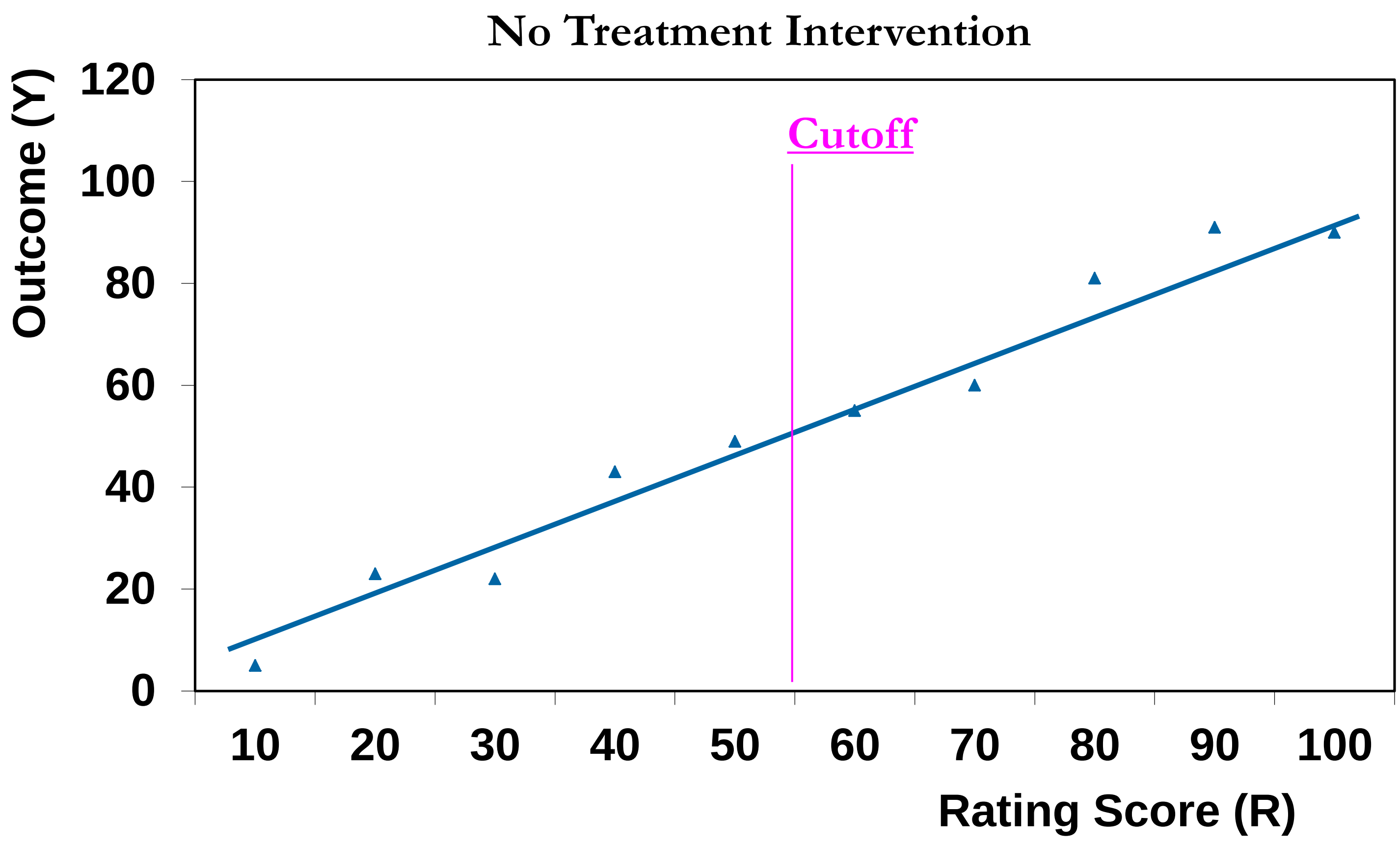
Estimating Impacts under RD Design

- **Regress Y on T, control for R**

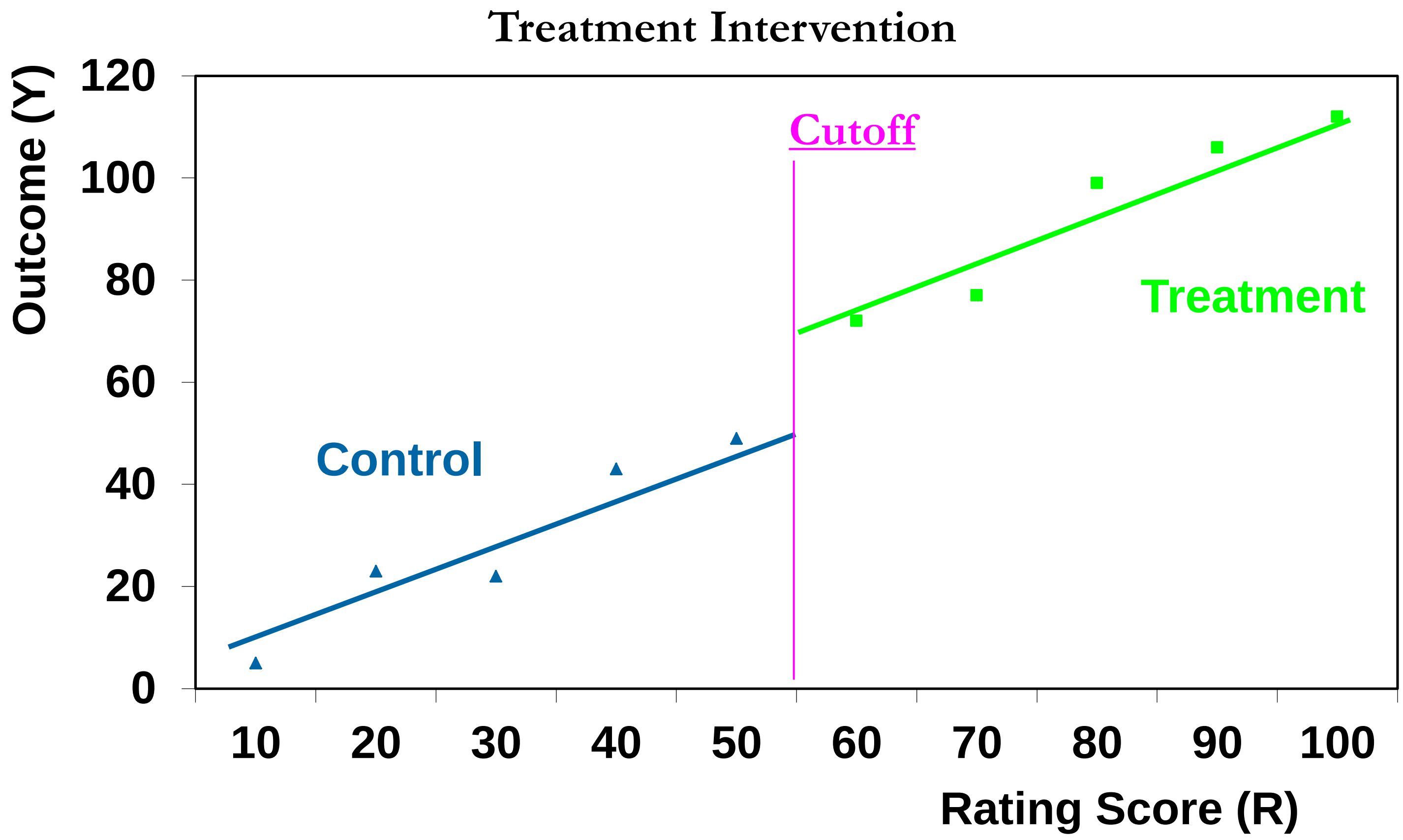
$$Y_i = \alpha + \beta * R_i + \delta * T_i + u_i$$

- **Discontinuity at cutoff = impact of treatment (δ)**
- **Impact estimate generalizes to population at cutoff**
  - **"local" treatment effect**
  - **Measures impact for marginal participant**
- **Conduct specification/robustness checks**

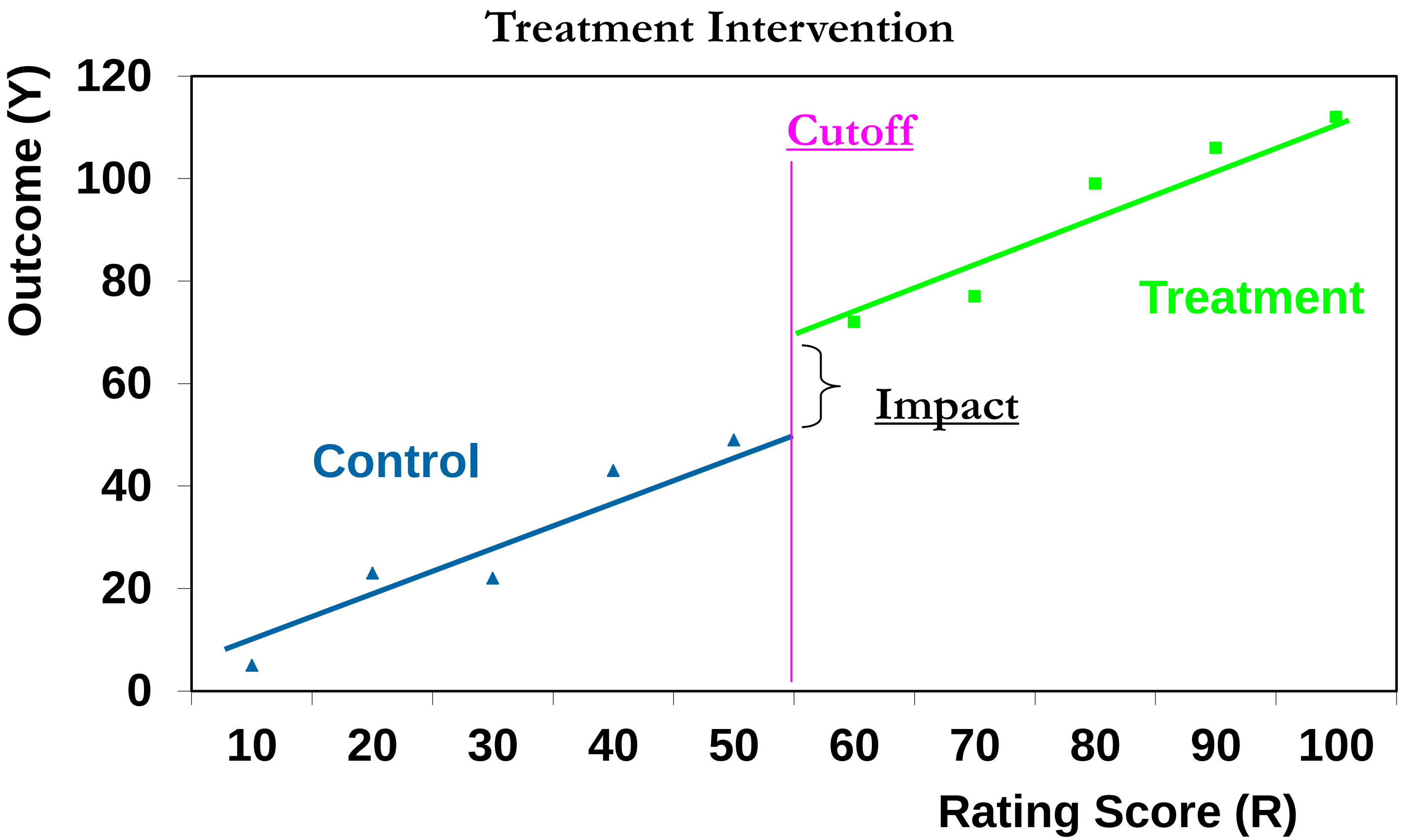
RD Design: Hypothetical Example



RD Design: Hypothetical Example



RD Design: Hypothetical Example



Wrinkles to the Basic RD Design

- **Fuzzy RD**
  - When there is noncompliance to treatment assignment
- **Multiple rating variables**
  - When multiple continuous variables determine participation
- **Different cutoffs for different subpopulations**
  - When treatment assignment rule varies by site/time
- **Supplementing data with pretest version of outcome**
  - Can boost power and allow estimates of impact away from cutoff

Challenges to Successful RD Implementation

- **Manipulation of rating variable**
- **Functional form of relationship between rating variable and outcome (R-Y)**
- **Statistical power**

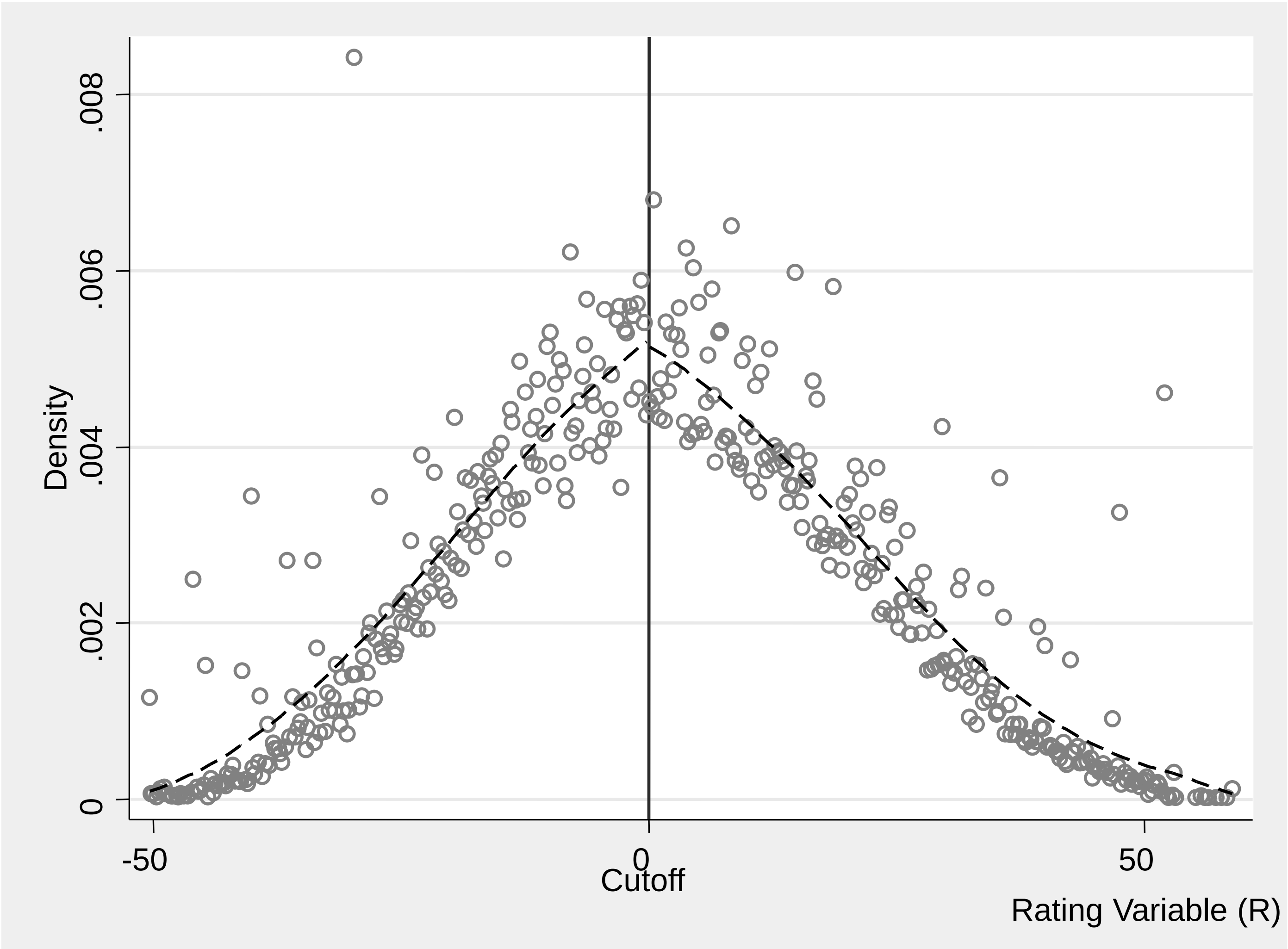
RD Replication Study

- IES methods study comparing RD and RCT designs for estimating impact of same program (Gleason et al. 2012)
- Used “synthetic” RD design: Data from actual RCT used to create synthetic RD data set
  - Choose R from among continuous baseline variables (pretest)
  - Arbitrarily select cutoff (median)
  - Retain only treatment group members with pretest values above cutoff; control group members with values below cutoff
- Conducted using 2 studies, 3 outcomes

Manipulation of Rating Variable

- **Key to RD success is integrity of treatment assignment**
  - Violated if manipulation of R
- **Manipulation occurs when values of R for some units "systematically changed from their true values to influence treatment assignment" (WWC) e.g.,**
  - Free school meal applicants underreport income to become eligible
  - Scholarship sponsor adds points to test scores of favored students
- **WWC calls for evidence of no manipulation**
  - Institutional: no opportunity or incentive to change scores
  - Statistical: no discontinuity in density of R at cutoff (McCrary 2008)

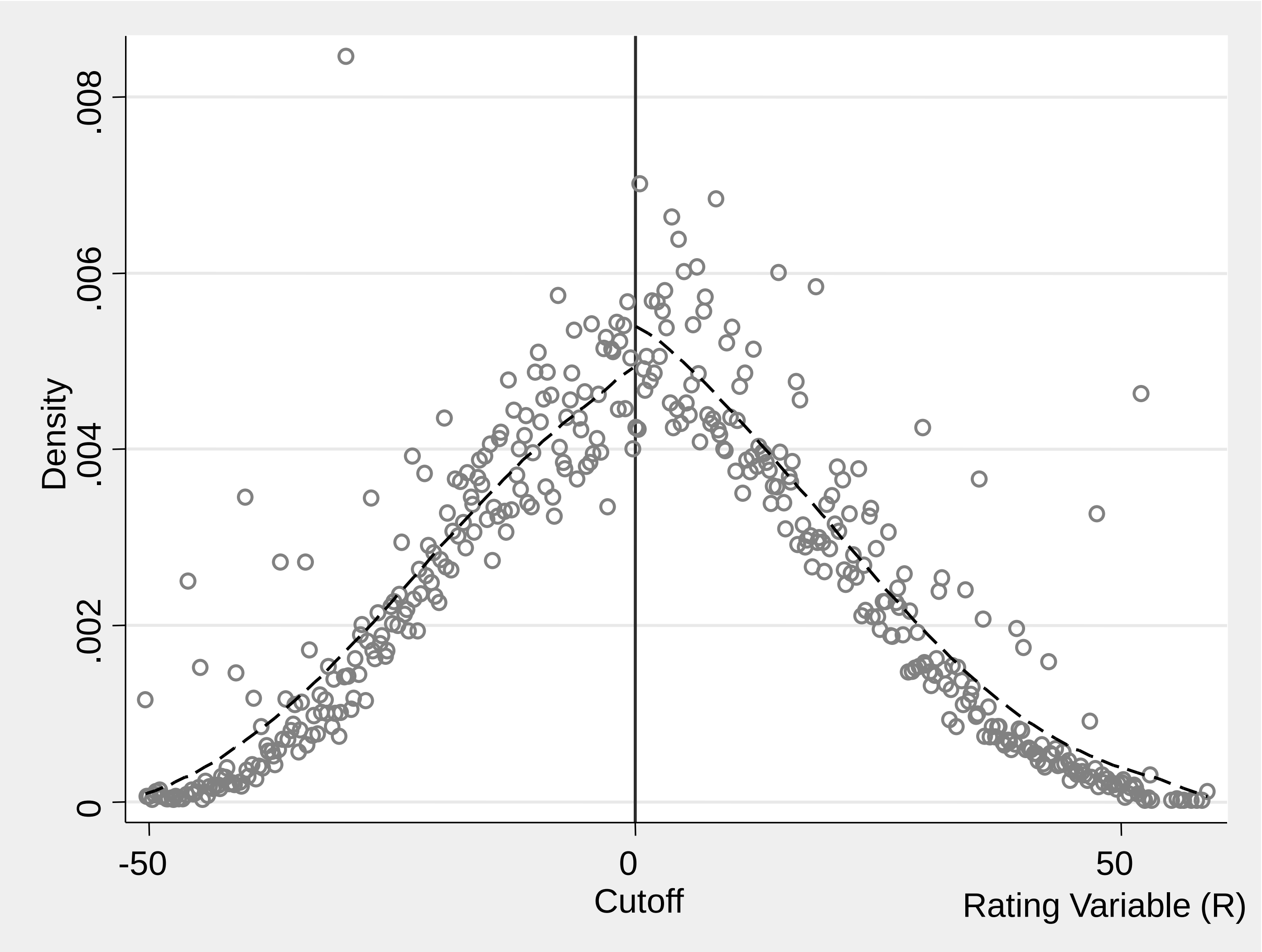
Example: No Manipulation



McCrary Test Stat = -0.30

Gleason, Resch, and Berk 2012

Example: Manipulation Present



McCrary Test Stat = 1.86*

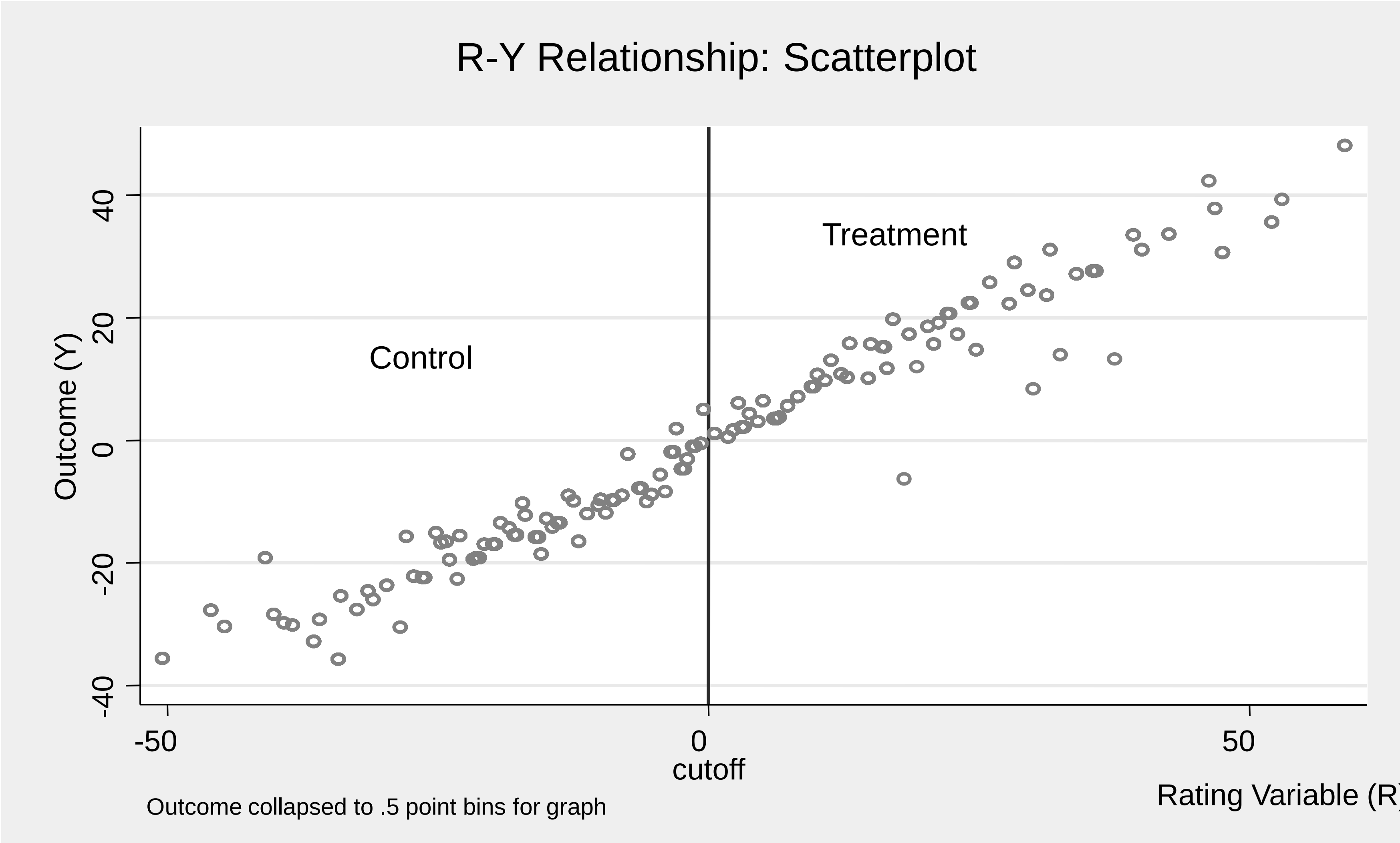
Gleason, Resch, and Berk 2012

MATHEMATICA Policy Research

Properly Modeling Functional Form

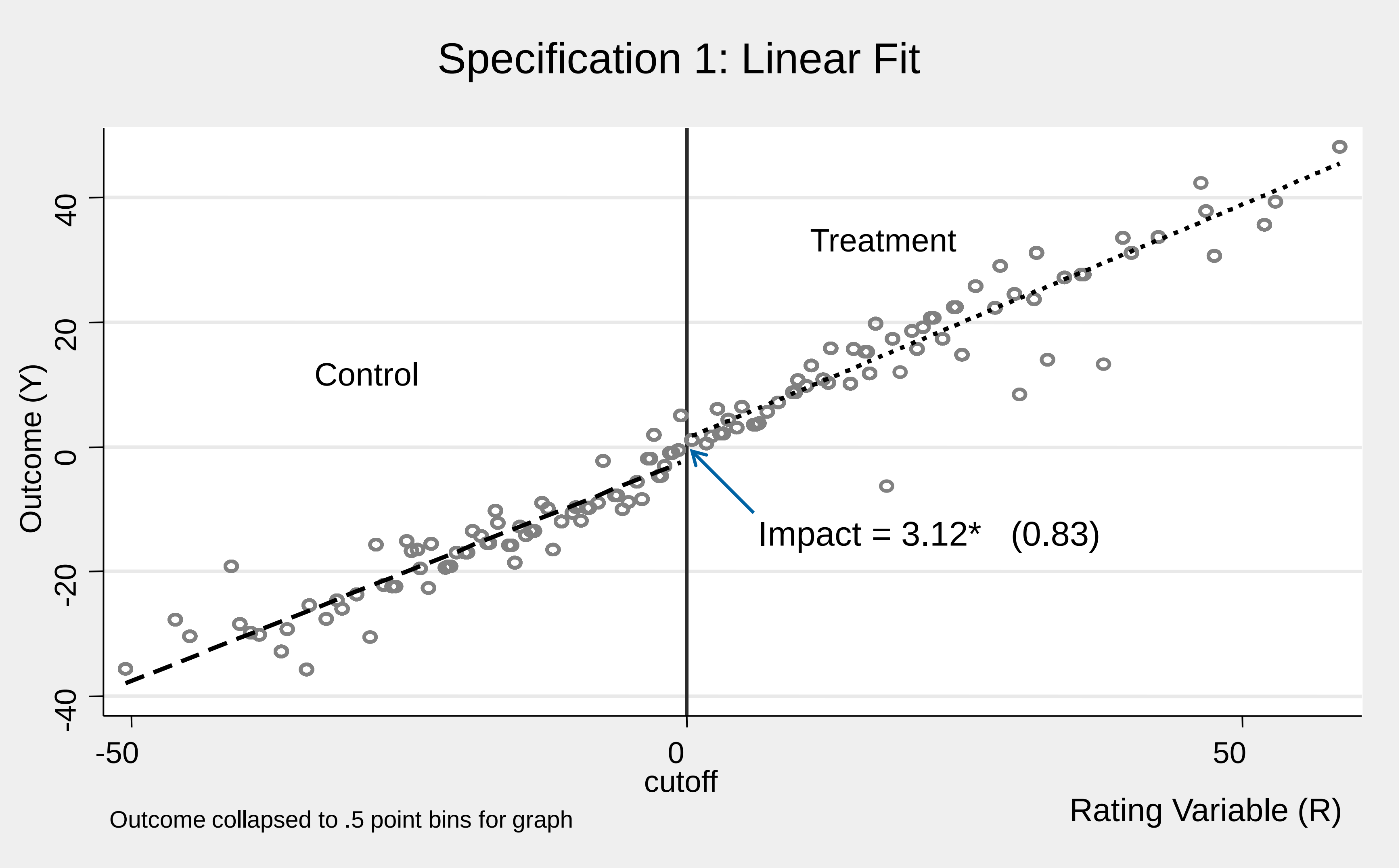
- **RD success hinges on properly modeling R-Y relationship**
- **Approaches**
  - **Graphical**
  - **Parametric**
    - Use all data; assume particular parametric form
    - Potentially allow linear, quadratic, cubic terms (allow to differ on either side of cutoff)
  - **Nonparametric**
    - Estimate range of data (bandwidth) for which relationship is linear
    - Assume linear functional form within that bandwidth

RD High Graphical Analysis



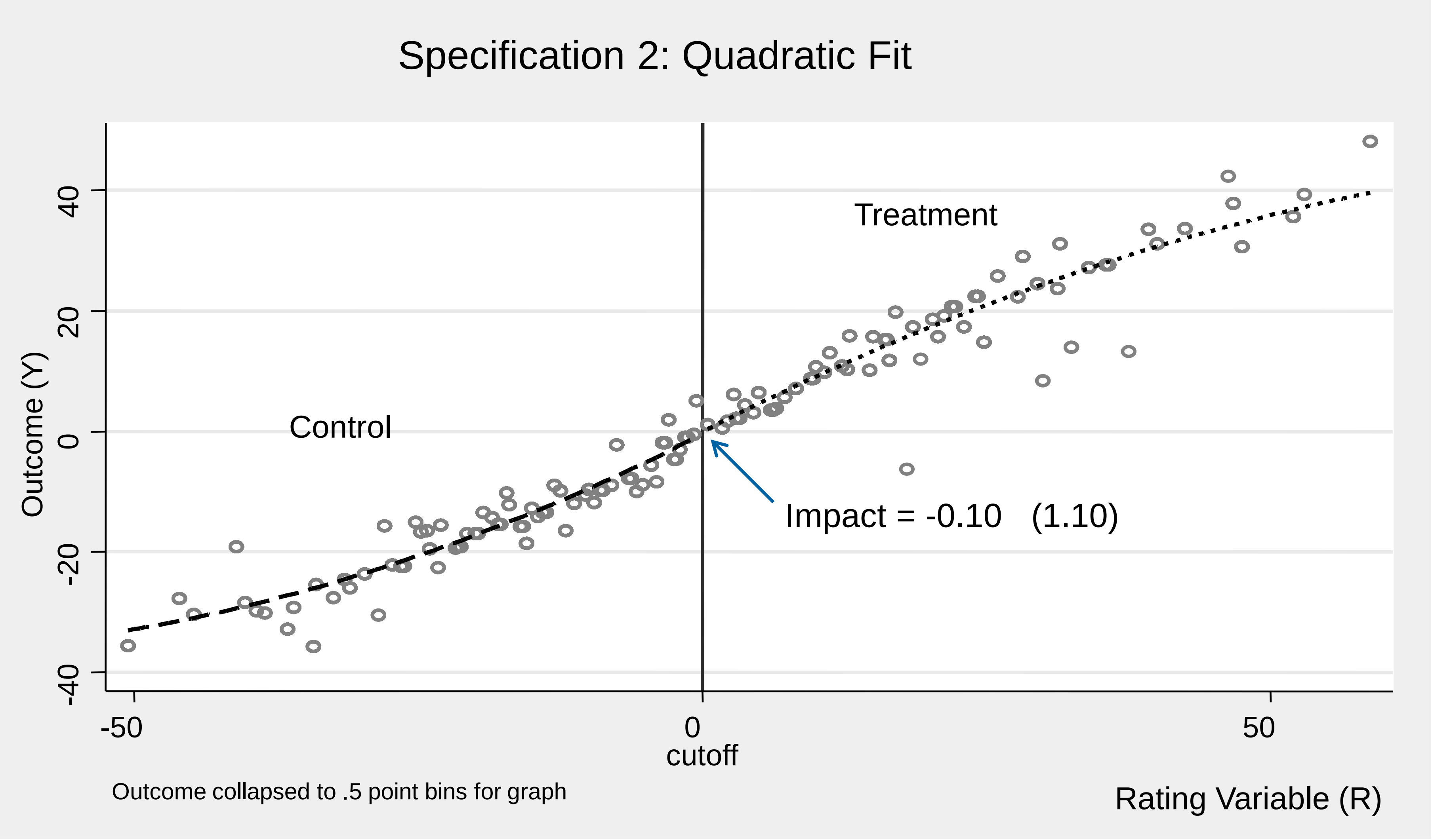
Gleason, Resch, and Berk 2012

RD High Graphical Analysis



Gleason, Resch, and Berk 2012

RD High Graphical Analysis



Gleason, Resch, and Berk 2012

Statistical Power of RD Designs

- **RD has less power than RCT with same sample size**
  - **Correlation between R and T leads to RD design effect (DE) of 2.75 to 4.00 for linear spec (Goldberger 1972, Schochet 2008)**
  - **Need sample 2.75 to 4 times as large**
- **Power loss may be even larger depending on functional form**
  - **Parametric: possible inclusion of quadratic/cubic terms leads of DE of 2.75 to 9 (Bloom 2012)**
  - **Nonparametric: optimal bandwidth may be narrow, potentially leading to DE over 10 (Deke and Dragoset 2012)**

Implications of Using Underpowered RD Design

- **RD replication study**
  - **12 RD-RCT comparisons**
  - **For any single comparison, RD has much less power**
- **The good news:**
  - **On average, RD produces similar impact estimate as RCT**
  - **Similar RD/RCT estimates in some individual comparisons:**
    - RD impact = 0.08 (.05)   RCT impact = 0.07* (.03)
- **The bad news:**
  - **RD estimate rarely significant (1/12); RCT significant in 8/12**
  - **Some individual RD estimates very different from RCT**
    - RD impact = -0.06 (.12)   RCT impact = 0.07* (.03)

RD Promising, but Challenges Remain

Issue	Implication
Treatment must be assigned based on one or more continuous rating variables	RD can only be used in some situations
RD produces “local” impact estimate (at cutoff value)	Not appropriate if impact estimate covering all participants is needed
Limited statistical power	Need large data set
Hard to know in advance how much statistical power RD will have (i.e., how large sample needs to be)	Makes prospective RD studies more challenging
Requires sophisticated statistical analysis	Lots of ways analysis can go wrong; Much less intuitive/transparent than RCT