Non-Randomized Evaluations: Regression Discontinuity Design

Alan de Brauw
Markets, Trade, and Institutions Division
IFPRI

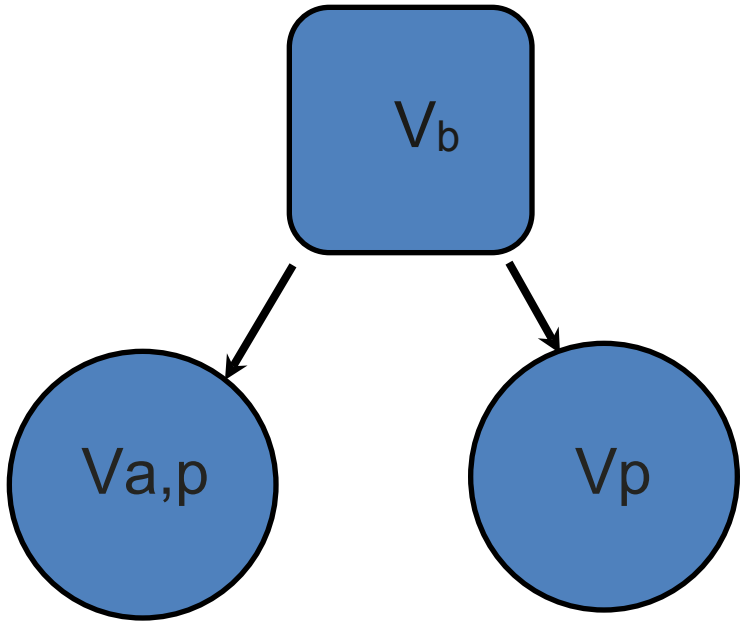
May 2013

Outline

- Basics of Program Evaluation (my terminology)
- Basics of RDD
- Testing Assumptions behind RDD
- More details on implementing RDD
  - Parametric
  - Non-Parametric

Basic Program Evaluation Problem

- Consider a program being run, and a potential participant, Vicente
- Outcome before program=V_b
- Outcome if receiving program=$V_{a,p}$
- Outcome if not receiving program=$V_{a,np}$



Impact=Va,p-Va,np

Problem with "ideal" measurement of impact

- We want to know the causal impact of the program on Vicente or his outcomes
  - We cannot observe Vicente (or anyone) in both "states of nature" (with and without the program, or treatment)
- So we need to instead estimate the impact
  - We're typically interested in the average treatment effect– what the average effect is from the program on individuals who receive it
- We are going to be concerned mostly about limiting **bias** in our estimate

Identification of Causal Impact

- *How do we know measured impacts are due to the program or treatment?*
  - Again, we want to claim that the effect of d on y is *causal*
  - Randomizing beneficiaries allows causal inference
  - But randomization not always possible
- What is the (main) theoretical problem with using non-randomized methods?

Problem of selection bias

- Assume we can break up our individuals into two groups: treatment and control, but not necessarily randomly allocated
  - Y_1 represents the outcome among those receiving program (treatment)
  - Y_0 represents outcome in the control group.
  - d indicates treatment group status. The average treatment effect on the treated is

$$\Delta^{ATT} = E\left[Y_1 | d = 1\right] - E\left[Y_1 | d = 0\right]$$

- If we add and subtract $E\left[Y_0 | d = 0\right]$
- We find…

Selection Bias

$$\Delta^{ATT} = \underbrace{E\left[Y_1|d=1\right] - E\left[Y_0|d=0\right]}_{observable} - \underbrace{E\left[Y_1|d=0\right] - E\left[Y_0|d=0\right]}_{bias}$$

- In words:
  - We can observe differences between the treatment and control groups
  - But if there are *any* sources of potential differences between the treatment and control groups, then estimates of the impacts of a program are likely to be biased

Selection bias (cont.)

- Most common sources of selection bias ...
  - Targeting or program placement bias
    - Sites for intervention selected based on characteristics (e.g., agroecology, poverty, infrastructure)
  - Self-selection bias
    - Whether a household receives treatment depends importantly on household (or community) decision to participate
    - E.g. agricultural extension demonstrations

Regression Discontinuity Design (RDD)

- RDD uses the fact that some programs are targeted using an eligibility score
  - Individuals above that score are eligible, below score are ineligible
  - Comparing beneficiaries *just* above and non-beneficiaries *just* below score gives a local estimate of the average treatment effect on the treated
    - Idea is that these individuals are quite comparable with one another

RDD (cont.)

- RD is possible when units can be ordered along a quantifiable dimension which is systematically related to the assignment of treatment
- The treatment effect is measured at the discontinuity – estimated impact around the cutoff may not generalize to entire population– Local Average Treatment Effect (LATE)

RDD Vocabulary

- **Forcing Variable:** The forcing variable is the variable that determines program eligibility (proxy means test, score, etc.). (*X*)
- **Threshold:** The value of the forcing variable that determines eligibility. Above the threshold, individuals or units become eligible for the program; below it they are not eligible. (*c*)
- **Bandwidth:** The width of the intervals around the threshold we use to compare the treatment and control groups. (*h*)

Review of Regression Discontinuity Estimator

- To estimate local avg. treatment effect δ:

$$\delta = Y^+ - Y^- = \lim_{\varepsilon \to 0} \begin{matrix} E(Y(1)|X = c + \varepsilon) \\ -E(Y(0)|X = c - \varepsilon) \end{matrix}$$

where:

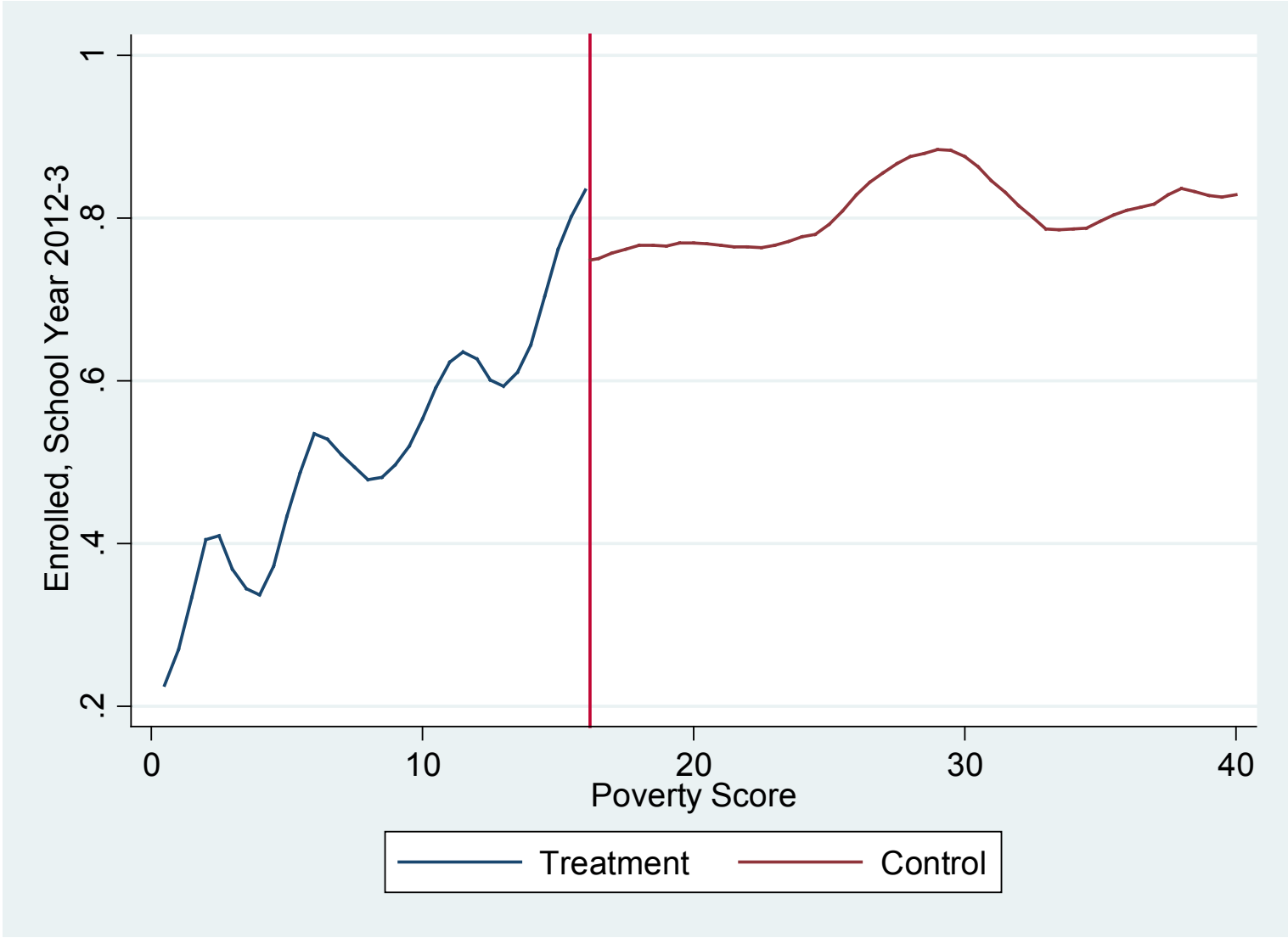
$$Y^+ = \frac{\sum_{X_i \geq c}^{i} Y_i K(\frac{X_i - c}{h})}{\sum_{X_i \geq c}^{i} K(\frac{X_i - c}{h})}$$

$$Y^- = \frac{\sum_{X_i < c}^{i} Y_i K(\frac{c - X_i}{h})}{\sum_{X_i < c}^{i} K(\frac{c - X_i}{h})}$$

Conceptual Discussion

- Should LATE=ATE? Depends upon the outcome of interest and the program.
  - If outcome is correlated with the forcing variable, then the answer is probably no (graph on next page is an example)
  - If outcome is NOT, then they should be equal
    - Example: Cash transfers, variables measuring women's empowerment

Example: Correlated Outcome (Pakistan BISP)



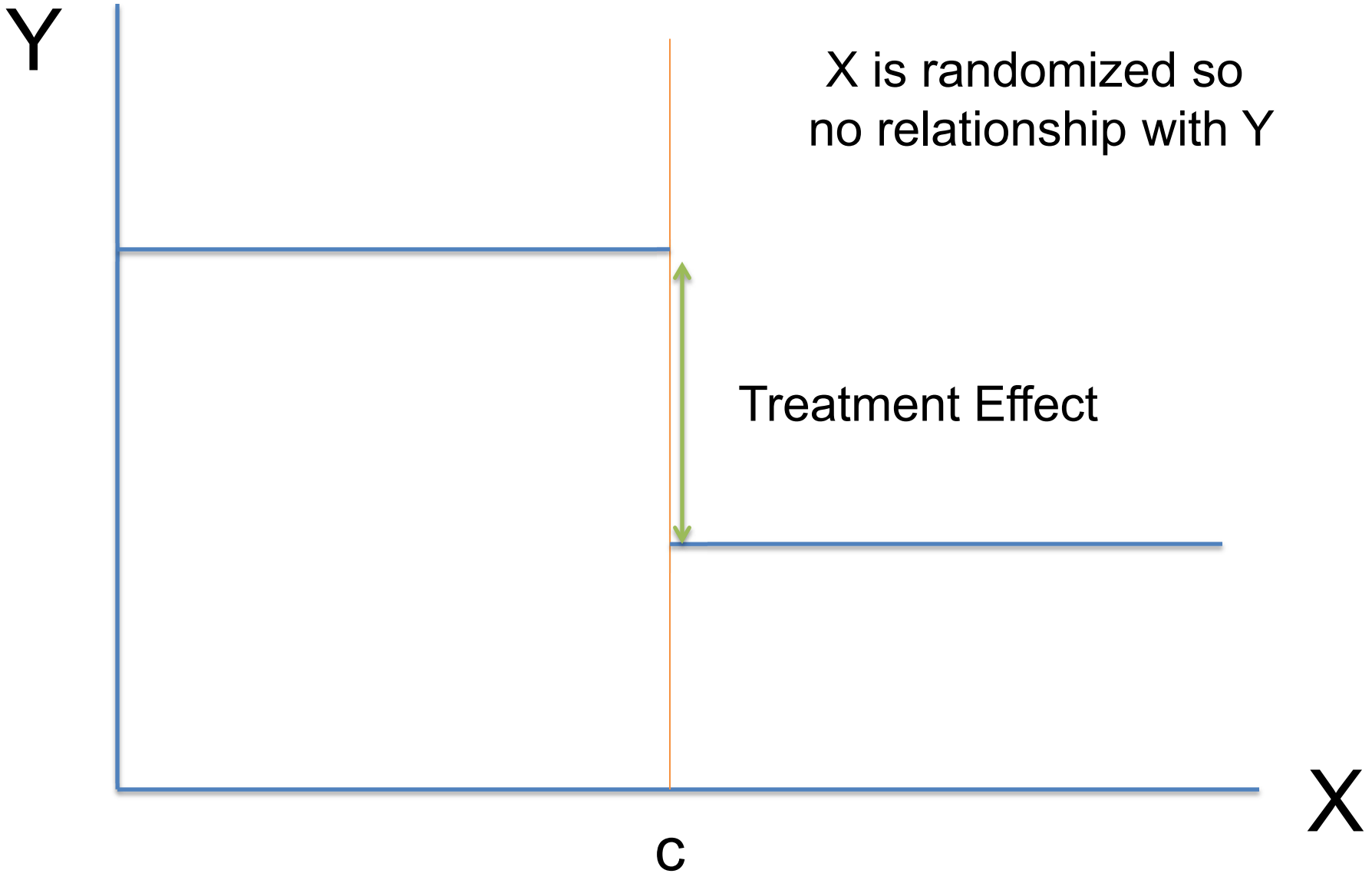
Thought Experiment

- What is relationship between RDD and randomized experiment?
  - What if we randomly assign a forcing variable in a randomized experiment?
  - Random cutoff c in interval of X between (0,1)

  Outcome among treated is E(Y(1)|X) and outcome among not treated is E(Y(0)|X)

  Treatment effect is E(Y(1)-Y(0)|X=c)

Thought Experiment



Assumptions necessary for RD to be asymptotically unbiased

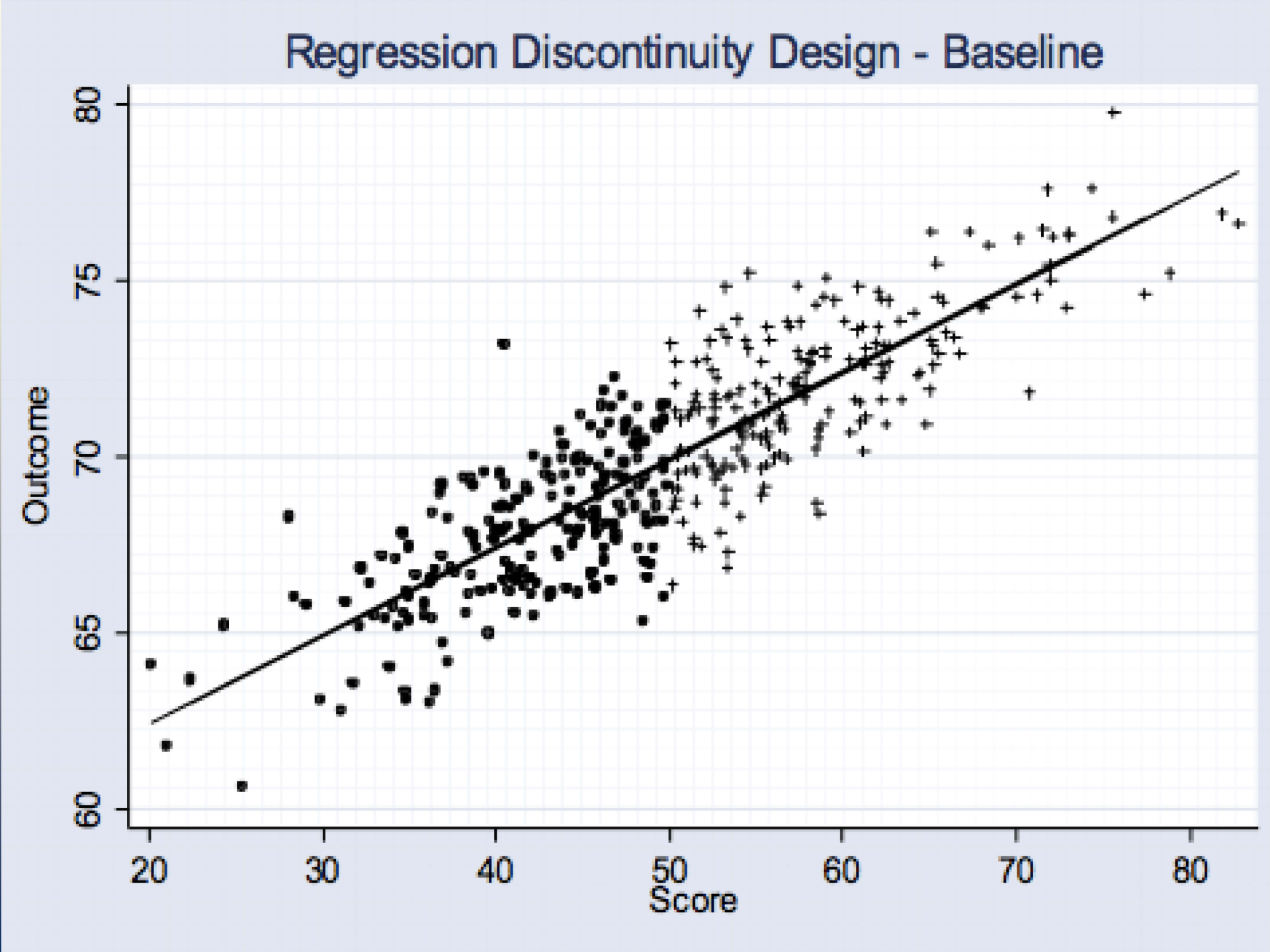
- Probability of treatment must vary discontinuously at the threshold
- Observations just above and below threshold must be similar in observed characteristics
- Must assume that if treatment did not occur, outcome would be continuous at threshold

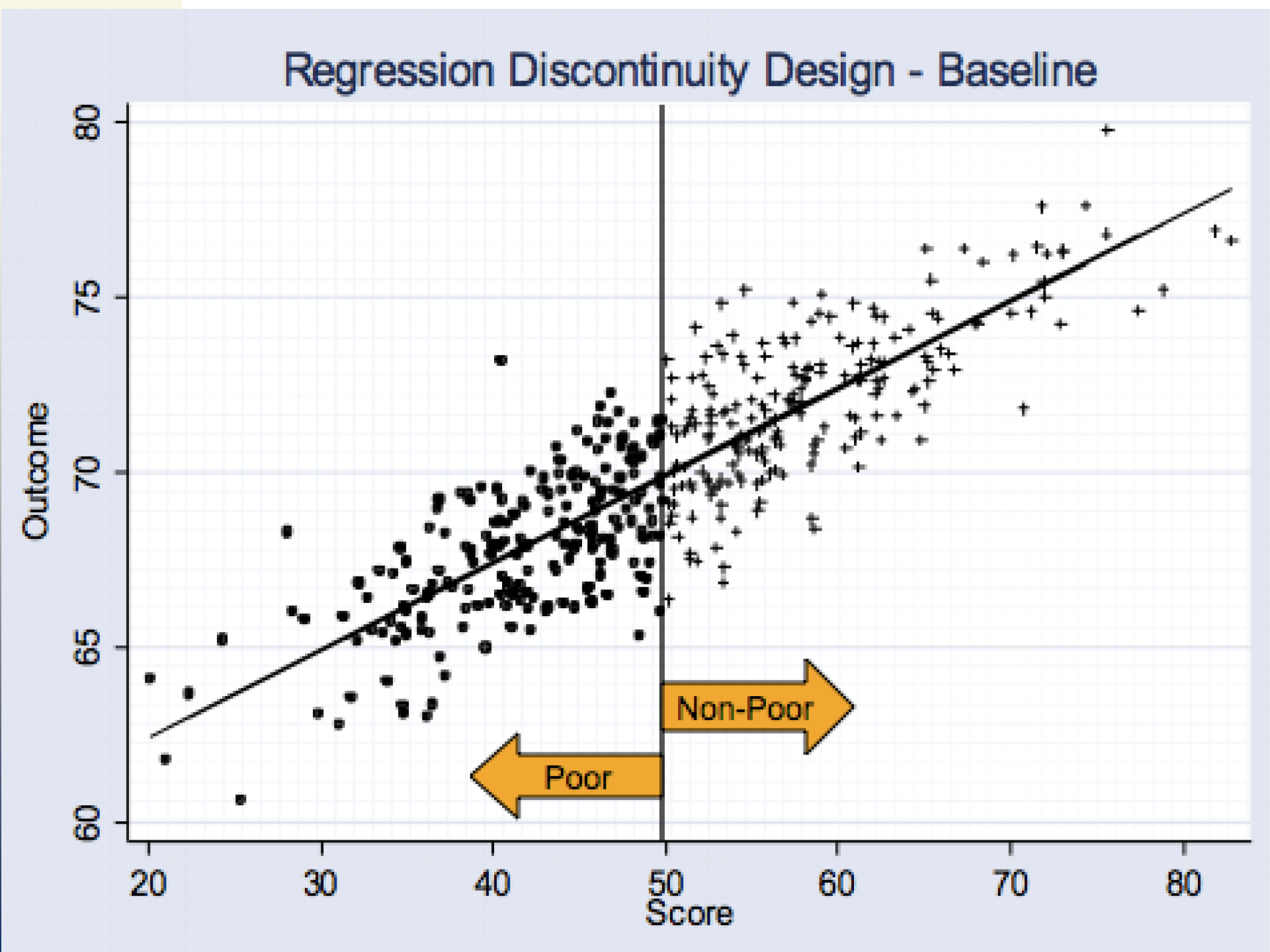
Examples of Programs with indexes used to determine eligibility

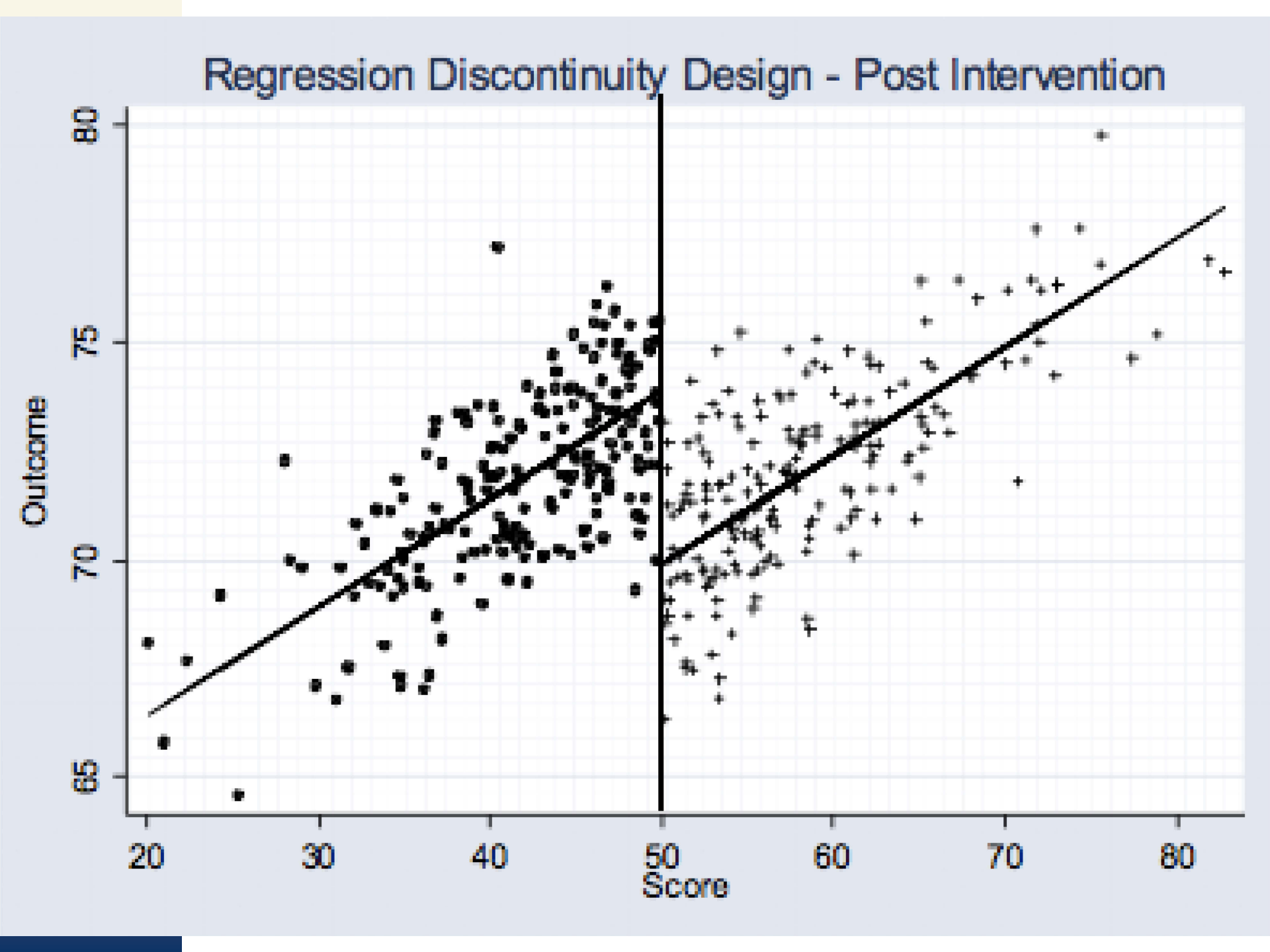
- Anti-poverty programs (cash, input subsidies)
  - Targeted to poor below specific poverty index score
- Pension Programs
  - Targeted to people at a specific age
- CDD Programs
  - Awarded to projects with highest scores

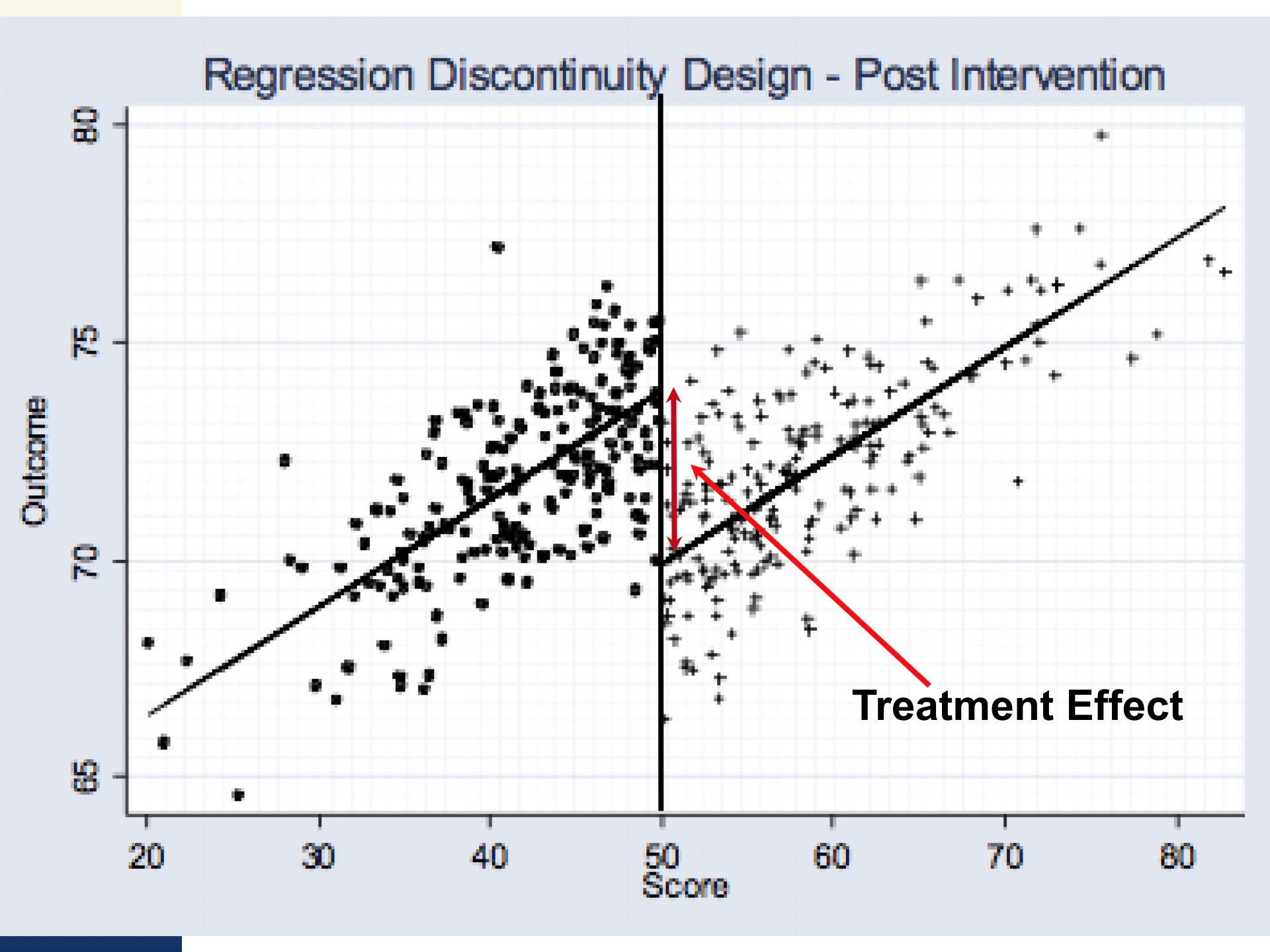
Example: Effect of Cash Transfer on Consumption (artificial)

- Target transfer to poorest households
- Construct poverty index from 1 to 100 with pre-intervention characteristics
- Households with a score <=50 are poor; Households with a score >50 are non-poor
- Cash transfer given to poor households
- Measure outcomes (i.e. consumption) before and after transfers begin


Regression Discontinuity Design - Baseline
Outcome
60
65
70
75
80
20
30
40
50
60
70
80
Score


Regression Discontinuity Design - Baseline
Outcome
60
65
70
75
80
20
30
40
50
60
70
80
Score
Poor
Non-Poor


Regression Discontinuity Design - Post Intervention
Outcome
65
70
75
80
20
30
40
50
60
70
80
Score


Regression Discontinuity Design - Post Intervention
Outcome
80
75
70
65
20
30
40
50
60
70
80
Score
Treatment Effect

Potential Disadvantages of RD

- Local average treatment effects – not technically generalizable
- Statistical Power: Effect estimated at discontinuity, so fewer observations than in a randomized experiment with the same sample size
- Specification can be sensitive to functional form or the way the relationship is modeled

Advantages of RDD

- RDD yields an unbiased estimate of treatment effect at the discontinuity
- Can many times take advantage of a known rule for assigning the benefit that are common in the designs of social policy
  - No need to "exclude" a group of eligible households/individuals from treatment
  - Can be used when rolling out a program
- Easy to illustrate impacts on a graph

Assumptions necessary for RD to be asymptotically unbiased

- Probability of treatment must vary discontinuously at the threshold
- Observations just above and below threshold must be similar in observed characteristics
- Must assume that if treatment did not occur, outcome would be continuous at threshold

Testing RDD Assumptions

- Probability of treatment must vary discontinuously at the threshold
  - McCrary (2008) density test is most rigorous
- Observations just above and below threshold must be similar in observed characteristics
  - Can check the distribution of observed characteristics around the threshold

Testing RDD Assumptions

- Must assume that if treatment did not occur, outcome would be continuous at threshold
  - Pre-program data available?
  - Can check for discontinuities in pre-program data or measure in diff-in-diff

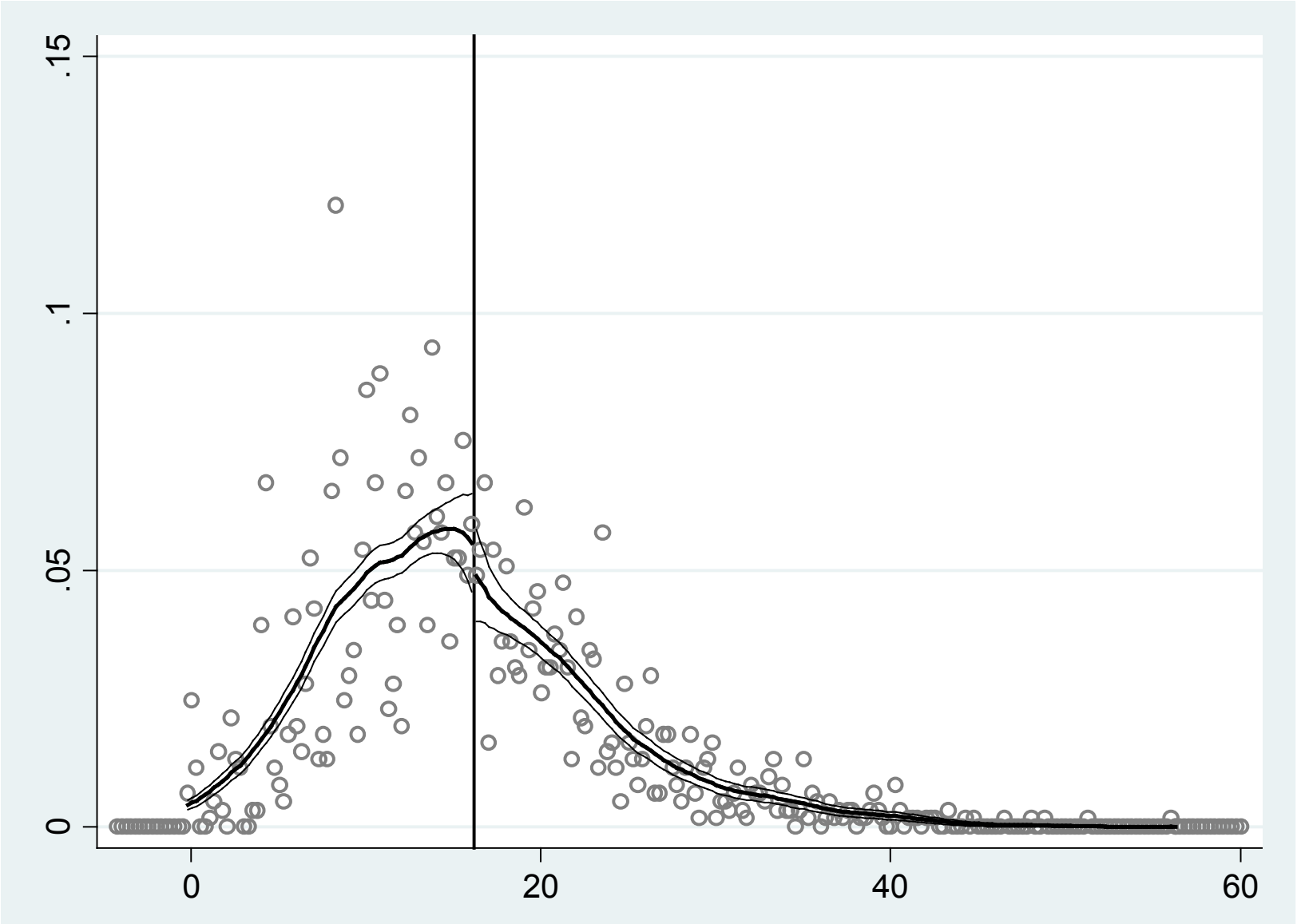
McCrary Density Test

- Really concerned with manipulation of the forcing variable
- Test: break up the forcing variable into "bins" of width *b* around the threshold
- Count frequency of observations in each bin, place more weight on bins near the threshold
- Interest is in the difference in (log) height of bins at the threshold

McCrary density test (cont.)

- Test statistic is whether there is a difference at the threshold
- Dependent upon the *sampling* (not in his article)
  - If you oversample beneficiaries (near the threshold), likely to reject even if no discontinuity

Example: Pakistan BISP



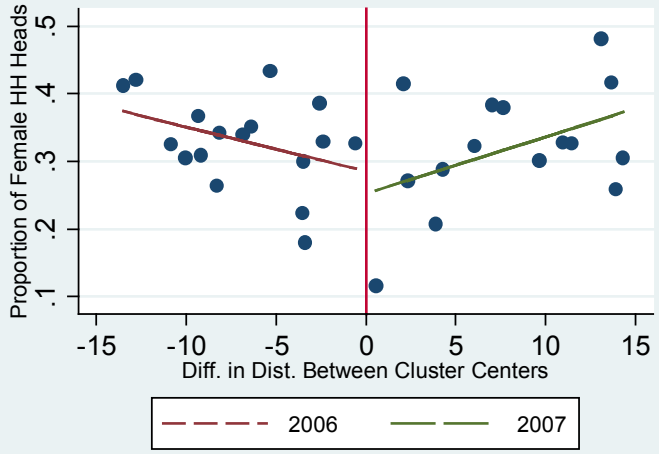
Testing for Differences in Other Characteristics

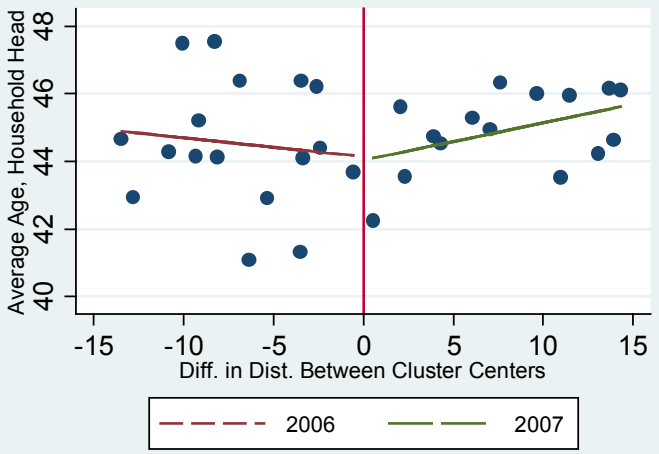
- Idea: Check for discontinuities in characteristics of observations that should not be affected by the treatment
  - Baseline characteristics are ideal
  - Like a randomization validity check

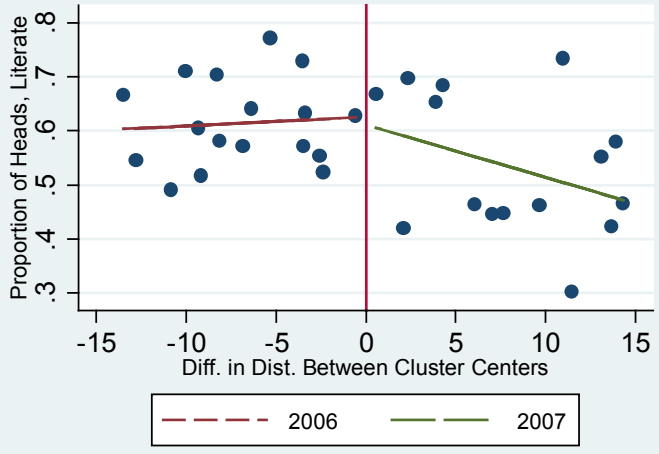
Testing for Differences in Other Characteristics

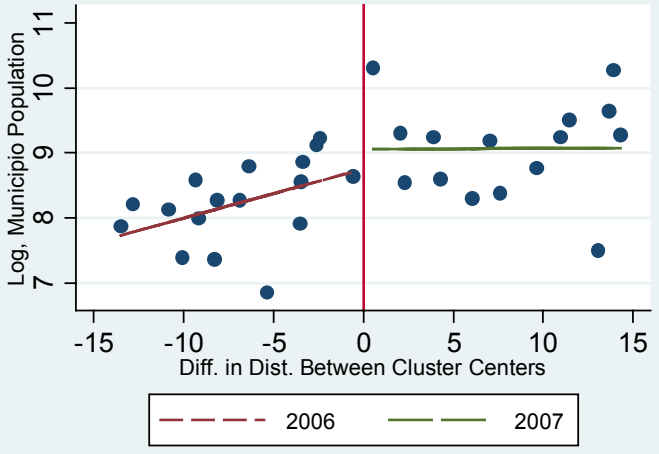
- Idea: Check for discontinuities in characteristics of observations that should not be affected by the treatment
  - Baseline characteristics are ideal
  - Like a randomization validity check
- Can use either a graphical method or a parametric method

Graphical Method: El Salvador









Last Quirk: Imperfect Compliance

- Perfect compliance:

$$\Pr(D|X > c) = 1;$$

$$\Pr(D|X < c) = 0$$

- Imperfect compliance:

$$\Pr(D|X > c) < 1;$$

$$\Pr(D|X < c) > 0;$$

$$\Pr(D|X > c) > \Pr(D|X < c)$$

- Can we use RDD?

Fuzzy RDD

- Yes, so long as

$$\lim_{x \downarrow c} \Pr(D = 1 | X = x) \neq \lim_{x \uparrow c} \Pr(D = 1 | X = x)$$

- Can we use RDD?
  - Yes, only need this jump in probability at the threshold
- New treatment effect estimator:

$$\delta_F = \frac{\lim_{x \downarrow c} \Pr(Y | X = x) - \lim_{x \uparrow c} \Pr(Y | X = x)}{\lim_{x \downarrow c} \Pr(D | X = x) - \lim_{x \uparrow c} \Pr(D | X = x)}$$

Sharp RDD Estimator

- To estimate local avg. treatment effect δ:

$$\delta = Y^+ - Y^- = \lim_{\varepsilon \to 0} \begin{matrix} E(Y(1)|X = c + \varepsilon) \\ -E(Y(0)|X = c - \varepsilon) \end{matrix}$$

where:

$$Y^+ = \frac{\sum_{X_i \geq c}^{i} Y_i K(\frac{X_i - c}{h})}{\sum_{X_i \geq c}^{i} K(\frac{X_i - c}{h})}$$

$$Y^- = \frac{\sum_{X_i < c}^{i} Y_i K(\frac{c - X_i}{h})}{\sum_{X_i < c}^{i} K(\frac{c - X_i}{h})}$$

Choice of Kernel

- Rectangular kernel
- Triangular kernel (linear decrease as farther away from threshold)
- Gaussian kernel
- Epanechnikov kernel (curve falling off to 0 at bandwidth)
- Others also
- Imbens and Lemieux (2009) suggest rectangular or triangular kernels

Parametric Regression

- For X<c:

$$Y = \alpha_l + f_l(X - c) + \varepsilon$$

- For X>c:

$$Y = \alpha_r + f_r(X - c) + \varepsilon$$

Where f(.) is a function. Then $\tau = \alpha_r - \alpha_l$

How to choose f(.)?

Choice of f(.)

- Linear

$$Y_i = \alpha + \beta D_i + \gamma(X_i - c) + \delta D_i(X_i - c) + \varepsilon_i$$

- Quadratic
  - Add quadratic term, interact with treatment variable
- Cubic
- Quartic, etc.
- How to choose?
  - Suggestion- Akaike Information Criterion

AIC

- Formula

$$AIC = Nln(\hat{\sigma}^2) + 2p$$

Where *p* is the number of parameters in the model

If explanatory power increases by adding terms, then AIC should decrease

Non-Parametric Estimation

- Can also estimate RDD impacts with a non-parametric Wald estimator (Hahn et al., 2001)
  - Programmed by Nichols (2007) in *Stata*
- Uses local polynomial estimation; standard errors either need to be bootstrapped "by hand" or in the Nichols program

Bandwidth

- How to choose the bandwidth?
- Recall

$$\delta = Y^+ - Y^- = \lim_{\varepsilon \to 0} \begin{matrix} E(Y(1)|X = c + \varepsilon) \\ -E(Y(0)|X = c - \varepsilon) \end{matrix}$$

- So the estimator is defined **at the limit**
- Bias increases as bandwidth increases, but…
- So does precision

Choice?

- Best recommendation- perhaps test several bandwidths
- Imbens and Kalayaranaman (2009) provide an "optimal bandwidth" calculation
  - Minimizes mean square error of the estimator
- Calonico et al (Working paper) are recommending even smaller bandwidth
- This is very context specific and you should look for consistency in estimates

Estimating Fuzzy RDD

- Treatment effect estimator:

$$\delta_F = \frac{\lim_{x \downarrow c} \Pr(Y|X = x) - \lim_{x \uparrow c} \Pr(Y|X = x)}{\lim_{x \downarrow c} \Pr(D|X = x) - \lim_{x \uparrow c} \Pr(D|X = x)}$$

- Can estimate as a system:

$$Y_i = \alpha + \beta D_i + \gamma(X_i - c) + \delta D_i(X_i - c) + \varepsilon_i$$
$$D_i = \alpha + \beta T_i + \gamma(X_i - c) + \delta T_i(X_i - c) + \varepsilon_i$$

Where *T* represents being beyond the threshold (2SLS)

Otherwise.. (estimating Fuzzy RDD)

- Can also use non-parametrics (Nichols program)
  - I am skeptical about this estimator- I think the results are too large
  - I'll show you later in Stata exercises, but…
- Duflo (2003)- impact on WHZ scores among girls
  - OLS, no controls: 0.24
  - OLS, controls: 0.61
  - 2SLS, controls: 1.19