

If you're a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact Marie Rienzo at marie.rienzo@nih.gov.

Methods: Mind the Gap Webinar Series

Regression Discontinuity Designs in Public Health Research

Presented by

Jacob Bor, Sc.D., S.M.

Boston University School of Public Health



Regression discontinuity designs in public health research

Jacob Bor

Departments of Global Health and Epidemiology
Boston University School of Public Health

September 27, 2018

NIH Office of Disease Prevention
“Mind the Gap” Webinar

Who said it?

“As is the case for any observational study, our results might ...be affected by unmeasured confounding factors.”

Who said it?

“As is the case for any observational study, our results might ...be affected by unmeasured confounding factors.”

Source:

- Me.

Who said it?

“As is the case for any observational study, our results might ...be affected by unmeasured confounding factors.”

Source:

- Me.
- You?

Who said it?

“As is the case for any observational study, our results might ...be affected by unmeasured confounding factors.”

Source:

- Me.
- You?
- Every epidemiologist and social scientist?

Who said it?

“As is the case for any observational study, our results might ...be affected by unmeasured confounding factors.”

Source:

- Me.
- You?
- Every epidemiologist and social scientist?
- When To Start Consortium. (2009). Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. *Lancet*, 373(9672), 1352.

Study designs in health research

Randomized Controlled Trials

- Internal validity
- Balance on both observed *and unobserved* factors
- Cost / ethics constraints

Observational Studies

Cohort, Case-Control,
Cross-Sectional

- External validity
- Balance only on observables
- Strong assumption: no residual confounding.

Study designs in health research

Randomized Controlled Trials

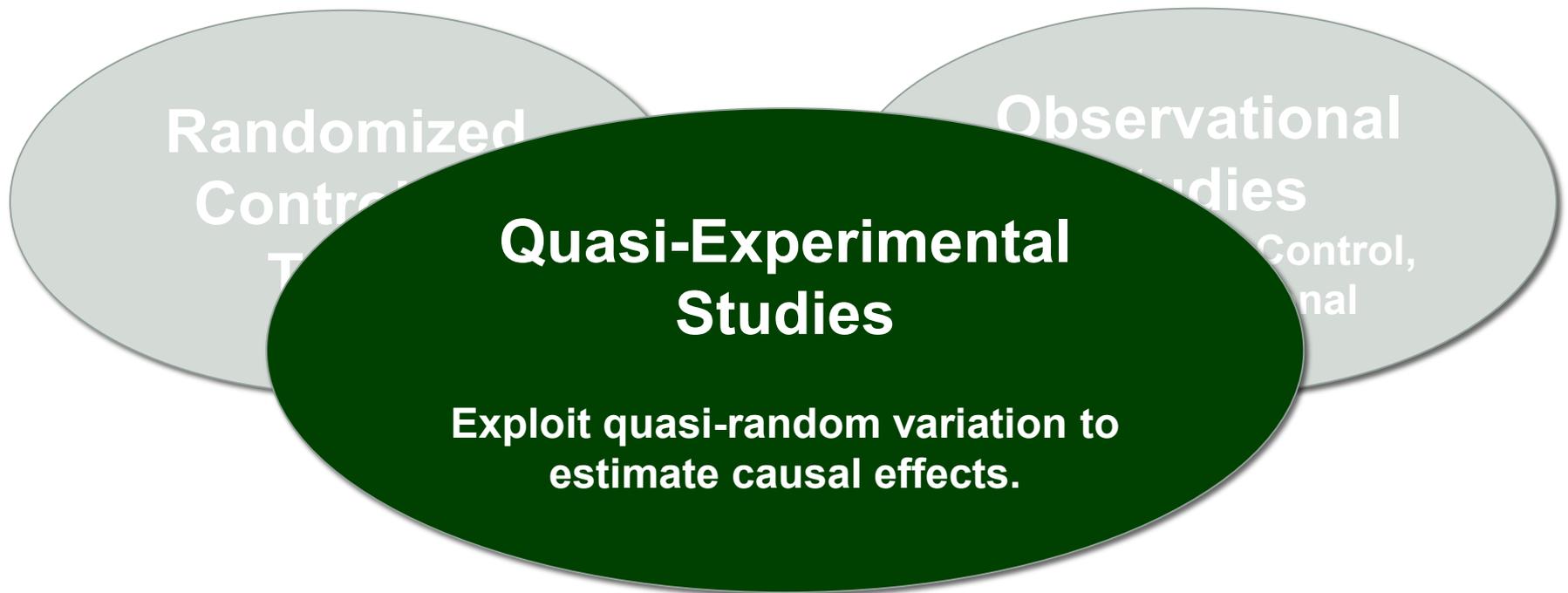
- Internal validity
- Balance on both observed *and unobserved* factors
- Cost / ethics constraints

Observational Studies

Cohort, Case-Control,
Cross-Sectional

- External validity
- Balance only on observables
- Strong assumption: no residual confounding.

Study designs in health research



- Balance on both observed *and* unobserved factors
- Observational data: fewer ethical, financial constraints
- Programs evaluated in “real life” not “controlled” setting

Agenda

- Overview of RDD
- Causal inference in RDD
- Estimating RDD treatment effects
- Examples of RDD in the health literature
- RDD with non-compliance

Agenda

- **Overview of RDD**
- Causal inference in RDD
- Estimating RDD treatment effects
- Examples of RDD in the health literature
- RDD with non-compliance

Regression discontinuity design (RDD)

- RDD can be implemented when treatment is assigned, in part, by a threshold rule on a continuous baseline variable
- Patients presenting just above vs. below 200 are similar on observed and unobserved characteristics...
...but assigned different exposures.

If CD4 < 200 or Stage IV,
initiate HIV therapy;
If CD4 ≥ 200 and no Stage IV,
return in 6 months

*SA National Treatment Guidelines for
Adults and Adolescents, 2004-2011*

RDD in clinical and public health research

ORIGINAL ARTICLE

OPEN

Regression Discontinuity Designs in Epidemiology *Causal Inference Without Randomized Trials*

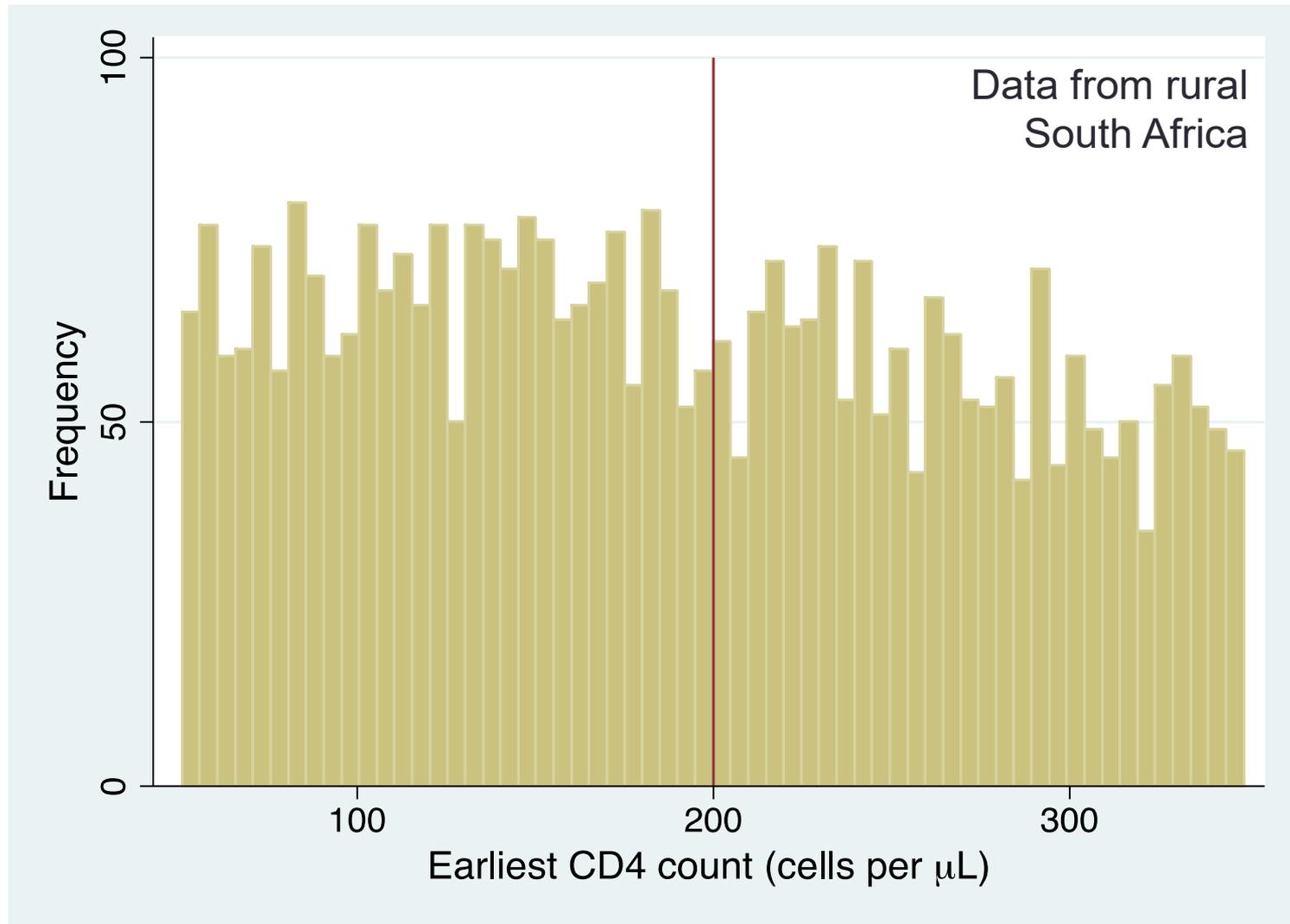
Jacob Bor,^{a,b,c} Ellen Moscoe,^c Portia Mutevedzi,^b Marie-Louise Newell,^{b,d} and Till Bärnighausen^{b,c}

Abstract: When patients receive an intervention based on whether they score below or above some threshold value on a continuously measured random variable, the intervention will be randomly assigned for patients close to the threshold. The regression discontinuity design exploits this fact to estimate causal treatment effects. In spite of its recent proliferation in economics, the regression discontinuity design has not been widely adopted in epidemiology. We describe regression discontinuity, its implementation, and the assumptions required for causal inference. We show that regression discontinuity is generalizable to the survival and nonlinear models that are mainstays of epidemiologic analysis. We then present an application of regression discontinuity to the much-debated epidemiologic question of when to start HIV patients on antiretroviral therapy. Using data from a large South African cohort (2007–2011), we estimate the causal effect of early versus deferred treatment eligibility on mortality. Patients whose first CD4 count was just below the 200 cells/ μ L CD4 count threshold had a 35% lower hazard of death (hazard ratio = 0.65 [95% confidence interval = 0.45–0.94]) than patients presenting with CD4 counts just above the threshold. We close by discussing the strengths and limitations of regression discontinuity designs for epidemiology.

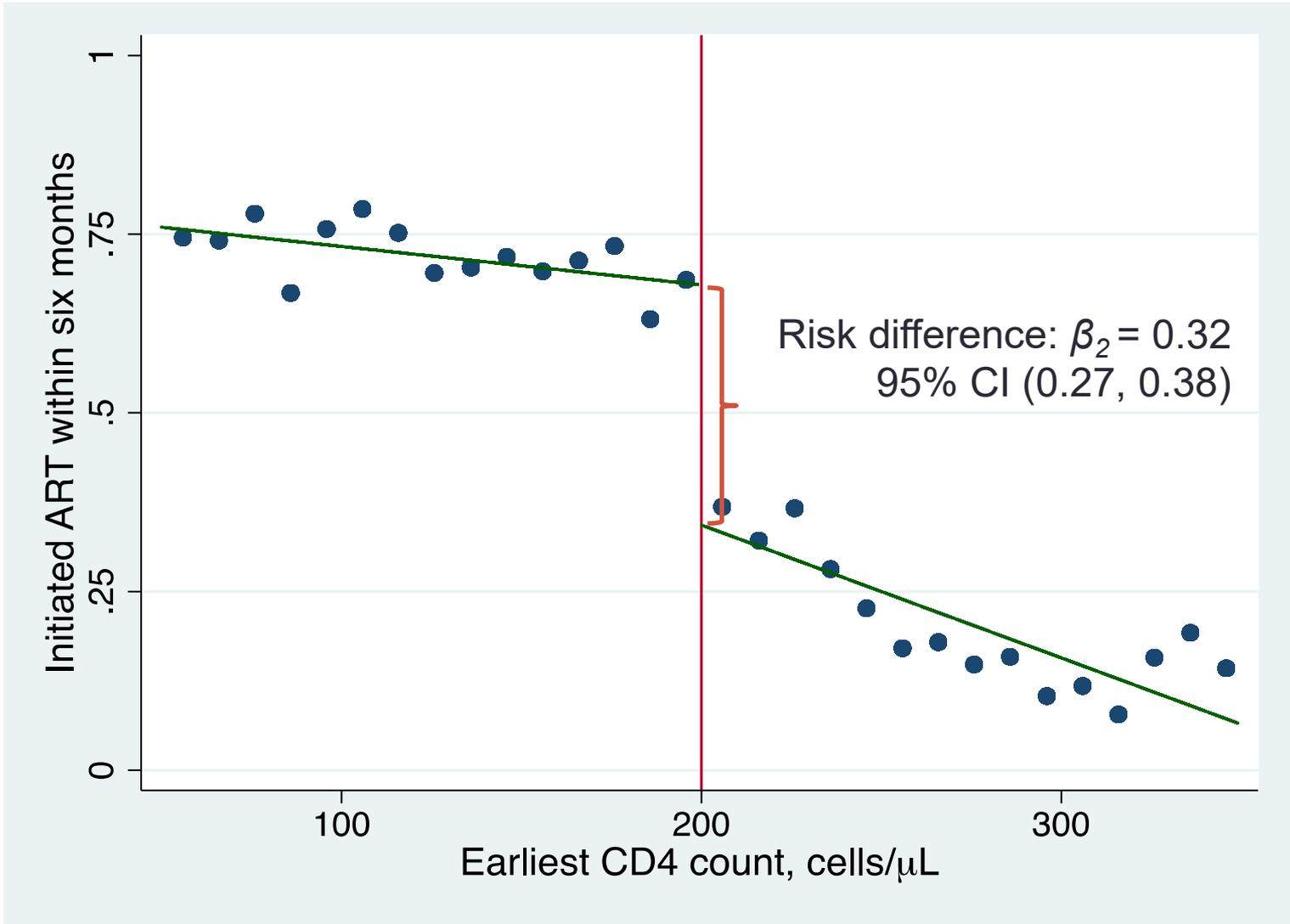
(*Epidemiology* 2014;25: 729–737)

- Primer for epidemiologists
- RDD with non-linear & survival models
- First application of RDD to a clinical threshold rule in epidemiology / clinical sciences: effect of immediate vs. deferred ART on survival

CD4 counts at first clinic visit



Eligibility affects treatment uptake



Source: Bor et al. 2014 *Epidemiology*

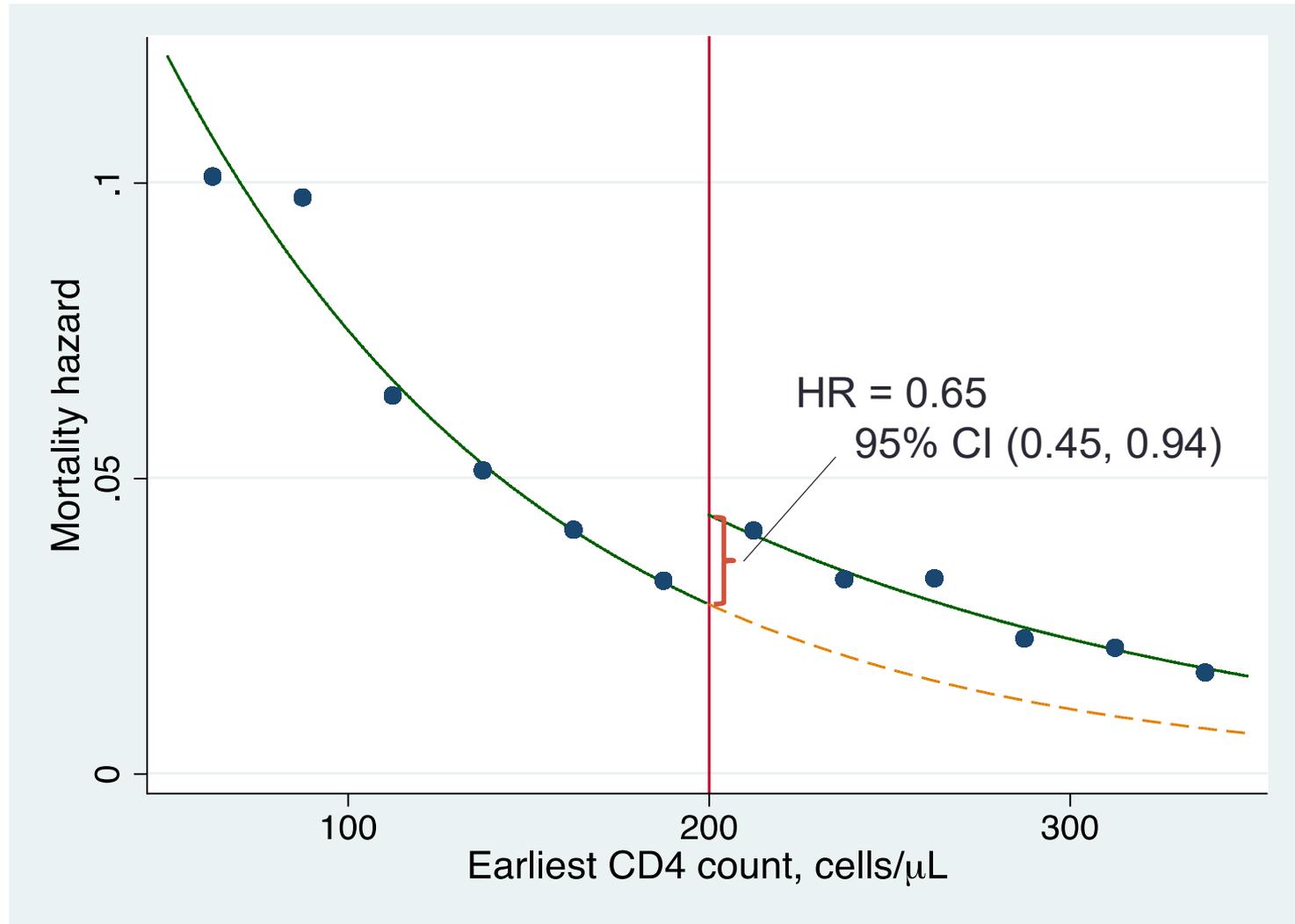
Regression discontinuity design (RDD)

- RDD can be implemented when treatment is assigned, in part, by a threshold rule on a continuous baseline variable
- Patients presenting just above vs. below 200 are similar on observed and unobserved characteristics...
...but assigned different exposures.
- Causal effect estimated as difference or ratio in predictions at threshold.
- In certain settings, no assumptions about “unmeasured confounding factors” are required

If CD4 < 200 or Stage IV,
initiate HIV therapy;
If CD4 ≥ 200 and no Stage IV,
return in 6 months

*SA National Treatment Guidelines for
Adults and Adolescents, 2004-2011*

Eligibility affects survival



Thistlethwaite & Campbell 1960

experimental comparison. In such studies the groups are presumed, as a result of matching, to have been equivalent prior to the exposure of the experimental group to some potentially change inducing event (the "experimental treatment"). If the groups differ on subsequent measures and if there are no plausible rival hypotheses which might account for the differences, it is inferred that the experi-

¹ This study is a part of the research program of the National Merit Scholarship Corporation. This research was supported by the National Science Foundation, the

much the same nature as would be produced by increasing magnitudes of that variable, examination of the details of the regression may be used to assess experimental effects. The experimental treatment should provide an additional elevation to the regression of dependent variables on the exposure determiner, providing a steplike discontinuity at the cutting score.

The argument—and the limitations on generality of the result—can be made more specific by considering a "true" experiment for which the regression-discontinuity analysis may be regarded as a substitute. It would

the same data; and, third, it qualifies interpretations of the ex post facto study recently reported in this journal (Thistlethwaite, 1959).

Two groups of near-winners in a national scholarship competition were matched on several background variables in the previous study in order to study the motivational effect of public recognition. The results suggested that such recognition tends to increase the favorableness of attitudes toward intellectualism, the number of students planning to seek the MD or PhD

Intervention: "Certificate of Merit" awarded to high school students competing for National Merit Scholarship who scored above a threshold on a standardized test. Outcomes: winning scholarships, career plans, intellectualism

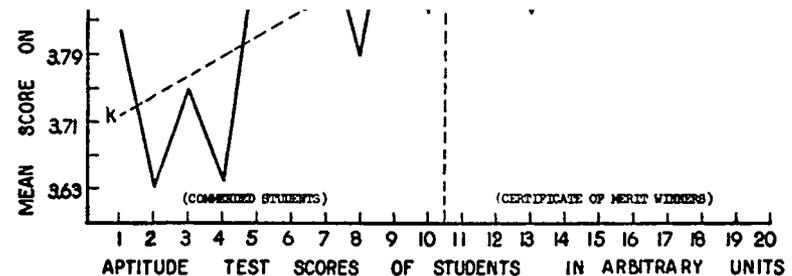


FIG. 4. Regression of attitudes toward intellectualism on exposure determiner.

actually in the wrong direction. On the other hand, it is confirming of the hypothesis of effect that all of the observed Points 11 through 20 lie above the extrapolated line of best fit for Points 1 to 10, in both II' and JJ'. But this could well be explained by the rival hypothesis of an uninterrupted curvilinear regression from Points 1 to

ful. A simple *t* test between Points and 11 is excluded, because it would show significance in an instance I DD' if the overall slope were great enough. That is, such a test ignores the general regression obtained independently of the experimental treatment. Such a test between adjacent points is likewise ruled out on the consideration

RDD in historical perspective

- Program evaluation
 - Thistlethwaite & Campbell 1960
 - Trochim 1984, 1990
 - Shadish, Cook, Campbell 2002
 - Cook, Shadish, Wong 2008
- Economics
 - Hahn, Todd & Van der Klaauw 2001
 - McCrary 2007
 - Imbens & Lemieux 2008
 - Lee 2008
 - Lee & Lemieux 2010

RDD in clinical and public health research



ELSEVIER



CrossMark

Journal of Clinical Epidemiology 68 (2015) 122e-133

Journal of
Clinical
Epidemiology

Systematic review
of empirical RDD
literature in health

Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice

Ellen Moscoe^{a,*}, Jacob Bor^{b,c}, Till Bärnighausen^{a,c}

^aDepartment of Global Health and Population, Harvard School of Public Health, 665 Huntington Avenue Building 1, room 1104, Boston, MA, USA

^bDepartment of Global Health, Boston University School of Public Health, 801 Massachusetts Ave, 3rd Floor, Boston, MA, USA

^cAfrica Centre for Health and Population Studies, University of KwaZulu-Natal, P.O. Box 198, Mtubatuba, 3935, South Africa

Accepted 1 June 2014

Just 32 empirical
RDD papers in
PubMed

Abstract

Objectives: Regression discontinuity (RD) designs allow for rigorous causal inference when patients receive a treatment based on scoring above or below a cutoff point on a continuously measured variable. We provide an introduction to the theory of RD and a systematic review and assessment of the RD literature in medicine, epidemiology, and public health.

Study Design and Setting: We review the necessary conditions for valid RD results, provide a practical guide to RD implementation, compare RD to other methodologies, and conduct a systematic review of the RD literature in PubMed.

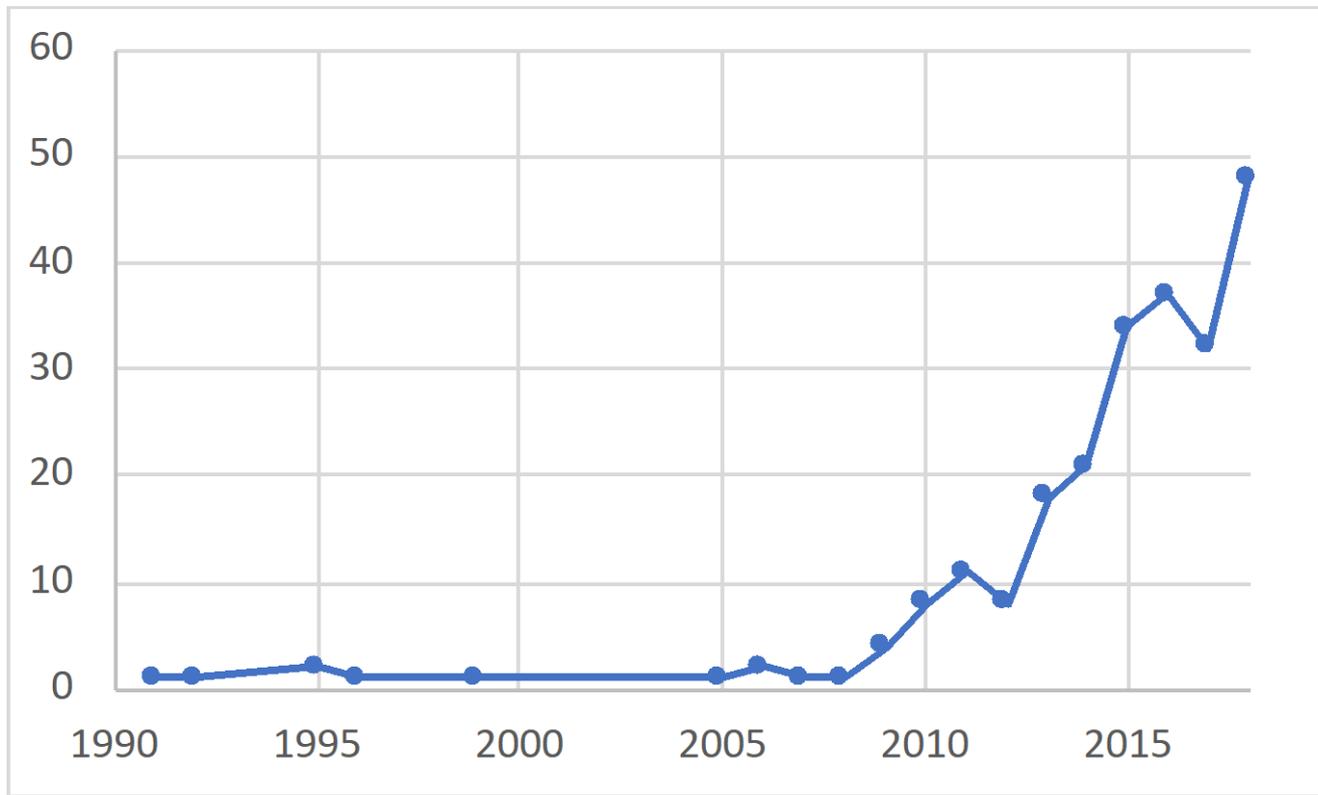
Results: We describe five key elements of analysis all RD studies should report, including tests of validity conditions and robustness checks. Thirty two empirical RD studies in PubMed met our selection criteria. Most of the 32 RD articles analyzed the effectiveness of social policies or mental health interventions, with only two evaluating clinical interventions to improve physical health. Seven out of the 32 studies reported on all the five key elements.

Conclusion: Increased use of RD provides an exciting opportunity for obtaining unbiased causal effect estimates when experiments are not feasible or when we want to evaluate programs under “real-life” conditions. Although treatment eligibility in medicine, epidemiology, and public health is commonly determined by threshold rules, use of RD in these fields has been very limited until now. © 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Just 2 studies
looking at clinical
threshold rules &
physical health:
Almond et al. 2009,
Bor et al. 2014

RDD in clinical and public health research

Increasing interest and use of RDD in health literature



PubMed search for “regression discontinuity” in Sept 2018 yielded 232 results, with 48 already this year.

Agenda

- Overview of RDD
- **Causal inference in RDD**
- Estimating RDD treatment effects
- Examples of RDD in the health literature
- RDD with non-compliance

Causal inference in RD designs

Set-up

- Potential outcomes: $Y_i(1)$ if treatment-eligible, $Y_i(0)$ if not eligible
- Continuous assignment variable Z_i , treatment-eligible if $Z_i < c$
- Potential outcome conditional expectation functions (**POCEF**s):
what are the mean potential outcomes for different values of Z ?
 $E(Y_i(1)|Z_i)$ and $E(Y_i(0)|Z_i)$

Identification in RD designs

$$E[Y_i(0)|Z_i = z]$$

Solid lines are the
observed data:

$$E[Y_i | Z_i = z]$$



$$E[Y_i(1)|Z_i = z]$$

Causal inference in RD designs

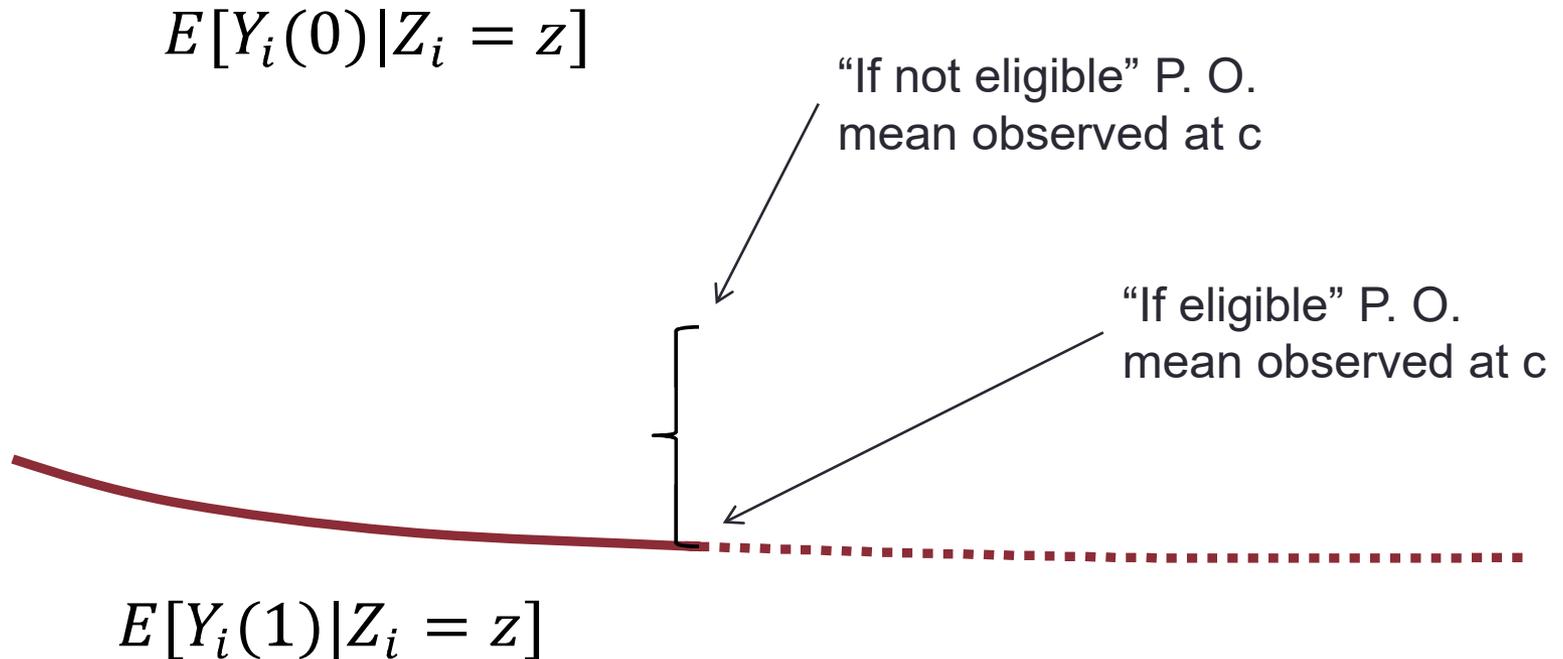
Set-up

- Potential outcomes: $Y_i(1)$ if treatment-eligible, $Y_i(0)$ if not eligible
- Continuous assignment variable Z_i , treatment-eligible if $Z_i < c$
- Potential outcome conditional expectation functions (**POCEF**s):
what are the mean potential outcomes for different values of Z ?
 $E(Y_i(1)|Z_i)$ and $E(Y_i(0)|Z_i)$

Identification

- In the limit as Z approaches c from below,
$$E(Y_i|Z_i = z) = E(Y_i(1)|Z_i=c)$$
- In the limit as Z approaches c from above,
$$E(Y_i|Z_i = z) = E(Y_i(0)|Z_i=c)$$

Identification in RD designs



Conditions required for identification

1. Threshold rule exists and c is known
2. Z is continuous near c
3. Key assumption: continuity in $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at c

Conditions required for identification

1. Threshold rule exists and c is known
2. Z is continuous near c
3. Key assumption: continuity in $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at c

Is condition 3 a strong assumption?

Conditions required for identification

1. Threshold rule exists and c is known
2. Z is continuous near c
3. Key assumption: continuity in $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at c

Is condition 3 a strong assumption?

- Case 1: geographical boundary determines cigarette tax

Conditions required for identification

1. Threshold rule exists and c is known
2. Z is continuous near c
3. Key assumption: continuity in $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at c

Is condition 3 a strong assumption?

- Case 1: geographical boundary determines cigarette tax
- Case 2: laboratory measure on clinical biomarker determines Rx

Conditions required for identification

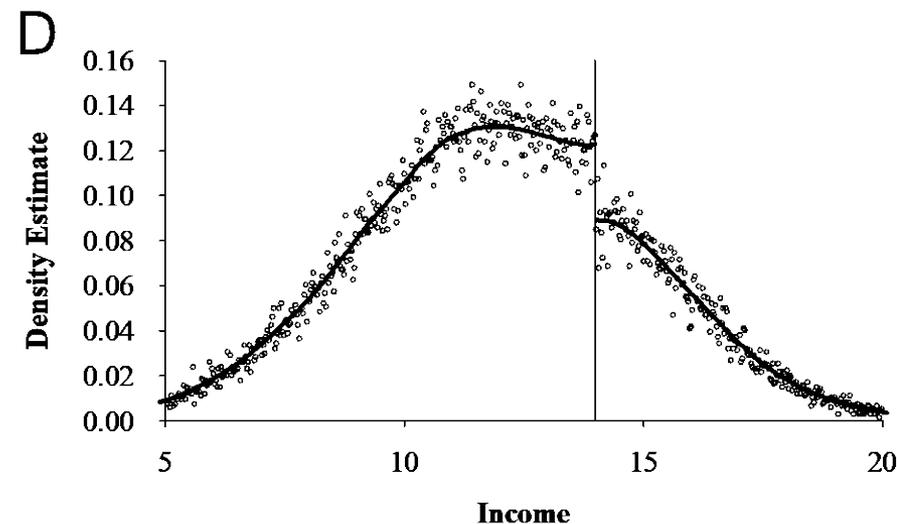
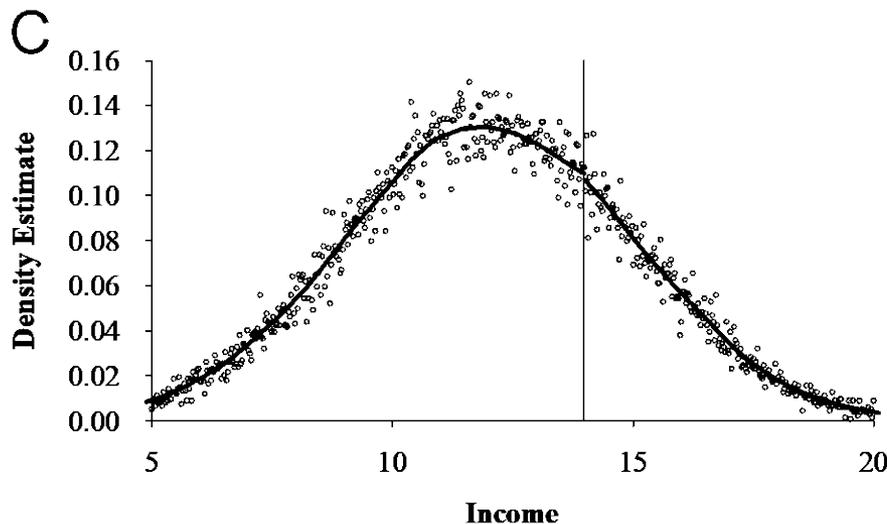
1. Threshold rule exists and c is known
2. Z is continuous near c
3. Key assumption: continuity in $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at c

Identification off of measurement error (or other random noise)

- Random noise in measured Z guarantees continuity in potential outcomes, so long as no direct manipulation (Lee 2008)
 - Suppose $CD4_{i,OBSERVED} = CD4_{i,TRUE} + e_i$, e_i random noise
 - Observations with $CD4_{i,TRUE} = 200$ randomly assigned to be $</> 200$
- Manipulation can be assessed in data (McCrary 2007)
- Support for continuity in POCEFs from baseline observables

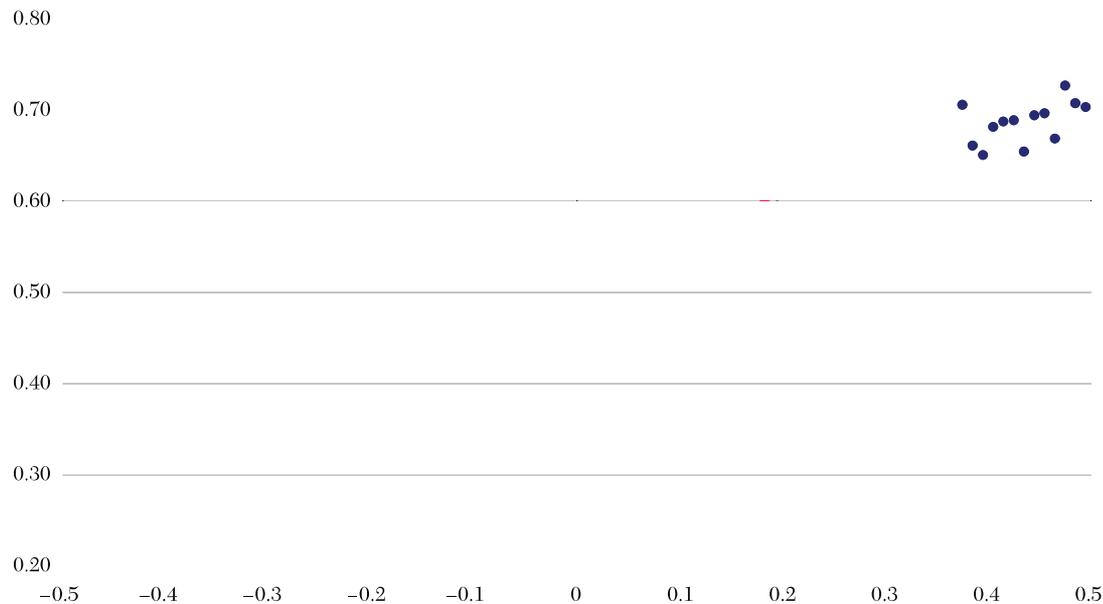
Tests for validity of identifying assumption

- Test 1: The presence of systematic manipulation of the assignment variable can be tested.
 - If patients (or providers) change their values of Z to gain (or avoid) access to treatment, this will result in bunching on one side of the threshold and a discontinuity in $f(Z)$ at c .
 - Test for continuity in density of assignment variable (McCrary 2008)



Tests for validity of identifying assumption

- Test 2: Continuity in baseline observables can be assessed
 - Just like assessment of balance in RCT
 - Systematic imbalance would suggest that treatment assignment was in fact non-random.



Lee & Lemieux 2010

Figure 17. Discontinuity in Baseline Covariate (Share of Vote in Prior Election)

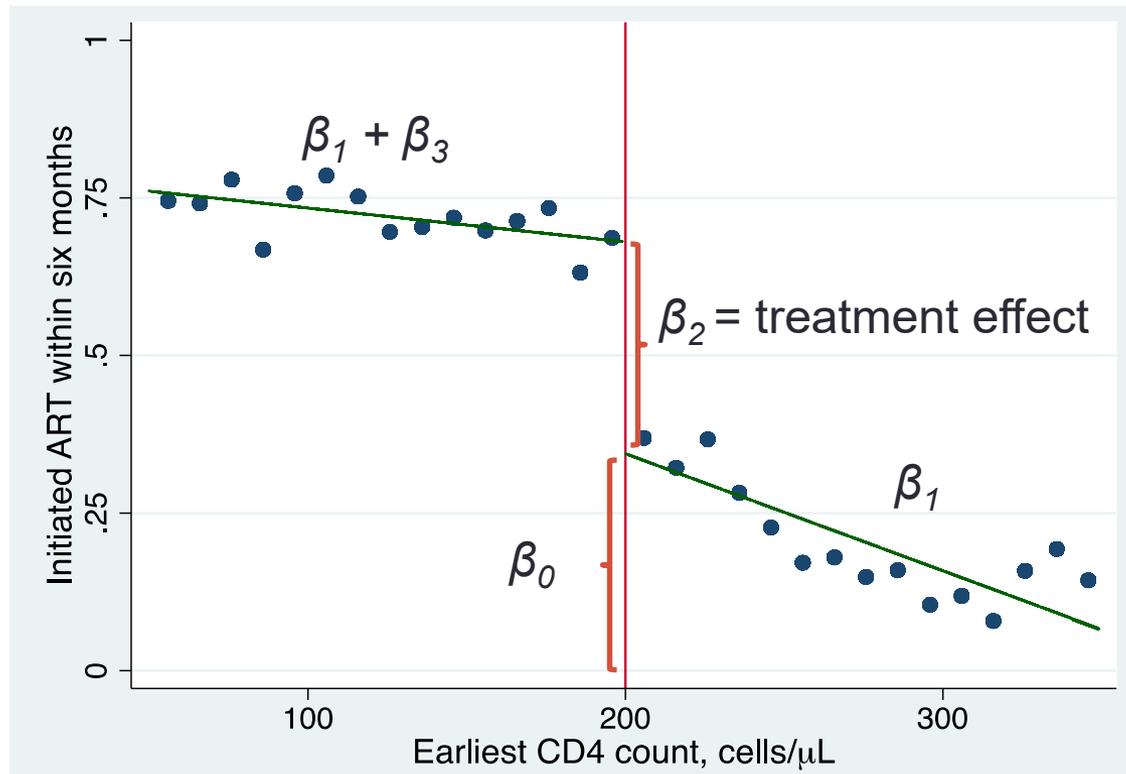
Agenda

- Overview of RDD
- Causal inference in RDD
- **Estimating RDD treatment effects**
- Examples of RDD in the health literature
- RDD with non-compliance

Estimation

- Goal: estimate $E[Y|Z]$ around c , predict $E[Y|Z \uparrow c]$ & $E[Y|Z \downarrow c]$

$$E[Y_i | Z_i] = \beta_0 + \beta_1(Z_i - c) + \beta_2 1[Z_i < c] + \beta_3(Z_i - c) * 1[Z_i < c]$$



Estimation

- Goal: estimate $E[Y|Z]$ around c , predict $E[Y|Z \uparrow c]$ & $E[Y|Z \downarrow c]$

$$E[Y_i | Z_i] = \beta_0 + \beta_1(Z_i - c) + \beta_2 1[Z_i < c] + \beta_3(Z_i - c) * 1[Z_i < c]$$

- Local linear regression
 - Modeling slopes reduces bias at boundary (Fan & Gijbels 1996)
 - Consistent (Hahn, Todd, van der Klaauw 2001)
 - Data-driven bandwidth selection (Imbens & Kalyanaraman 2012; Calonico, Cattaneo, Titiunik, 2015); show lots of bw's
 - Better than higher order polynomials (Gelman & Imbens 2014)

How should we interpret RDD effect estimates?

- RDD effect = causal effect “at the threshold”
- RDD identifies same causal effect as RCT if:
 - Constant treatment effects (common assumption in epidemiology)
 - Effects heterogeneous, but independent of Z (e.g. random number)
- If TE heterogeneous, *local* average causal effect
 - Weighted average across true Z^* with $Z=c$
 - Local effect often of policy interest: should we change the threshold?

Agenda

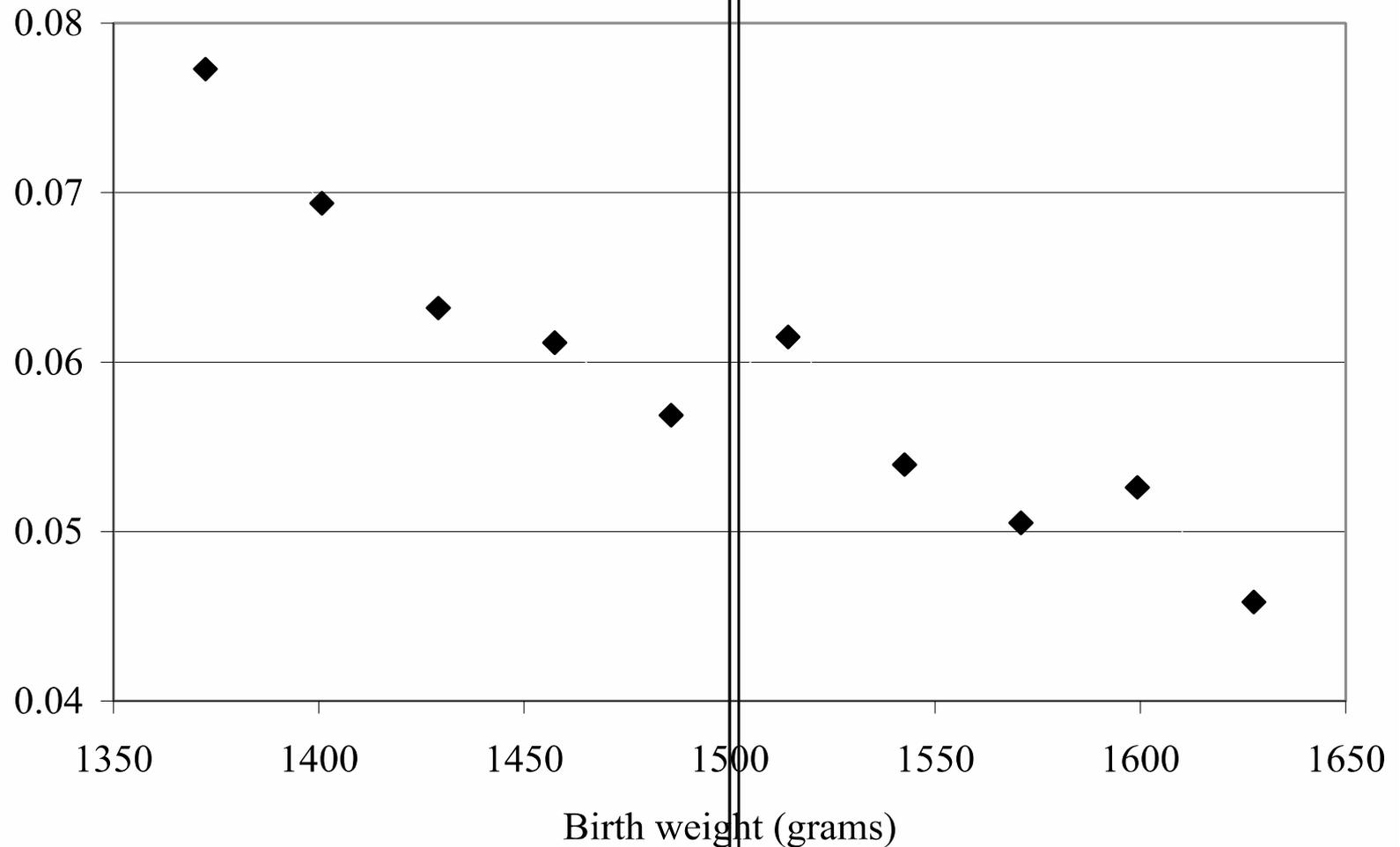
- Overview of RDD
- Causal inference in RDD
- Estimating RDD treatment effects
- **Examples of RDD in public health**
- RDD with non-compliance

RDD and health - examples

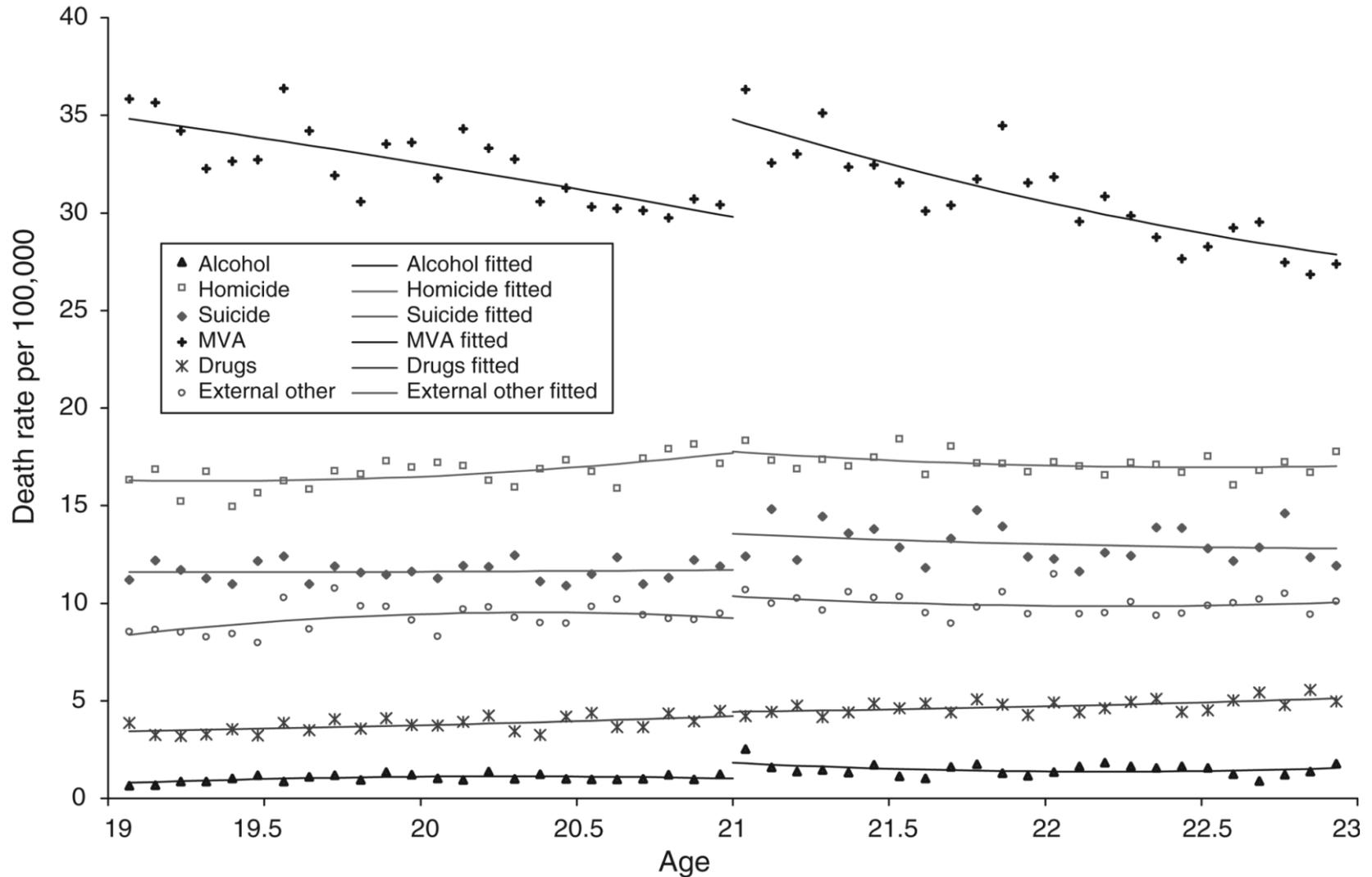
1. [Almond](#) D, Doyle JJ, Kowalski AE, Williams H. Estimating marginal returns to medical care: evidence from at-risk newborns. *Q J Econ.* 2010;125(2):591-634.
2. [Carpenter](#) C, Dobkin C. The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *Am Econ J Appl Econ.* 2009;1(1):164-182.
3. [Chen](#) Y, Ebenstein A, Greenstone M, Li H. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *PNAS.* 2013, doi/10.1073/pnas.1300018110.
4. [Ludwig](#) J, Miller DL. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Q J Econ.* 2007; 122(1):159-208.
5. [Zhao](#) M, Konishi Y, Glewwe P. Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption. *J Health Econ.* 2013;32(2):367-85.
6. [Anderson](#) S. Legal Origins and Female HIV. *Am. Econ. Review.* 2018;108(6):1407.
7. [Chen](#) H, Li Q, Kaufman JS, Wang J, Copes R, Su Y, Benmarhnia T. Effect of air quality alerts on human health: a regression discontinuity analysis in Toronto, Canada. *The Lancet Planetary Health.* 2018 Jan 31;2(1):e19-26.

Almond et al. (2010), Low Birth Weight

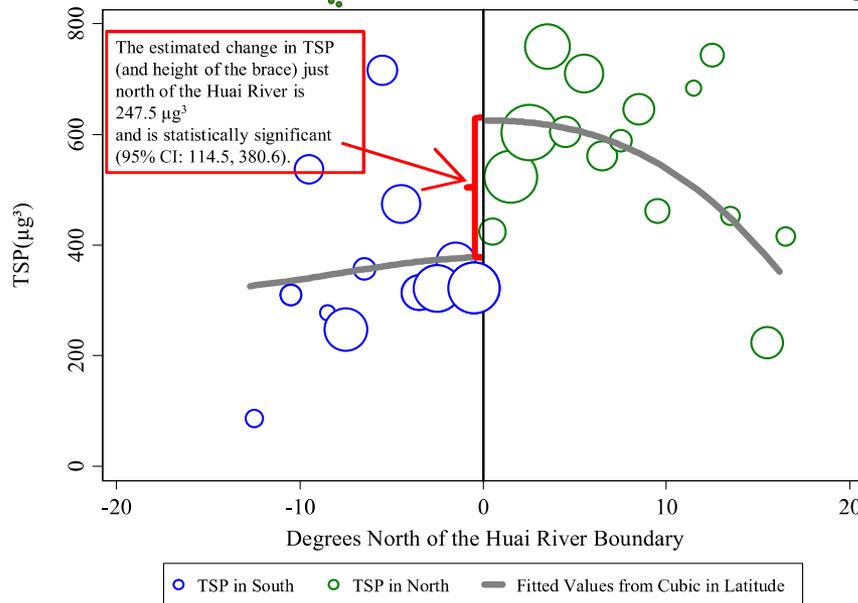
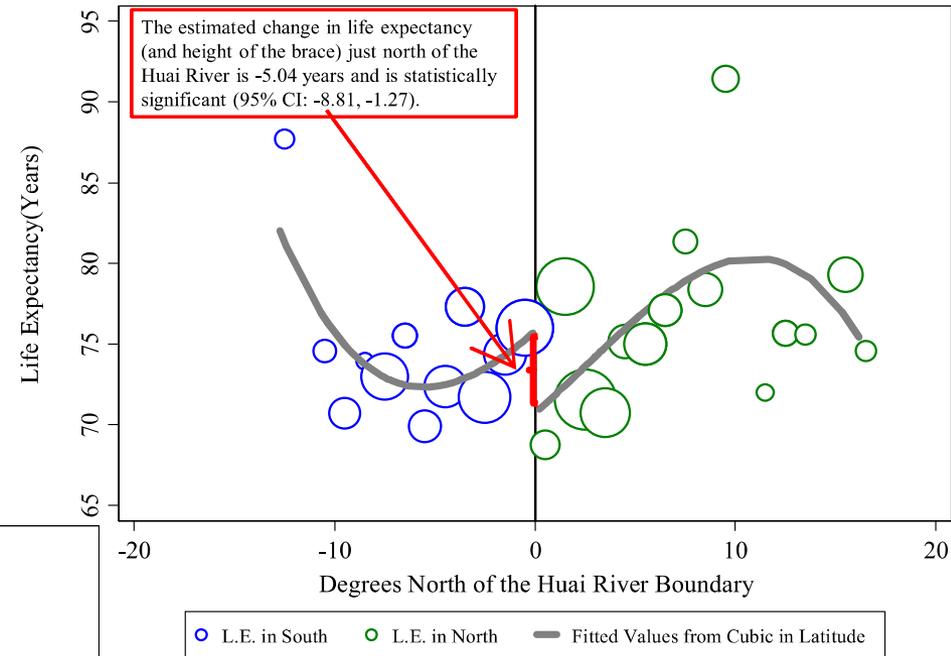
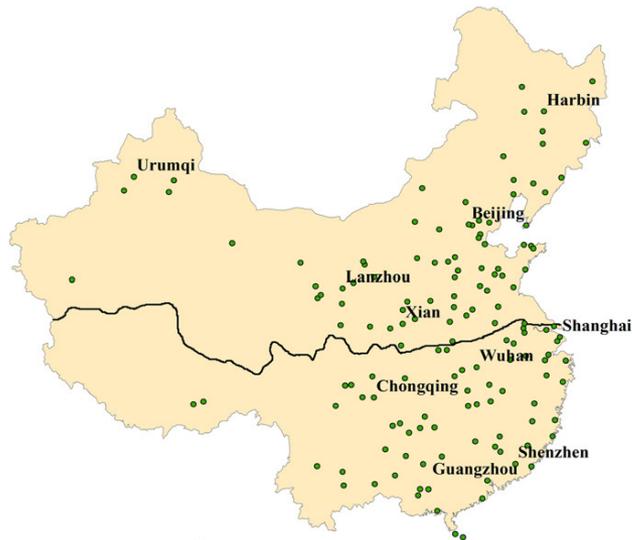
A: One-year mortality



Carpenter & Dobkin (2009), Drinking Age

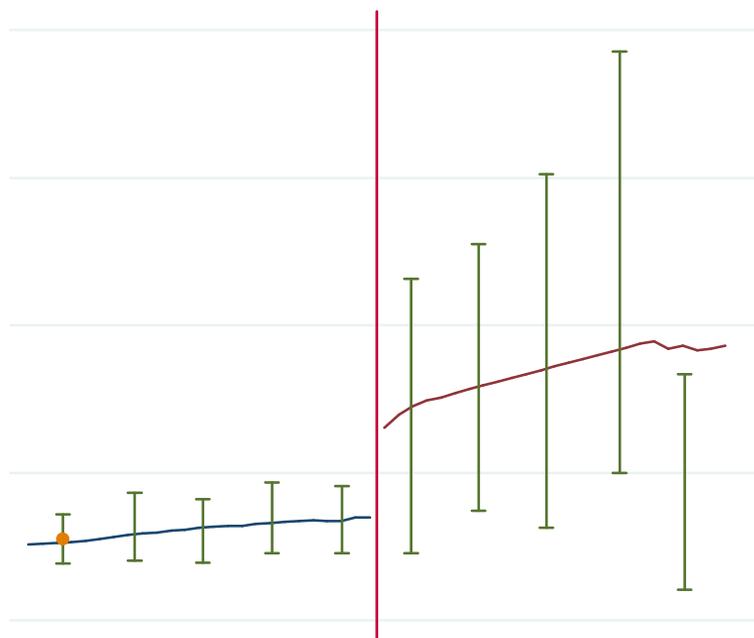


Chen et al. (2013), Huai River Policy



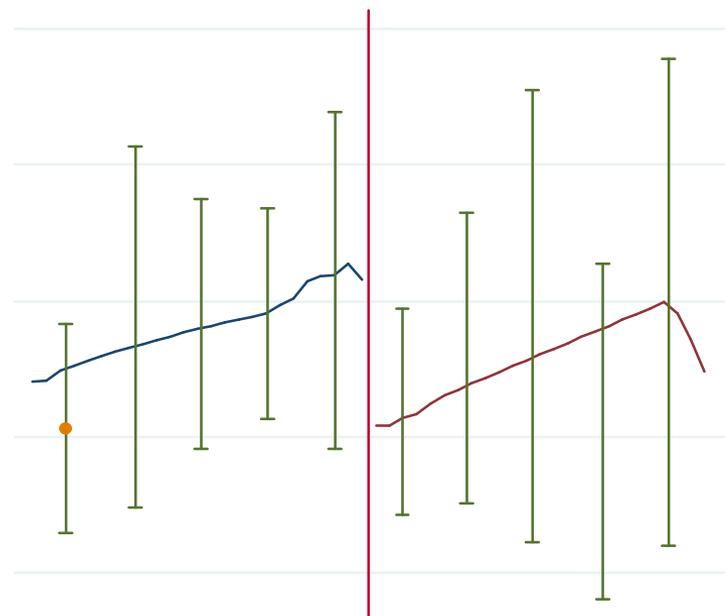
Ludwig & Miller (2007), Head Start

Panel A: 1968 Head Start funding per 4 year old



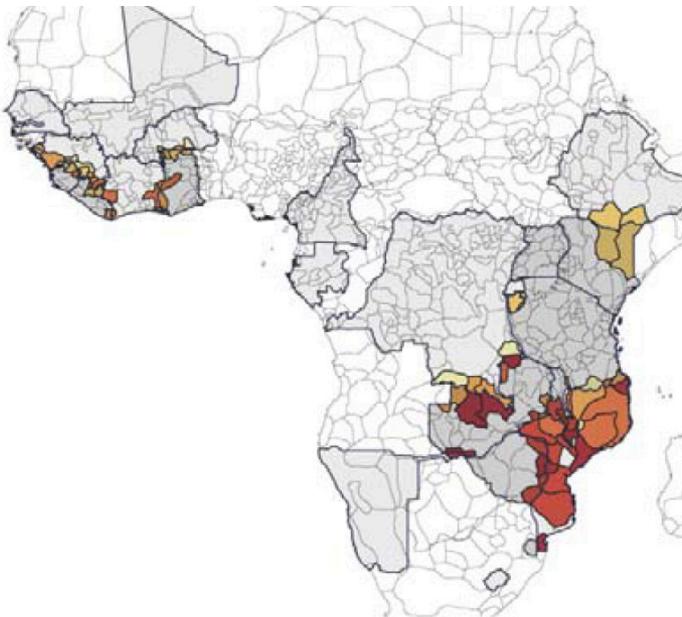
Panel A: Children 5-9

Causes susceptible to Head Start

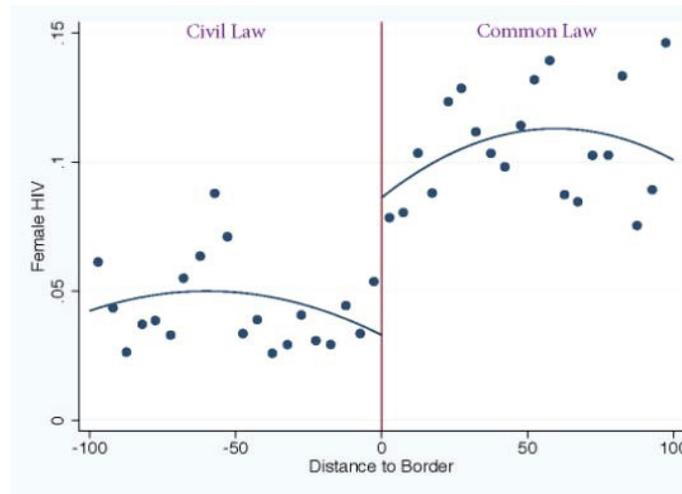


Anderson (2018), women's property rights and HIV risk

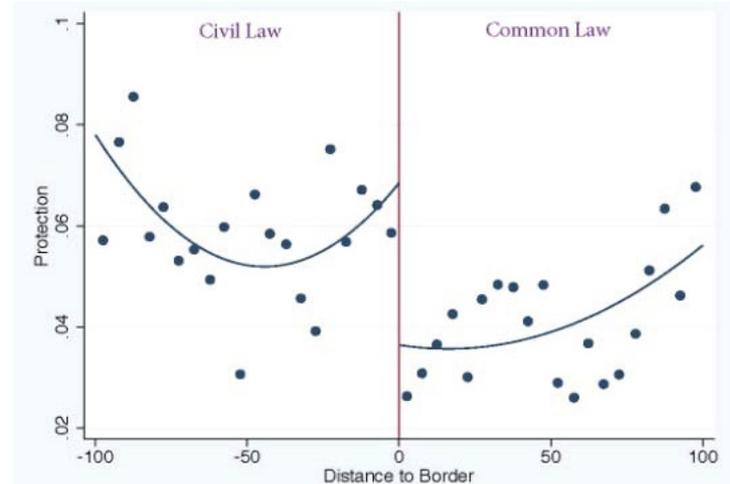
Ethnic groups that split national boundaries with common v. civil law legal origins



Female HIV prevalence

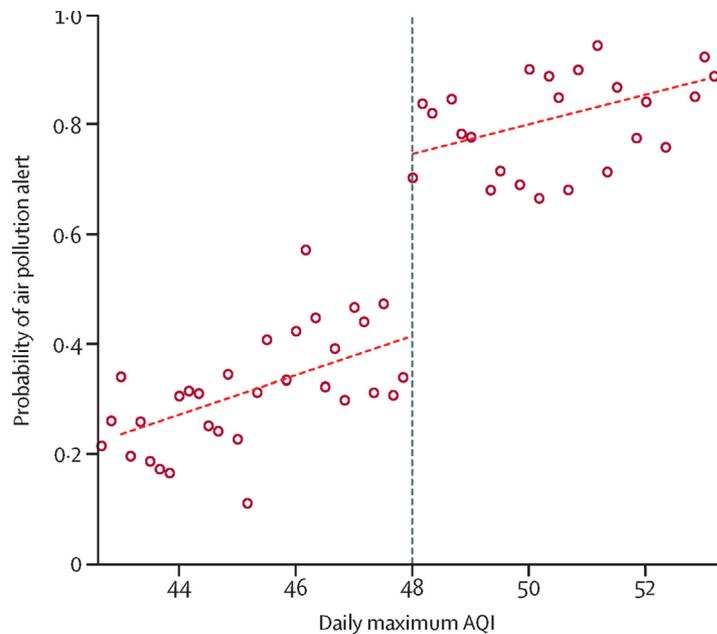


Condom use

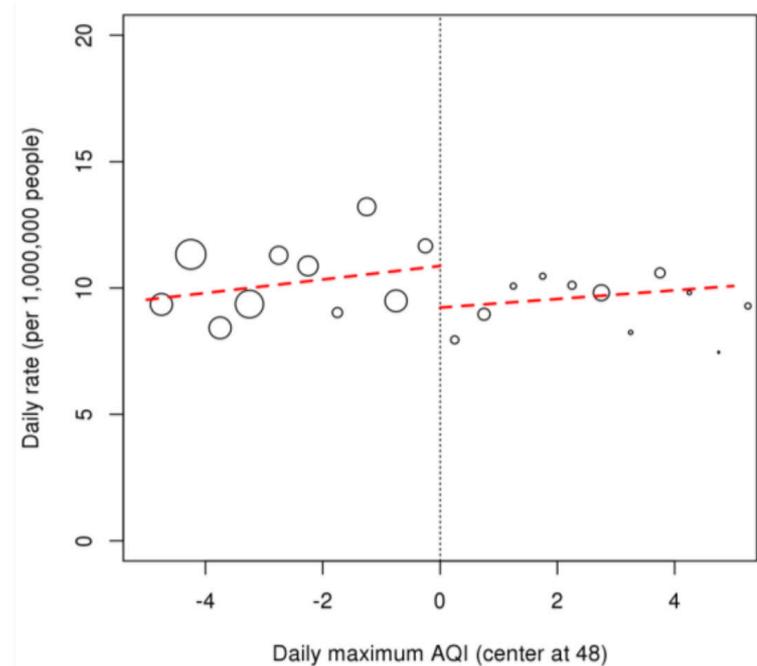


Chen (2018), air pollution alerts and ER visits

Air pollution alerts

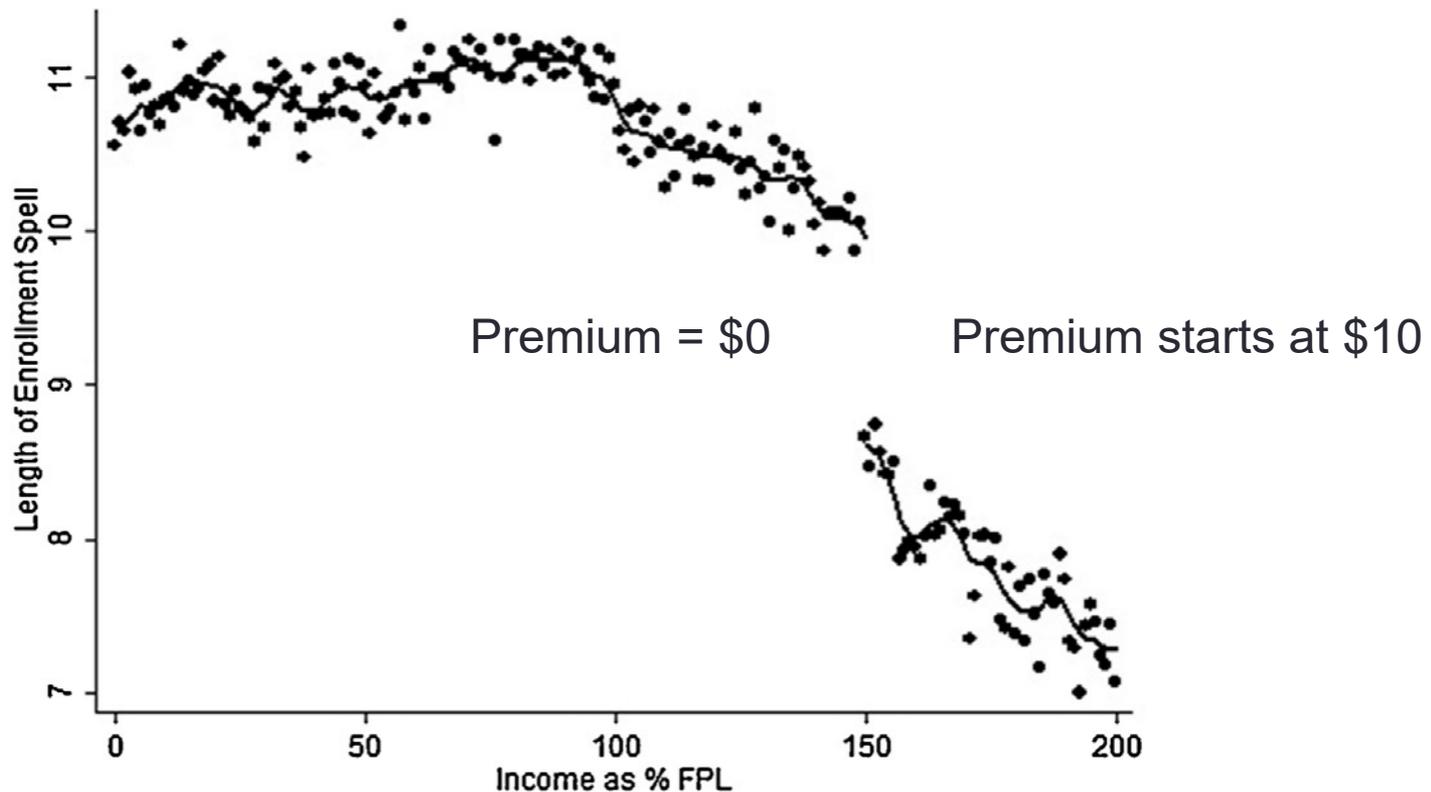


Asthma ER visits



Dague (2017), Medicaid premiums and rates of enrollment

Adults are more likely to stay enrolled if they don't have to pay a premium



Agenda

- Overview of RDD
- Causal inference in RDD
- Estimating RDD treatment effects
- Examples of RDD in the health literature
- **RDD with non-compliance**

RDD with non-compliance

- What if the threshold rule only applies to some patients?
 - What if there are other indications or contra-indications for treatment?
 - What if some patients opt out despite being eligible? Or vice-versa
 - Similar to a clinical trial with non-compliance
 - Very common. Known as “fuzzy RDD”

Example: HIV treatment eligibility and retention in care



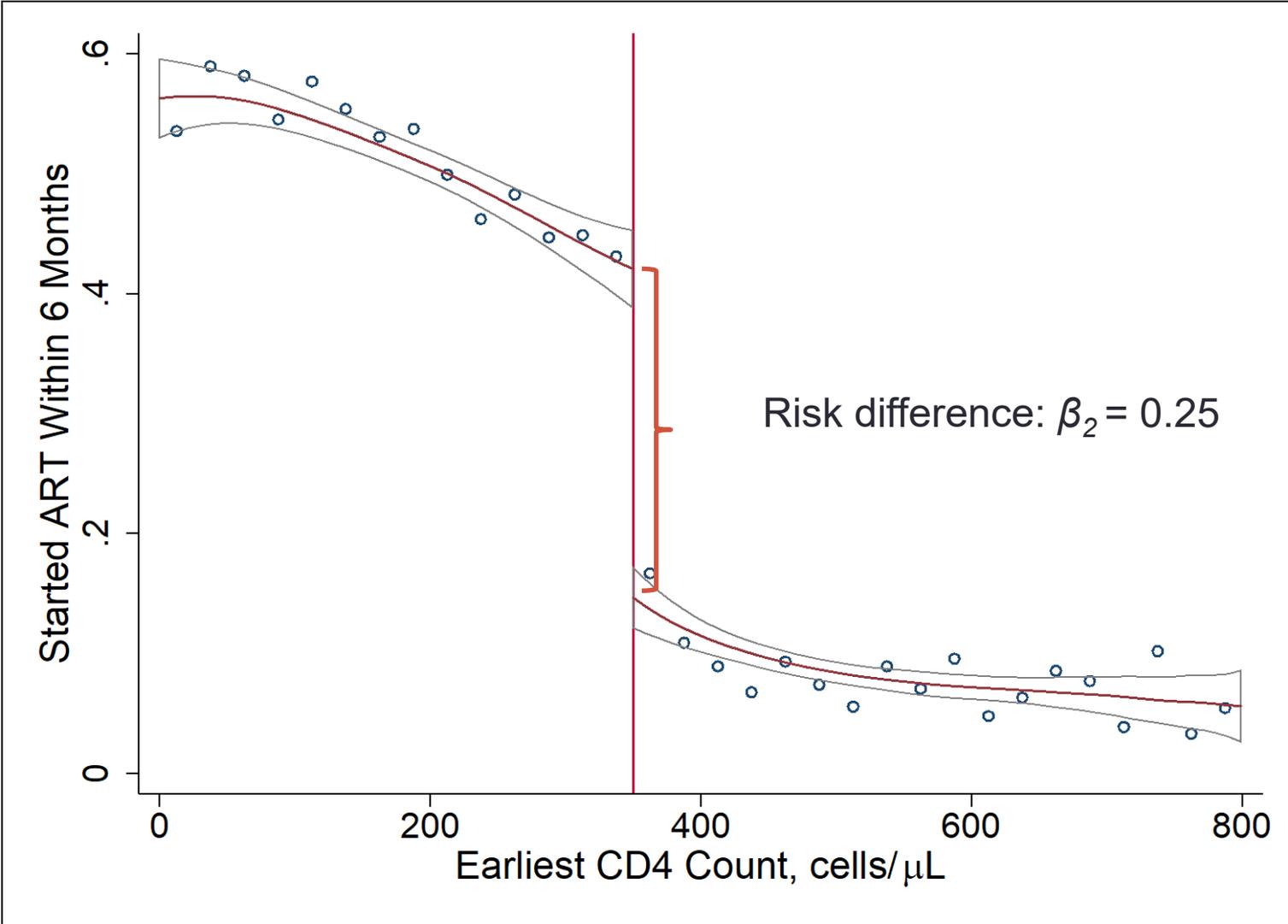
RESEARCH ARTICLE

Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa

Jacob Bor^{1,2,3,4*}, Matthew P. Fox^{1,2,4}, Sydney Rosen^{1,4}, Atheendar Venkataramani⁵, Frank Tanser^{3,6,7,8}, Deenan Pillay^{3,9}, Till Bärnighausen^{3,8,10,11}

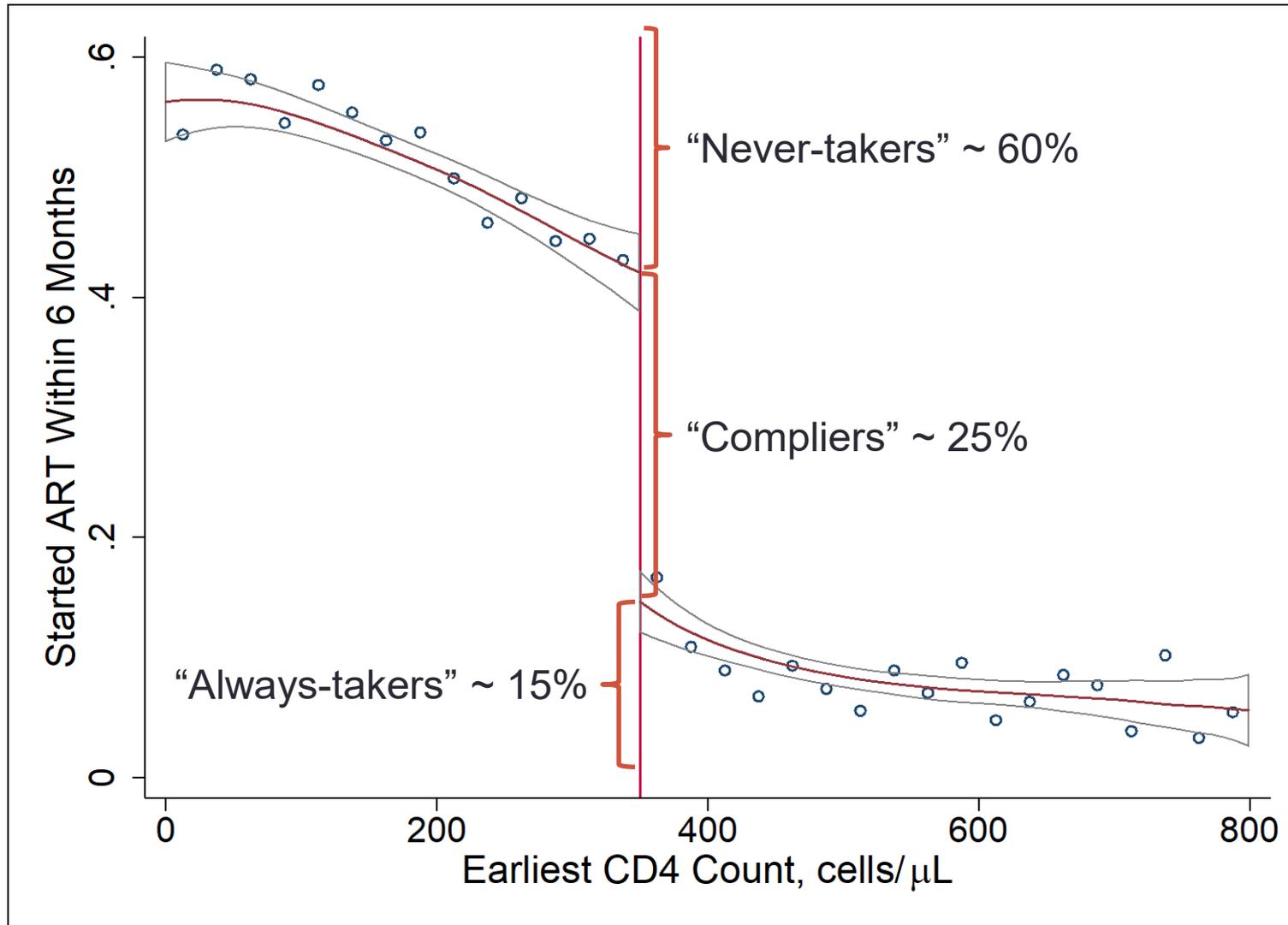
Citation: Bor J, Fox MP, Rosen S, Venkataramani A, Tanser F, Pillay D, et al. (2017) Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. PLoS Med 14(11): e1002463. <https://doi.org/10.1371/journal.pmed.1002463>

Eligibility affects treatment uptake



Source: Bor et al. PLOS Medicine 2017

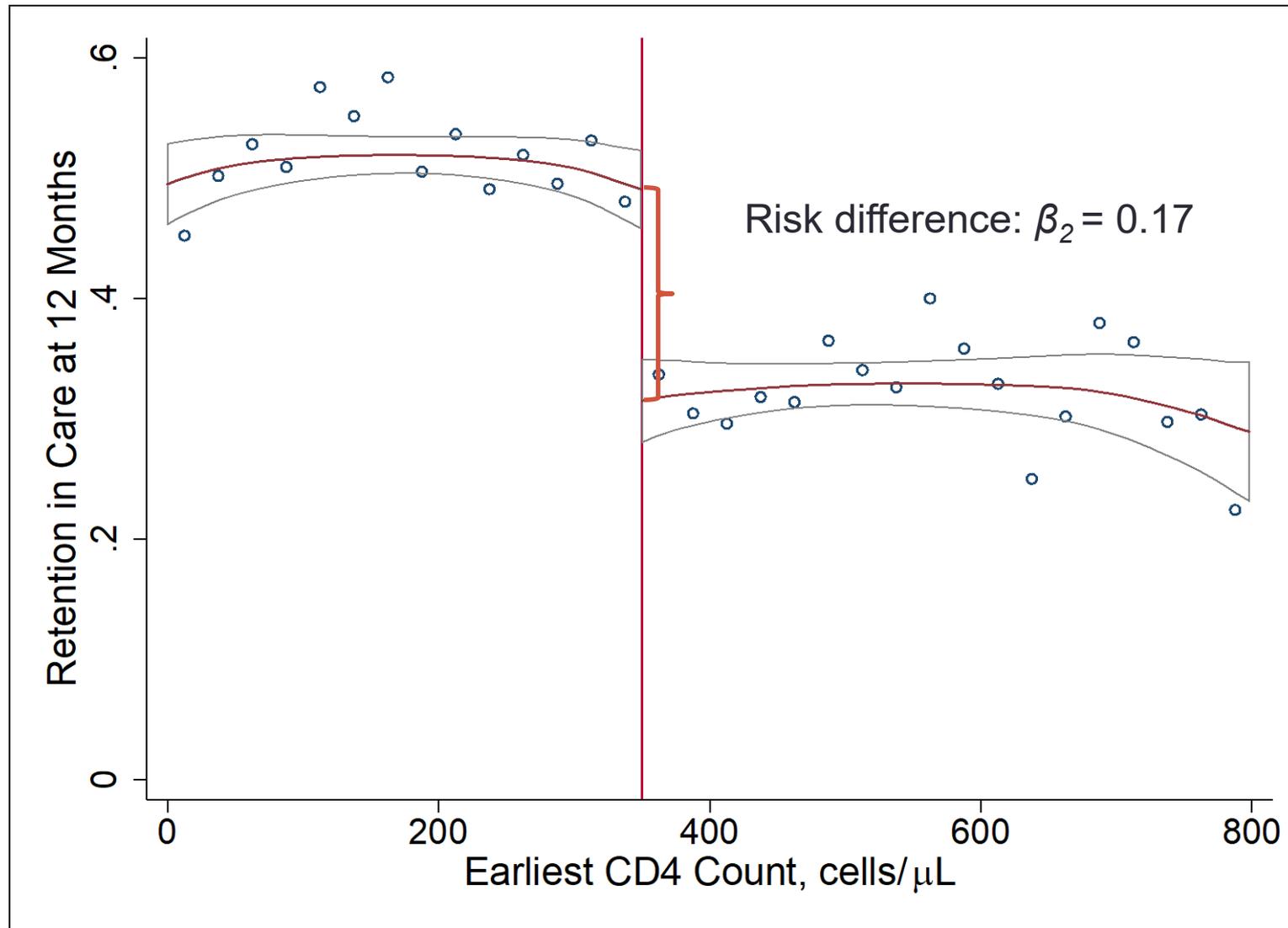
Change in treatment at the threshold is not 0 to 100%



RDD with non-compliance

- What if the threshold rule only applies to some patients?
 - What if there are other indications or contra-indications for treatment?
 - What if some patients opt out despite being eligible? Or vice-versa
 - Similar to a clinical trial with non-compliance
 - Very common. Known as “fuzzy RDD”
- Interpretation
 - Effect of being below threshold has an “intention to treat” interpretation
 - Effect of treatment itself can be recovered using the threshold rule as an instrumental variable, i.e. by scaling the ITT effect by the share of patients whose treatment status was determined by the threshold rule, so-called “compliers” (Angrist & Imbens 1994).

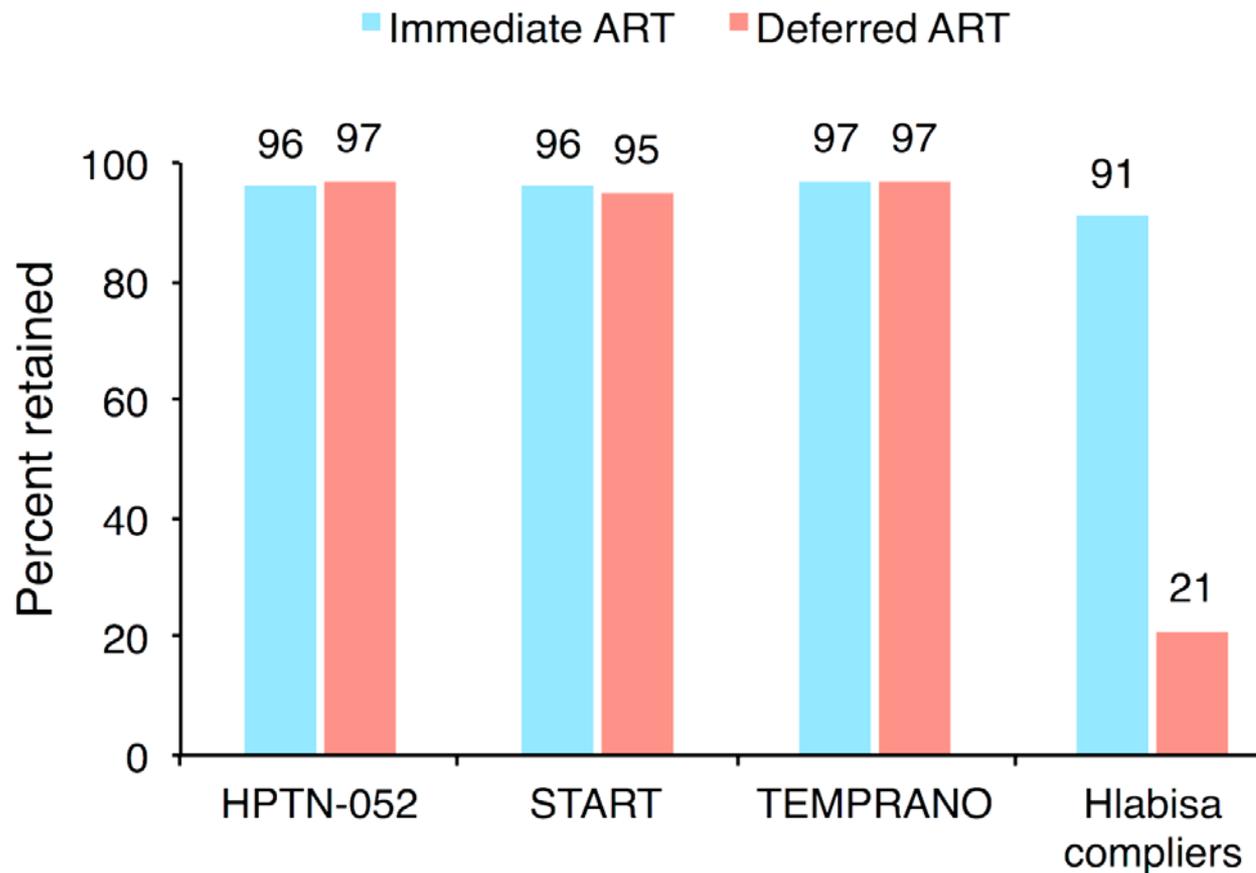
ITT effect of eligibility on retention in care



What was the effect on compliers?

- Compliers: patients whose treatment decision was based on the eligibility threshold
- Complier Average Causal Effect (CACE)
$$\text{CACE} = \text{ITT}/\text{FS} = .17/.25 = .70$$
- Interpretation: HIV treatment eligibility increased 12-month retention by **70 percentage points** among so-called “compliers”, i.e. those patients whose treatment decision was based on the CD4<350 threshold.
- Under excludability and monotonicity assumptions.
- Among “compliers”, immediate eligibility increased retention from 21% to 91%.

Effect on retention completely missed in RCTs



Benefits of immediate ART eligibility are due to behavior, not just biology

Recap

- RDD offers rigorous approach to causal inference when an exposure is assigned by a threshold rule
 - Second only to RCT; “local randomization”
- Increasing use of RDD in public health and medicine
 - Clinical thresholds = classic case; other applications too
- RDDs sometimes have benefits over RCTs
 - Lower cost, evaluations of difficult-to-randomize interventions
 - Population representative data; no opt-in consent
 - Real world settings (e.g. control receives true standard of care)
- Key limitation: RDD is not always available
 - But threshold rules are more common than you would think

Thank you.

Jacob Bor

jbor@bu.edu

Appendix

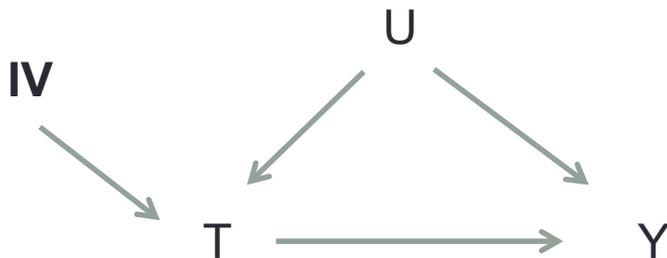
- Alternate approaches to causal inference
- Details on fuzzy RDD and instrumental variables

Alternate approaches to causal inference

1. Global ACE is identified under the assumption that functional forms of POCEFs are known across full range (Rubin 1977).
 - Linearity under joint normality (Vandenbrouke & Le Cessie, 2014)
 - Strong, untestable assumption.
2. Local ACE identified under much weaker assumption that POCEFs are continuous
 - Unconfoundedness assumption but a weak one.
3. If Z is a random variable and patients cannot precisely manipulate Z , local ACE identified without assumptions.
 - “Local randomization” interpretation

Bor J, Moscoe E, Bärnighausen T. (2015). Three approaches to causal inference in regression discontinuity designs (Letter). *Epidemiology*.

An aside: instrumental variables 101



Directed Acyclic Graph (DAG)

Simple case: binary IV, binary T

- RCT with non-compliance
- Threshold rule in fuzzy RDD

LATE IV Assumptions

1. First stage: IV causally affects T
2. Exchangeability: IV as good as randomly assigned
3. Excludability: IV only affects Y through T
4. Monotonicity: no “defiers”

Then: LATE (CACE) Identified

Imbens & Angrist (1994)

*Note: economists say “local average treatment effect” (LATE); epidemiologists say “complier average causal effect” (CACE); identical concept

LATE (CACE) Theorem in RDD

1. $ITT = E[Y|Z\uparrow c] - E[Y|Z\downarrow c]$

2. Consider four latent types.

$$ITT = \{E[Y|Z\uparrow c, AT] - E[Y|Z\downarrow c, AT]\} * Pr(AT) \\ + \{E[Y|Z\uparrow c, NT] - E[Y|Z\downarrow c, NT]\} * Pr(NT) \\ + \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C) \\ + \{E[Y|Z\uparrow c, Def] - E[Y|Z\downarrow c, Def]\} * Pr(Def)$$

Latent Type	Potential Treatment Status, $Z\uparrow c$	Potential Treatment Status, $Z\downarrow c$
Always-taker	$T = 1$	$T = 1$
Never-taker	$T = 0$	$T = 0$
Complier	$T = 1$	$T = 0$
Defier	$T = 0$	$T = 1$

LATE (CACE) Theorem in RDD

1. $ITT = E[Y|Z\uparrow c] - E[Y|Z\downarrow c]$

2. Consider four latent types.

$$ITT = \{E[Y|Z\uparrow c, AT] - E[Y|Z\downarrow c, AT]\} * Pr(AT) \\ + \{E[Y|Z\uparrow c, NT] - E[Y|Z\downarrow c, NT]\} * Pr(NT) \\ + \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C) \\ + \{E[Y|Z\uparrow c, Def] - E[Y|Z\downarrow c, Def]\} * Pr(Def)$$

Latent Type	Potential Treatment Status, $Z\uparrow c$	Potential Treatment Status, $Z\downarrow c$
Always-taker	$T = 1$	$T = 1$
Never-taker	$T = 0$	$T = 0$
Complier	$T = 1$	$T = 0$
Defier	$T = 0$	$T = 1$

LATE (CACE) Theorem in RDD

1. $ITT = E[Y|Z\uparrow c] - E[Y|Z\downarrow c]$

2. Consider four latent types.

$$\begin{aligned}
 ITT = & \{E[Y|Z\uparrow c, AT] - E[Y|Z\downarrow c, AT]\} * Pr(AT) \\
 & + \{E[Y|Z\uparrow c, NT] - E[Y|Z\downarrow c, NT]\} * Pr(NT) \\
 & + \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C) \\
 & + \{E[Y|Z\uparrow c, Def] - E[Y|Z\downarrow c, Def]\} * Pr(Def)
 \end{aligned}$$

3. $ITT = \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C)$
 $= \{E[Y|T=1, C, Z=c] - E[Y|T=0, C, Z=c]\} * Pr(C)$
 $= \{E[Y(1)|C, Z=c] - E[Y(0)|C, Z=c]\} * Pr(C)$

Latent Type	Potential Treatment Status, $Z\uparrow c$	Potential Treatment Status, $Z\downarrow c$
Always-taker	T = 1	T = 1
Never-taker	T = 0	T = 0
Complier	T = 1	T = 0
Defier	T = 0	T = 1

LATE (CACE) Theorem in RDD

1. $ITT = E[Y|Z\uparrow c] - E[Y|Z\downarrow c]$

2. Consider four latent types.

$$\begin{aligned}
 ITT = & \{E[Y|Z\uparrow c, AT] - E[Y|Z\downarrow c, AT]\} * Pr(AT) \\
 & + \{E[Y|Z\uparrow c, NT] - E[Y|Z\downarrow c, NT]\} * Pr(NT) \\
 & + \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C) \\
 & + \{E[Y|Z\uparrow c, Def] - E[Y|Z\downarrow c, Def]\} * Pr(Def)
 \end{aligned}$$

3. $ITT = \{E[Y|Z\uparrow c, C] - E[Y|Z\downarrow c, C]\} * Pr(C)$
 $= \{E[Y|T=1, C, Z=c] - E[Y|T=0, C, Z=c]\} * Pr(C)$
 $= \{E[Y(1)|C, Z=c] - E[Y(0)|C, Z=c]\} * Pr(C)$

4. $CACE = ITT / Pr(C)$

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

Latent Type	Potential Treatment Status, $Z\uparrow c$	Potential Treatment Status, $Z\downarrow c$
Always-taker	$T = 1$	$T = 1$
Never-taker	$T = 0$	$T = 0$
Complier	$T = 1$	$T = 0$
Defier	$T = 0$	$T = 1$