Principles of Empirical Analysis

Lecture 10: Regression discontinuity design

Spring 2024
Miri Stryjan

Observational alternatives to experiments

1. **Selection on observables: treatment and control groups differ from each other only w.r.t. observable characteristics**
   - Lecture 8
2. **Selection on unobservables: treatment and control groups differ from each other in unobservable characteristics**
   - Exogenous variable induces variation in treatment – instrumental variables (IV) Lecture 9
   - Selection mechanism is known – regression discontinuity designs (RDD)
   - Treatment and controls are observed before and after treatment – difference-in-differences (DID)

Outline

- **Basic idea of regression discontinuity designs**
  - Recap
- **Applications**
- **Geographic boundary as regression discontinuity**

RDD – the setup

- **RDD has three fundamental components: running variable, cut-off, and treatment**
- **Individuals become treated after crossing some cutoff in the running (or forcing or score) variable**
  - Sharp RDD: treatment received with probability zero below the cut-off (or threshold) and probability one above cut-off
  - Fuzzy RDD: The probability of receiving the treatment increases discontinuously at the threshold (imperfect compliance)
- **Assumption: the potential outcomes evolve smoothly across the cutoff.** In other words:
  - If there is no precise manipulation of the running variable, observations just below the threshold are very similar to those just above the threshold and therefore constitute a valid control group.

Example:
Minimum legal drinking age in the US

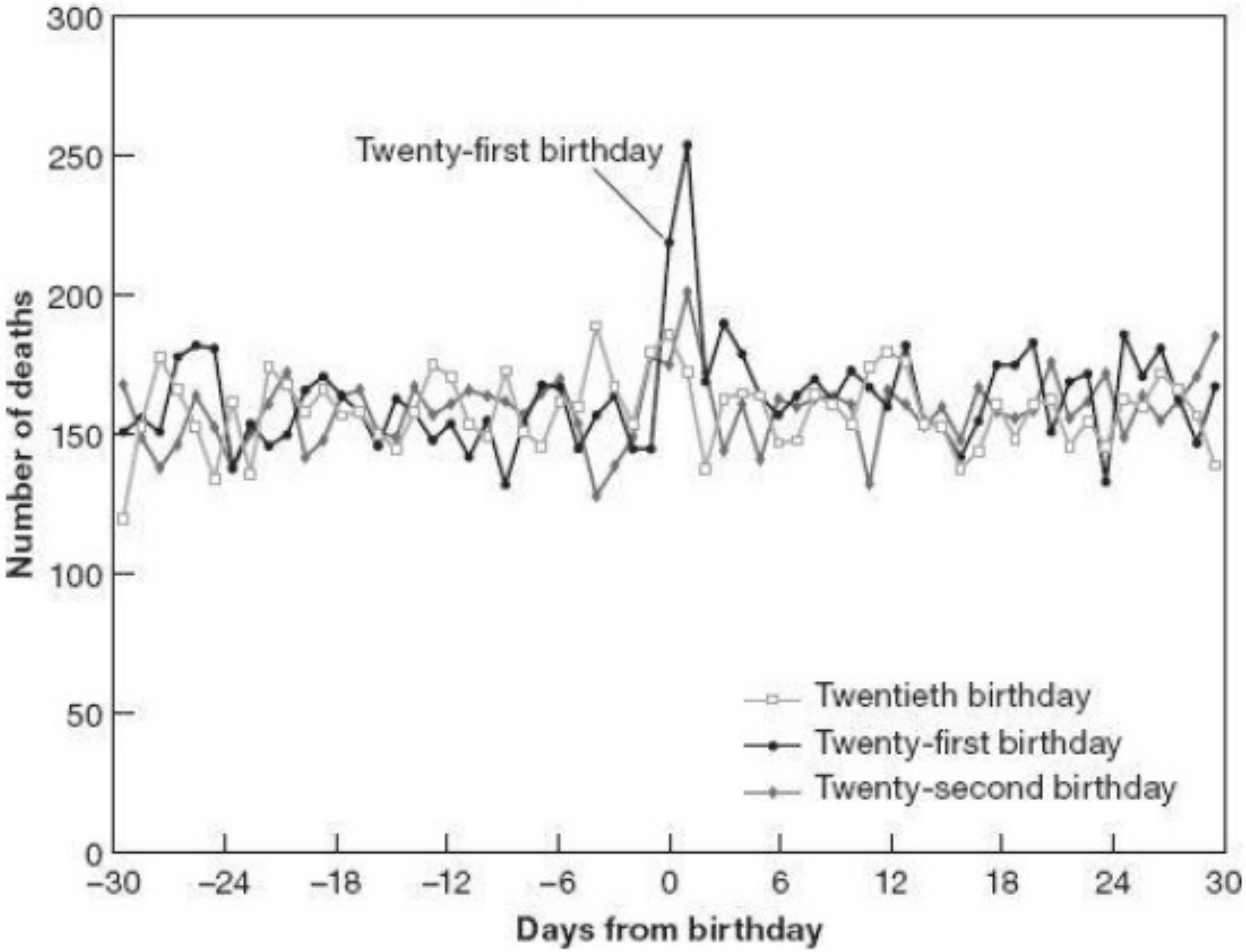
American Economic Journal: Applied Economics 2009, 1:1, 164–182
http://www.aeaweb.org/articles.php?doi=10.1257/app.1.1.164

The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age[†]

By CHRISTOPHER CARPENTER AND CARLOS DOBKIN[*]

We estimate the effect of alcohol consumption on mortality using the minimum drinking age in a regression discontinuity design. We find large and immediate increases in drinking at age 21, including a 21 percent increase in recent drinking days. We also find a discrete 9 percent increase in the mortality rate at age 21, primarily due to motor vehicle accidents, alcohol-related deaths, and suicides. We estimate a 10 percent increase in the number of drinking days for young adults results in a 4.3 percent increase in mortality. Our results suggest policies that reduce drinking among young adults can have substantial public health benefits. (*JEL* I12, I18)

FIGURE 4.1
Birthdays and funerals



Source: Angrist & Pischke (2015): Mastering Metrics.

What happens at age 21?

-> Legal drinking age in the US

T: legal access to alcohol **(here denoted by D_a using Mastering Metrics notation)**

Y: likelihood of dying (and specific cause of death)

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21. \end{cases}$$

- **Running variable?**
- **Cutoff?**
- **Treatment?**

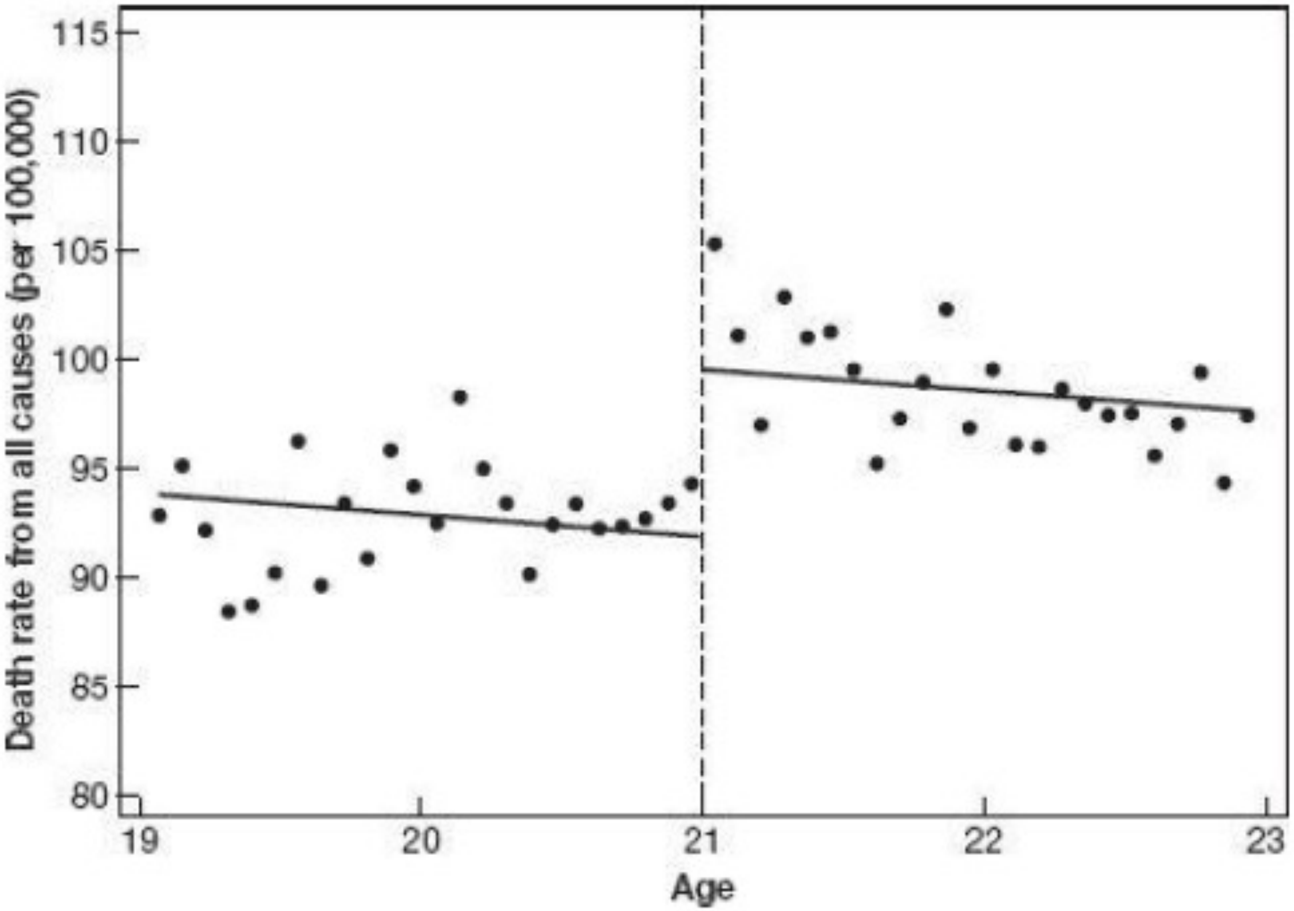
Alcohol and deaths

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21. \end{cases}$$

Two important things to notice:

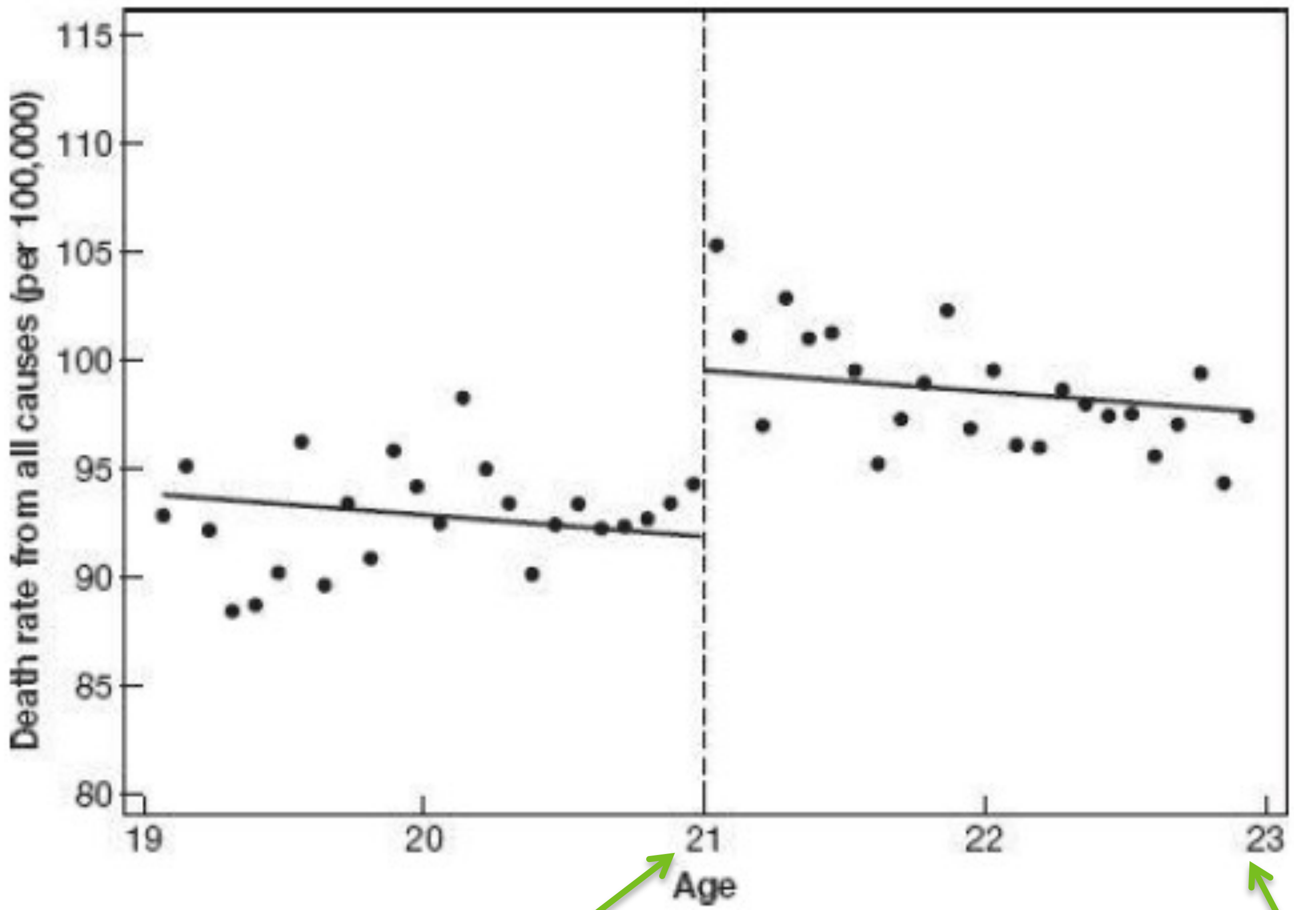
- Treatment status is a **deterministic function of a**, so once we know a, we know D_a (treatment status)
- Treatment status is a **discontinuous function of a**, because no matter how close a gets to the cut-off, D_a remains unchanged until the cut-off is reached.

“Although treatment is not randomly assigned, we know where it comes from”!



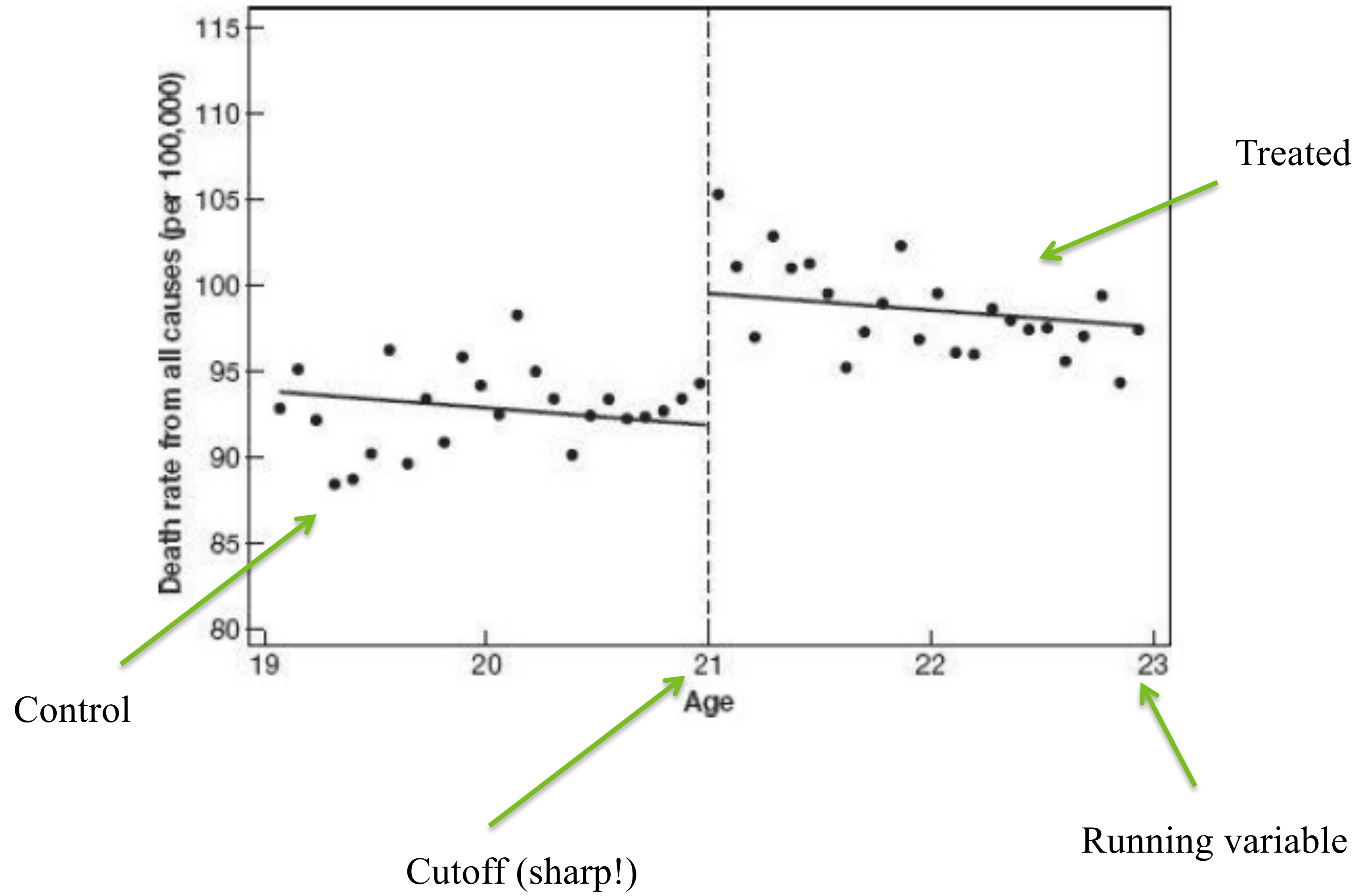
Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

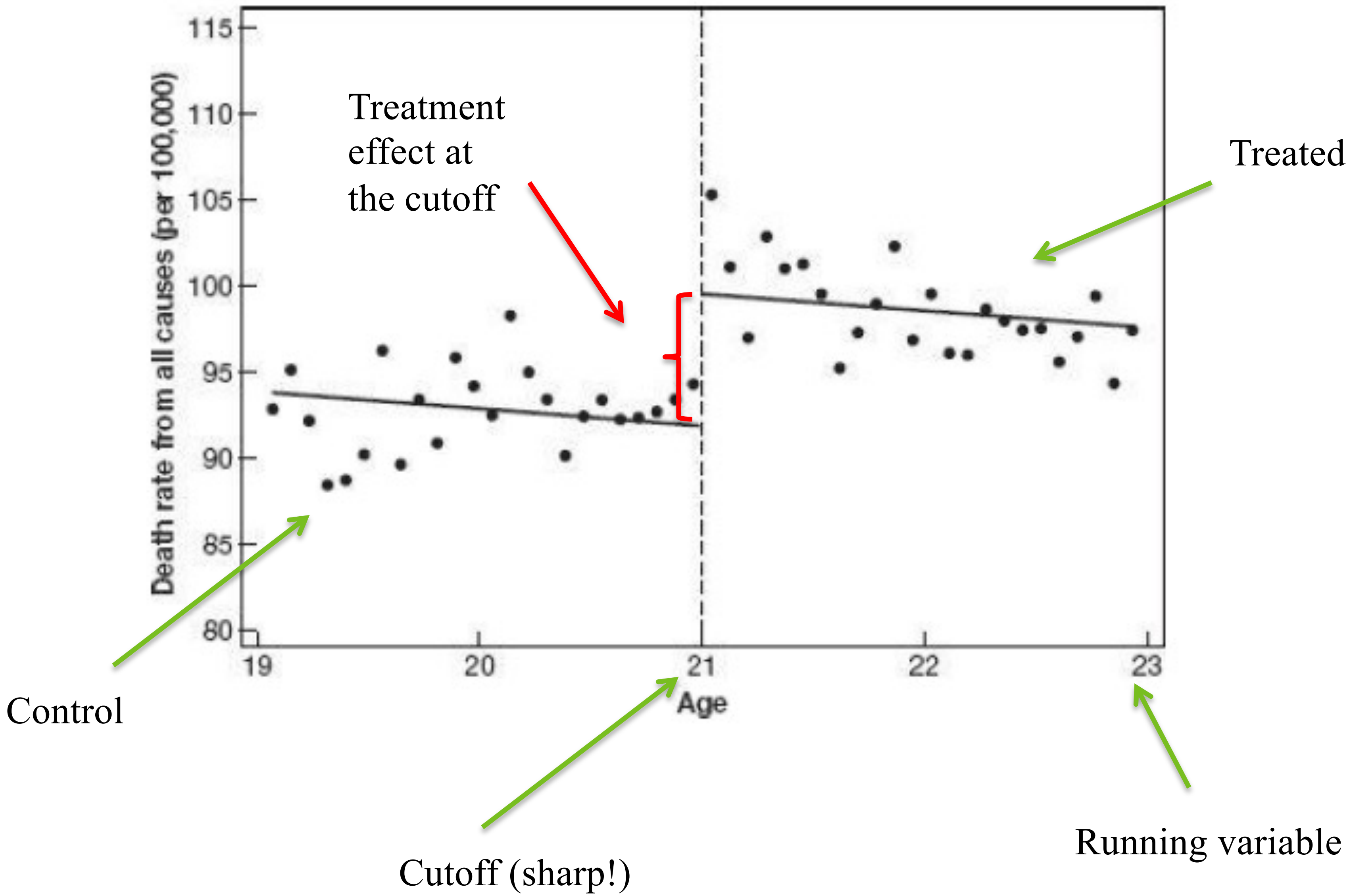
Source for the figure: Angrist & Pischke (2015): Mastering Metrics.

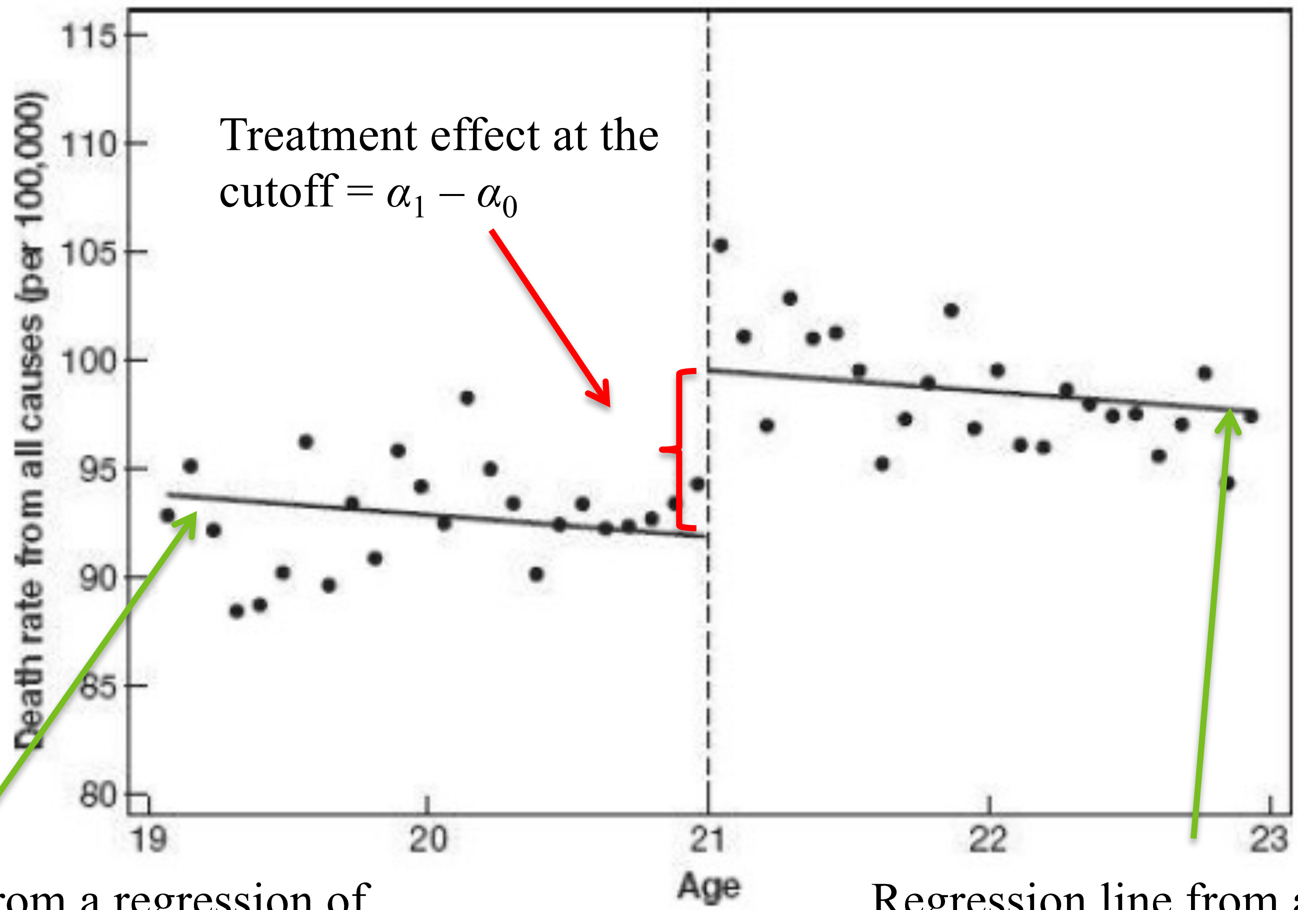


Cutoff (sharp!)

Running variable


Death rate from all causes (per 100,000)
115
110
105
100
95
90
85
80
19
20
21
22
23
Age
Treated
Control
Cutoff (sharp!)
Running variable


Treatment
effect at
the cutoff
Treated
Control
Cutoff (sharp!)
Running variable
Death rate from all causes (per 100,000)
115
110
105
100
95
90
85
80
19
20
21
22
23
Age



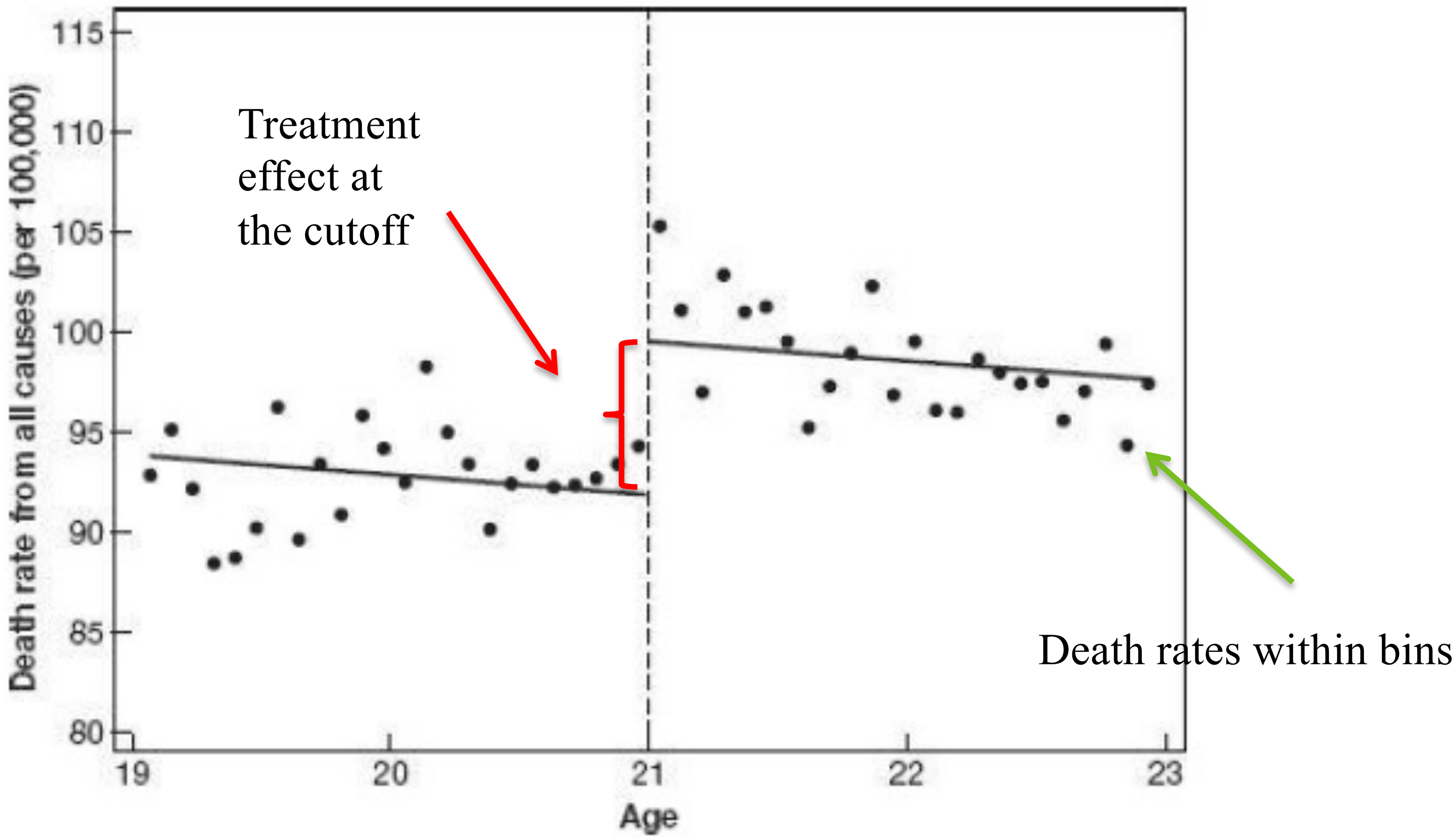
Treatment effect at the cutoff = $\alpha_1 - \alpha_0$

Regression line from a regression of death rates on age:

$E[death\ rate|age,\ age > 19\ \&\ age < 21\ ]$
$= \alpha_0 + \beta_0{*}age,$

Regression line from a regression of death rates on age:

$E[death\ rate|age,\ age \geq 21\ \&\ age < 23]$
$= \alpha_1 + \beta_1{*}age$



- Instead of showing a scatter plot on individual level data, papers often show a scatter plot where the data is binned into smaller number of groups along the variable on the x axis.
- The regression lines are **fitted separately for each side of the cutoff** using individual level micro data

Mechanisms

There seems to be a jump in death rates after age 21 in the US data. But how do we know that this jump is due to alcohol access and consumption?

We may need additional data. What data?

- Data on alcohol consumption by age
- Data on the causes of death by age

More results – alcohol consumption

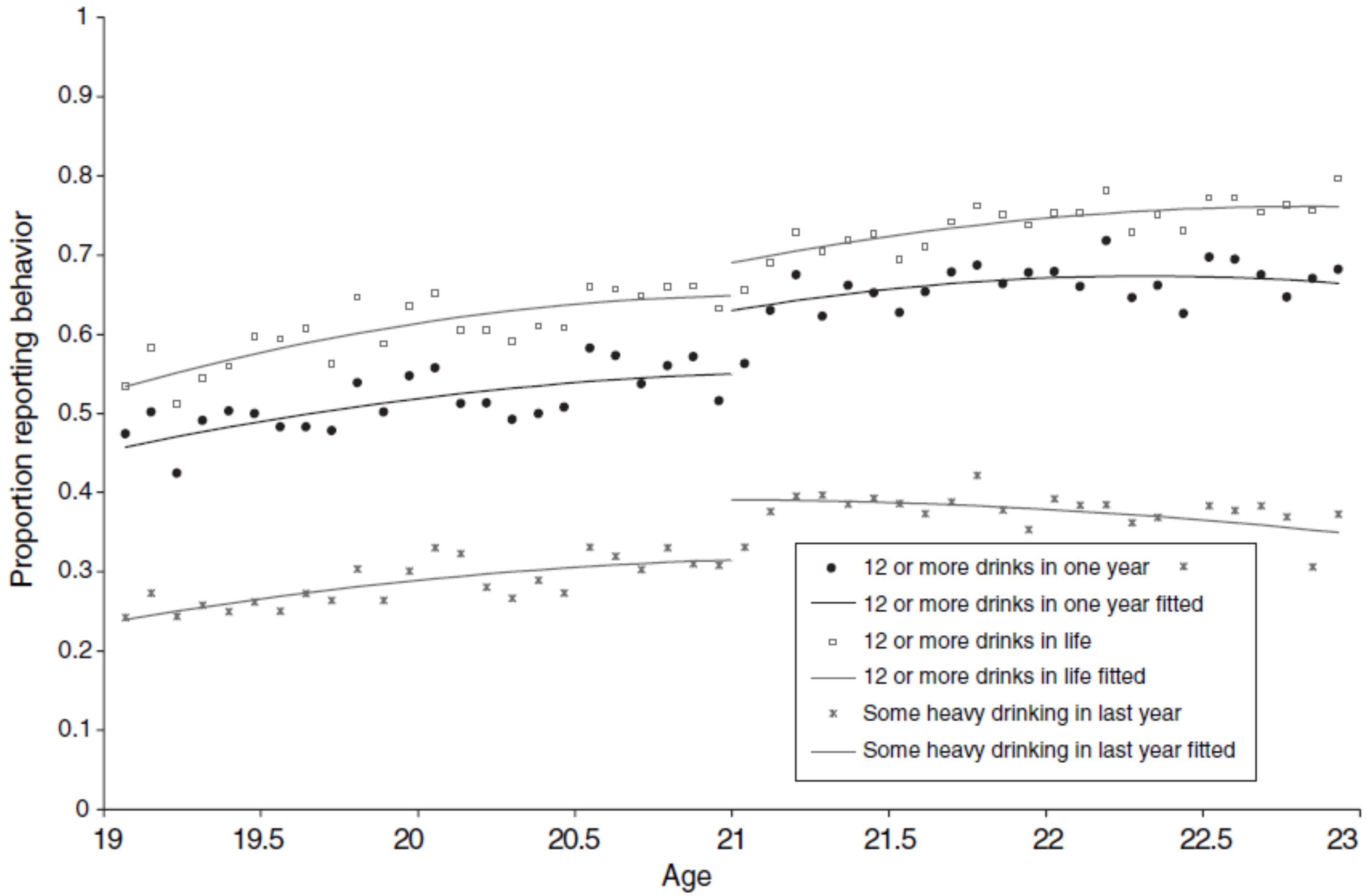


FIGURE 1. AGE PROFILE OF DRINKING PARTICIPATION

Notes: NHIS Sample Adult 1997–2005. Cells are the proportion of people in a 30-day block that report the behavior. The regression line is a second-order polynomial fitted on unweighted individual observations on either side of the age 21 cutoff.

More results – alcohol consumption

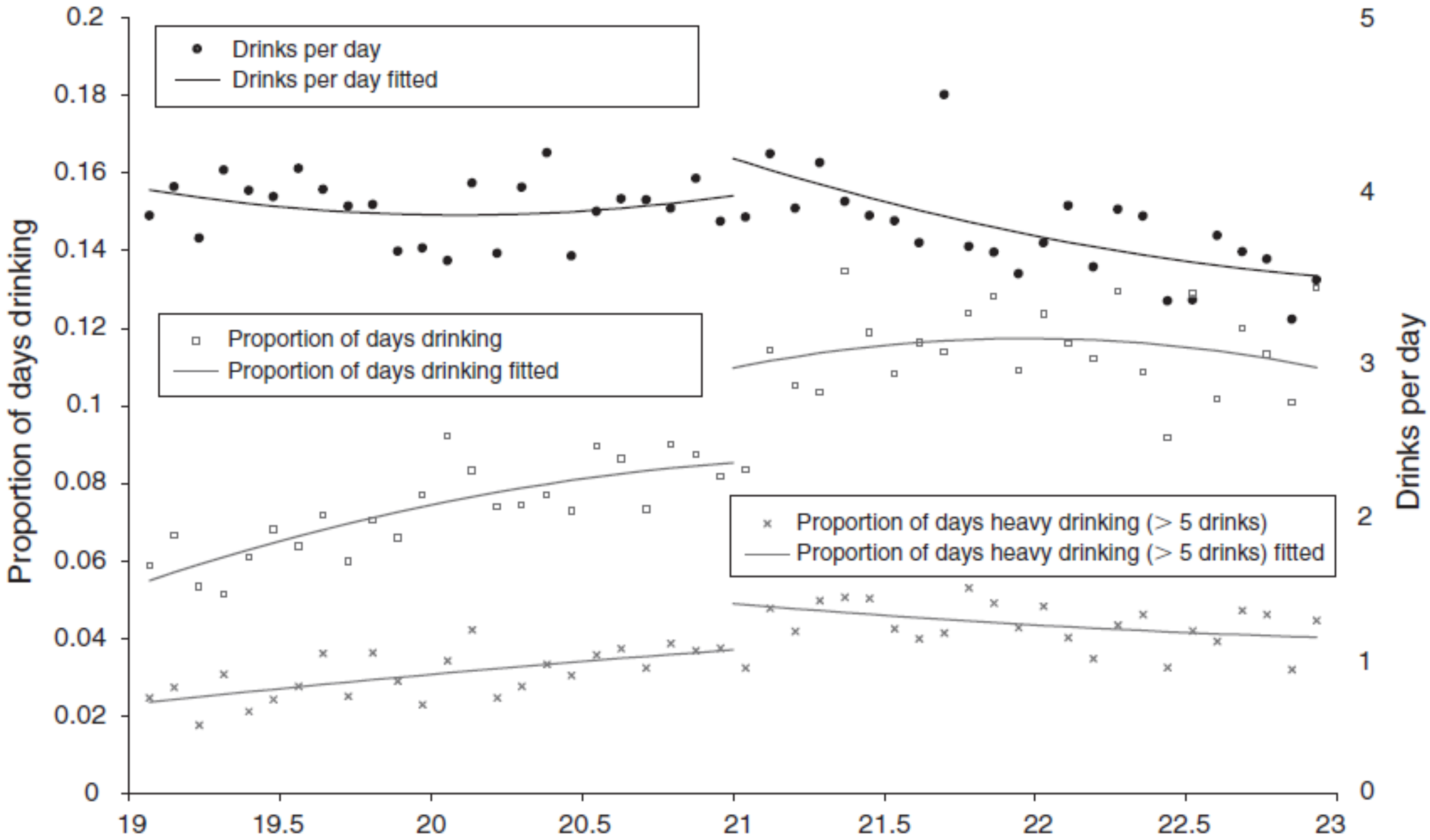


FIGURE 2. AGE PROFILE OF DRINKING INTENSITY

Notes: People can report their drinking for the last week, month, or year; 71 percent of respondents used a reference period of one week or one month. Average number of drinks per day is for people who reported some drinking.

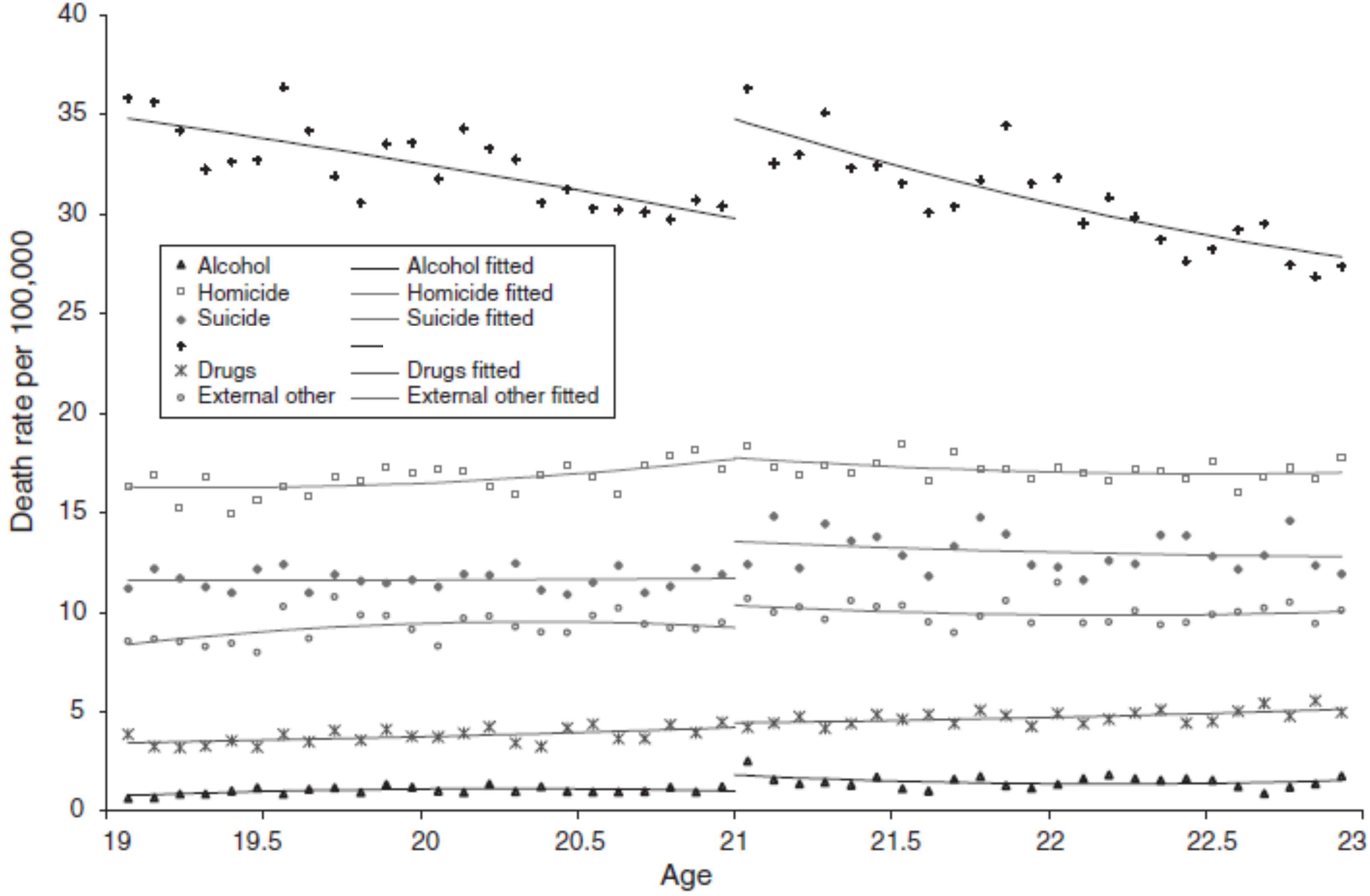


Figure 4. Age Profiles for Death Rates by External Cause

Notes: See notes to Figure 3. The categories are mutally exclusive. The order of precedence is homicide, suicide, ., deaths with a mention of alcohol, and deaths with a mention of drugs. The ICD-9 and ICD-10 Codes are in Appendix C.

Testing for RDD assumptions

Testing RDD assumptions

- **The underlying assumption in RDD is that units do not have the ability to precisely manipulate their own value of the running variable**

Ways to test if this assumption is fulfilled:

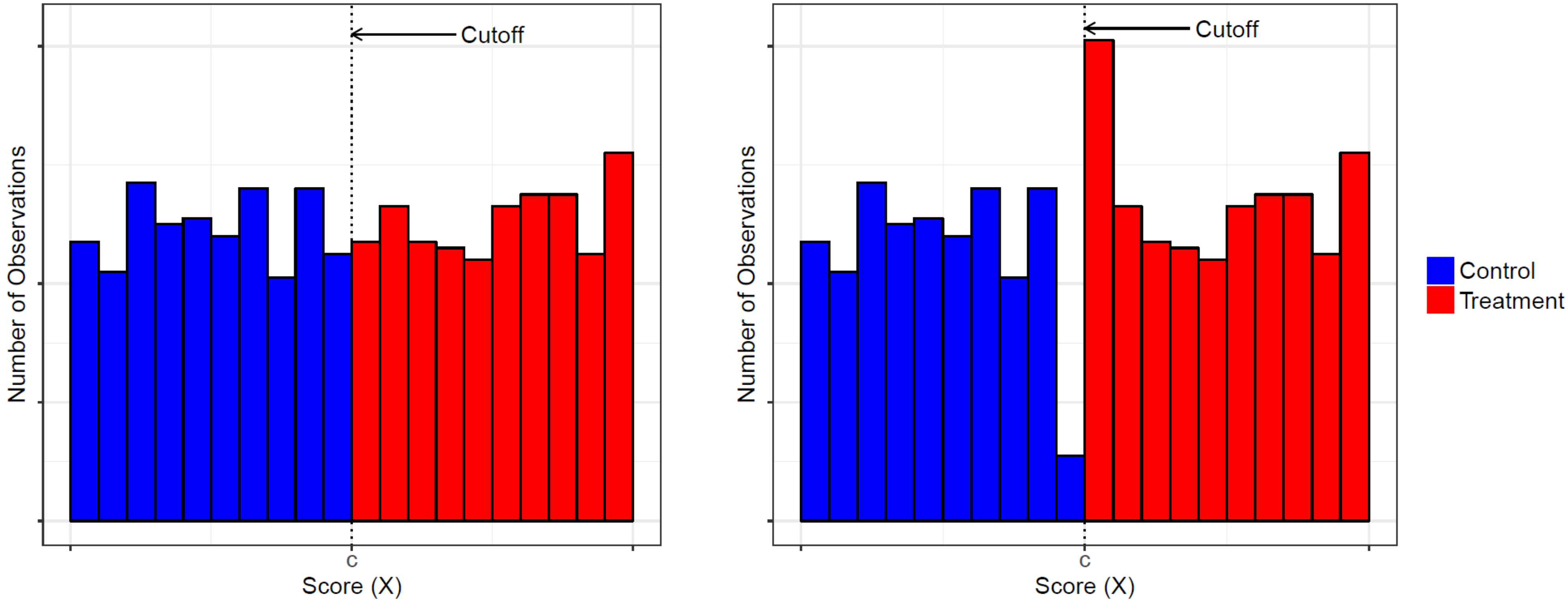
- Testing for sorting around the threshold: Plotting the histogram of the running variable to see if observations are evenly distributed around the threshold.
- Checking if observations just above and just below the threshold are indeed similar with respect to other observables.
- Placebo tests using other cutoffs that are not actually affecting the treatment, such as other ages than 21 for the drinking age paper.

Sorting or "manipulation" of the running variable – distribution of units around the threshold

- **The underlying assumption in RDD is that units do not have the ability to precisely manipulate the value of the running variable**
  - If they could and the treatment is something beneficial, units would want to receive the treatment and sort on the right side of the cutoff
- **With no precise manipulation, the number of treated observations just above the cutoff should be approximately the same as the number of control observations below it**
  - Test: plotting the histogram of the running variable and inspecting whether the number of observations are similar near the cutoff
  - Also, a formal statistical density test (McCrary test)

Test for sorting or “manipulation” of the running variable

Which graph below shows signs of “sorting”



(a) (b)

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

Thinking more about testing RDD assumptions

Suppose you want to estimate the effect of a property tax increase on profits and survival of businesses in a municipality by comparing businesses near the border in bordering municipalities with different property tax rates.

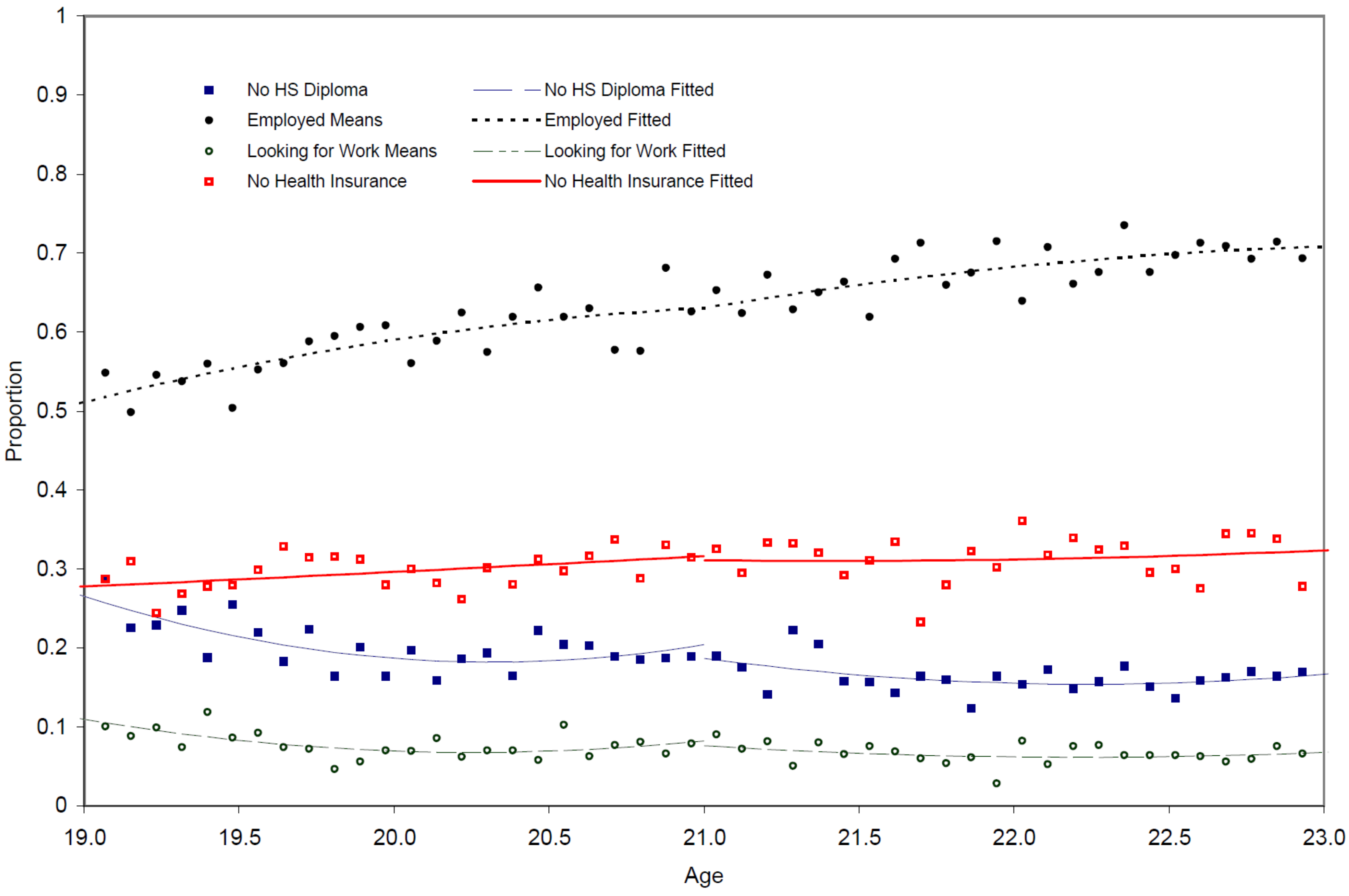
- **Your running variable is distance to the municipal border**
- **the cutoff is the border**
- **The treatment is facing higher property taxes**

Would you be worried that businesses manipulate their running variable? If yes, what would you be worried about? Explain!

Test of observable variables – are units above and below threshold comparable?

- **One of the most important RDD falsification tests involves examining whether, near the cutoff, treated units are similar to control units in terms of observable characteristics**
- **Idea: if units lack the ability to precisely manipulate their value of the running variable, there should be no systematic differences between units with similar values of the running variable**
  - Thus, except for their treatment status, units just above and just below the cutoff should be similar in all variables that could not have been affected by the treatment ("predetermined")
- **Implementation: all predetermined variables should be analyzed using RDD in the same way as the outcome of interest**

Test of predetermined covariates



Placebo tests – using “fake” cutoffs

Placebo test 1: to replace the true cutoff value with a fake cutoff value in the running variable

- A value at which the treatment status does not really change and perform estimation and inference using this “fake” cutoff
- A significant treatment effect should occur only at the true cutoff value and not at other values where the treatment status is constant
- No jumps in death rates at 18, 19 or 25 etc.

Placebo test 2: run placebos at the true cutoff but replace the outcome Y with other outcomes that should not be affected by the treatment

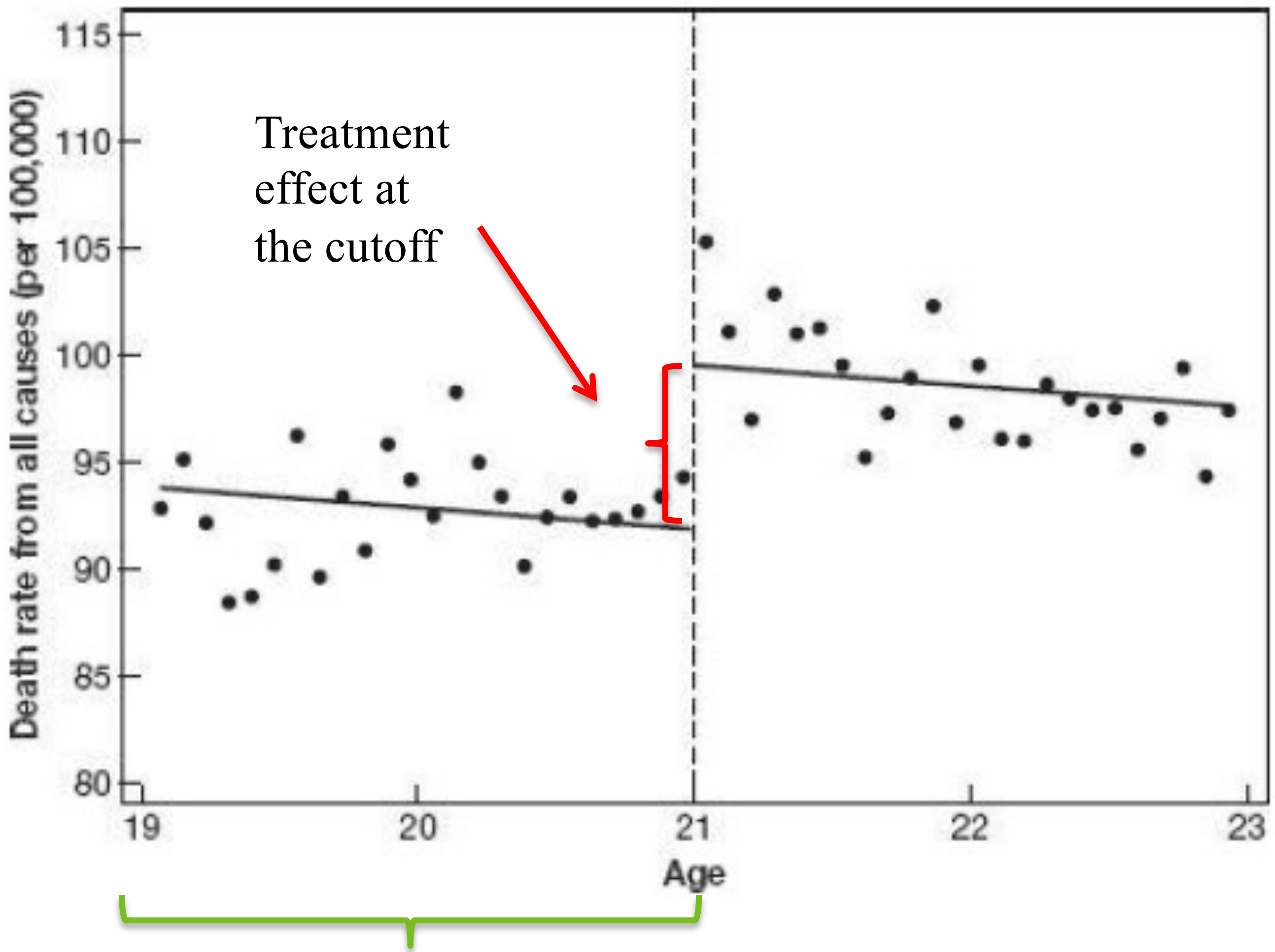
Limitations of RDD

Local randomization interpretation

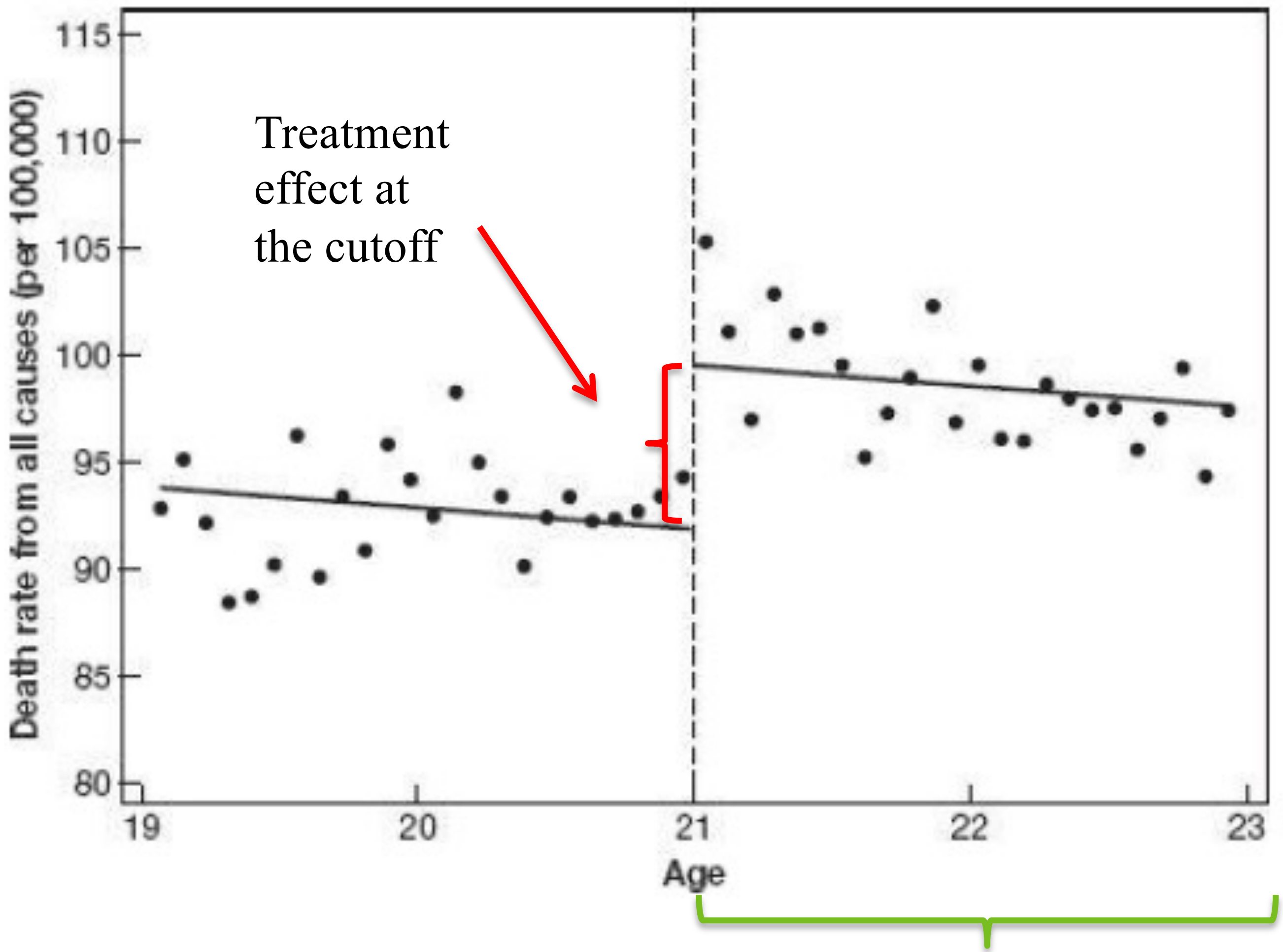
- **Given that units are unable to precisely manipulate the running variable, the RDD can be interpreted as a randomized experiment inside a window around the cutoff**
  - That is, the treatment assignment is locally random
  - Results can be generalized to a narrow segment of the running variable: in the case of the MLDA study: young people near age 21.
- "zoom in" on points close to the cut-off.
- BUT: this requires a lot of data near the cutoff.

Technical issues

- **In principle, what we do in RDD is comparing means for those just above to those just below the cut-off.** Often, we do not have enough data to estimate the treatment effect simply by comparing means at the cutoff so we need to use data away from the cutoff
- **How much data away from the cutoff should we use?**
  - In other words, how large a bandwidth should we use?
- **The choice involves a bias-variance trade-off:**
  - The closer to the cutoff you are, more likely it is that you are able estimate an unbiased causal effect.
  - But at the same time variance or the standard error of your estimate is larger as you are using fewer data points – more noise.



The bandwidth: the share of observations used in estimating the local linear regression:
$E[death\ rate|age,\ age > 19\ \&\ age < 21] = \alpha_0 + \beta_0{*}age$



The bandwidth: the share of observations used in estimating the local linear regression:
$E[death\ rate|age\ ,\ age \geq 21\ \&\ age < 23] = \alpha_1 + \beta_1{*}age$

How to address limitations - robustness

In an RDD paper, the authors need to show that their results are robust to different modelling and data choices, for example:

- different choices of bandwith around the cutoff (within reasonable values)
- different specification of the relationship between the running variable and the outcome Y.

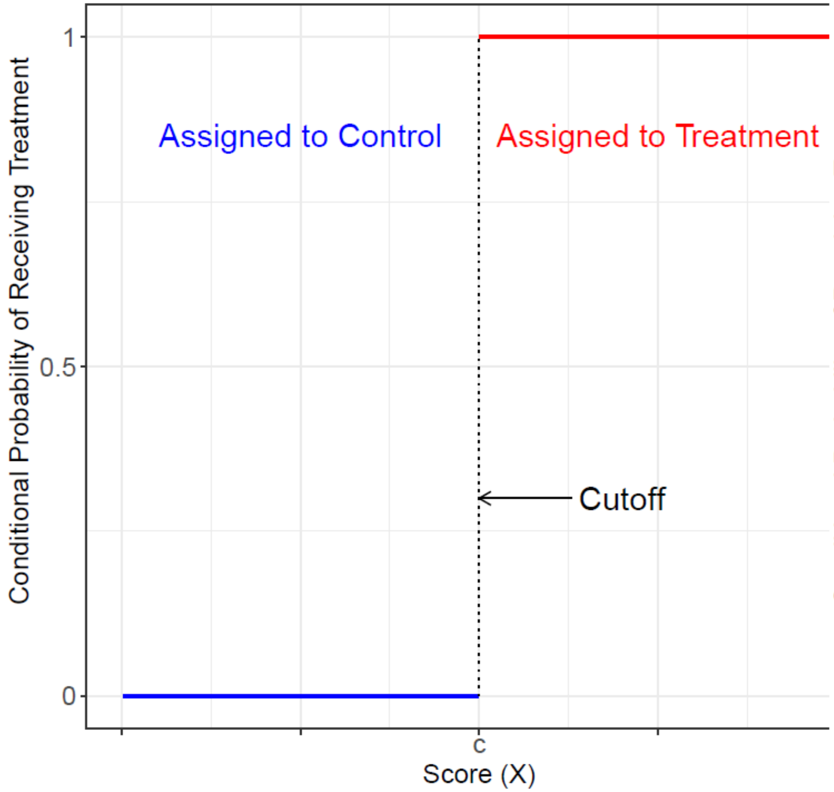
Fuzzy RD

When passing the cutoff creates a jump in treatment probabilities or treatment intensity, rather than switching the treatment on or off completely, the resulting RD design is said to be fuzzy.

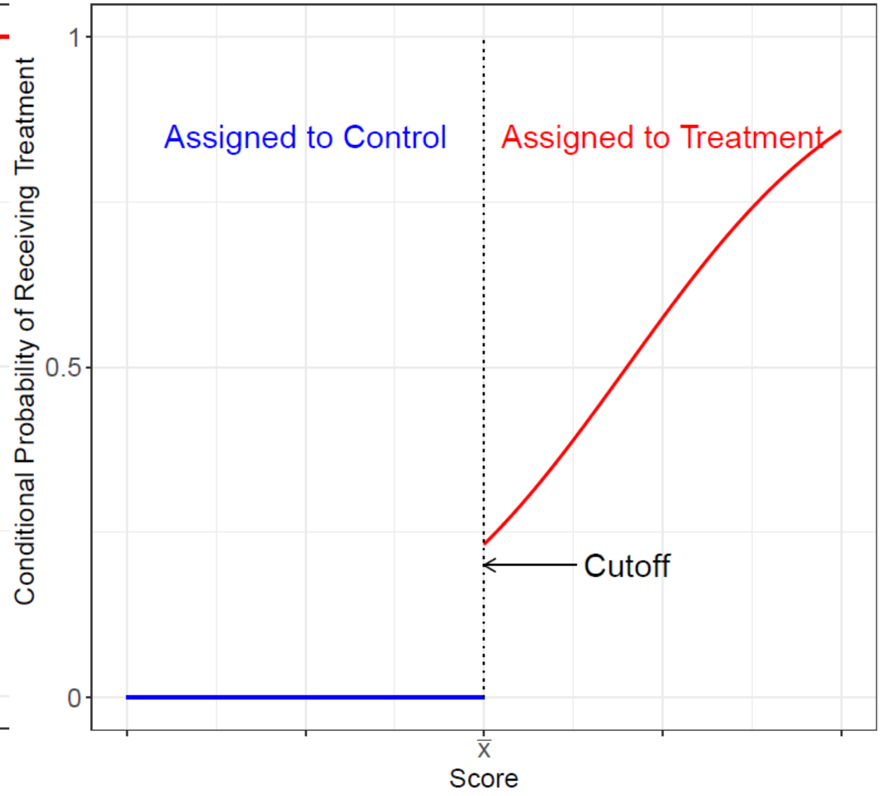
Example: "exam schools" in the US. Admission if you score high enough on Independent Schools Entrance Exam (ISEE). Different schools have different cut-offs. Suppose there are 3 schools, and school 1 has the highest cut-off for intake.

- Scoring > cut-off for school 1: increases your probability of attending an exam school, but not to p=1; some still choose to go elsewhere.
- Scoring < cut-off for school 1: decreases your probability of attending an exam school, but not to 0: you cannot go to *school 1* but you can still apply to the two other exam schools!

Sharp and fuzzy RDD



(a) Sharp RD



(b) Fuzzy RD (One-Sided)

”the conditional probability of actually receiving treatment given the score changes discontinuously at the cut-off.”
Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Extensions.

Fuzzy RD

Situations where fuzzy RD can be used are situations with ***imperfect compliance***.

- Compliance is a topic we will get back to next lecture.

non-compliance introduces complications and typically requires stronger assumptions to learn about treatment effects of interest.

With fuzzy RDD, being above the cutoff is an *instrument* for being treated: it increases probability of treatment, but not to 1. We will discuss instrumental variables (IV) more next week.

Other applications of RDD

Sarvimäki, Uusitalo & Jäntti (2021)

http://www.aalto-econ.fi/sarvimaki/forced.pdf

Sarvimäki et al. (2021)

- **After World war II, 11% of the population was forced to migrate and resettled into the remaining parts of Finland**
  - For those working in agriculture – roughly one half of the population – the government attempted to reconstruct the pre-war conditions as closely as possible
  - Displaced farmers were given land and assistance to establish new farms in areas that had similar soil and climate as the origin regions
  - Former neighbors were resettled close to each other in order to preserve social networks
- **Once the resettlement was completed in 1948, the displaced farmers were not subject to any special policies**
  - They received no further subsidies and, like everyone else, were free to sell and buy land and to move across locations and sectors

Sarvimäki et al. (2021)

Idea of paper:

treated group: those who were on the side of the border ceded to the Soviet Union (black area) and were moved into the white areas on the map.

control group: those who were on the Finnish side of the border and could stay on their farms.

What do you think is the:

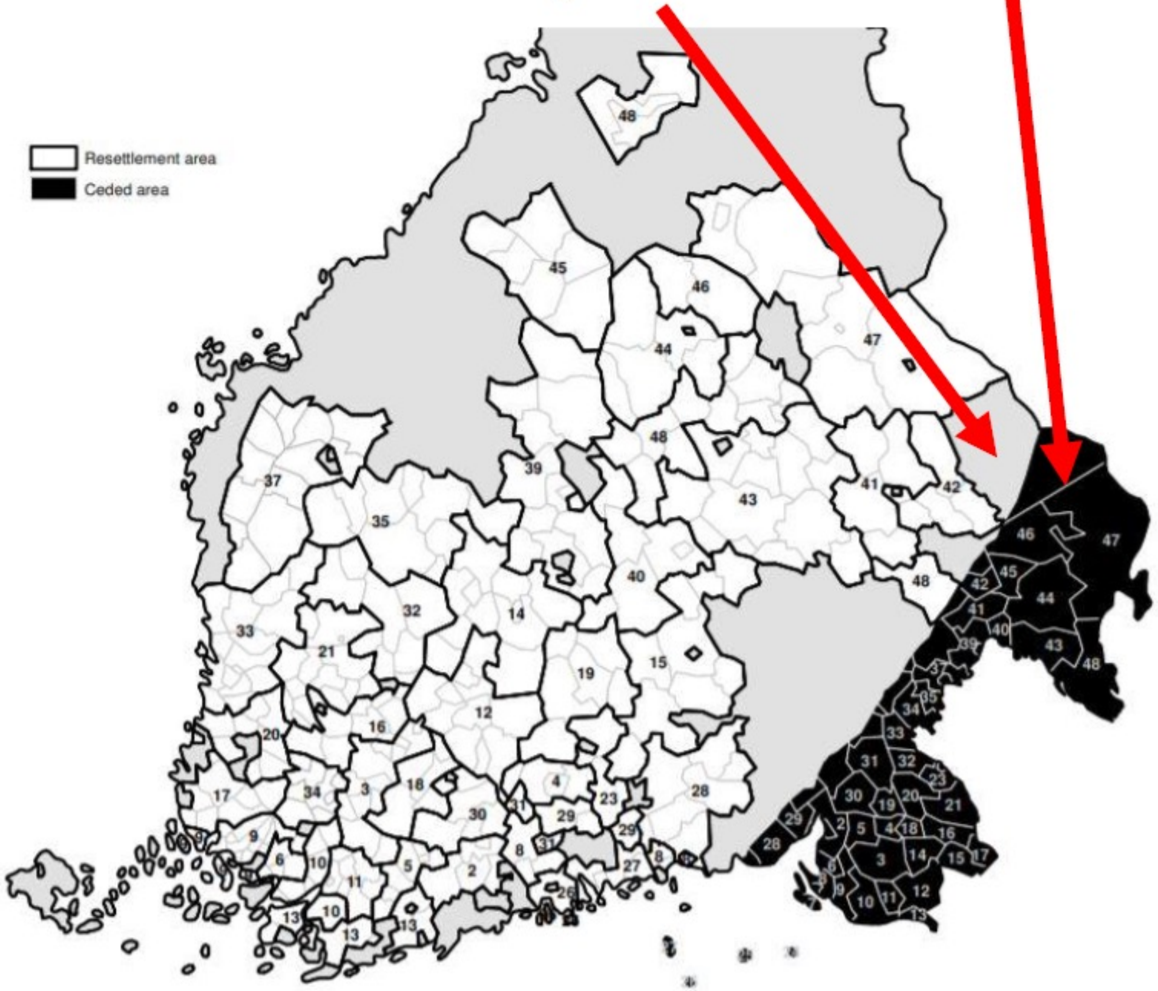
Running variable?

Cutoff?

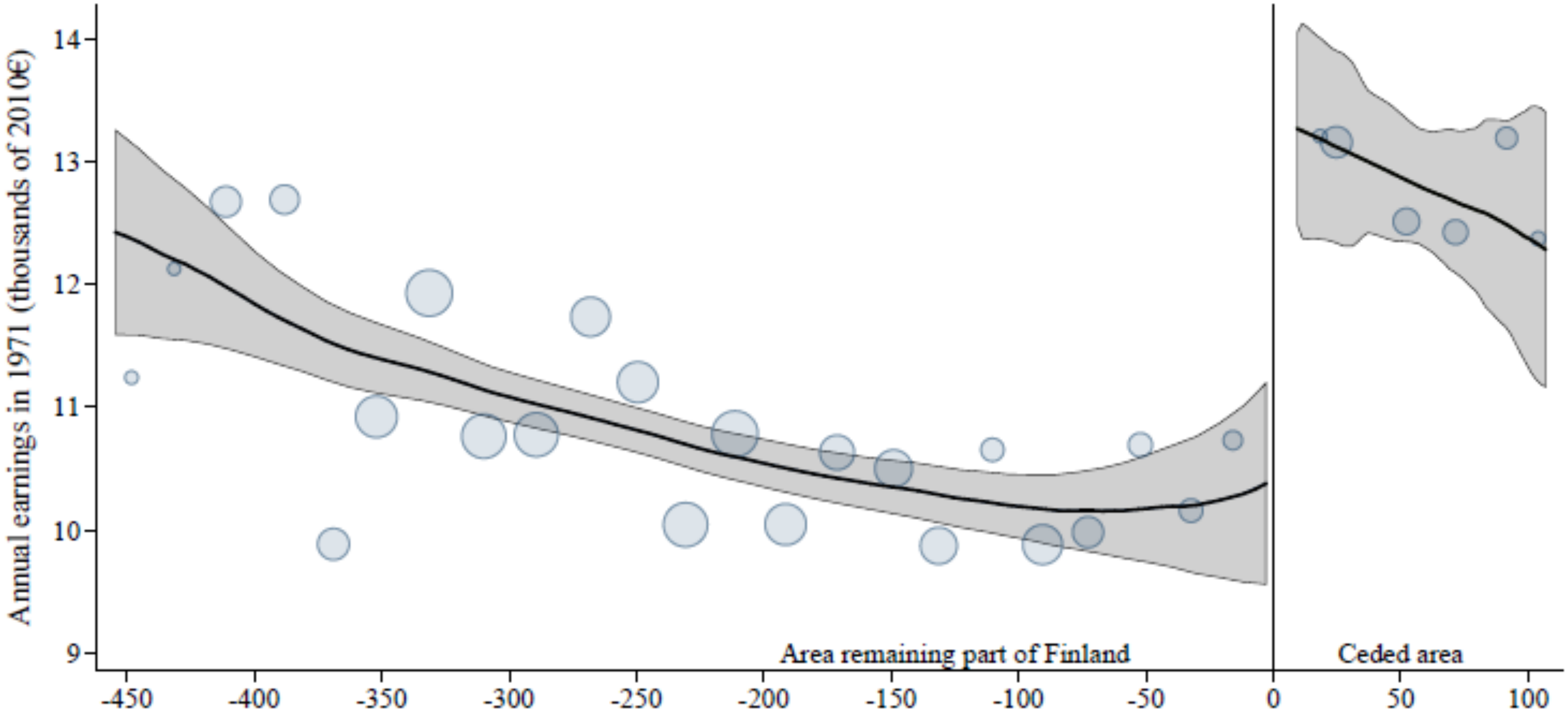
Treatment?

Families in these areas displaced.

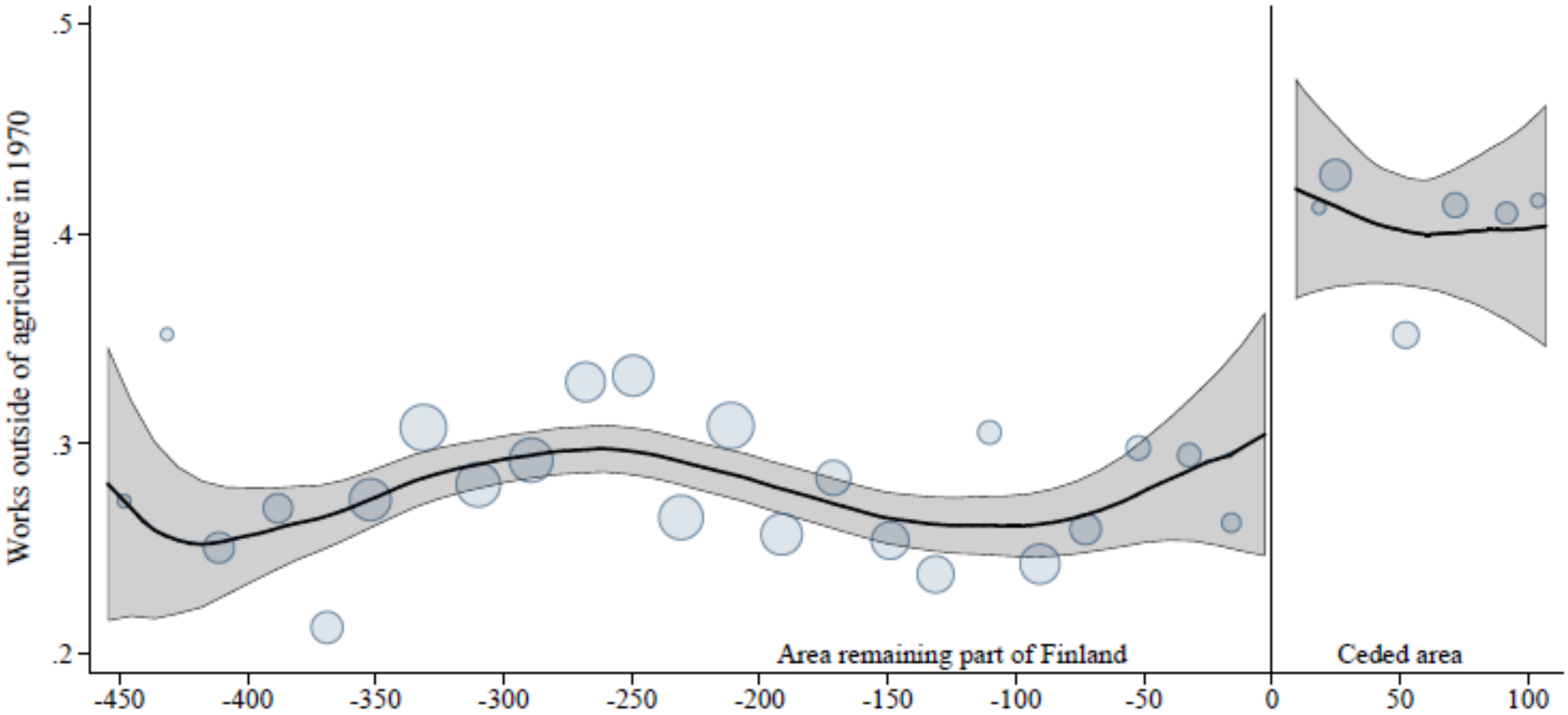
Families in these areas can stay where they are.



Sarvimäki et al. (2021) – main results



Sarvimäki et al. (2021) – main results



Conclusions – Sarvimäki et al. (2021)

- **The post war difference between displaced and non-displaced farmers suggests that forced migration increased long term income by 10–29% among men working in agriculture before the war**
  - Forced migration increased the likelihood of leaving agriculture between 1939 and 1970 by 10–16 %-points from a baseline of 28%
  - Increased the likelihood of moving to a city and to complete secondary education among the displaced farmers
- **These results suggest that the positive impact of forced migration on the income of farmers can be attributed to an increased likelihood of leaving agriculture**

Huttunen et al. 2023

Journal of Public Economics 218 (2023) 104804

ELSEVIER

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Lost boys? Secondary education and crime

Kristiina Huttunen[a,*], Tuomas Pekkarinen[b], Roope Uusitalo[c], Hanna Virtanen[d]

[a] *Aalto University, VATT Institute for Economic Research, Helsinki GSE and IZA, Finland*
[b] *VATT Institute for Economic Research, Aalto University, and IZA, Finland*
[c] *University of Helsinki, VATT Institute for Economic Research, Helsinki GSE and IZA, Finland*
[d] *ETLA Economic Research, Finland*

ARTICLE INFO

Article history:
Received 7 May 2021
Revised 16 December 2022
Accepted 20 December 2022
Available online 2 January 2023

JEL:
K42
I2

Keywords:
Crime
School admission
Secondary education

ABSTRACT

We study the effect of secondary education on criminal behavior of young men in Finland. We exploit admission cut-offs in over-subscribed programs and estimate the effect of gaining access to a) any secondary school vs no access, b) general vs vocational school, and c) selective vs less selective general school. Our results show that admission to any secondary school has a sizeable negative effect on the propensity to commit crime. There are no effects at the other two margins. The negative effects at the extensive margin are largest in the years following school admission and result in a reduction of the probability of ever committing crime rather than simply delaying the onset of crime. Our results suggest that keeping youth at school at a critical age has effects that last beyond years where effects on enrollment are observed.

Huttunen et al: the effect of secondary education on criminal behavior.

Focus: Eight cohorts of Finnish male students who graduated from compulsory schooling between 1996 and 2003 and apply to secondary education immediately upon graduation.

Running variable: Grade point average from compulsory school (grade 9)

Cutoff: minimum score for being accepted into sec education, depends on how many apply each year so not known ex ante

Treatment: attending secondary education

Y: Criminal behavior

Huttunen et al. explanation for how they measure "attending any secondary school"

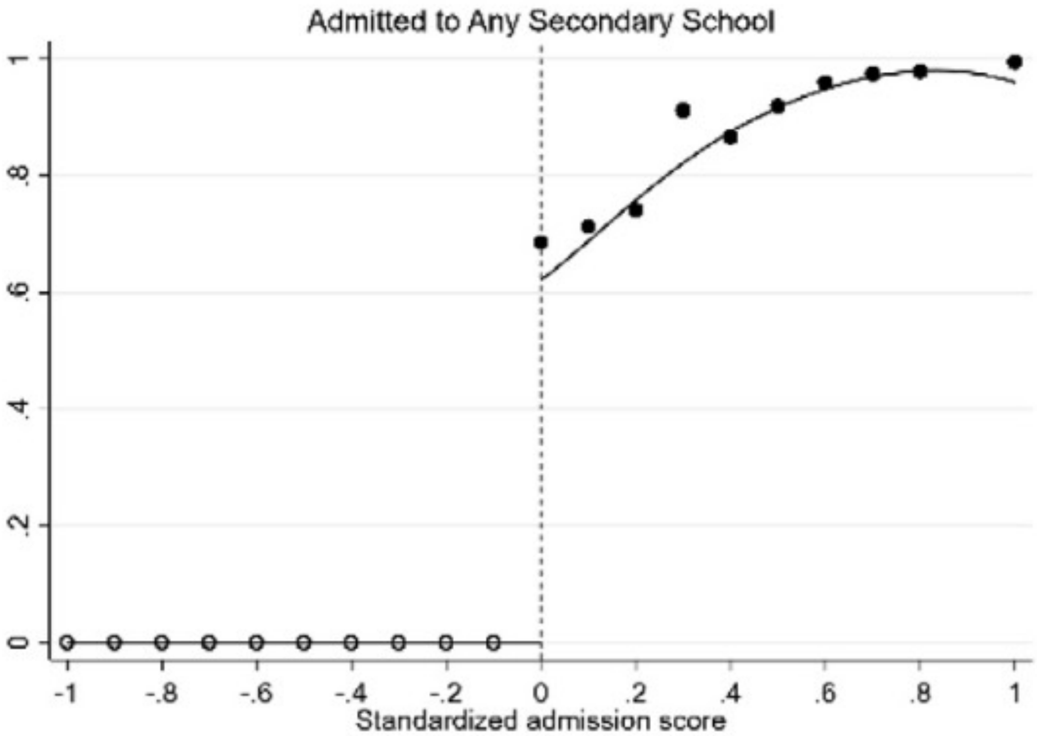
”we use admission cut-offs that are critical in determining the access to any type of secondary education. For each applicant we pick the program that has the lowest cut-off of the programs the applicant listed in his application.

The applicants who are rejected at this margin have been rejected by all secondary schools that they applied to.

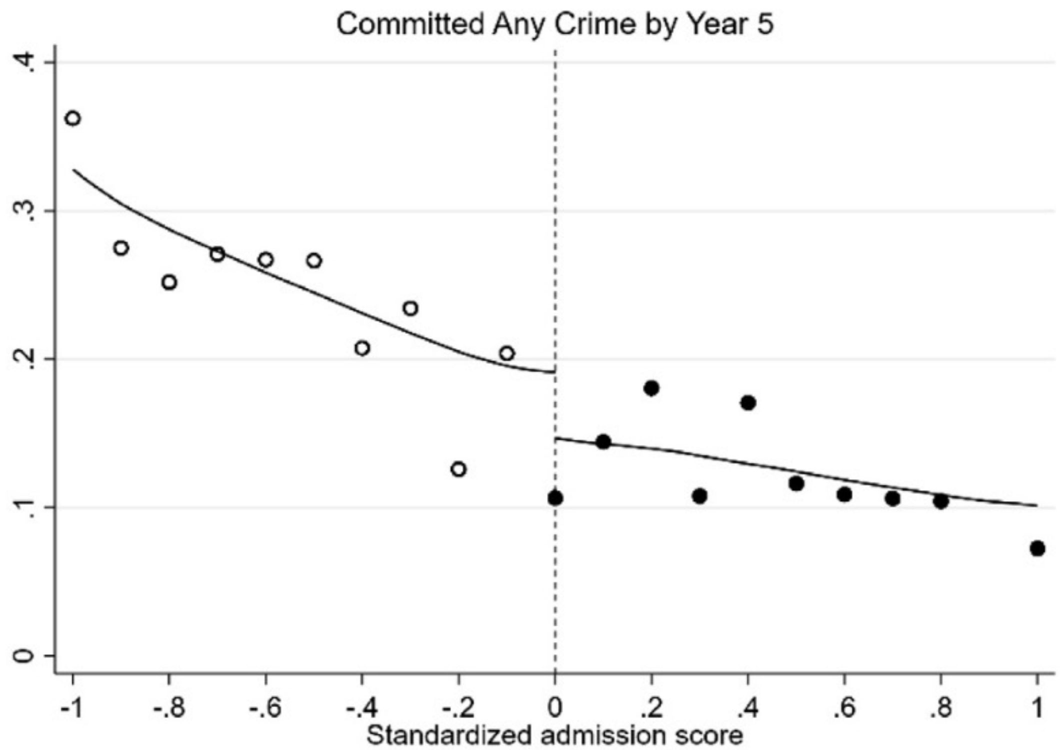
They may enroll in optional 10th grade of comprehensive school or in preparatory training, or they may opt out of education altogether.

Rejected applicants can also apply again to secondary education in the following years.“

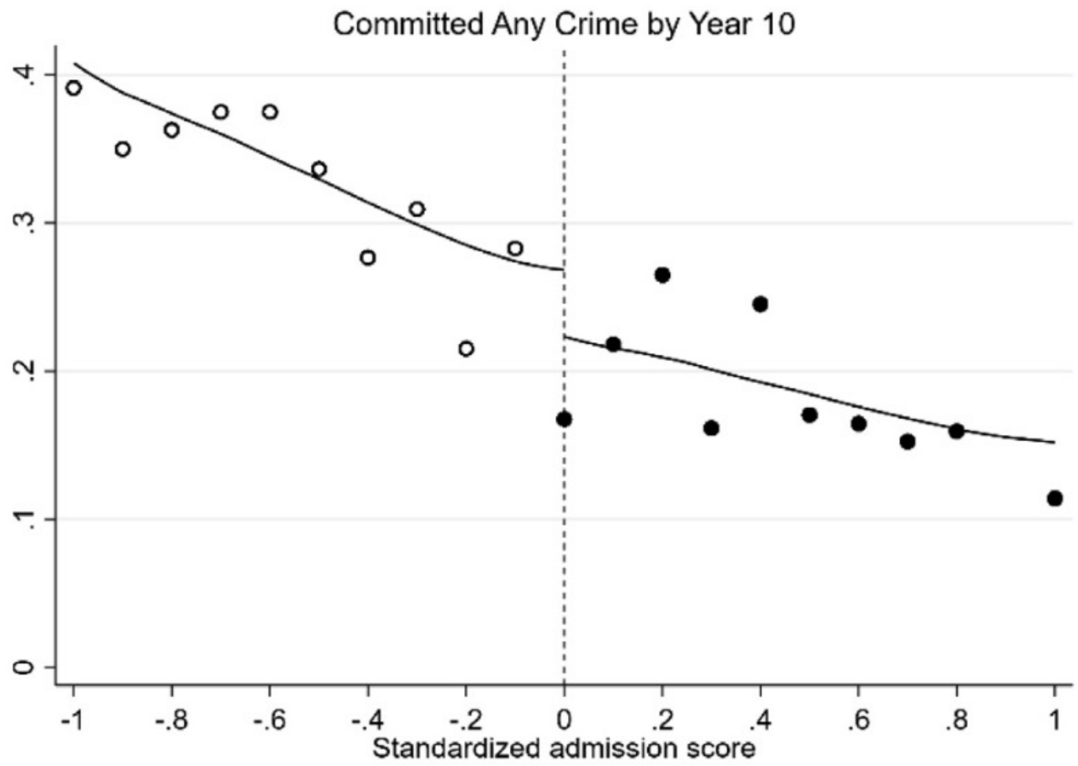
grade score and secondary school admission



grade score and crime by year 5 after finishing 9th grade



grade score and crime by year 10 after finishing 9th grade



Aalto University
School of Business

Huttunen et al. conclusion

"Our results show that being successful in gaining access to secondary education decreases the likelihood of committing crimes among young men."

- **Men admitted to secondary schools are 52 % less likely to be convicted within 10 years after admission than men that were not admitted.**
- **The crime reducing effect of secondary education is restricted to the extensive margin (= secondary school vs no secondary school)**
  - no effect on crime when examining admission to the general track vs. vocational track, or to the more selective general secondary schools.

Terms: extensive vs intensive margin

Meyersson (2014)

ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Islamic Rule and the Empowerment of the Poor and Pious

Erik Meyersson

First published: 05 February 2014 | https://doi.org/10.3982/ECTA9878 | Citations: 81

"Using a regression discontinuity design, I compare municipalities where this Islamic party barely won or [barely] lost elections. **Despite negative raw correlations, the RD results reveal that, over a period of six years, Islamic rule increased female secular high school education"**

Units: Turkish municipalities

running variable: the margin of victory of the (largest) Islamic party in the 1994 Turkish mayoral elections.

Treatment: Islamic party's electoral victory

Cut-off: zero, since municipalities elect an Islamic mayor when the Islamic vote margin is above zero, and elect a secular mayor otherwise.

Meyersson (2014)

238 E. MEYERSSON

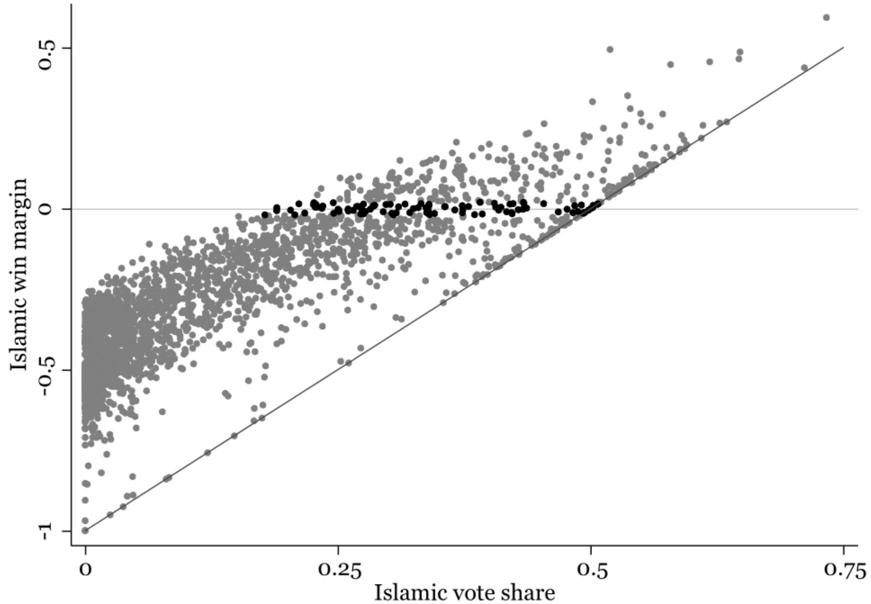


FIGURE 1.—Islamic win margin and Islamic vote share in 1994. The graph shows the total vote share for the Islamic party plotted against the Islamic win margin—the difference between the Islamic party's vote share and the largest secular party's vote share—both in 1994. Observations within 2 percentage points of the threshold at zero are in black. The diagonal line is the hypothetical one-to-one relationship between the two variables in an election with only two parties.

The figure shows that in the municipalities where the Isamic party won the election (win margin >0) there is still a large variation in the Islamic vote share (x axis).
This is interesting because we can then estimate the effect of Islamic rule on municipalities in a range of "preferences" for the Islamic party.
The black dots are the munici-palities close to the cutoff which are the focus of the RCC analsyis.

RDD recap

- **Idea:**
  - If a rule determines treatment due to a hard to predict cut-off, we can use the rule to estimate a causal effect without an RCT
- **Necessary criteria for using RDD:**
  - the running variable, treatment, and cutoff must exist and the probability of treatment assignment as a function of the running variable changes discontinuously at the cut-off
- **Assumption:**
  - Units just below and just above the cut-off are very similar and comparable (Potential outcomes develop smoothly across the cut-off)
- **Testing for design validity:**
  - Density tests, covariate balance test, placebo tests
- **Challenges:**
  - Requires a lot of observations near the cut-off
  - We cannot extrapolate results to units far from the cut-off (local causal effects!)