What variation can you exploit, and how far does it travel?

Francisco Villamil

Research Design for Social Sciences

MA Computational Social Science, UC3M

Fall 2026

Where we are

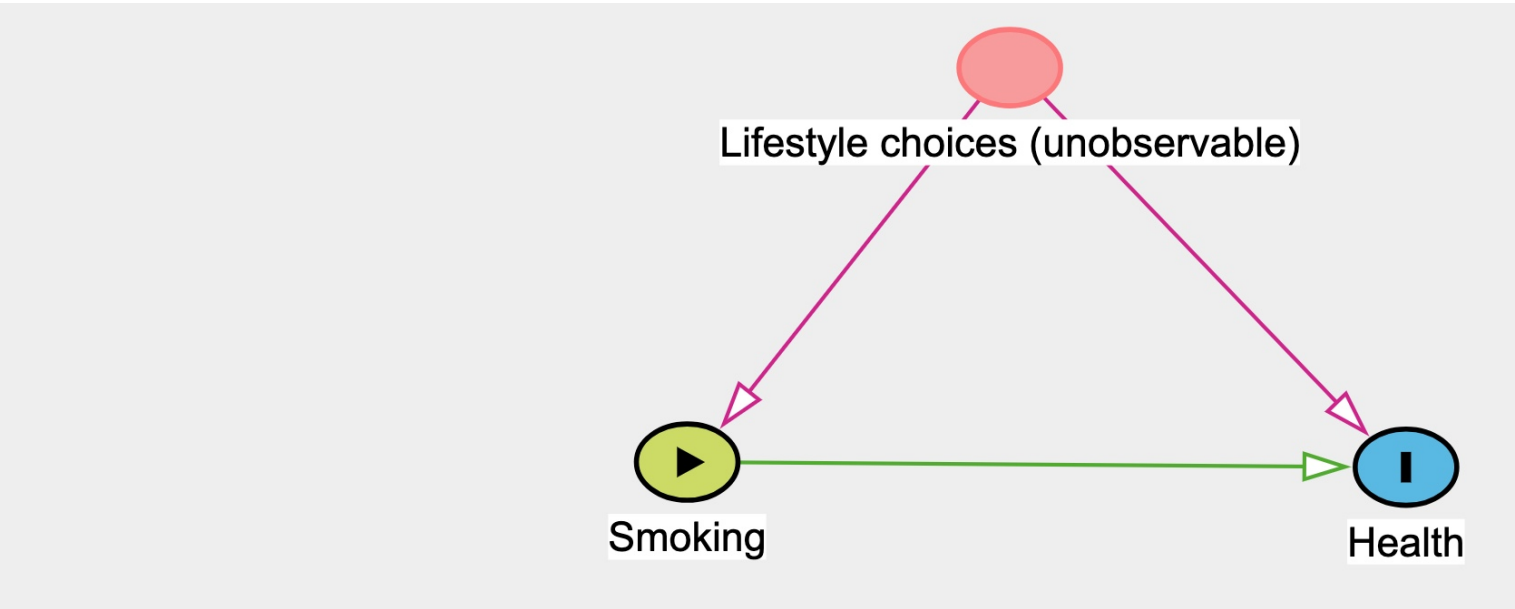
1. Why design at all?
2. What is the question, and where would the data come from?
3. Do the numbers mean what you think they mean?
4. What comparisons can we use to establish causal claims?
5. **What variation can you exploit, and how far does the answer travel?**

Where we left it

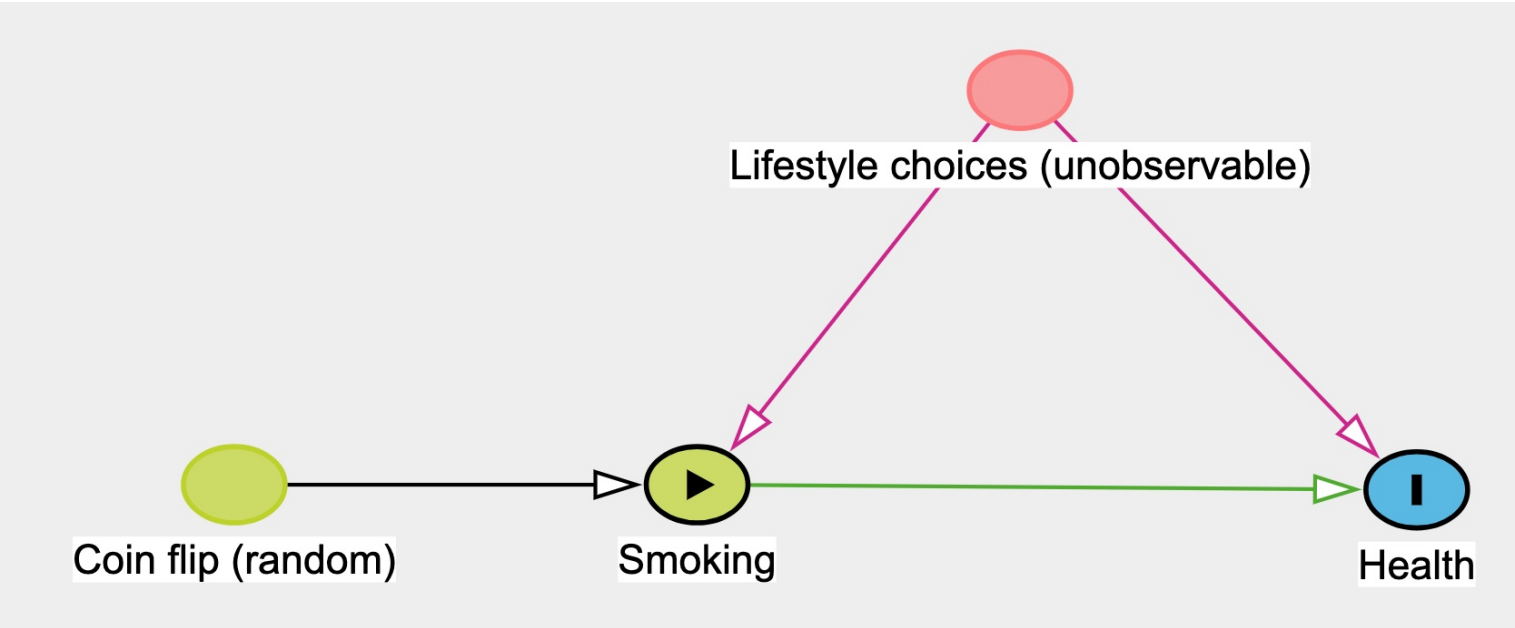
- Every causal claim is a comparison
- The effect on Gary needs a stand-in for Gary
- Controlling closes back doors, **if** you can observe what is behind them

Where we left it

- Every causal claim is a comparison
- The effect on Gary needs a stand-in for Gary
- Controlling closes back doors, **if** you can observe what is behind them



The experimental fix



A coin flip moves smoking, and nothing else.

The back door is still there, but it no longer matters.

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

Two ways of finding a stand-in

1. **Close the back doors**
   - → controlling, matching, fixed effects
   - → works if you can name and measure (or hold constant) the confounders

Two ways of finding a stand-in

1. **Close the back doors**
   - → controlling, matching, fixed effects
   - → works if you can name and measure (or hold constant) the confounders

2. **Find variation where they are closed by design**
   - → difference-in-differences, regression discontinuity, instrumental variables
   - → something moves the treatment for reasons unrelated to the outcome

Two ways of finding a stand-in

1. **Close the back doors**
   - → controlling, matching, fixed effects
   - → works if you can name and measure (or hold constant) the confounders

2. **Find variation where they are closed by design**
   - → difference-in-differences, regression discontinuity, instrumental variables
   - → something moves the treatment for reasons unrelated to the outcome

- In practice, most good papers do both

Back to Gary: who is his stand-in?

- A non-smoking neighbour of the same age, income, diet...
  - → **controlling, matching**

Back to Gary: who is his stand-in?

- A non-smoking neighbour of the same age, income, diet...
  → **controlling, matching**
- Gary compared with himself, or with his own family
  → **fixed effects**

Back to Gary: who is his stand-in?

- A non-smoking neighbour of the same age, income, diet...
  → **controlling, matching**
- Gary compared with himself, or with his own family
  → **fixed effects**
- Gary before he started, plus how non-smokers changed meanwhile
  → **difference-in-differences**

Back to Gary: who is his stand-in?

- A non-smoking neighbour of the same age, income, diet...
  → **controlling, matching**
- Gary compared with himself, or with his own family
  → **fixed effects**
- Gary before he started, plus how non-smokers changed meanwhile
  → **difference-in-differences**
- A Gary just old enough to buy cigarettes, or nudged into smoking by chance
  → **regression discontinuity, instrumental variables**

Three questions for every design

1. What **variation** does it use?
2. What has to be **true** for that variation to be as good as random?
3. How could you **check**, at least partly?

No estimation today. That is for the causal inference course.

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

Controlling, again

- When you control for Z, you compare units with the **same value of** Z
  - $\rightarrow$ (Technically: you remove variation due to the other variable)
- Subsetting is controlling too
  - $\rightarrow$ studying education $\rightarrow$ income only among people raised in big cities holds urban/rural constant

Controlling, again

- When you control for Z, you compare units with the **same value of** Z
  - $\rightarrow$ (Technically: you remove variation due to the other variable)
- Subsetting is controlling too
  - $\rightarrow$ studying education $\rightarrow$ income only among people raised in big cities holds urban/rural constant
- **Matching** pushes this idea further
  - $\rightarrow$ for each treated unit, find untreated units that look the same on Z
  - $\rightarrow$ compare only within those matched pairs

Two ways to match

- **Distance matching**: pair units with similar values on each Z
- **Propensity score**: estimate the probability of being treated given Z, then match on that one number
- Usually you match controls to treated units $\rightarrow$ the **ATT**

Two ways to match

- **Distance matching**: pair units with similar values on each Z
- **Propensity score**: estimate the probability of being treated given Z, then match on that one number
- Usually you match controls to treated units $\rightarrow$ the **ATT**
- Treated units with no comparable control are **dropped**
  - $\rightarrow$ the estimate is now about a different population

What matching cannot do

- Matching is **selection on observables**, just like regression

What matching cannot do

- Matching is **selection on observables**, just like regression
- If lifestyle is unobserved, two matched "identical" people still differ in lifestyle

What matching cannot do

- Matching is **selection on observables**, just like regression
- If lifestyle is unobserved, two matched "identical" people still differ in lifestyle
- Matching makes the comparison look clean.
  It does not make the confounder go away

Variation used: treated vs. untreated, *among units that look alike*

Must be true: **no unobserved confounders**

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

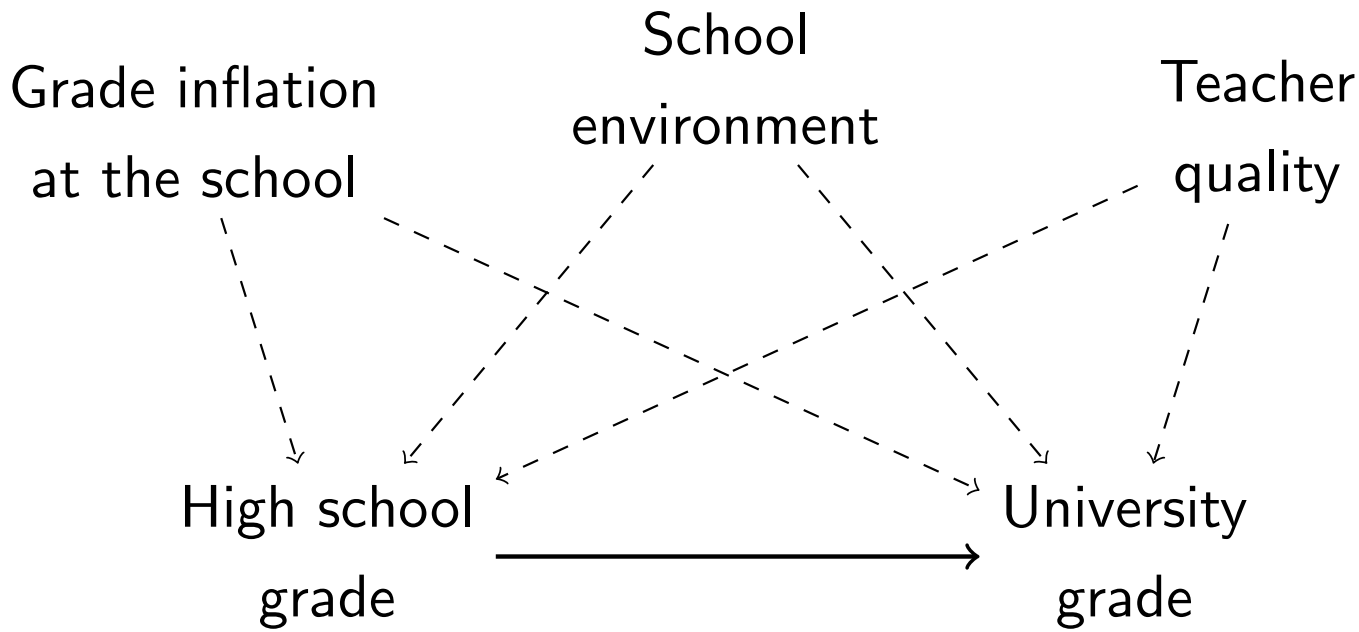
Fixed effects: compare within groups

- Some unobserved confounders are **constant within a group**
  - → country history, in a cross-national panel
  - → school, in a study of students
  - → the person, in a panel survey
  - → the user, in repeated platform data

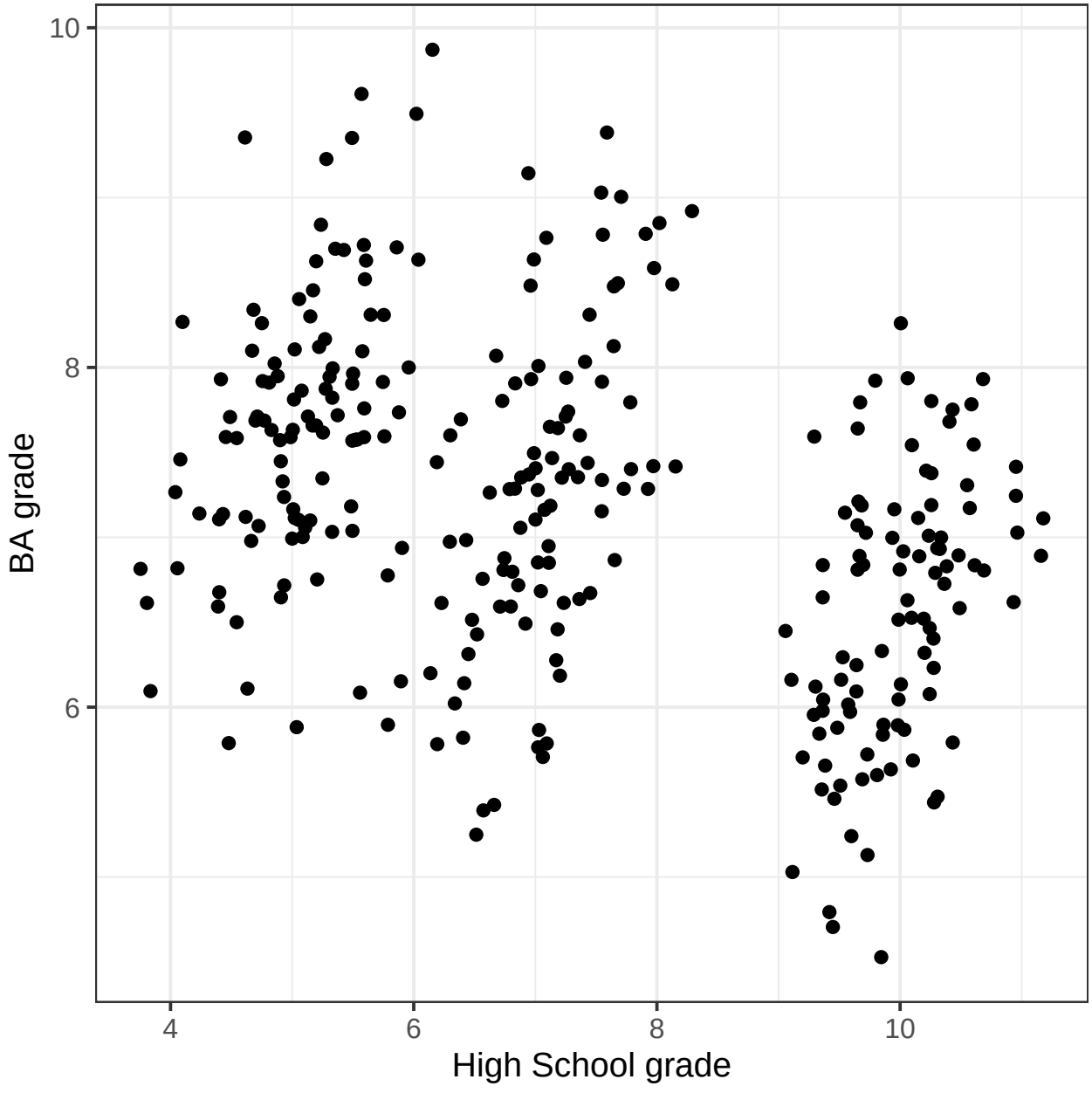
Fixed effects: compare within groups

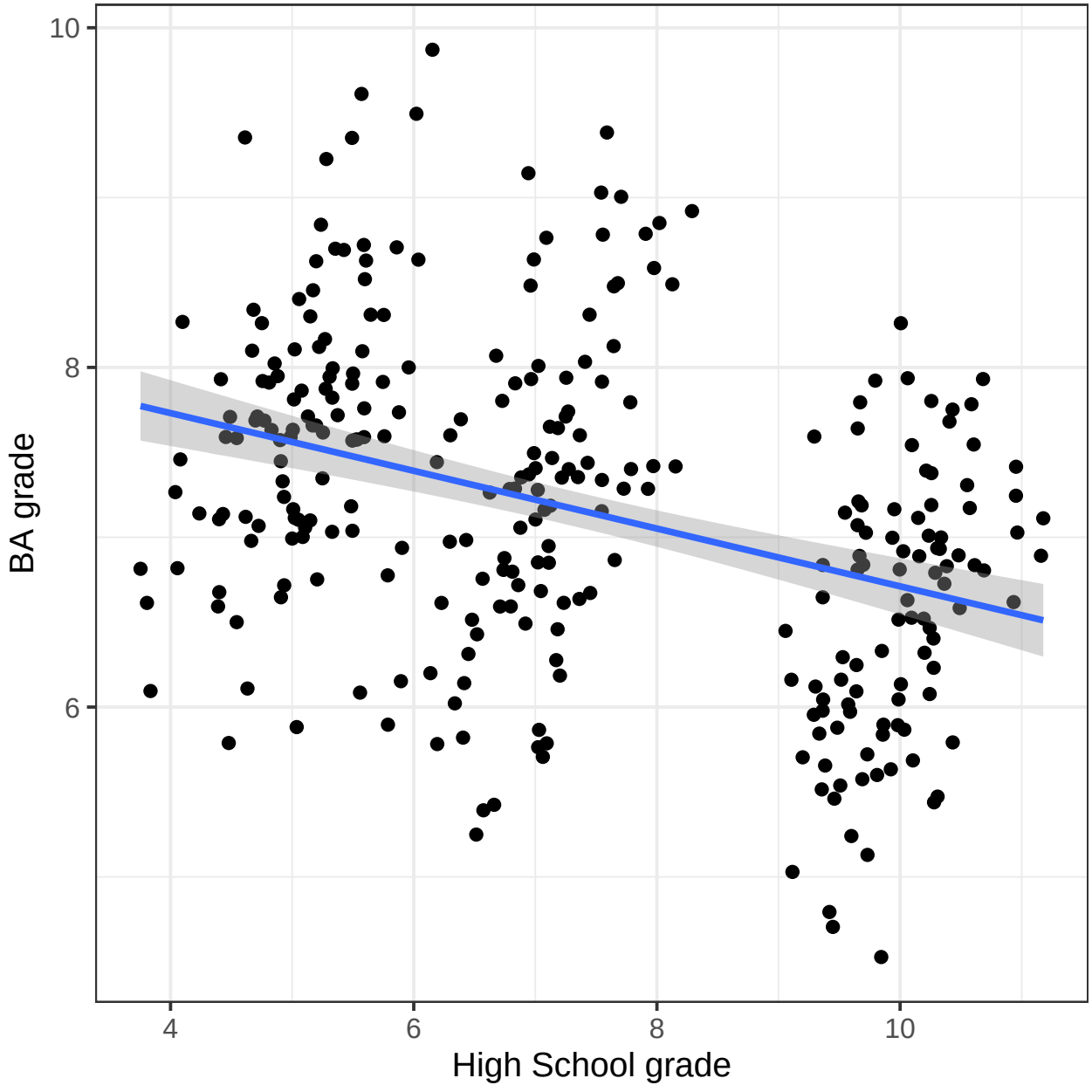
- Some unobserved confounders are **constant within a group**
  - → country history, in a cross-national panel
  - → school, in a study of students
  - → the person, in a panel survey
  - → the user, in repeated platform data
- Then compare only **within** each group
  - → every group-level confounder is held constant, *observed or not*

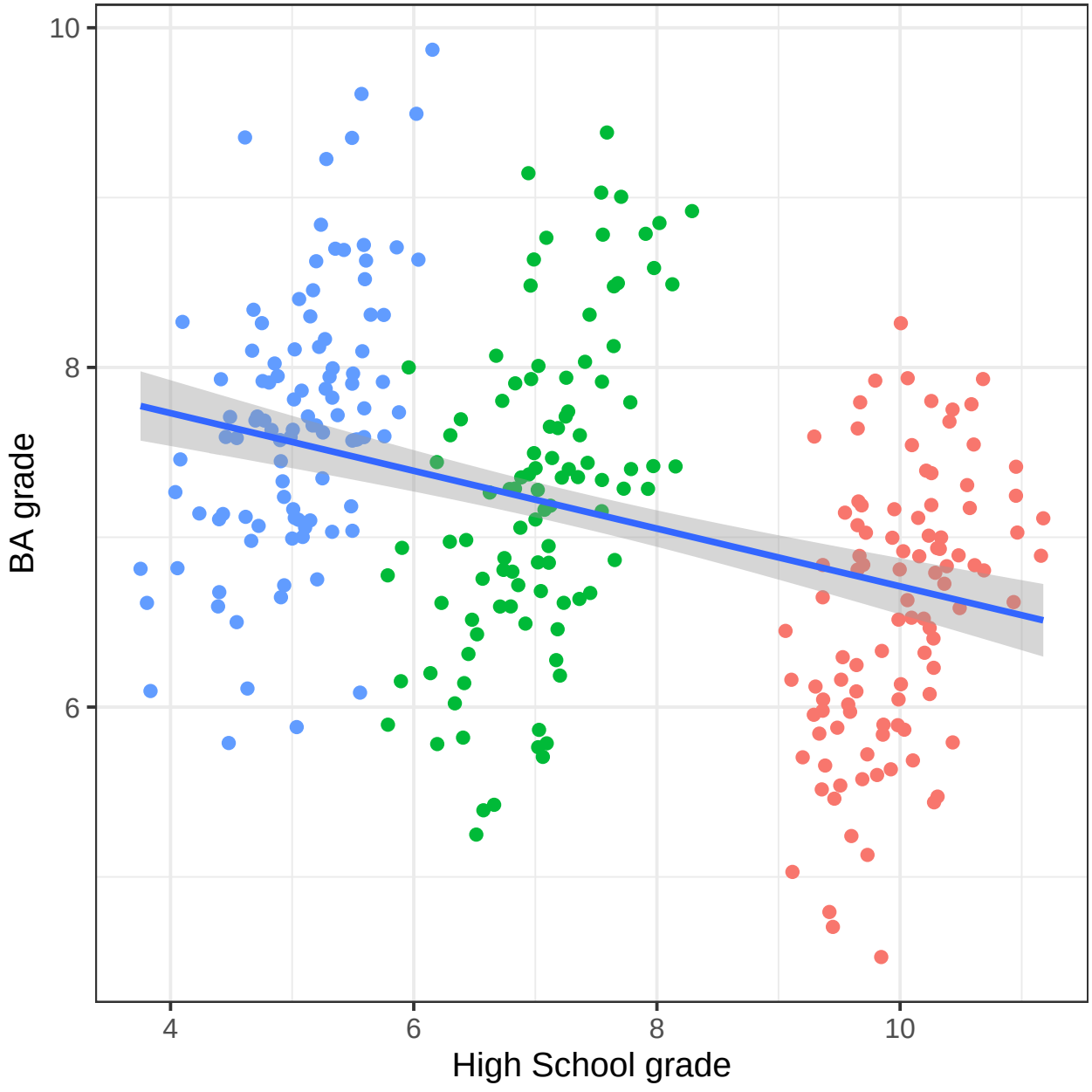
Example: high school and university grades

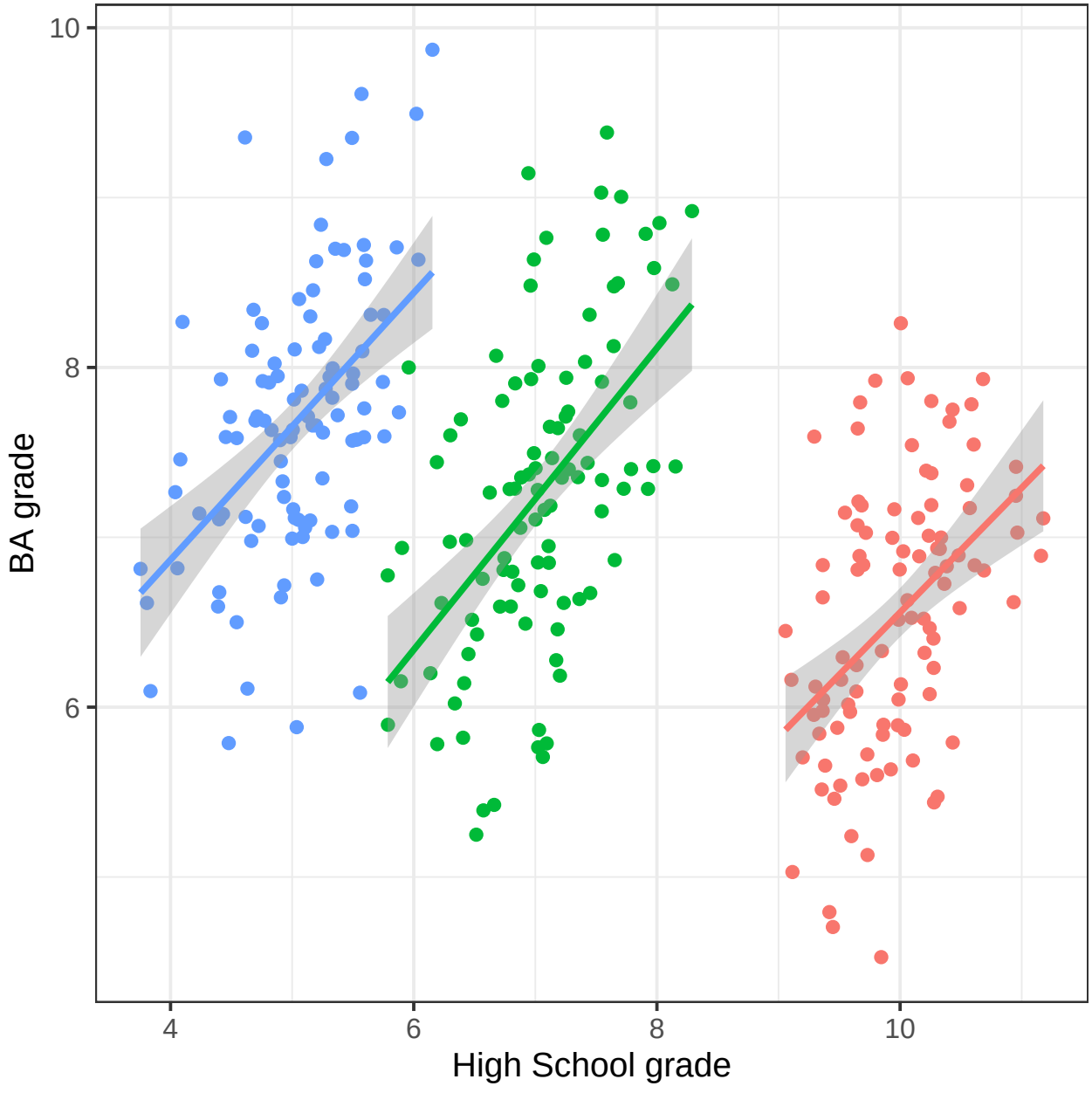


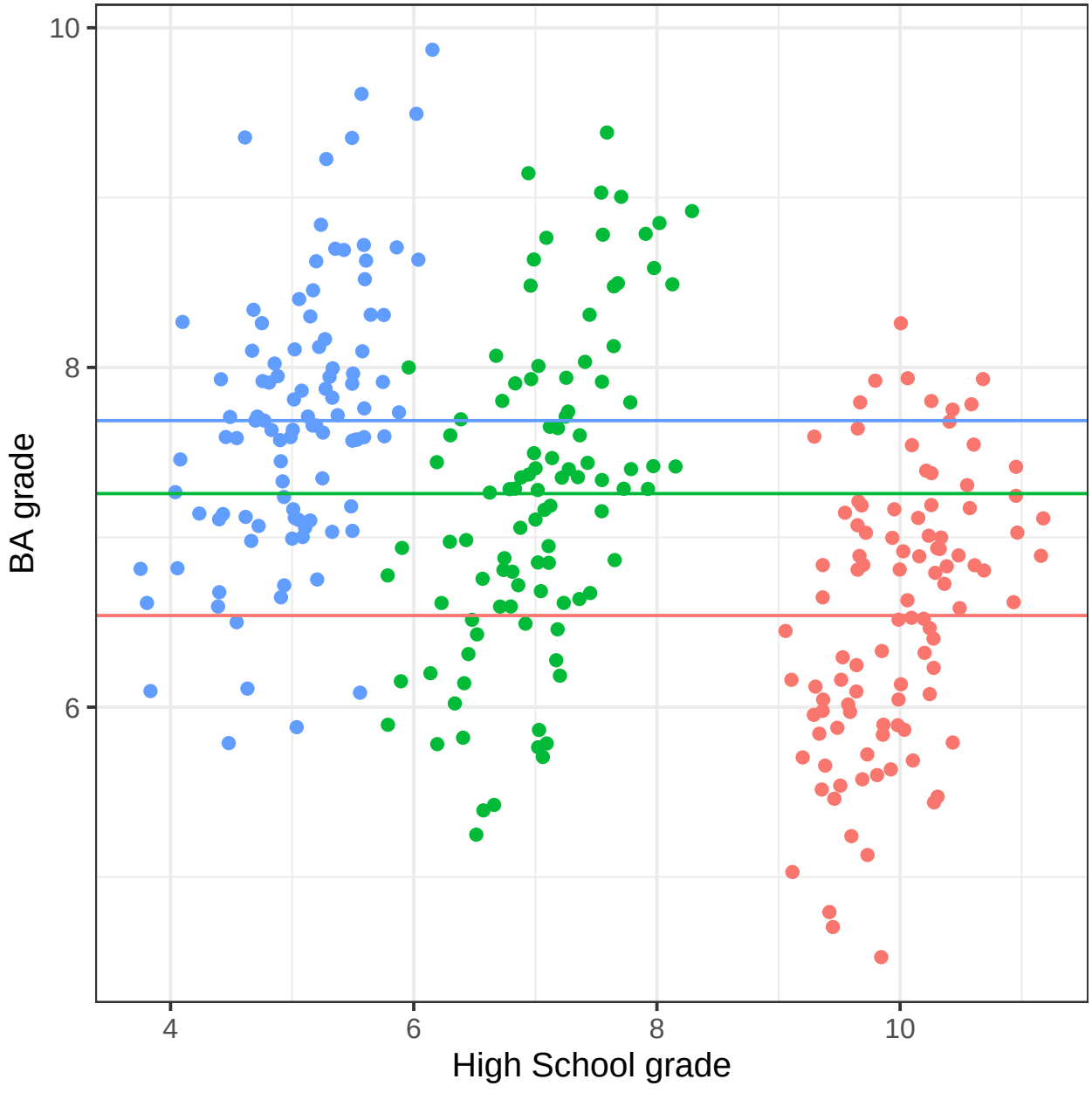
All the dashed arrows come from the **school**. We can't measure them, but we know which school each student went to

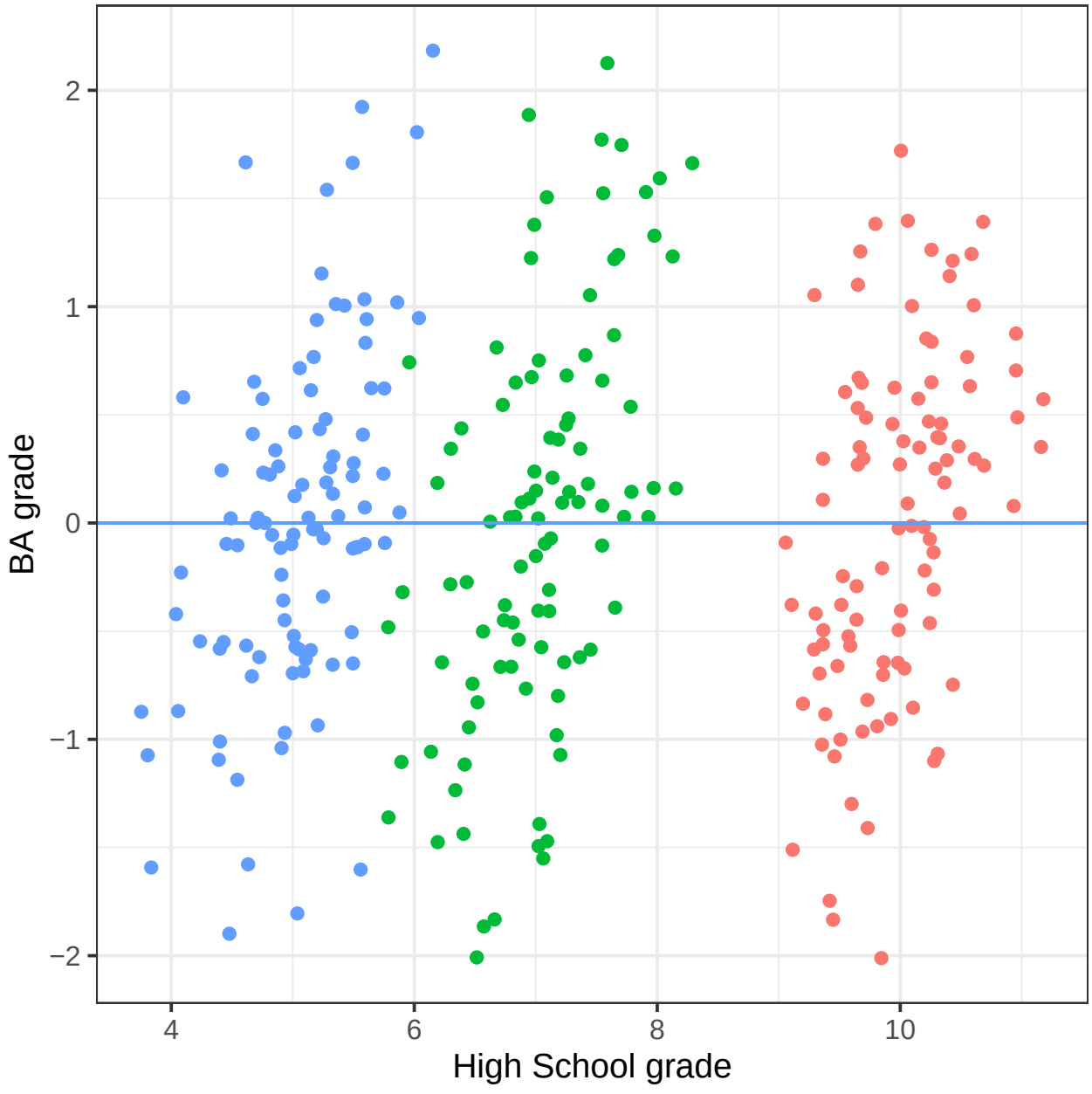

10
8
6
BA grade
4
6
8
10
High School grade

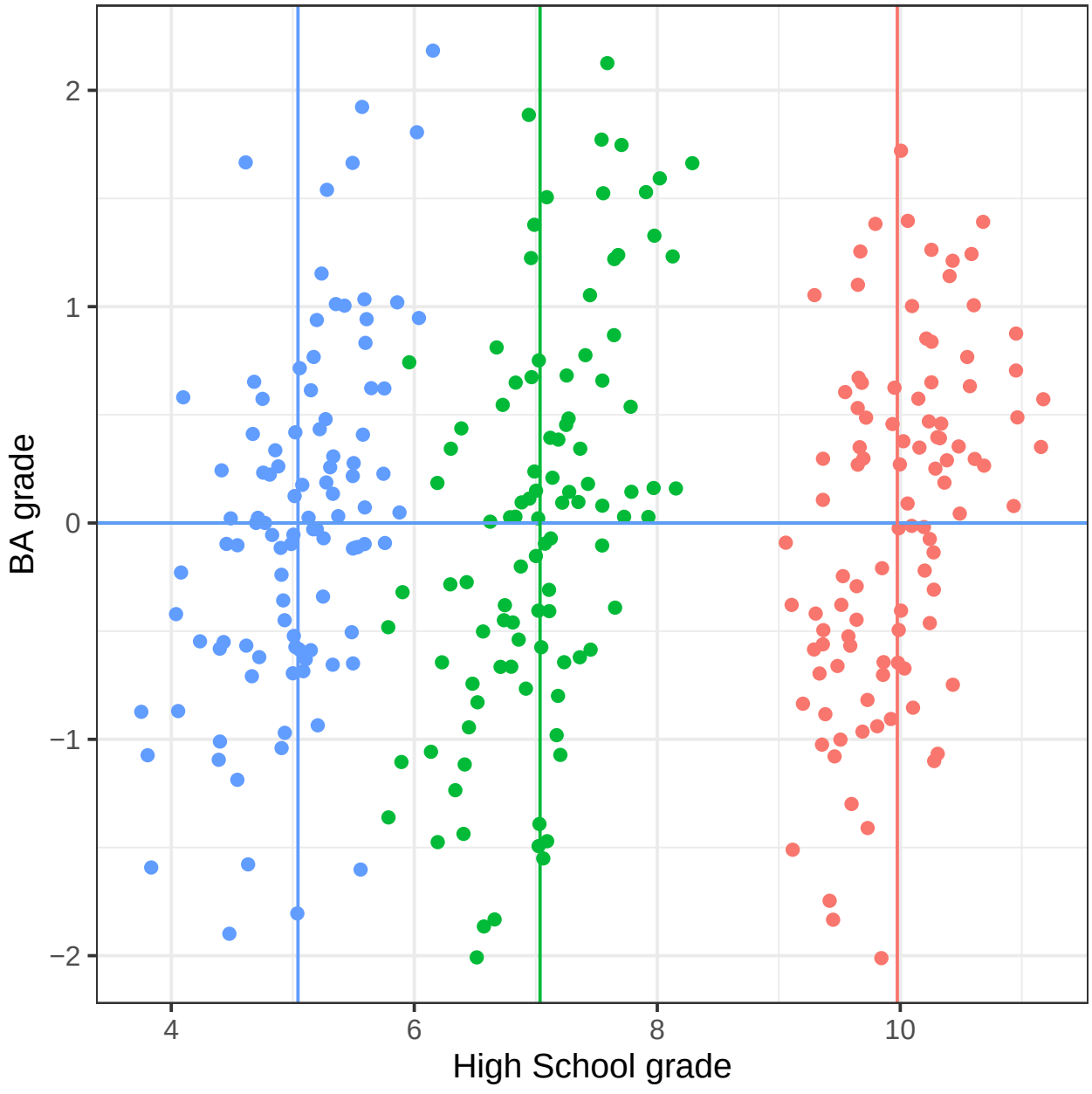

10
8
6
BA grade
4
6
8
10
High School grade

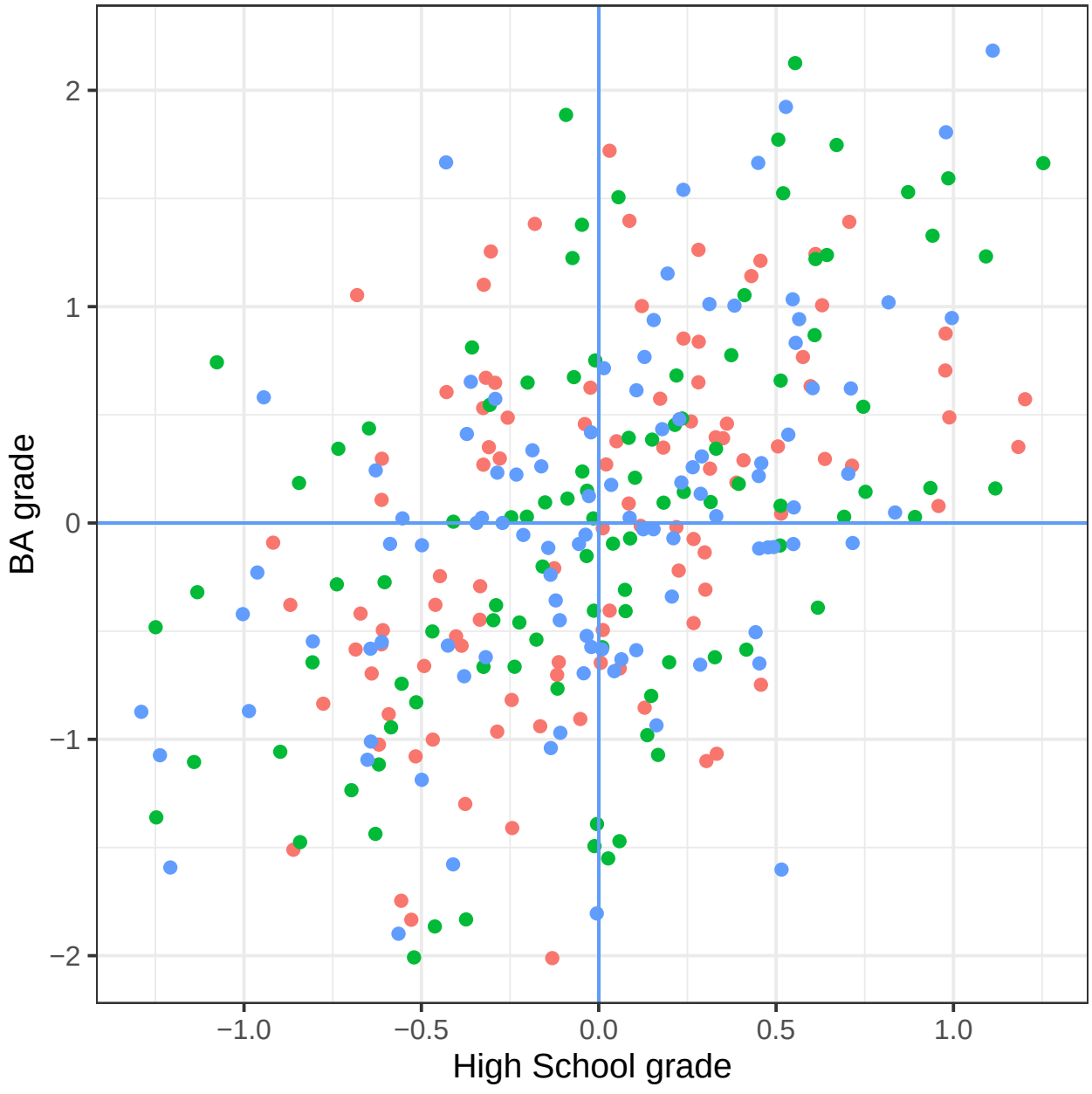

10
8
6
BA grade
4
6
8
10
High School grade

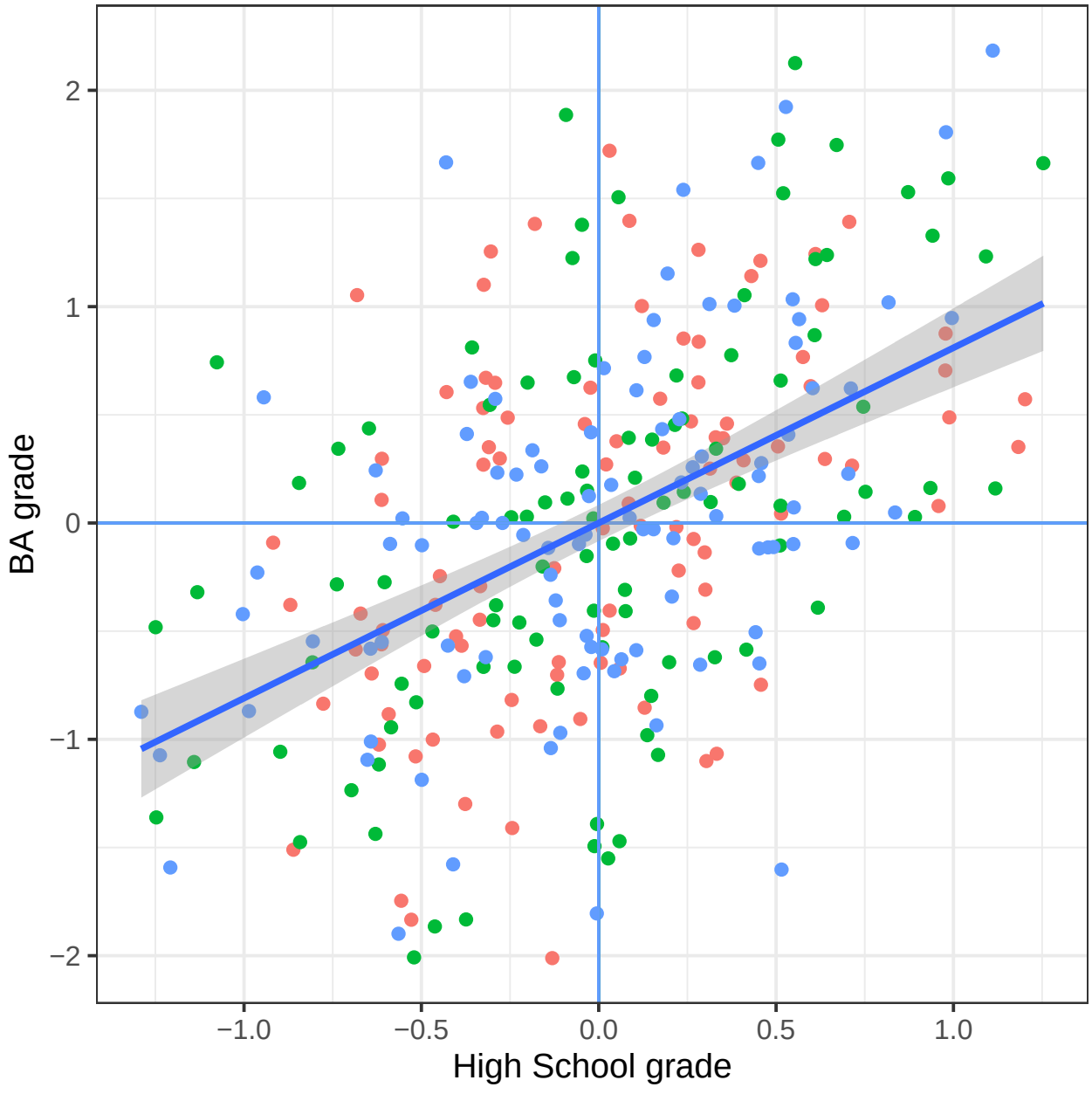

10
8
6
BA grade
4
6
8
10
High School grade


10
8
6
BA grade
4
6
8
10
High School grade


BA grade
2
1
0
−1
−2
4
6
8
10
High School grade


BA grade
2
1
0
−1
−2
4
6
8
10
High School grade


BA grade
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
High School grade


BA grade
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
High School grade

What fixed effects cannot do

- Confounders that **change within the group**

What fixed effects cannot do

- Confounders that **change within the group**
  - $\rightarrow$ a new head teacher, a policy change mid-panel

Variation used: differences *within* groups

Must be true: **no confounders that vary within groups**

What fixed effects cannot do

- Confounders that **change within the group**
  - $\rightarrow$ a new head teacher, a policy change mid-panel
- Treatments that **do not vary within the group**

Variation used: differences *within* groups

Must be true: **no confounders that vary within groups**

What fixed effects cannot do

- Confounders that **change within the group**
  - → a new head teacher, a policy change mid-panel
- Treatments that **do not vary within the group**
  - → all between-group variation is thrown away

Variation used: differences *within* groups

Must be true: **no confounders that vary within groups**

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

Something changes at a moment in time

- A minimum wage increase, a terrorist attack, a platform changes its algorithm, a town renames its streets

Something changes at a moment in time

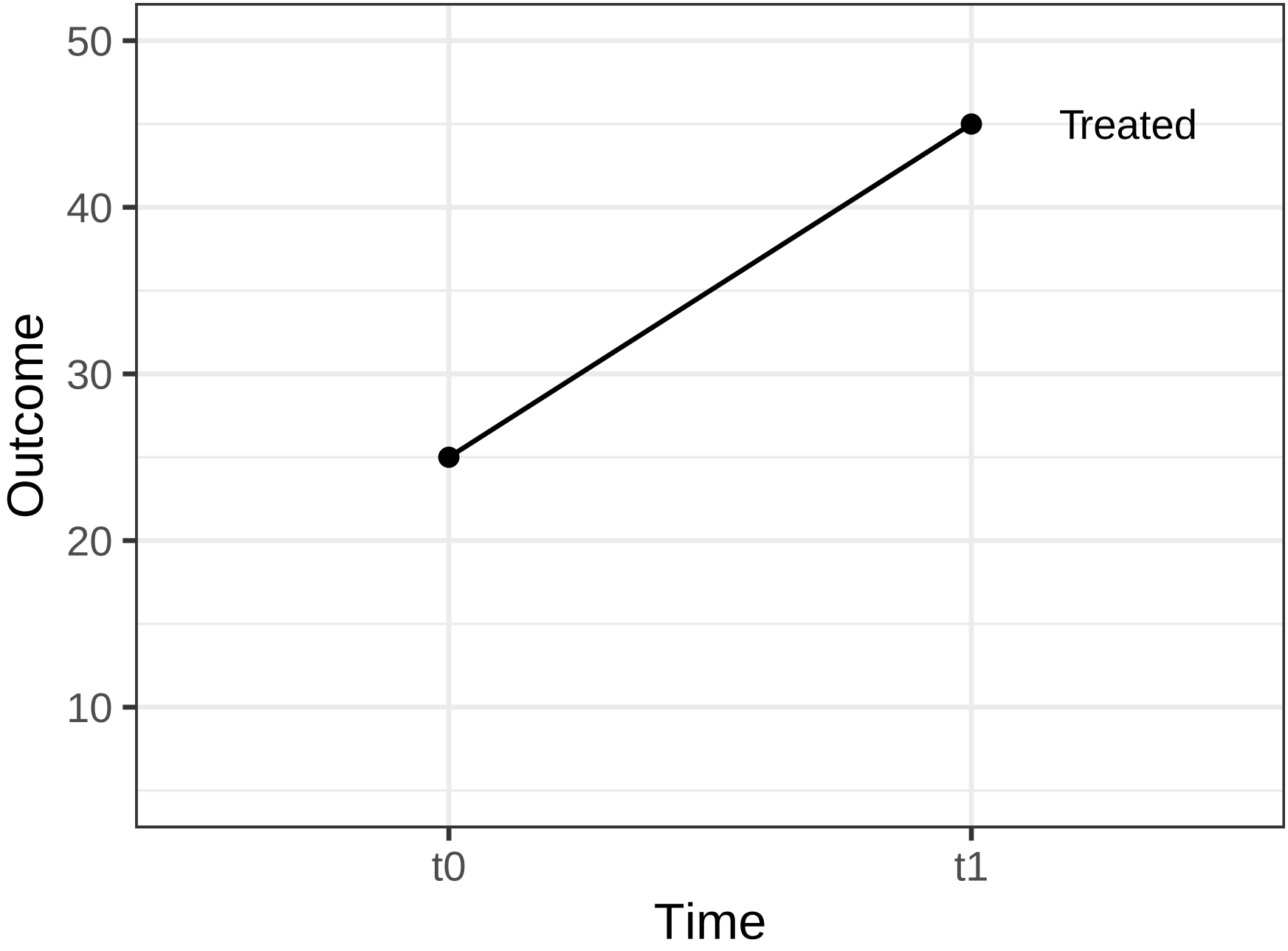
- A minimum wage increase, a terrorist attack, a platform changes its algorithm, a town renames its streets
- Naive comparison 1: **before vs. after**, in treated units
  - $\rightarrow$ but other things changed over time too

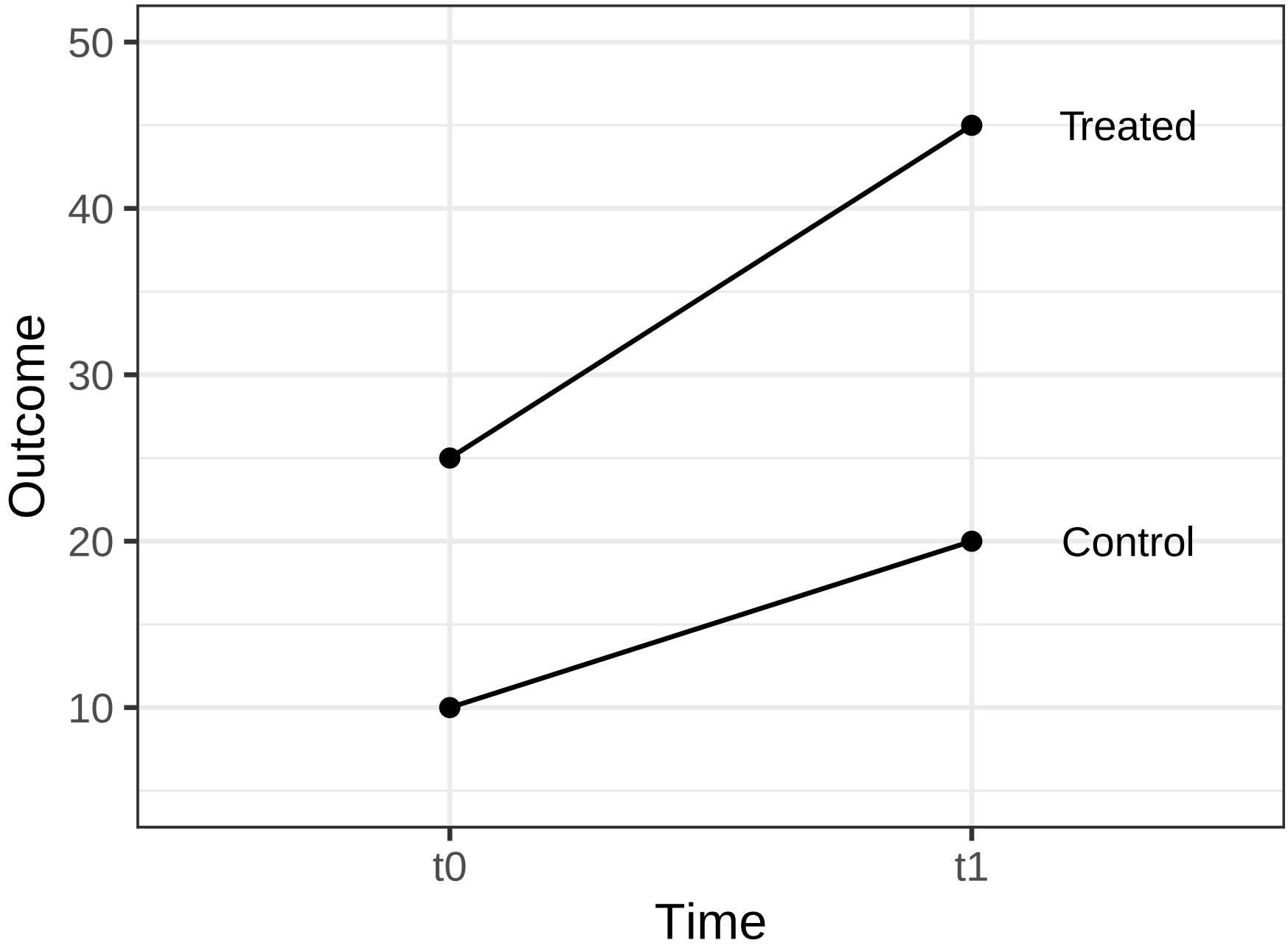
Something changes at a moment in time

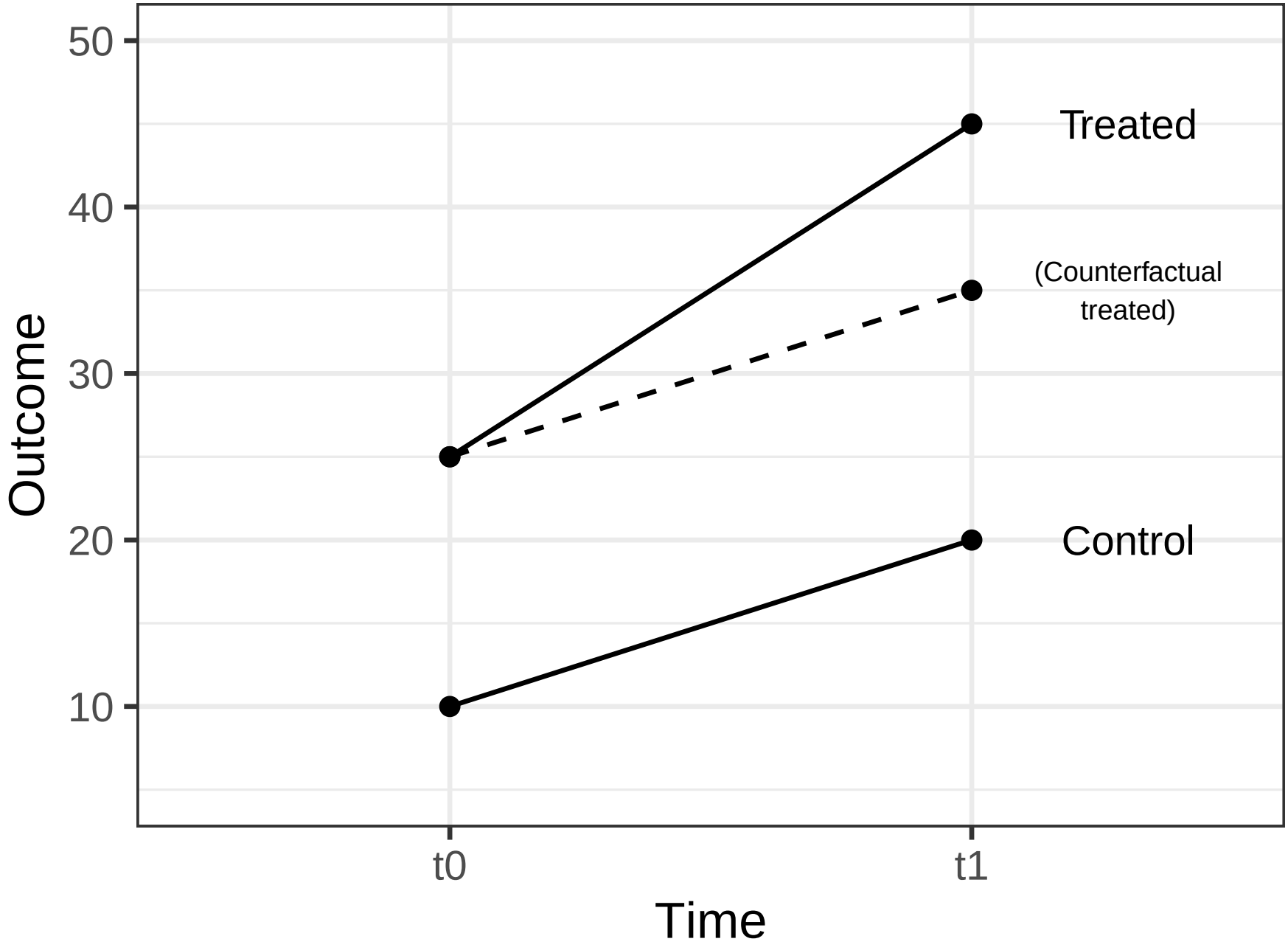
- A minimum wage increase, a terrorist attack, a platform changes its algorithm, a town renames its streets
- Naive comparison 1: **before vs. after**, in treated units
  - → but other things changed over time too
- Naive comparison 2: **treated vs. untreated**, after
  - → but they were different to begin with

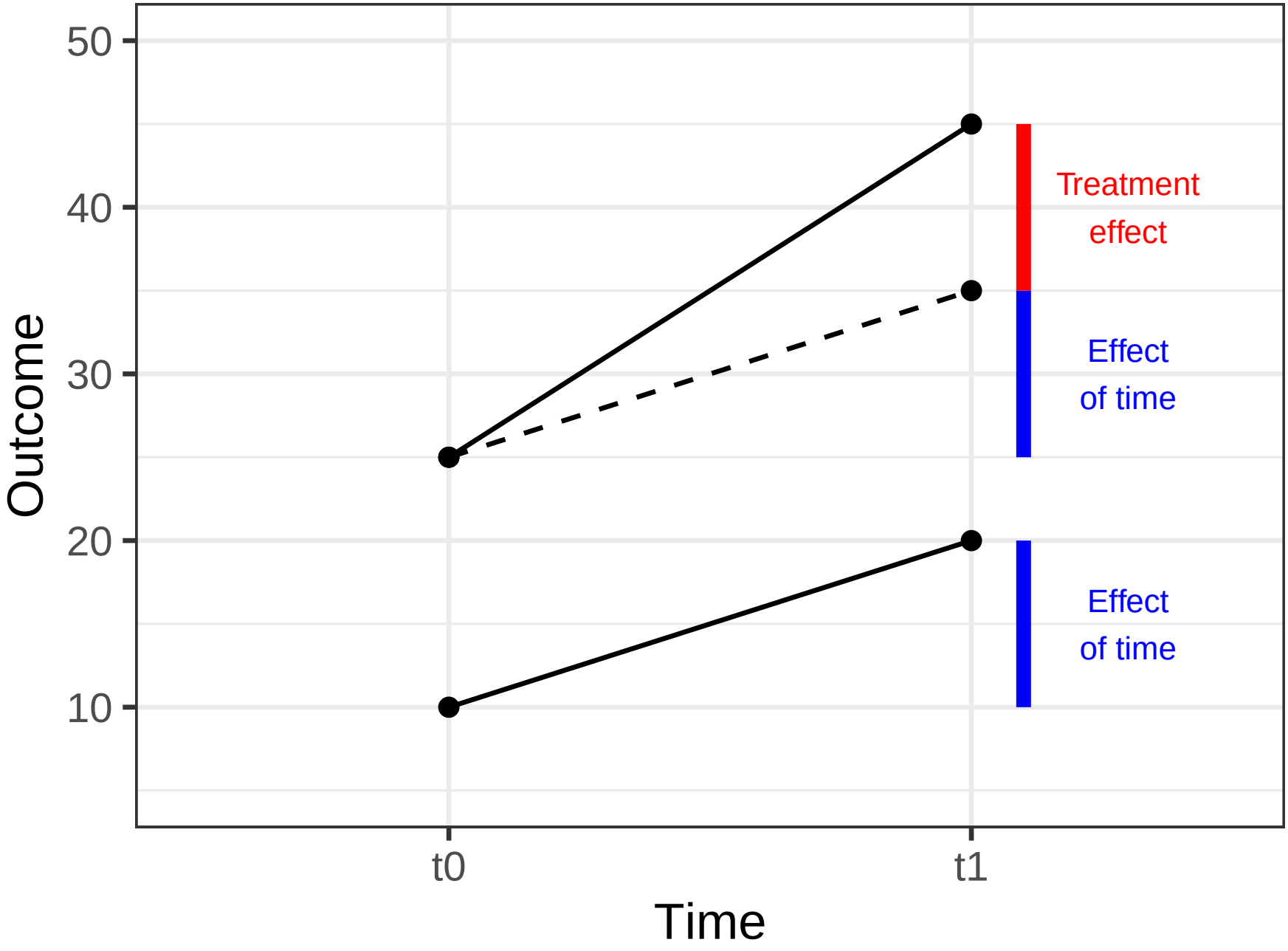
Something changes at a moment in time

- A minimum wage increase, a terrorist attack, a platform changes its algorithm, a town renames its streets
- Naive comparison 1: **before vs. after**, in treated units
  - → but other things changed over time too
- Naive comparison 2: **treated vs. untreated**, after
  - → but they were different to begin with
- **DiD**: use the untreated group's change over time as the counterfactual for the treated


50
40
30
20
10
Outcome
Treated
t0
t1
Time


50
40
30
20
10
Outcome
Treated
Control
t0
t1
Time


50
40
30
20
10
Outcome
Treated
(Counterfactual
treated)
Control
t0
t1
Time


50
40
30
20
10
Outcome
t0
t1
Time
Treatment
effect
Effect
of time
Effect
of time

The first DiD? Cholera in London, 1850s



- The Lambeth company moved its water intake upriver in 1852. Southwark & Vauxhall did not
- John Snow compared cholera deaths in their customers' homes before and after

The key assumption: parallel trends

- Without treatment, both groups would have **changed in the same way**
- Not testable after treatment (that's the counterfactual). But partly checkable:

The key assumption: parallel trends

- Without treatment, both groups would have **changed in the same way**
- Not testable after treatment (that's the counterfactual). But partly checkable:
  - → look at **pre-treatment trends**: were they parallel before?

The key assumption: parallel trends

- Without treatment, both groups would have **changed in the same way**
- Not testable after treatment (that's the counterfactual). But partly checkable:
  - → look at **pre-treatment trends**: were they parallel before?
  - → ask **why** these units were treated now

The key assumption: parallel trends

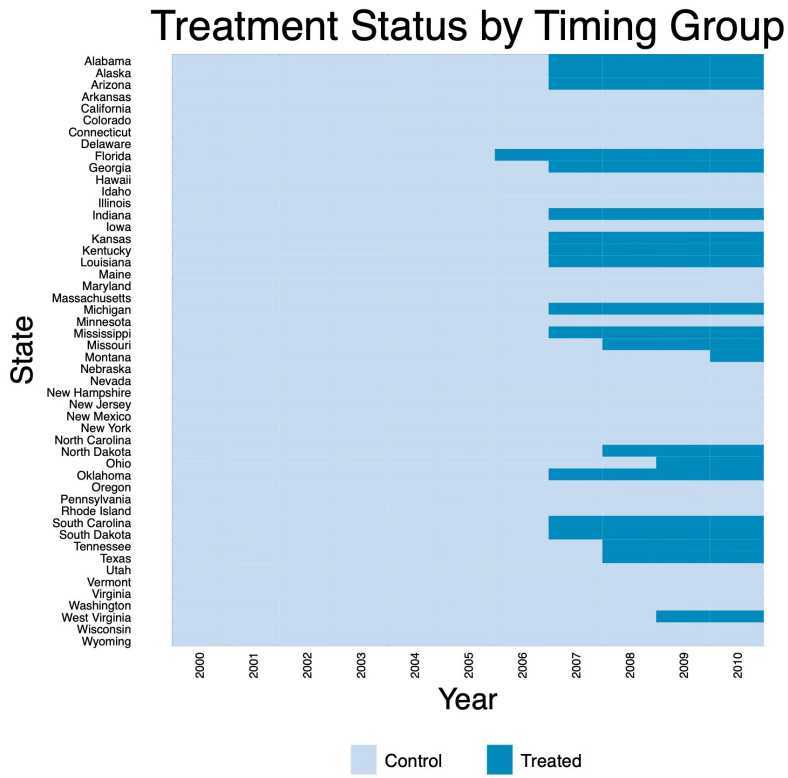
- Without treatment, both groups would have **changed in the same way**
- Not testable after treatment (that's the counterfactual). But partly checkable:
  - → look at **pre-treatment trends**: were they parallel before?
  - → ask **why** these units were treated now
- Threats: anticipation, other changes at the same time, selection into treatment based on trends, **spillovers** to the control group

Variation used: changes over time, treated vs. untreated

Must be true: **parallel trends**

When units are treated at different times

- **Staggered** adoption is the normal case
- Standard two-way fixed effects can then compare late-treated units with early-treated ones
- If effects change over time, that comparison is biased
- See Goodman-Bacon (2021), Callaway & Sant'Anna (2021), Roth et al. (2023)



Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

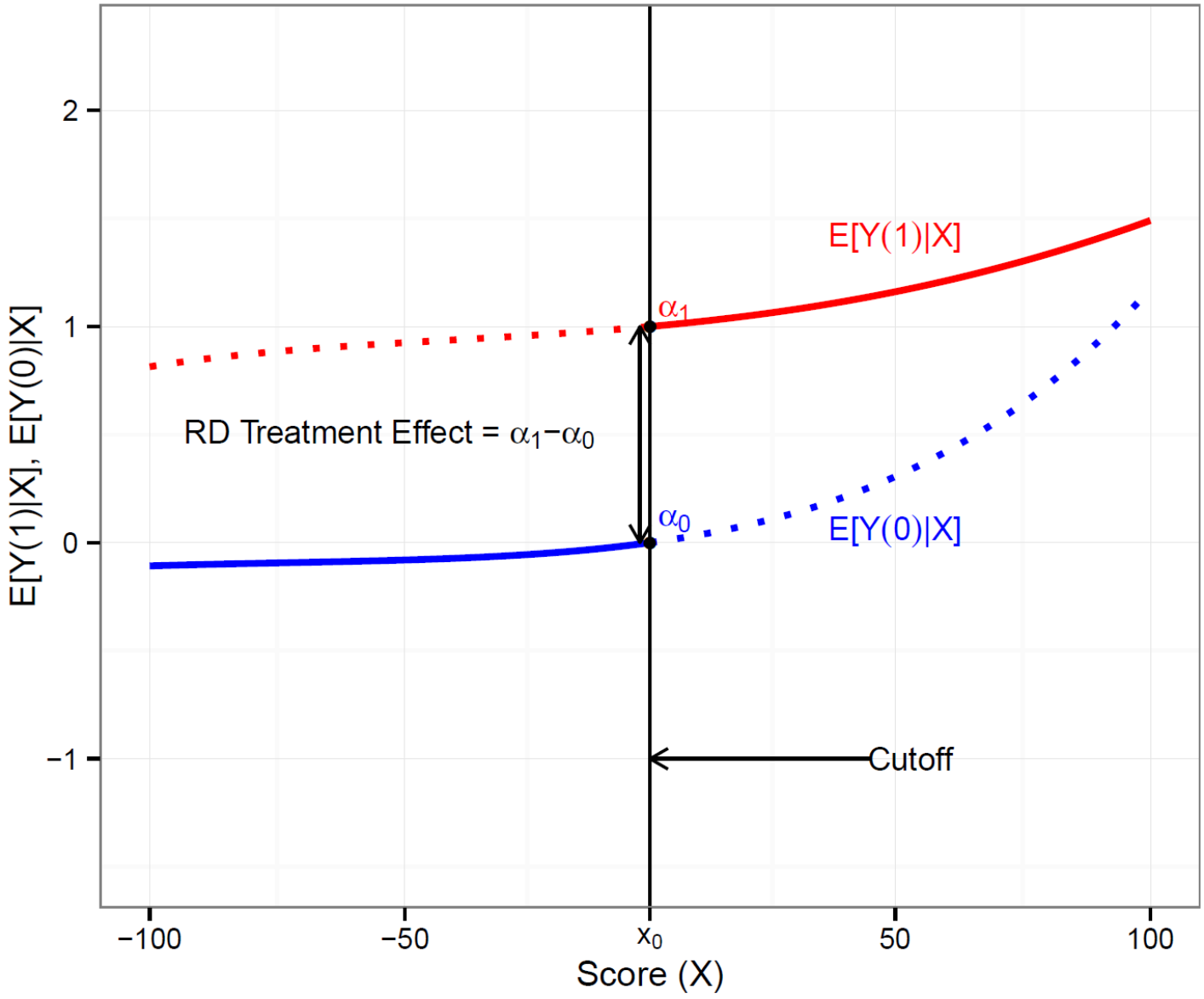
Treatment assigned by a cutoff

- Treatment switches on when a **running variable** crosses a threshold
  - → winning an election (vote share above 50%)
  - → a scholarship (test score above x)
  - → living on one side of a border (distance)
  - → being old enough to vote or be drafted (birth date)

Treatment assigned by a cutoff

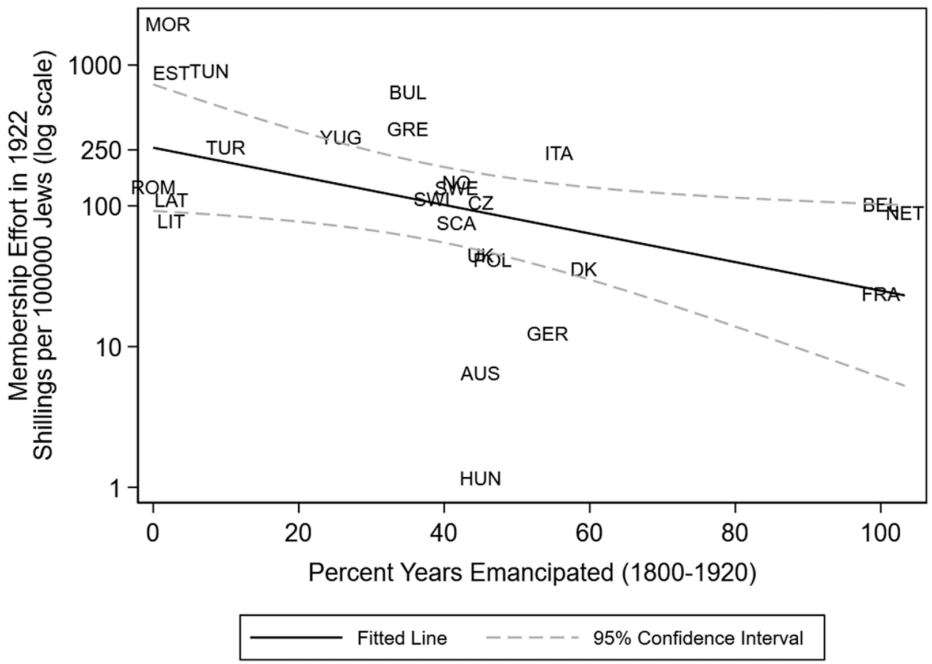
- Treatment switches on when a **running variable** crosses a threshold
  - → winning an election (vote share above 50%)
  - → a scholarship (test score above x)
  - → living on one side of a border (distance)
  - → being old enough to vote or be drafted (birth date)
- Far from the cutoff, units differ in many ways
- **Right at the cutoff**, they should be almost identical

Figure 2: RD Treatment Effect in Sharp RD Design



Example: emancipation and Jewish national identity

FIGURE 1. Emancipation and Zionism



Boix (2025), *APSR*

FIGURE 4. Location of Zionist Schools in Interwar Poland's Towns

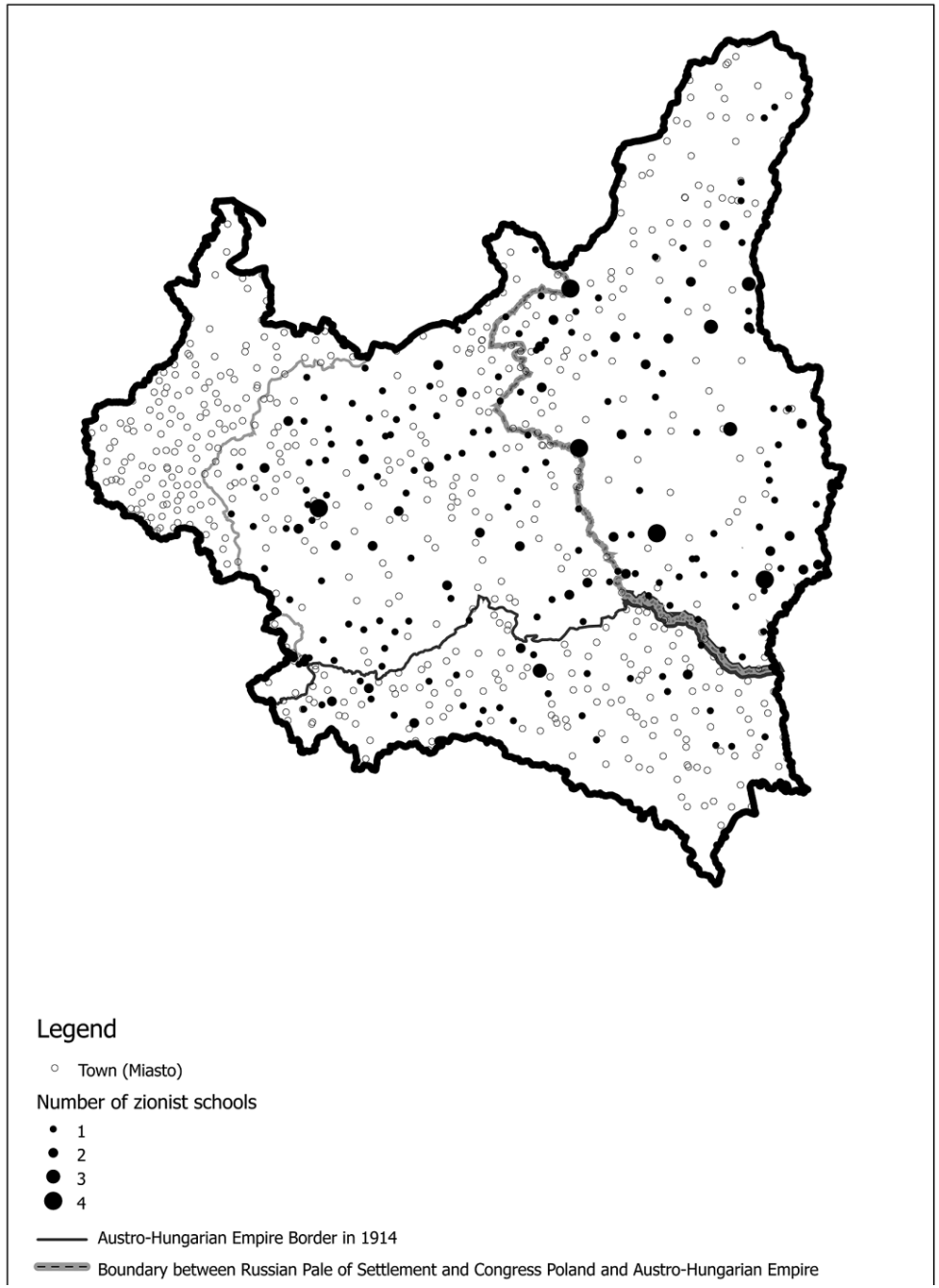
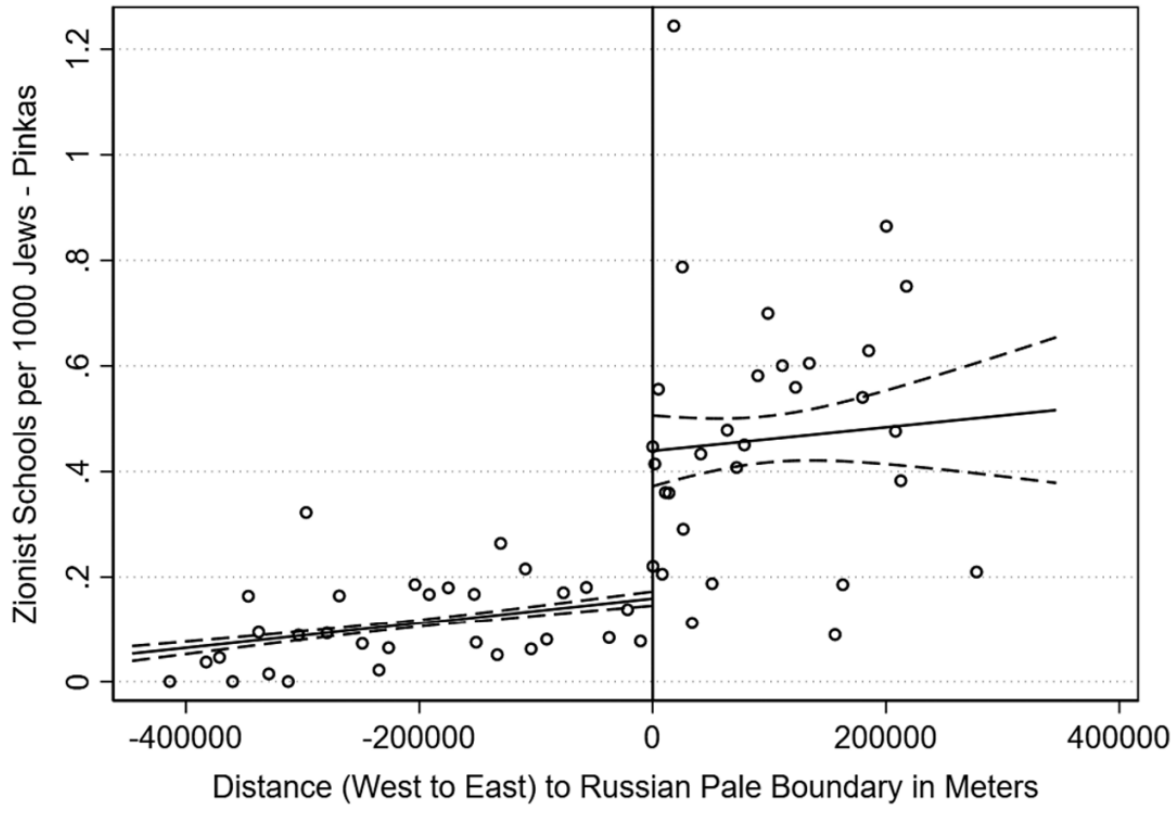
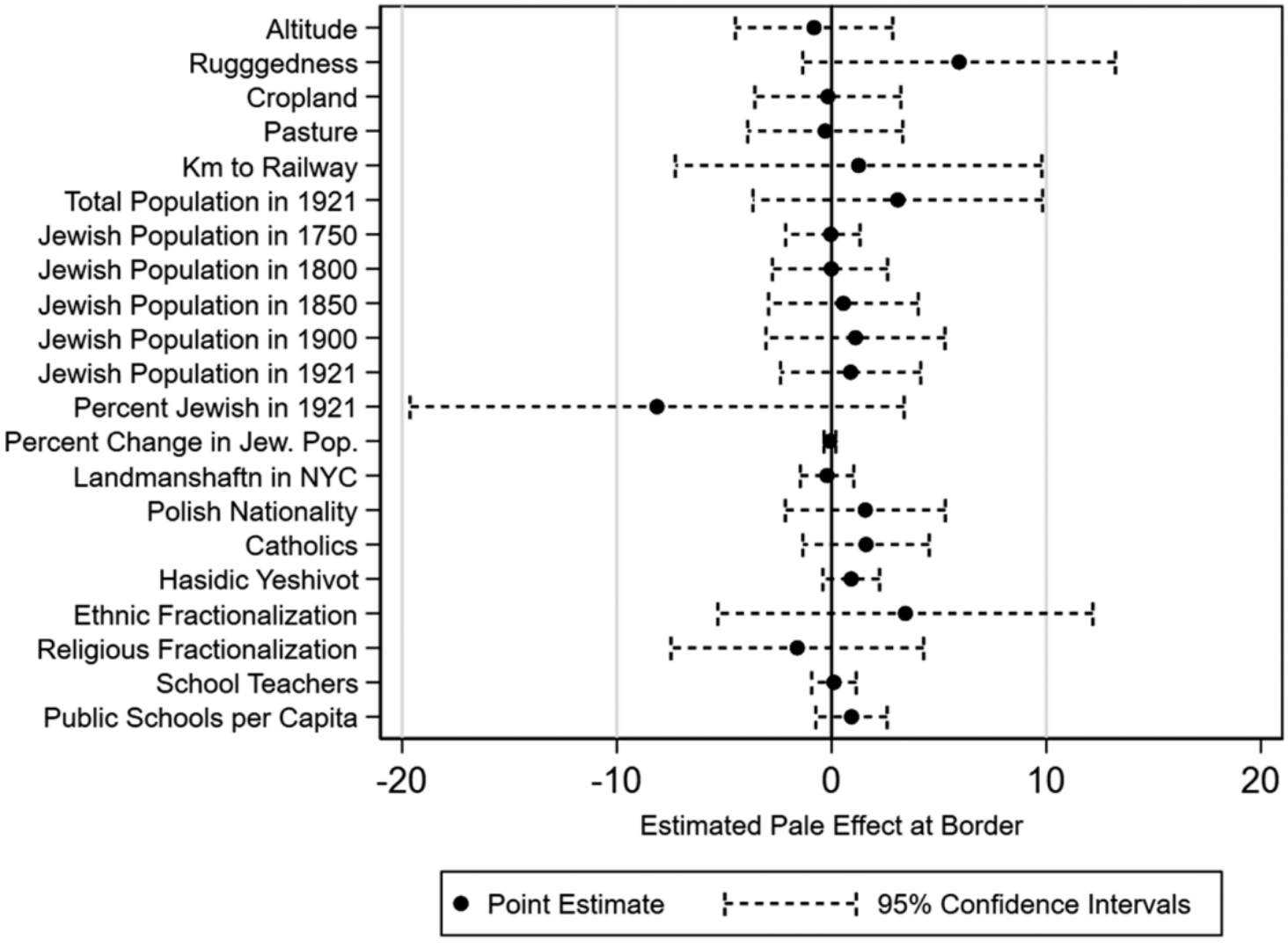


FIGURE 7. Zionist Schools in Towns



Note: Schools aggregated in bins with an equal number of observations. Source: Pinkas Ha-kehilot.

FIGURE 6. Spatial and Demographic Discontinuities at the Border



Two more cutoffs

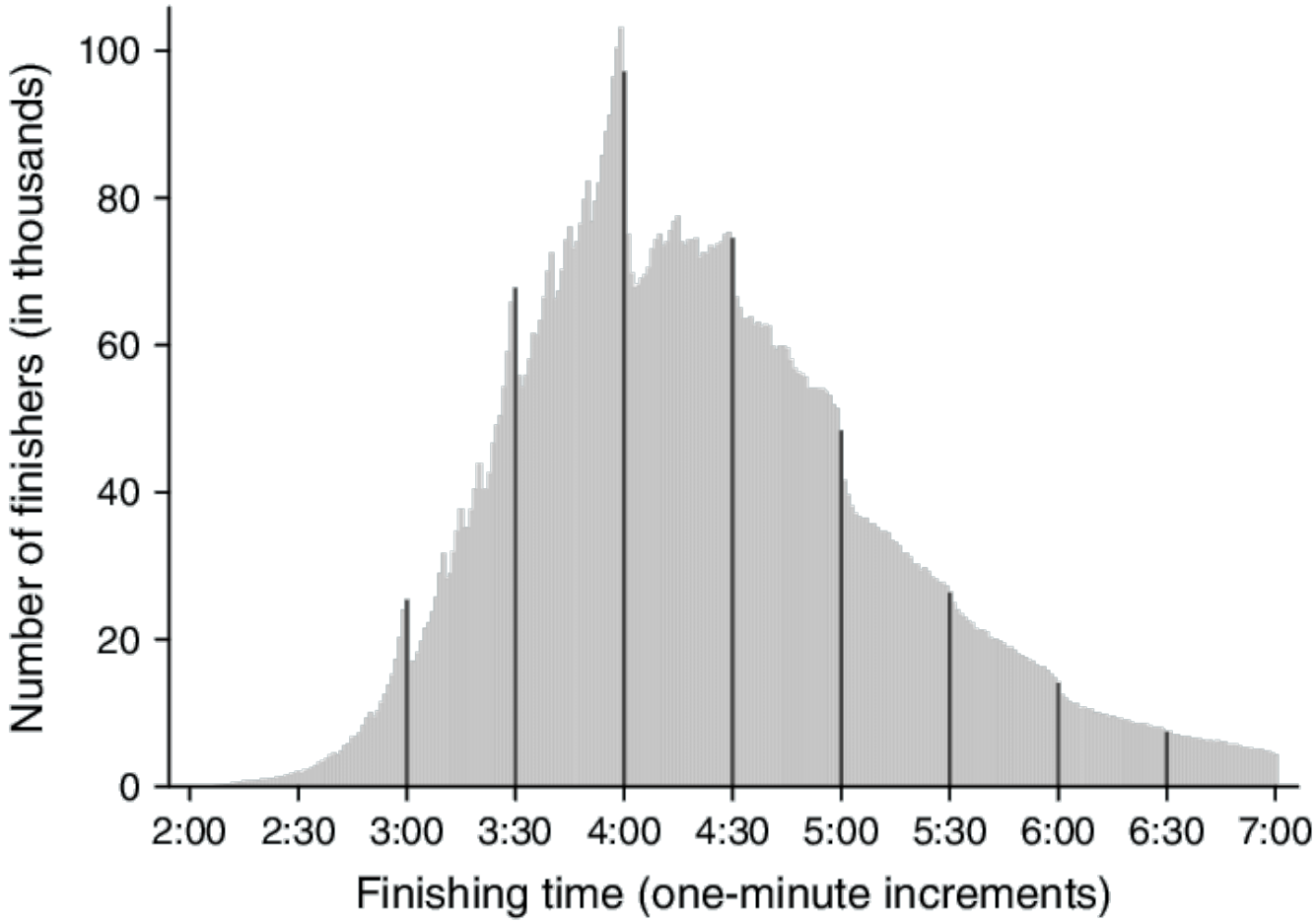
- **Military retirement age** in Burundi
  - → does serving in an ethnically integrated army reduce prejudice?
  - → soldiers just below vs. just above the age for mandatory retirement
  - → Samii (2013), *APSR*

Two more cutoffs

- **Military retirement age** in Burundi
  - → does serving in an ethnically integrated army reduce prejudice?
  - → soldiers just below vs. just above the age for mandatory retirement
  - → Samii (2013), *APSR*

- **Close elections** and the incumbency advantage
  - → does holding office help you win the next election?
  - → parties that barely won (50.1%) vs. barely lost (49.9%)
  - → Lee (2008), *Journal of Econometrics*

Threats: what if units can choose their side?



Marathon finishing times. Allen et al. (2017), *Management Science*

Are close elections really coin flips?

- **Sorting**: in postwar US House races, incumbents win close races far more often than they lose them
  - → better resources and information at the margin?
  - → Caughey & Sekhon (2011), *Political Analysis*

Are close elections really coin flips?

- **Sorting**: in postwar US House races, incumbents win close races far more often than they lose them
  - → better resources and information at the margin?
  - → Caughey & Sekhon (2011), *Political Analysis*
- But across 40,000 close races in other countries and offices, no such sorting
  - → Eggers et al. (2015), *AJPS*

Are close elections really coin flips?

- **Sorting**: in postwar US House races, incumbents win close races far more often than they lose them
  - → better resources and information at the margin?
  - → Caughey & Sekhon (2011), *Political Analysis*
- But across 40,000 close races in other countries and offices, no such sorting
  - → Eggers et al. (2015), *AJPS*
- **What is the treatment?** Winning a close race also changes the type of candidate who wins
  - → comparing winners with a trait (e.g. women, incumbents) mixes that trait with whatever let them win a close race
  - → Marshall (2024), *AJPS*

What RDD needs

- No **precise sorting** around the cutoff

What RDD needs

- No **precise sorting** around the cutoff
- Nothing else changes at the same cutoff
  - $\rightarrow$ borders separate many things at once: compound treatments

What RDD needs

- No **precise sorting** around the cutoff
- Nothing else changes at the same cutoff
  - $\rightarrow$ borders separate many things at once: compound treatments
- The estimate is **local** (LATE): it is the effect for units near the cutoff

Variation used: the jump at the cutoff

Must be true: **units just above and below are comparable**

Roadmap

Two strategies

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Putting it together

How far does it travel?

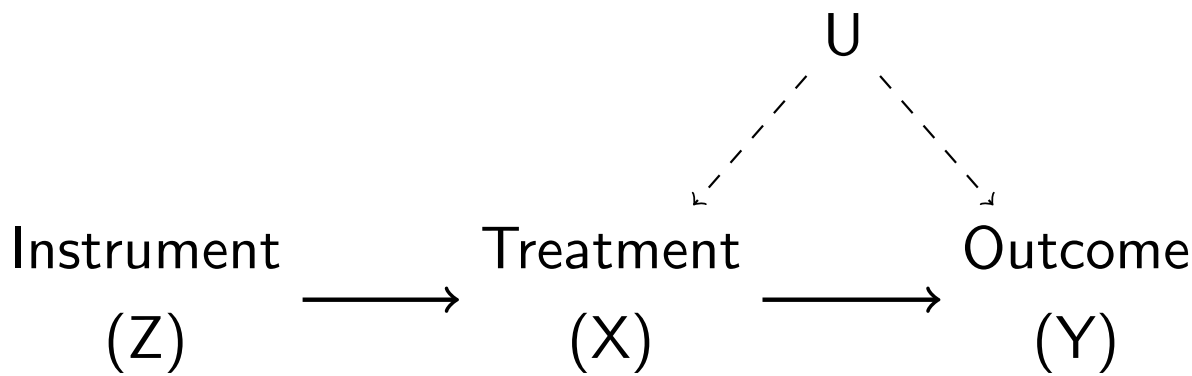
Keeping yourself honest

Ethics for computational work

Wrapping up

Memo 5, workshop and essay

Find something that moves the treatment



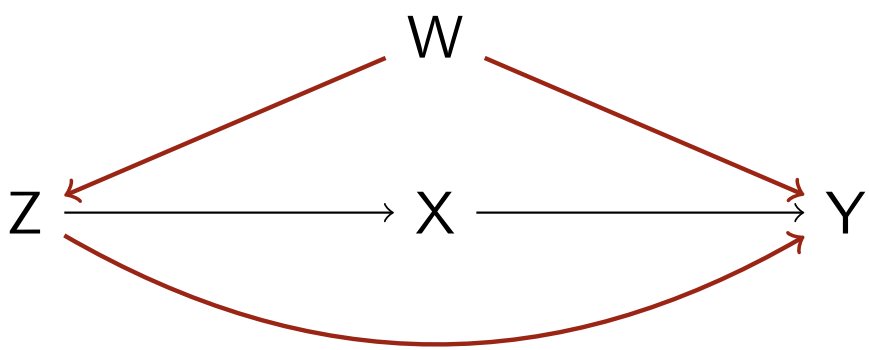
- A natural experiment: the coin flip, found in the wild
- Use only the part of X that is driven by Z

Two assumptions

- **Relevance**: Z actually moves X
  - $\rightarrow$ checkable: look at the first stage

Two assumptions

- **Relevance**: Z actually moves X
  - $\rightarrow$ checkable: look at the first stage
- **Exclusion restriction**: Z affects Y **only** through X
  - $\rightarrow$ no back doors from Z to Y, and no other paths
  - $\rightarrow$ **not testable**. An argument, not a statistic



Example: are ethnic riots self-perpetuating?

Studying whether ethnic riots are self-perpetuating is challenging given that the same underlying conditions that ignited riots initially may persist and therefore also cause subsequent riots, creating a classic confounding problem (Collier and Sambanis 2002; Hazlett 2019).[1] Given our imperfect knowledge of what all the relevant confounding factors are, and how to measure them sufficiently precisely, and given the fact that randomly assigning ethnic violence ourselves is both unethical and impossible, causal identification has to effectively come from a credible quasi-experiment, which accounts for confounding by design.

van Noort & Goyal (2025), *JCR*

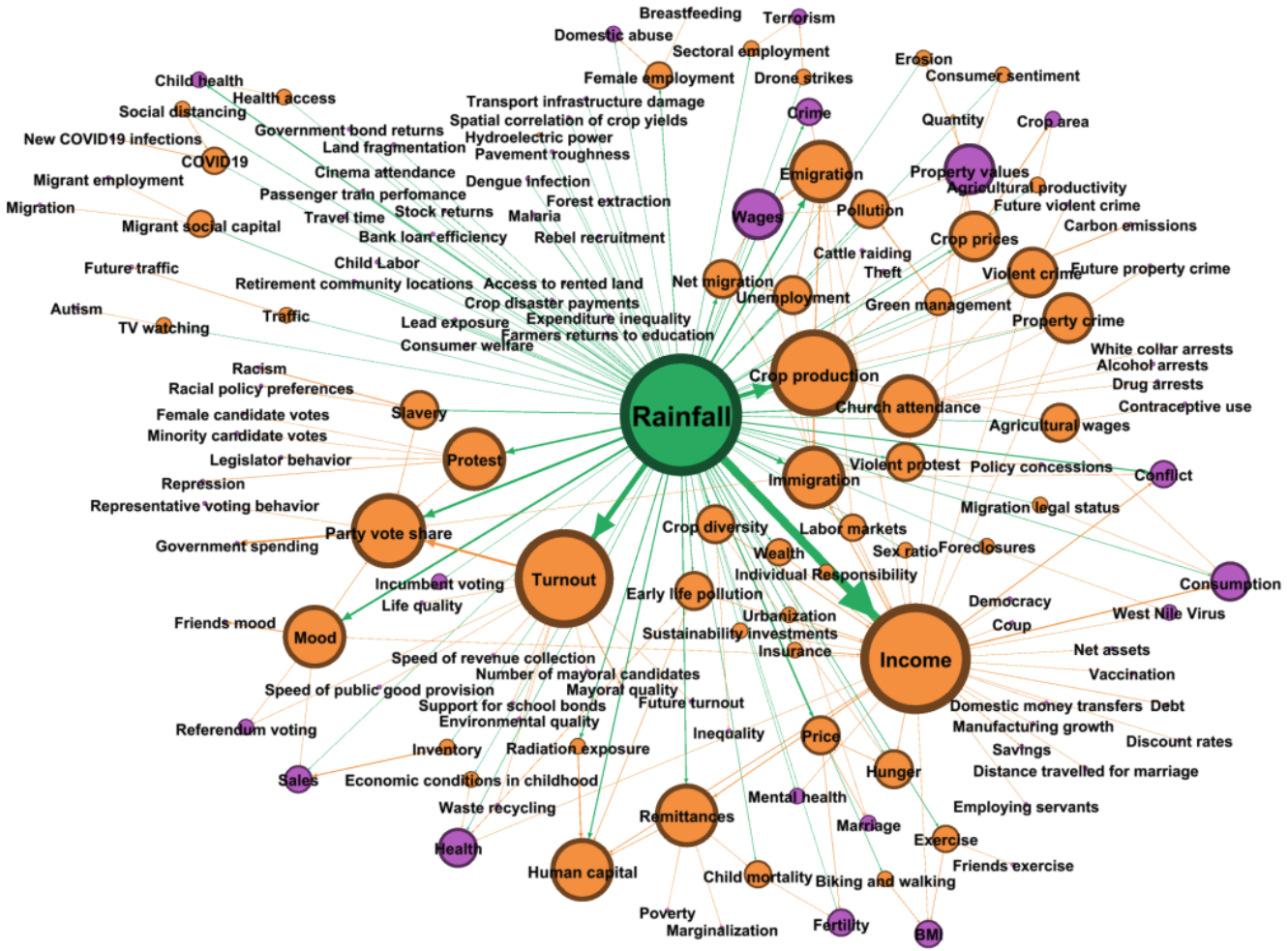
Instrument

To identify the causal effect of ethnic riots over time we follow the innovative identifications strategy of Iyer and Shrivastava (2018) by regressing future Hindu-Muslim riots on past Hindu-Muslim riots, while instrumenting the latter with whether in that same year an important Hindu festival occurred on a Friday, or not.

This identification strategy relies on five key assumptions. First, Friday festivals are correlated with Hindu-Muslim riots (relevance assumption). Second, Hindu festivals do not fall on Friday because of other factors that affect Hindu-Muslim riots, and Hindu festivals do not fall on Friday because of past Hindu-Muslim riots (exogeneity assumption). Third, Friday festivals do not affect future Hindu-Muslim riots through any other channel than present Hindu-Muslim riots (excludability assumption). Fourth, Friday festivals do not lead to a decrease in the probability of Hindu-Muslim riots in any district within our sample (monotonicity assumption). Last, Friday festivals in one area do not affect the probability of Hindu-Muslim riots in any other area within the sample (stable unit treatment value assumption).

We now discuss why we believe that these five identifying assumptions are likely to hold in our specific case.

When everyone uses the same instrument



Mellon (working paper, SSRN), "Rain, Rain, Go Away: 192 Potential Exclusion-Restriction Violations for Studies Using Weather as an Instrumental Variable"

What IV needs

- A **relevant** instrument
- A believable **exclusion restriction**

What IV needs

- A **relevant** instrument
- A believable **exclusion restriction**
- Also local: the effect is for units whose treatment **responds to the instrument**
  - $\rightarrow$ the LATE, or effect on "compliers"

Variation used: the part of X driven by Z

Must be true: Z **affects** Y **only through** X

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

The templates in one table

Design	**Variation used**	**Must be true**
Controls, matching	Treated vs. untreated who look alike	No unobserved confounders
Fixed effects	Within groups	No confounders varying within groups
DiD	Change over time, treated vs. untreated	Parallel trends
RDD	Jump at a cutoff	No sorting, nothing else at the cutoff
IV	Part of X moved by Z	Exclusion restriction

Combining: does pollution make people leave the city?

Chasing Clean Air: Pollution-Induced Travels in China

Shuai Chen, Yuyu Chen, Ziteng Lei, Jie-Sheng Tan-Soo

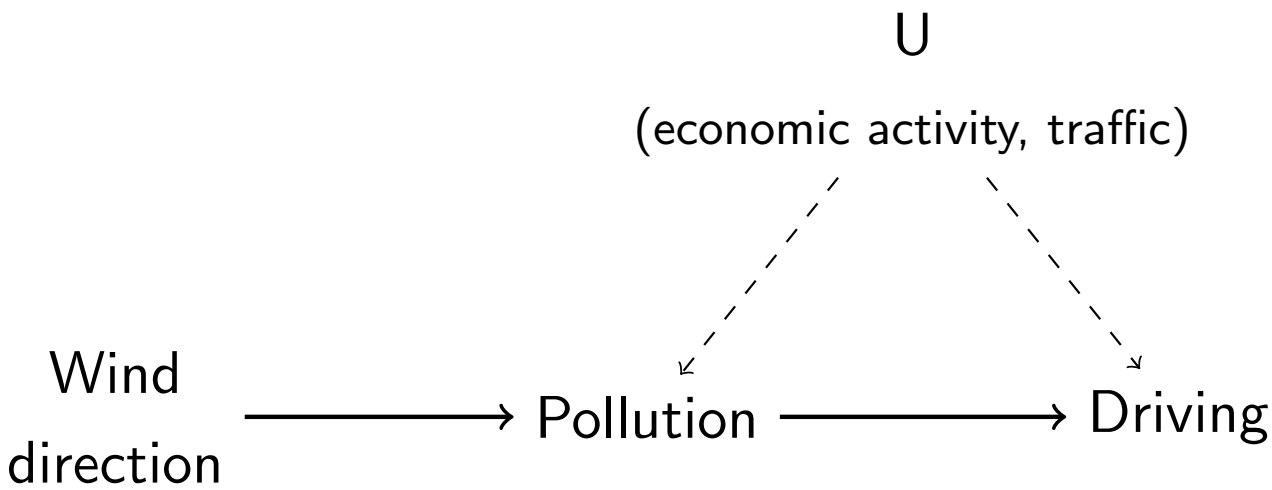
Abstract: This study uses "big" data to empirically investigate a highly effective, but underexamined way of reducing one's exposure toward air pollution—short-term travel. We determine subscribers' locations using mobile phones' signals and thereby establish linkages between air pollution and short-term population movements between cities in China. Using an instrumental variable based on daily variation in wind directions and pollution levels from distant upwind locations, we find that a one-unit increase in the origin city's air quality index (AQI) over the destination city's AQI increases short-term population flow from origin to destination by 0.15%. Further analyses provide richer characterizations of the decision-making process behind travel movements. Our findings add to the evidence base by examining in detail an understudied behavioral response toward air pollution.

JEL Codes: O15, Q53, Q56

Keywords: air pollution, short-term population movement, avoidance cost, averting strategy, China

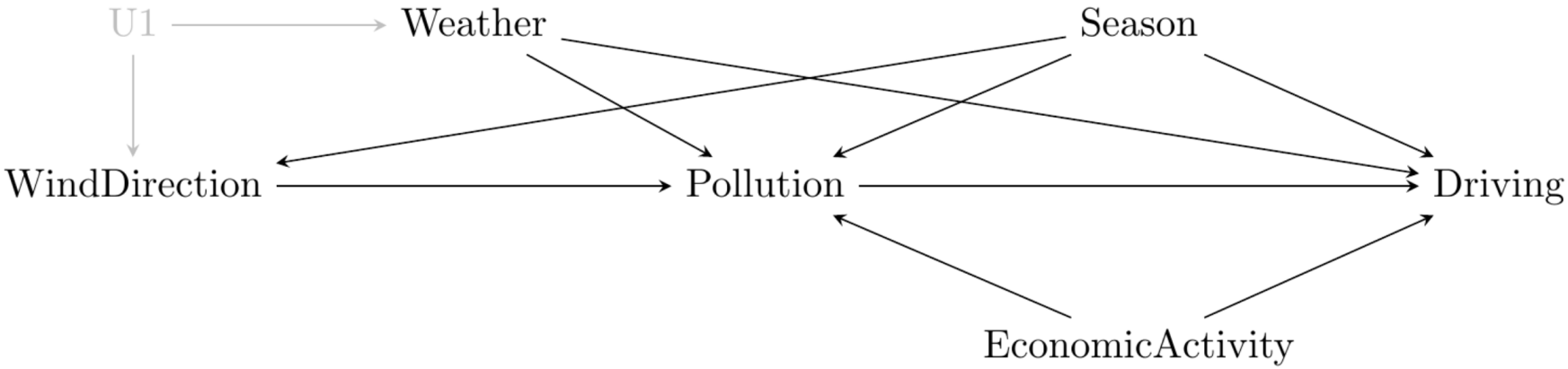
Chen, Chen, Lei & Tan-Soo (2021), *JAERE*

The naive version: wind direction alone

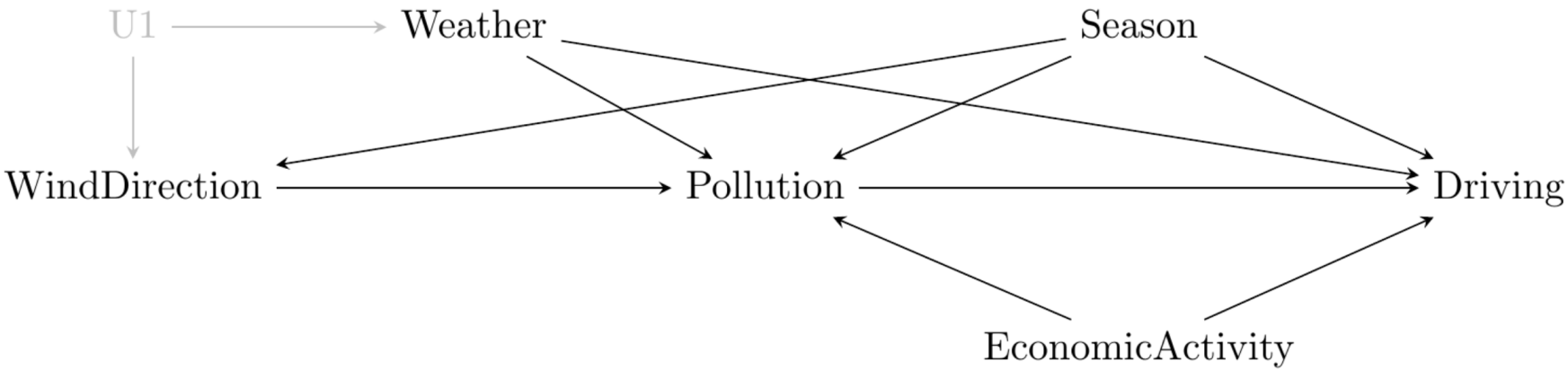


- Use only the variation in pollution that comes from the wind
- Valid **if** wind direction affects driving only through pollution

Combining: does pollution make people leave the city?



Combining: does pollution make people leave the city?



- Wind direction is linked to **weather and season**, which also affect travel
- So: instrument with wind, **and** control for weather, season and city

Try with these questions, which template?

1. Do Airbnb listings raise rents in a neighbourhood?
2. Does content moderation reduce hate speech on a platform?
3. Does having a female mayor change local spending?
4. Does exposure to cross-cutting news reduce polarization?

For each: what variation would you use, and what has to be true?

Roadmap

Two strategies

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Putting it together

How far does it travel?

Keeping yourself honest

Ethics for computational work

Wrapping up

Memo 5, workshop and essay

Credible here, but where else?

- **Internal validity**: is the effect real *here*?
- The **credibility revolution**: designs with strong internal validity, usually in a specific place and time

Credible here, but where else?

- **Internal validity**: is the effect real *here*?
- The **credibility revolution**: designs with strong internal validity, usually in a specific place and time
- The cost: the cleanest designs answer narrow questions
  - → close elections, compliers, one border, one platform, one year

Credible here, but where else?

- **Internal validity**: is the effect real *here*?
- The **credibility revolution**: designs with strong internal validity, usually in a specific place and time
- The cost: the cleanest designs answer narrow questions
  - → close elections, compliers, one border, one platform, one year
- Every step from topic to question to measure to design **loses something**

Credible here, but where else?

- **Internal validity**: is the effect real *here*?
- The **credibility revolution**: designs with strong internal validity, usually in a specific place and time
- The cost: the cleanest designs answer narrow questions
  - → close elections, compliers, one border, one platform, one year
- Every step from topic to question to measure to design **loses something**
- So: what have we learned about the **big question**?

Three kinds of validity beyond "here"

- **Construct validity**: is the treatment really the cause you name? Is the outcome really what you mean?
- **External validity**: would it hold for other people, treatments, outcomes, places?
- **Temporal validity**: would it still hold now?

Construct validity

- About **mislabeling** the cause or the outcome
- Of the cause: **bundled treatments**
  - → Guess et al.: the chronological feed changed political content, untrustworthy content, incivility, *and* time on platform
  - → a border is laws, language, markets and media at once

Construct validity

- About **mislabeling** the cause or the outcome
- Of the cause: **bundled treatments**
  - → Guess et al.: the chronological feed changed political content, untrustworthy content, incivility, *and* time on platform
  - → a border is laws, language, markets and media at once
- Of the outcome: session 3 again
  - → does your measure of polarization capture polarization?

External validity

- Unlikely that any effect holds under all conditions. The question is **under which**
- Generalize to other. . .
  - → **units** (who was studied?)
  - → **treatments** (a different dose or version?)
  - → **outcomes** (a different measure?)
  - → **contexts** (another country, platform, period?)

Egami & Hartman (2023), *APSR*

External validity

- Unlikely that any effect holds under all conditions. The question is **under which**
- Generalize to other. . .
  - → **units** (who was studied?)
  - → **treatments** (a different dose or version?)
  - → **outcomes** (a different measure?)
  - → **contexts** (another country, platform, period?)
- Generalizing the **sign** is easier than the **size**

Egami & Hartman (2023), *APSR*

Temporal validity

- "All of the knowledge contained in published papers comes from the past, and all action is in the future"

Munger (2023), *Research & Politics*

Temporal validity

- "All of the knowledge contained in published papers comes from the past, and all action is in the future"
- Knowledge **decays**, at a rate that depends on how fast the world changes

Munger (2023), *Research & Politics*

Temporal validity

- "All of the knowledge contained in published papers comes from the past, and all action is in the future"
- Knowledge **decays**, at a rate that depends on how fast the world changes
- In computational social science, that rate is high
  - → what do we know about Twitter that holds for X?
  - → platforms change their algorithms, users, and rules continuously

Munger (2023), *Research & Politics*

Learning across studies: meta-analysis

Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments

GRAEME BLAIR *University of California, Los Angeles*
DARIN CHRISTENSEN *University of California, Los Angeles*
AARON RUDKIN *University of California, Los Angeles*

Scholars of the resource curse argue that reliance on primary commodities destabilizes governments: price fluctuations generate windfalls or periods of austerity that provoke or intensify civil conflict. Over 350 quantitative studies test this claim, but prominent results point in different directions, making it difficult to discern which results reliably hold across contexts. We conduct a meta-analysis of 46 natural experiments that use difference-in-difference designs to estimate the causal effect of commodity price changes on armed civil conflict. We show that commodity price changes, on average, do not change the likelihood of conflict. However, there are cross-cutting effects by commodity type. In line with theory, we find price increases for labor-intensive agricultural commodities reduce conflict, while increases in the price of oil, a capital-intensive commodity, provoke conflict. We also find that price increases for lootable artisanal minerals provoke conflict. Our meta-analysis consolidates existing evidence, but also highlights opportunities for future research.

- Do commodity price shocks cause conflict? 46 studies, same design

How are meta-analyses done?

1. Literature review $\rightarrow$ 3,300 records

How are meta-analyses done?

1. Literature review $\rightarrow$ 3,300 records
2. Keep studies with **the same design and topic**:

   "the 46 included studies (102 estimates) quantitatively analyze the effect of plausibly exogenous variation in world commodity prices on armed civil conflict by using a generalized difference-in-difference design with unit and time fixed effects"

How are meta-analyses done?

1. Literature review $\rightarrow$ 3,300 records
2. Keep studies with **the same design and topic**:

   > "the 46 included studies (102 estimates) quantitatively analyze the effect of plausibly exogenous variation in world commodity prices on armed civil conflict by using a generalized difference-in-difference design with unit and time fixed effects"

3. Make studies comparable $\rightarrow$ standardize the coefficients

How are meta-analyses done?

1. Literature review $\rightarrow$ 3,300 records
2. Keep studies with **the same design and topic**:

   > "the 46 included studies (102 estimates) quantitatively analyze the effect of plausibly exogenous variation in world commodity prices on armed civil conflict by using a generalized difference-in-difference design with unit and time fixed effects"

3. Make studies comparable $\rightarrow$ standardize the coefficients
4. Pool them: fixed-effects and random-effects meta-analysis models

What does theory predict?

H1: resources are the **prize** to be won by fighting (+)

H2: higher prices change the **opportunity cost** of fighting
(– for labour-intensive, + for capital-intensive)

H3: **lootable** resources can be captured by fighting (+)

TABLE 1. Commodity Classifications and Predicted Effect Direction from Each Hypothesis

	Characteristics		Predicted direction		
Commodity type	Labor-intensive	Lootable	(H1)	(H2)	(H3)
Pooled (average of commodities)	Mix	Mix	+	+/–	+/0
Agriculture	✓		+	–	0
Artisanal minerals	✓	✓	+	–	+
Commercial minerals			+	+	0
Oil & gas		✓	+	+	+/0
Bundle of multiple types	Mix	Mix	+	+/–	+/0

FIGURE 1. Effects of Commodity Prices on Conflict by Commodity Type

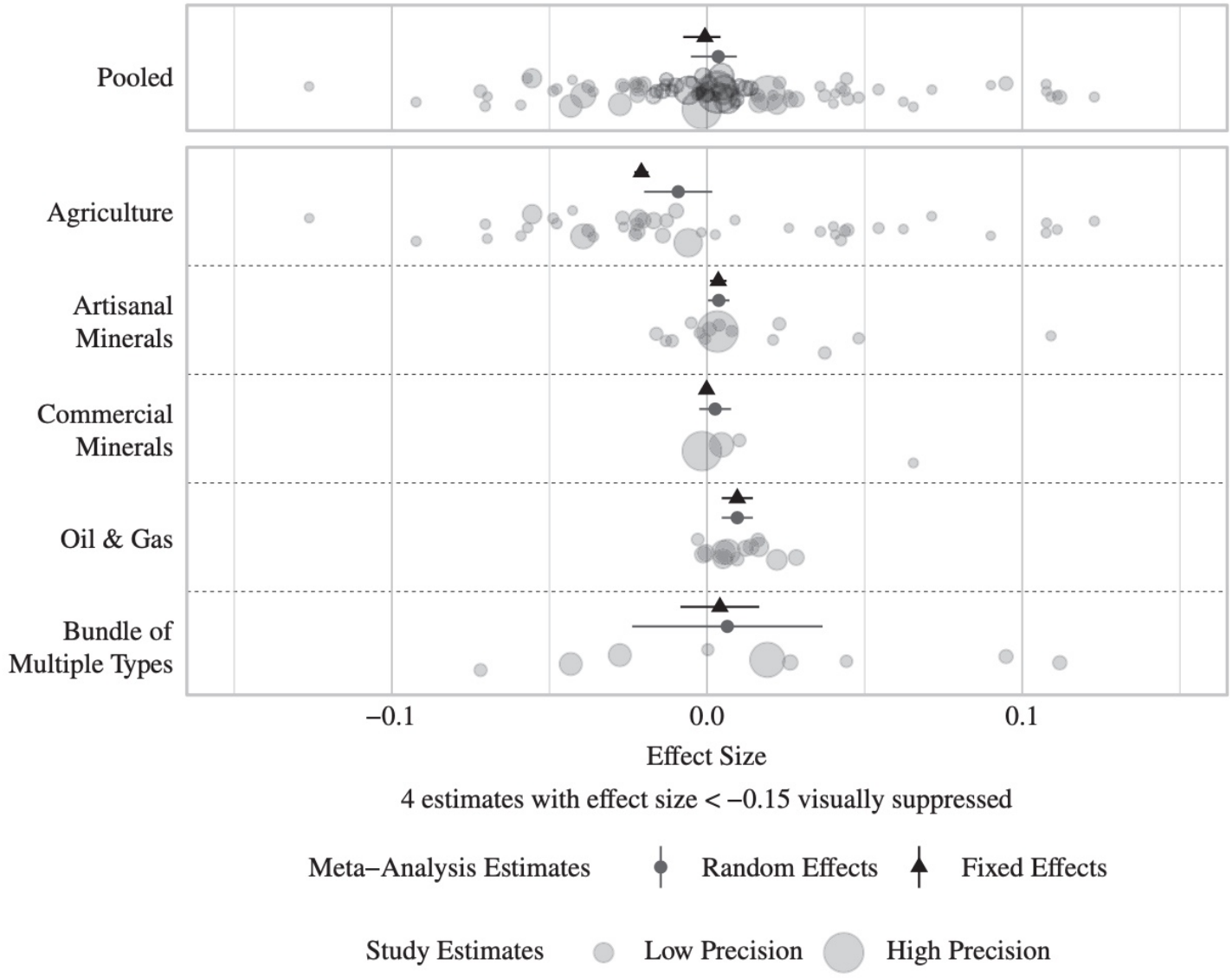
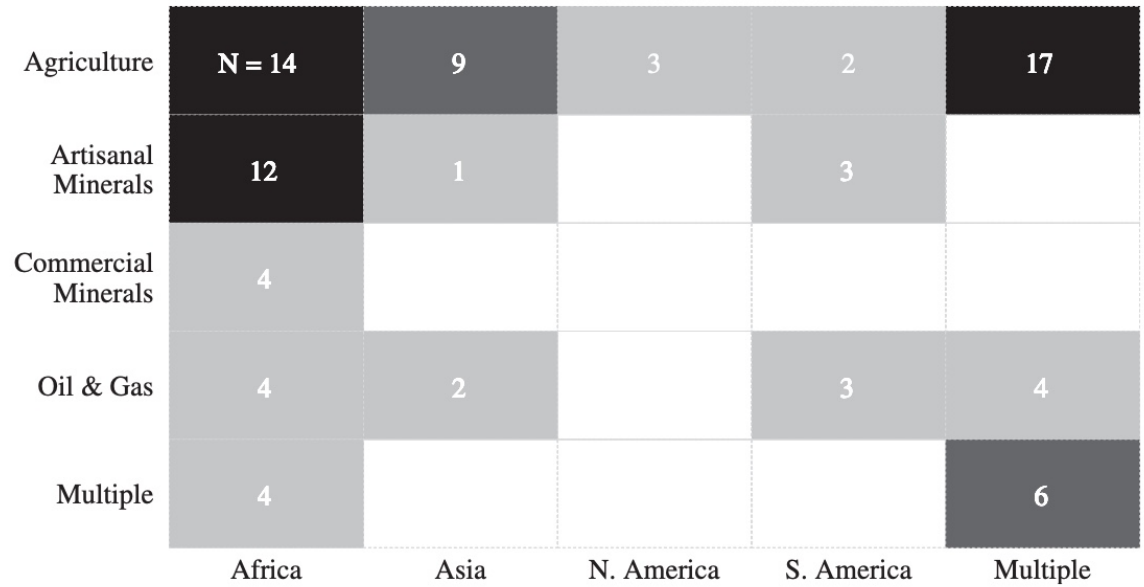


FIGURE 2. Evidence Gap Map (Number of Estimates) by Commodity Type and Continent



What the meta-analysis found

- Pooling all commodities: **no effect**

What the meta-analysis found

- Pooling all commodities: **no effect**
- Split by type, as theory predicts:
  - → agricultural prices up → **less** conflict (opportunity cost of fighting)
  - → oil and gas prices up → **more** conflict (bigger prize)

What the meta-analysis found

- Pooling all commodities: **no effect**
- Split by type, as theory predicts:
  - → agricultural prices up → **less** conflict (opportunity cost of fighting)
  - → oil and gas prices up → **more** conflict (bigger prize)
- Also shows the **gaps**: where and when the evidence comes from, and what goes unpublished

Roadmap

Two strategies

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Putting it together

How far does it travel?

Keeping yourself honest

Ethics for computational work

Wrapping up

Memo 5, workshop and essay

Replication crisis

18 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

This article is about an issue of scientific methodology. For the reproducibility crisis in humans, see Male infertility crisis.

The **replication crisis** (also called the **replicability crisis** and the **reproducibility crisis**) is an ongoing methodological crisis in which the results of many scientific studies are difficult or impossible to reproduce. Because the reproducibility of empirical results is an essential part of the scientific method,[2] such failures undermine the credibility of theories building on them and potentially call into question substantial parts of scientific knowledge.

The replication crisis is frequently discussed in relation to psychology and medicine, where considerable efforts have been undertaken to reinvestigate classic results, to determine both their reliability and, if found unreliable, the reasons for the failure.[3][4] Data strongly indicate that other natural, and social sciences are affected as well.[5]

The phrase *replication crisis* was coined in the early 2010s[6] as part of a growing awareness of the problem. Considerations of causes and remedies have given rise to a new scientific discipline, metascience,[7] which uses methods of empirical research to examine empirical research practice.

Considerations about reproducibility fall into two categories. *Reproducibility* in the narrow sense refers to re-examining and validating the analysis of a given set of data. *Replication* refers to repeating the experiment or study to obtain new, independent data with the goal of reaching the same or similar conclusions.


Why Most Published Research Findings Are False

It can be proven that most claimed research findings are false.

Ioannidis (2005): "Why Most Published Research Findings Are False".[1]

Background [edit]

Researcher degrees of freedom

- Every project involves dozens of reasonable decisions
  - → which outcome, which measure, which controls, which sample, which period, which estimator

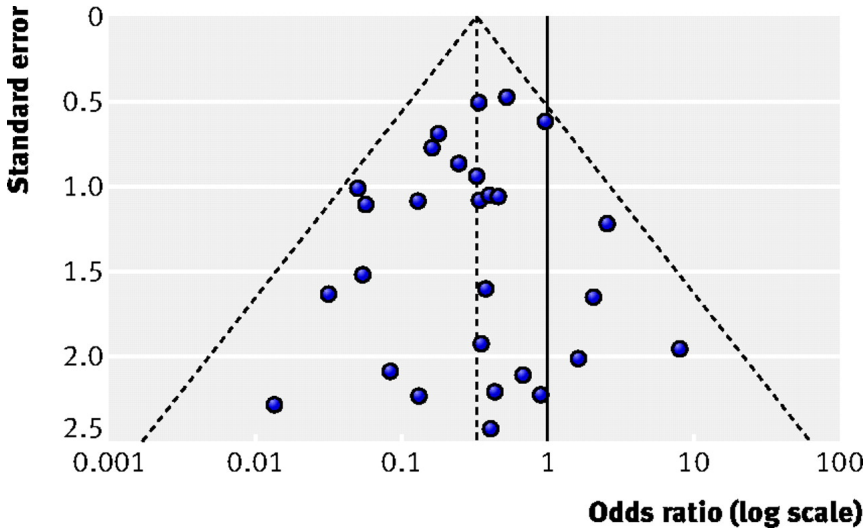
Researcher degrees of freedom

- Every project involves dozens of reasonable decisions
  - → which outcome, which measure, which controls, which sample, which period, which estimator
- Each choice made **after seeing the results** tilts them
  - → no bad intentions needed: the "garden of forking paths"

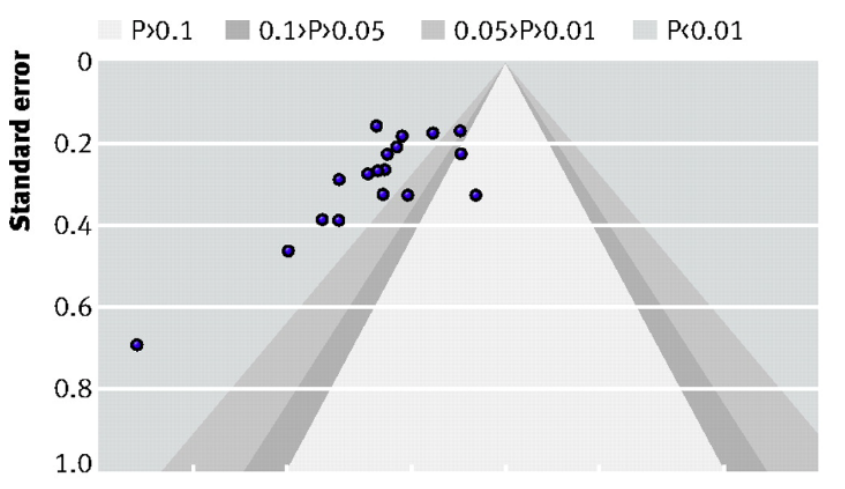
Researcher degrees of freedom

- Every project involves dozens of reasonable decisions
  - → which outcome, which measure, which controls, which sample, which period, which estimator
- Each choice made **after seeing the results** tilts them
  - → no bad intentions needed: the "garden of forking paths"
- Worse with big data: more variables, more subgroups, more ways to slice

Publication bias leaves a trace



No bias



Bias

Funnel plots: effect size vs. precision across studies.

Imprecise studies should scatter on **both** sides of the truth

Two responses

- **Pre-registration** and **pre-analysis plans**
  - → write down hypotheses, measures and analyses *before* seeing outcomes
  - → what you did not pre-register is still allowed, but labelled exploratory

Two responses

- **Pre-registration** and **pre-analysis plans**
  - → write down hypotheses, measures and analyses *before* seeing outcomes
  - → what you did not pre-register is still allowed, but labelled exploratory
- **Robustness** as design logic
  - → which reasonable choices could change the result? Show them all

Two responses

- **Pre-registration** and **pre-analysis plans**
  - → write down hypotheses, measures and analyses *before* seeing outcomes
  - → what you did not pre-register is still allowed, but labelled exploratory
- **Robustness** as design logic
  - → which reasonable choices could change the result? Show them all
- Your final essay can be a pre-analysis plan

Placebo tests: look where there should be nothing

- If your design is right, some things should **not** happen
- Example: Russian TV signal strength and voting in Ukraine

Peisakhin & Rozenas (2018), *AJPS*; Eggers, Tuñón & Dafoe (2024), *AJPS*

Placebo tests: look where there should be nothing

- If your design is right, some things should **not** happen
- Example: Russian TV signal strength and voting in Ukraine

1. **Placebo outcome**: signal should not predict things it can't affect (pre-existing share of Russian speakers)

Peisakhin & Rozenas (2018), *AJPS*; Eggers, Tuñón & Dafoe (2024), *AJPS*

Placebo tests: look where there should be nothing

- If your design is right, some things should **not** happen
- Example: Russian TV signal strength and voting in Ukraine

1. **Placebo outcome**: signal should not predict things it can't affect (pre-existing share of Russian speakers)
2. **Placebo treatment**: signal of Russian *sports* channels should not predict voting

Peisakhin & Rozenas (2018), *AJPS*; Eggers, Tuñón & Dafoe (2024), *AJPS*

Placebo tests: look where there should be nothing

- If your design is right, some things should **not** happen
- Example: Russian TV signal strength and voting in Ukraine

1. **Placebo outcome**: signal should not predict things it can't affect (pre-existing share of Russian speakers)
2. **Placebo treatment**: signal of Russian *sports* channels should not predict voting
3. **Placebo population**: no effect among satellite TV users, for whom signal is irrelevant

Peisakhin & Rozenas (2018), *AJPS*; Eggers, Tuñón & Dafoe (2024), *AJPS*

Roadmap

Two strategies

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Putting it together

How far does it travel?

Keeping yourself honest

Ethics for computational work

Wrapping up

Memo 5, workshop and essay

An experiment on 689,000 people

- Facebook, 2012: reduced positive or negative posts in users' feeds for a week
- Outcome: the emotional tone of what users then posted
- Published in *PNAS* (Kramer, Guillory & Hancock 2014)

An experiment on 689,000 people

- Facebook, 2012: reduced positive or negative posts in users' feeds for a week
- Outcome: the emotional tone of what users then posted
- Published in *PNAS* (Kramer, Guillory & Hancock 2014)
- No specific consent. Justified by the terms of service

An experiment on 689,000 people

- Facebook, 2012: reduced positive or negative posts in users' feeds for a week
- Outcome: the emotional tone of what users then posted
- Published in *PNAS* (Kramer, Guillory & Hancock 2014)
- No specific consent. Justified by the terms of service
- Same platform, same kind of intervention as Guess et al. What is different?

Questions that design raises

- **Consent**: accepting terms of service is not consenting to research
- **Harm**: would anyone be worse off because of your study?
- **Re-identification**: "anonymised" traces often aren't
- **Legality**: scraping, platform rules, data protection

Salganik, *Bit by Bit*, ch. 6: respect for persons, beneficence, justice, respect for law and public interest

Roadmap

- Two strategies
- Controlling and matching
- Fixed effects
- Difference-in-differences
- Regression discontinuity
- Instrumental variables
- Putting it together
- How far does it travel?
- Keeping yourself honest
- Ethics for computational work
- Wrapping up
- Memo 5, workshop and essay

What credibility means for each goal

- **Describe**: coverage, a defensible denominator, a clear population
- **Explain**: a comparison whose assumption you can defend
- **Predict**: performance on data the model has not seen, *from the future*

What credibility means for each goal

- **Describe**: coverage, a defensible denominator, a clear population
- **Explain**: a comparison whose assumption you can defend
- **Predict**: performance on data the model has not seen, *from the future*

All three: **how far does it travel?**

First steps

1. Say your question in one sentence. Which goal?
2. Which template(s) could you use?
3. What has to be true? How would you check it?
4. Which threat will you **not** be able to solve?

Roadmap

Two strategies

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Putting it together

How far does it travel?

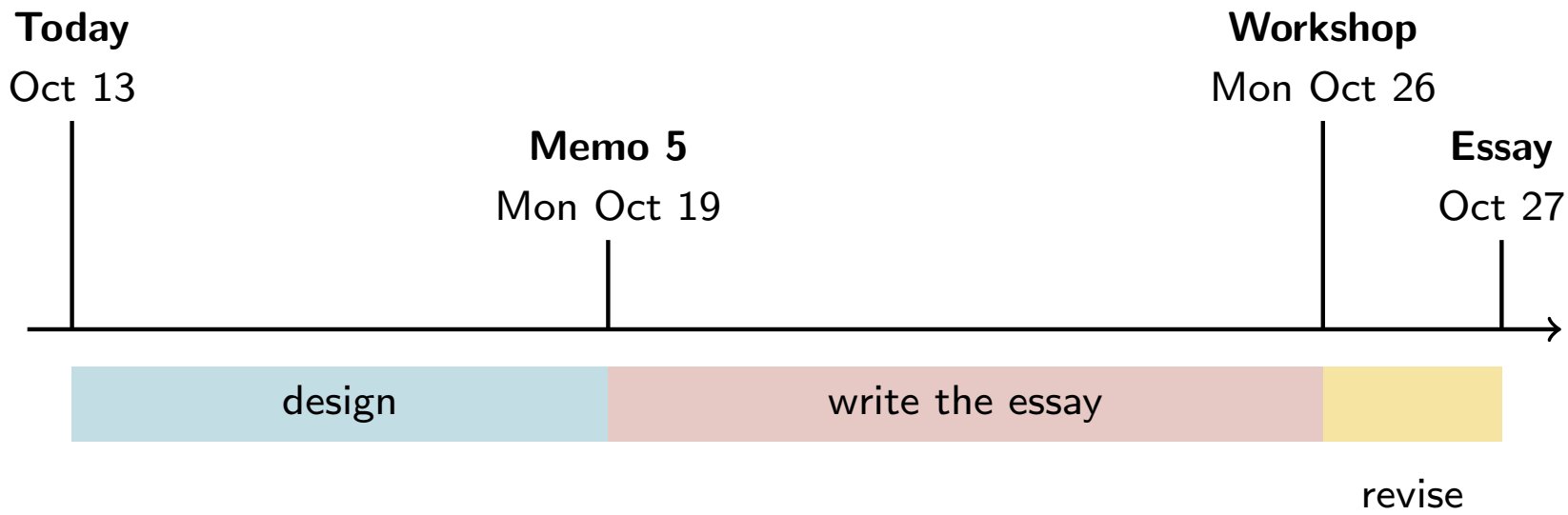
Keeping yourself honest

Ethics for computational work

Wrapping up

Memo 5, workshop and essay

Housekeeping



Design memos (5)	15%	the pair (Memo 5: the group)
Workshop presentation	35%	the group
Workshop feedback	10%	the group
Final essay	40%	the group

Memo 5: the design on one page

- Written by your **workshop group** (3–4 people), not in pairs
- Due **Monday Oct 19**, Aula Global, by midnight

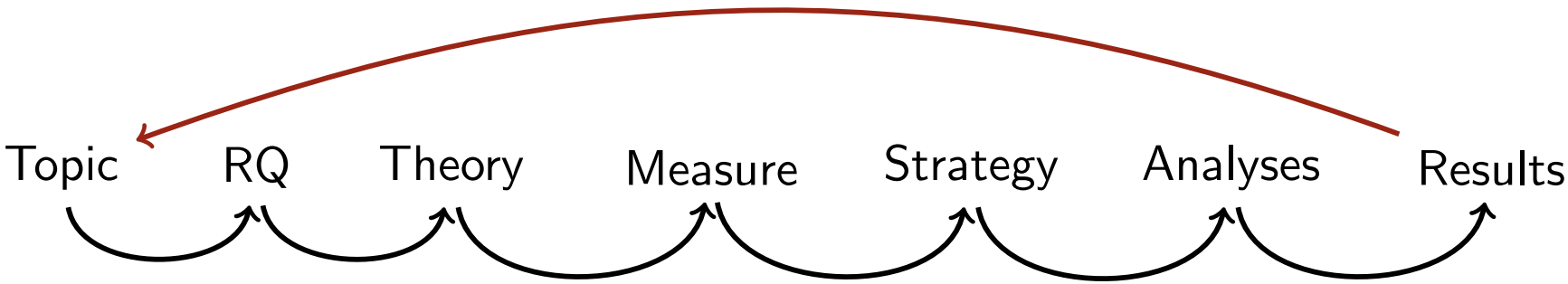
Memo 5: the design on one page

- Written by your **workshop group** (3–4 people), not in pairs
- Due **Monday Oct 19**, Aula Global, by midnight
- One page, six items:
  - → goal · question · units and variation
  - → measure · comparison · the threat you cannot solve

Memo 5: the design on one page

- Written by your **workshop group** (3–4 people), not in pairs
- Due **Monday Oct 19**, Aula Global, by midnight
- One page, six items:
  - → goal · question · units and variation
  - → measure · comparison · the threat you cannot solve
- Whoever submits together **is a group**: binding for the presentation and the essay
- It goes to your **paired group**, who use it to prepare

The research process



Same diagram as in session 1. Your project covers everything up to **Strategy**

What your presentation should cover

- The **goal**: describe, explain, or predict?
- The **question**, and why it is worth asking
- The **units** and the **variation** you would exploit
- The key **concept** and how you would measure it
- The **comparison** you would make
- **The threat you cannot solve**

Twelve minutes to present, eight minutes of discussion

Workshop day: Monday Oct 26, 15–21h

- Double session, around **twelve slots**
- Each group: **12 minutes** presenting + **8 minutes** discussion
- Attendance for the whole session is mandatory

Workshop day: Monday Oct 26, 15–21h

- Double session, around **twelve slots**
- Each group: **12 minutes** presenting + **8 minutes** discussion
- Attendance for the whole session is mandatory
- Presenting:
  - → a few slides, one idea each. The design, not a literature review
  - → spend your time on the **variation** and the **comparison**
  - → rehearse: twelve minutes is shorter than you think

Workshop feedback (10%)

- Each group is **paired** with another group
- You get their Memo 5 in advance, and come with **written comments**:
  - → what is **unclear**
  - → the **main threat** to the design
  - → what you would **do differently**

Workshop feedback (10%)

- Each group is **paired** with another group
- You get their Memo 5 in advance, and come with **written comments**:
  - → what is **unclear**
  - → the **main threat** to the design
  - → what you would **do differently**
- Feedback should be based on the **presentation**
  - → Memo 5 is a primer: draft from it, update after you hear them
- Handed in and graded on the group

The final essay (40%)

- A research design, **not** a data analysis. Two formats:
  - → a **pre-analysis plan** for a study
  - → an overview of how a topic could be studied quantitatively, e.g. **comparing two strategies** for the same question

The final essay (40%)

- A research design, **not** a data analysis. Two formats:
  - → a **pre-analysis plan** for a study
  - → an overview of how a topic could be studied quantitatively, e.g. **comparing two strategies** for the same question

- Required:
  - → the **goal** (describe, explain, predict) in the **first paragraph**
  - → a paragraph on **data provenance**: who created it, why, what's missing, can you get it?
  - → a section on **threats**, including one you **cannot solve**

The final essay (40%)

- A research design, **not** a data analysis. Two formats:
  - → a **pre-analysis plan** for a study
  - → an overview of how a topic could be studied quantitatively, e.g. **comparing two strategies** for the same question
- Required:
  - → the **goal** (describe, explain, predict) in the **first paragraph**
  - → a paragraph on **data provenance**: who created it, why, what's missing, can you get it?
  - → a section on **threats**, including one you **cannot solve**
- Max **5000 words** (excluding appendix and references). Same group, same grade

A structure that works

1. **Introduction**: question, goal, why it matters
2. **Argument**: the mechanism, and what it predicts
3. **Data**: sources, units, provenance
4. **Measurement**: concepts and how you would measure them
5. **Design**: the comparison, what must be true, how you would check
6. **Threats**: what could go wrong, and what you cannot solve
7. **How far it travels**: scope, validity, ethics

A suggestion, not a template. One section per session, roughly

Common pitfalls

- A **topic** instead of a question

Common pitfalls

- A **topic** instead of a question
- A **list of methods** instead of a comparison

Common pitfalls

- A **topic** instead of a question
- A **list of methods** instead of a comparison
- "We will control for everything"

Common pitfalls

- A **topic** instead of a question
- A **list of methods** instead of a comparison
- "We will control for everything"
- A predictive model presented as **causal**

Common pitfalls

- A **topic** instead of a question
- A **list of methods** instead of a comparison
- "We will control for everything"
- A predictive model presented as **causal**
- Data you **cannot actually get**

Common pitfalls

- A **topic** instead of a question
- A **list of methods** instead of a comparison
- "We will control for everything"
- A predictive model presented as **causal**
- Data you **cannot actually get**
- No threat left unsolved

Questions?