Overview of Quasi-Experimental Designs

Advanced Social Epidemiology PhD Course

Sam Harper

University of Copenhagen
2024-12-02 to 2024-12-06

Quasi-Experiments

1. Motivation
2. Randomization and Observation
3. Quasi-Experimental Designs
4. Final Thoughts

Quasi-Experiments

1. Motivation
2. Randomization and Observation
3. Quasi-Experimental Designs
4. Final Thoughts

Stylized "forms" of questions asked in social epidemiology

What question do most studies in social epidemiology answer?

- Do individuals who are disadvantaged with respect to social position have worse health than those who are advantaged?

Other kinds of questions that could be asked:

- Would individuals who are disadvantaged with respect to social position have better health if they were to become advantaged?
- Would individuals who are advantaged with respect to social position have worse health if they were to become disadvantaged?

These are **causal** questions.

"Normal" etiological science in social epidemiology

1. Follow-up of individuals in different social groups for various health outcomes (incidence, mortality, risk factors)

2. Adjustment for various confounders/mediators (are inequalities "explained" by....A, B, C?).

- "Our results demonstrate that"...we should:

    - *raise* education levels
    - *increase* economic assistance to the poor
    - *remove* noxious exposures from the environment
    - *reduce* psychosocial workplace hazards
    - *eliminate* hierarchies, and the like.

- These statements are based on making **causal** inferences.

What's the problem?

- We are mainly (though not exclusively) interested in causal effects.
- We want to know:
  - Should we intervene to reduce exposure to X?, or
  - Did the program work? If so, for whom? If not, why not?, or
  - If we implement the program elsewhere, should we expect the same result?
- These questions involve counterfactuals about what would happen **if** we intervened to do something.
- These are causal questions.

How to interpret statistical associations of health inequality?

We have lots of statistical associations between social exposures and health.

$$X - - - Y$$

Some possible situations *consistent* with statistical associations:

1. Causal $X \rightarrow Y$
2. Heterogeneity $X_a \;\; Y_a$ vs. $X_b \rightarrow Y_b$
3. Reverse causation $Y \rightarrow X$
4. Confounding $X \leftarrow C \rightarrow Y$
5. Selection bias $X \rightarrow S \leftarrow Y$

Quasi-Experiments

1. Motivation
2. Randomization and Observation
3. Quasi-Experimental Designs
4. Final Thoughts

Randomized Trials vs. Observational Studies

RCTs, Defined

RCTs involve:

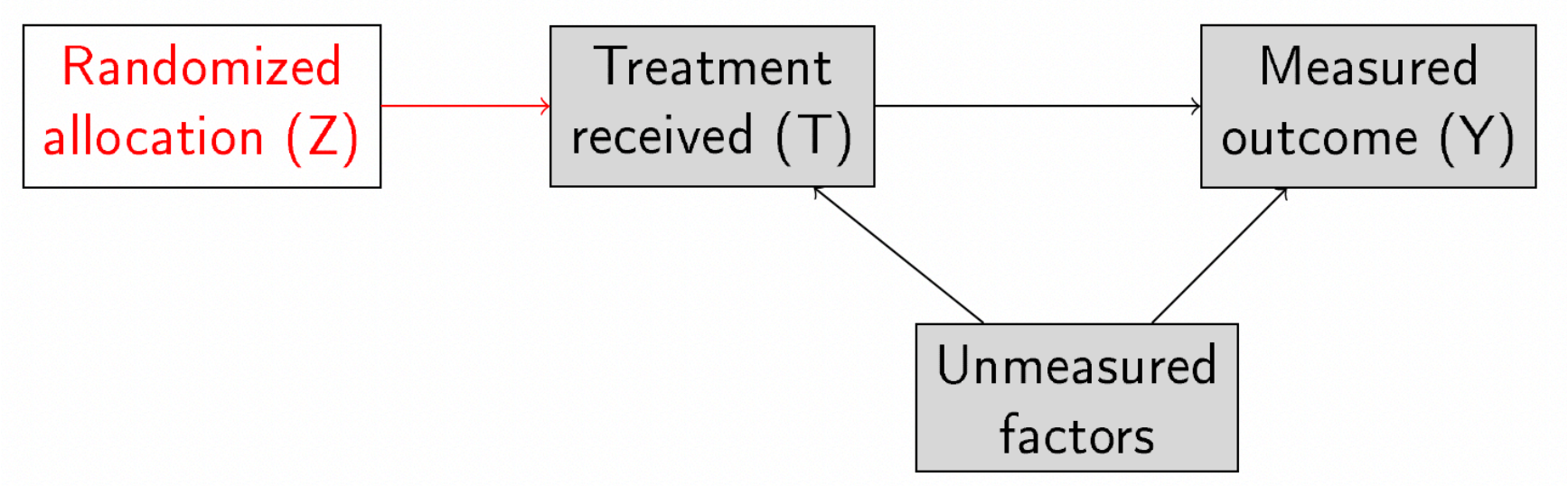
1. comparing treated and control groups;
2. the treatment assignment is random;
3. investigator does the randomizing.

In an RCT, treatment/exposure is assigned by the investigator

- In observational studies, exposed/unexposed groups exist in the source population and are selected by the investigator.
- Good natural experiments do (1) and (2), but not (3).
- Because there is no control over assignment, the credibility of natural experiments hinges on how good "as-if random" approximates (2).

Strength of randomized treatment allocation

- Recall that randomization means that we can generally estimate the causal effect *of being randomized* without bias.
- Randomization guarantees exchangeability on measured and unmeasured factors.



Randomize if you can

- Randomization leads to:
  - balance on measured factors.
  - balance on unmeasured factors.
- Unmeasured factors cannot bias the estimate of the exposure effect.
- Example from Home Injury Prevention Intervention cluster RCT
- What do you notice about Table 1?

(Keall et al. 2015)

	Treatment group (n=950)	Control group (n=898)
Female sex	541 (57%)	501 (56%)
Indigenous Māori	88 (9%)	86 (10%)
Mean (SD) age (years)*	45 (28·0)	43 (28·1)
Age range (years)	0–94	0–92
0–9	175 (18%)	187 (21%)
10–19	89 (9%)	82 (9%)
20–29	34 (4%)	37 (4%)
30–39	116 (12%)	112 (12%)
40–49	90 (9%)	96 (11%)
50–59	65 (7%)	51 (6%)
60–69	132 (14%)	105 (12%)
≥70	249 (26%)	228 (25%)
Number of injuries at home, excluding falls, in past year (per person)†	122 (0·129)	103 (0·115)
Number of fall injuries at home in past year (per person)‡	87 (0·092)	61 (0·068)
Number of specific injuries in past year (per person)§	23 (0·024)	24 (0·027)

Data are number of individual occupants (%), unless otherwise indicated.
*At Aug 3, 2010. †Injuries arising in the home during the 365-day period before the intervention date, obtained from matched insurance claim data. ‡Slips, trips, or fall injuries in the home during the 365-day period before the intervention date. §Injuries most specific to the package of home modifications, arising in the home during the 365-day period before the intervention date.

Table 1: **Characteristics of individual occupants at baseline**

Or maybe don't randomize?

RCT limitations

- Non-compliance.
- Attrition.
- Spillovers.
- Blinding (esp. in clinical trials).

Other trial challenges:

- Unethical (poverty, parental social class, job loss)
- Impossible (ethnic background, place of birth)
- Expensive (neighborhood environments)
- Long latency periods (many years before outcomes are observable).
- Effects may be produced by complex, intermediate pathways.

- We need alternatives to RCTs.

Unmeasured confounding is a serious challenge

- We often compare socially advantaged and disadvantaged on health.
- Key problem: people choose/end up in treated or untreated group for reasons that are difficult to measure and that may be correlated with their outcomes.
- So...**adjust.**
  - Measure and adjust (regression) for C confounding factors.
  - Conditional on C, we are supposed to believe assignment is "as good as random" = causal.

Key issue is credibility

- If we have a good design and assume that we have measured all of the confounders, then regression can give us exactly what we want: an estimate of the causal effect of exposure to T.
- Core issue: How credible is this assumption?

"Now, keep in mind that these numbers are only as accurate as the fictitious data, ludicrous assumptions and wishful thinking they're based upon!"

SEP and CVD in Australia. Many low p-values

Table 1 Characteristics of 38 355 subjects in the Melbourne Collaborative Cohort Study at baseline (1990–1994)

		Highest level of education attained				
		Completed tertiary* n=8588	Completed secondary† n=7882	Some secondary‡ n=14543	Primary only§ n=7342	p Value for trend¶
Male	n (%)	4025 (47%)	3776 (48%)	4680 (32%)	2780 (38%)	<0.001
Female	n (%)	4563 (53%)	4106 (52%)	9863 (68%)	4562 (62%)	<0.001
Age (years)	(Mean, SD)	51.6 (8.4)	54.5 (8.8)	55.7 (8.5)	57.8 (7.1)	<0.001
Country of birth, n (%)	Australia, New Zealand or northern Europe (n=28 835)	8263 (96%)	6814 (86%)	12696 (87%)	1062 (14%)	<0.001
	Southern Europe (n=9520)	325 (4%)	1068 (14%)	1847 (13%)	6280 (86%)	<0.001
Behavioural risk factors						
Current smoker	n (%)	574 (7%)	960 (12%)	1828 (13%)	947 (13%)	<0.001
Vegetable intake (times/day)	Mean (SD)	5.7 (3)	5.3 (3)	5.2 (3)	5.8 (4)	1.000
Fruit intake (times/day)	Mean (SD)	4.4 (3)	4.0 (3)	3.9 (3)	4.7 (4)	0.007
Saturated fat intake (g/day)	Mean (SD)	35.0 (15)	34.3 (16)	33.7 (16)	30.3 (18)	<0.001
Current drinker	n (%)	7061 (82%)	5883 (75%)	9397 (65%)	3666 (50%)	<0.001
Alcohol intake, current drinkers (g/day)	Median (IQR)	14 (5,26)	13 (4,26)	10 (3, 23)	15 (4,30)	<0.001
Physical activity (% inactive)	n (%)	1224 (14%)	1520 (19%)	3238 (22%)	2546 (35%)	<0.001
Social connection						
Living alone	n (%)	1514 (18%)	1274 (16%)	2250 (15%)	498 (7%)	<0.001

Beauchamp et al. (2010)

Why we worry about observational studies

Recent evaluation of "Workplace Wellness" program in US state of Illinois

Treatment: biometric health screening; online health risk assessment, access to a wide variety of wellness activities (e.g., smoking cessation, stress management, and recreational classes).

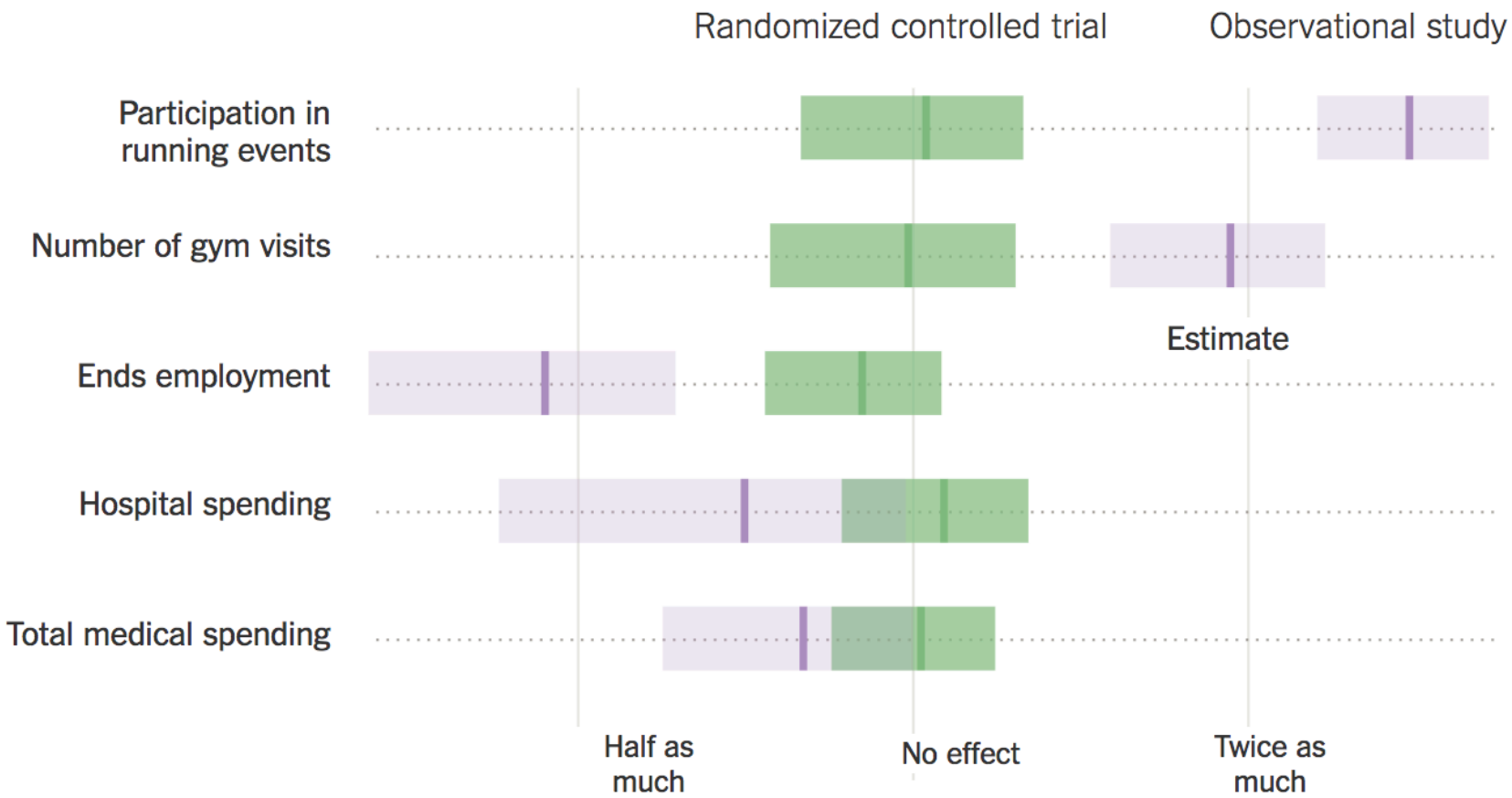
Randomized evaluation:

- 3,300 individuals assigned treated group.
- 1,534 assigned to control (could not access the program).

Also analyzed as an observational study comparing "participants" vs. non-participants in treated group.

Jones et al. (2018)

How the Illinois Wellness Program Affected . . .



Carroll, New York Times, Aug 6, 2018.

Quasi-Experiments

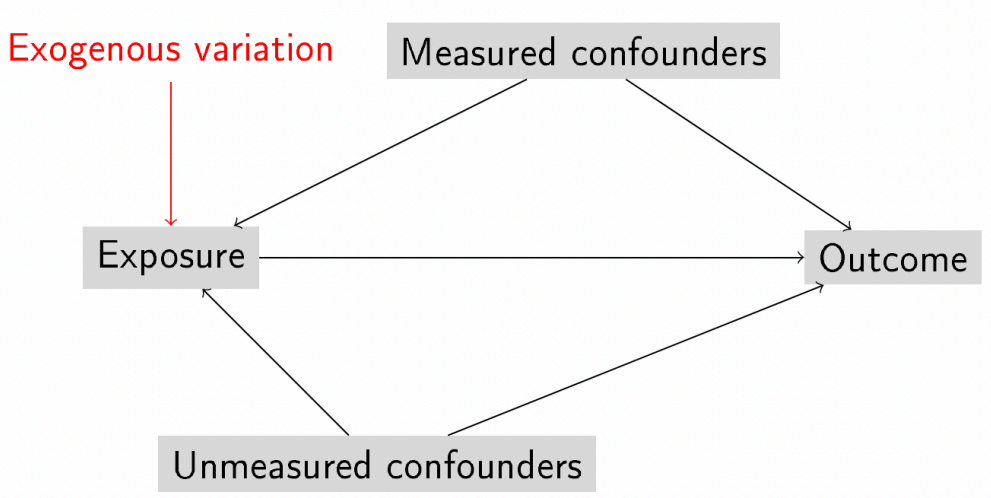
1. Motivation
2. Randomization and Observation
3. Quasi-Experimental Designs
4. Final Thoughts

How can quasi-experiments help?

- Quasi-experiments aim to mimic RCTs.
- "Accidents of chance" that create:
    1. Comparable treated and control units
    2. Random or "as-if" random assignment to treatment.
- Control for (some) sources of bias that cannot be adequately controlled using regression adjustment.
- More credible designs also help us to understand the relevance of other factors that may be implicated in generating inequalities.

Selection on "observables" and "unobservables"

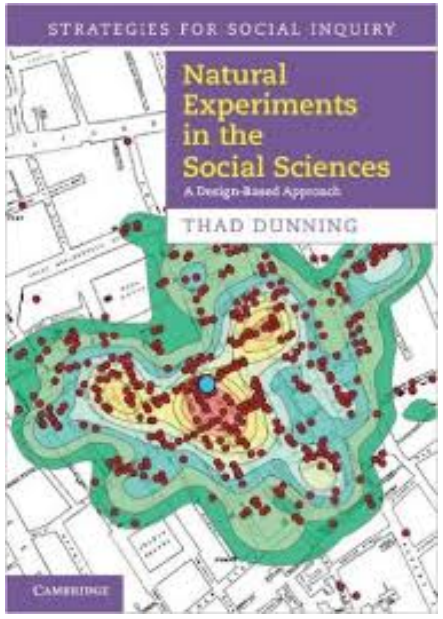
- Observables: Things you measured or can measure
- Unobservables: Things you can't measure (e.g., innate abilities, motivation)
- Exogenous variation: predicts exposure but (**we assume**) not associated with anything else [mimicking random assignment].



Strategies based on observables and unobservables

- Most observational study designs control for *measured* factors using:
  - Stratification
  - Regression adjustment
  - Matching (propensity scores, etc.)
- Quasi-experimental strategies aim to control for some *unmeasured* factors using:
  - Interrupted time series (ITS)
  - Difference-in-differences (DD)
  - Synthetic controls (SC)
  - Instrumental variables (IV)
  - Regression discontinuity (RD)

Some *potential* sources of natural experiments



- Law changes
- Eligibility for social programs (roll-outs)
- Lotteries
- Genes
- Weather shocks (rainfall, disasters)
- Arbitrary policy or clinical guidelines (thresholds)
- Business / factory closures
- Historical legacies (physical environment)
- Seasonality

Difference-in-Differences

Difference-in-Differences: Basic Idea

In the simplest DD setting, outcomes are observed for units in two groups and in two time periods.

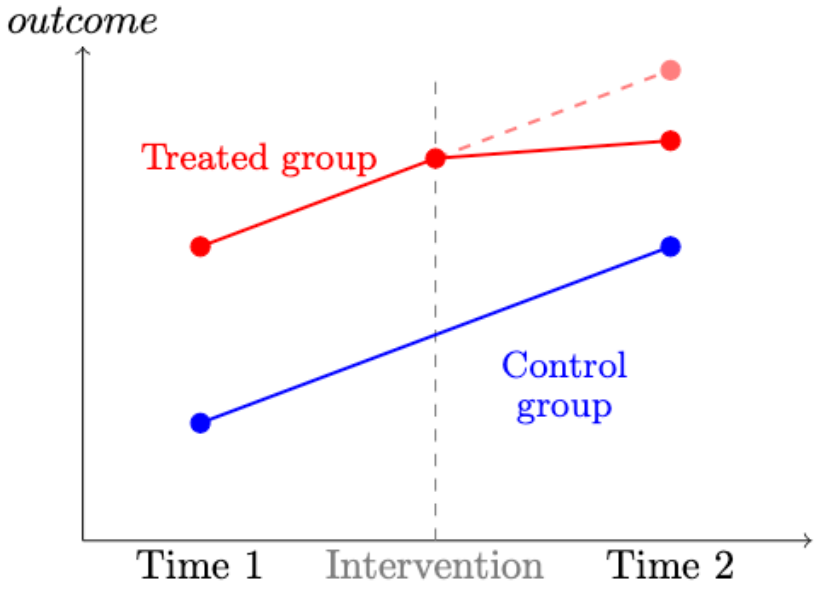
Treated:

- only units in one of the two groups are exposed to a treatment, in the second time period.

Control:

- Never observed to be exposed to the treatment.

Difference-in-Differences: Basic Idea



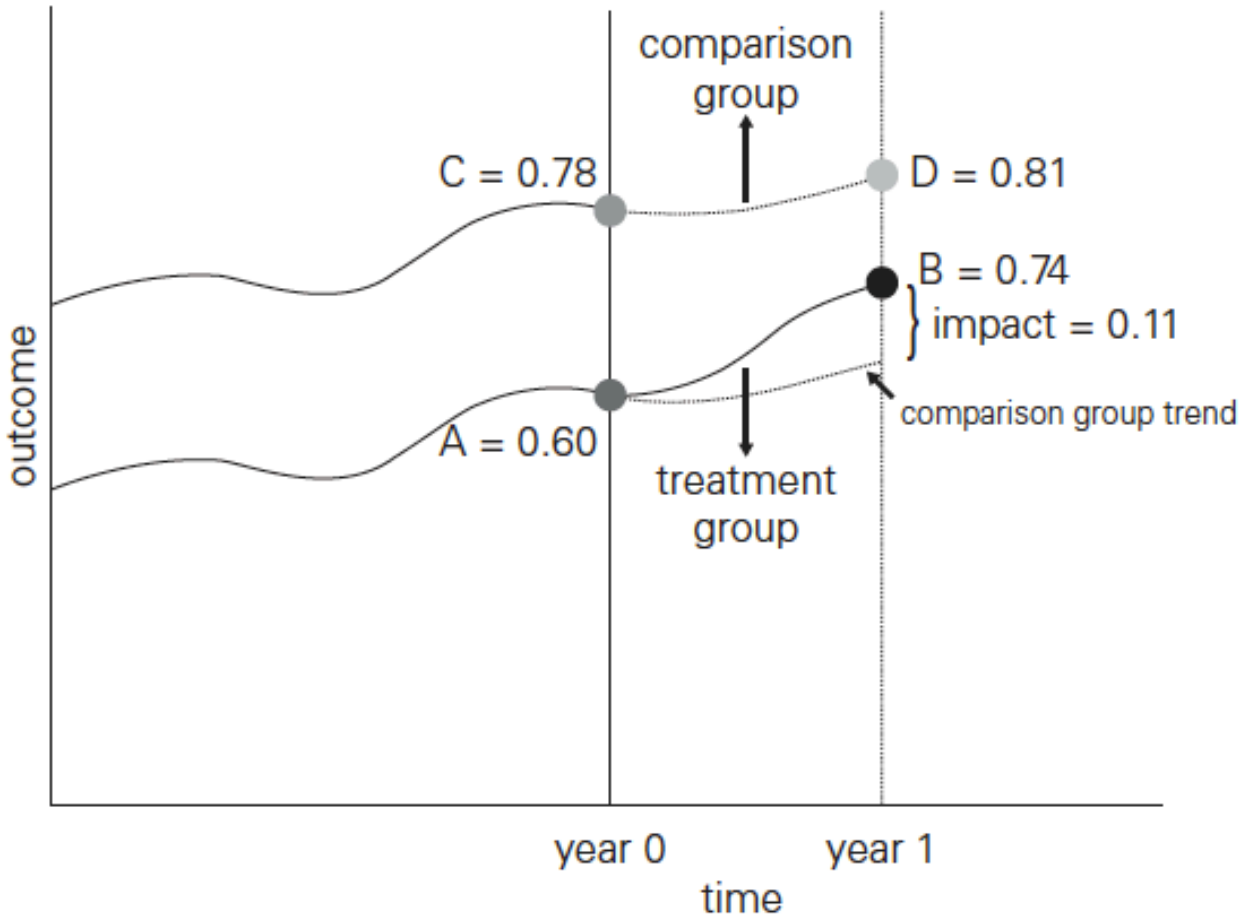
The average change over time in the non-exposed (control) group is subtracted from the gain over time in the exposed (treatment) group.

These are our two 'differences'.

Double differencing removes biases in second period comparisons between the treatment and control group that could be the result from

- permanent differences between those groups
- secular trends affecting both groups.

Visual Intuition of DD



Gertler et al. (2011)

Difference-in-Differences without Regression

DD is just differences in means! Let $\mu_{it} = E(Y_{it})$

- $i = 0$ is control group, $i = 1$ is treatment.
- $t = 0$ is pre-period, $t = 1$ is post-period.
- One 'difference' estimate of causal effect is: $\mu_{11} — \mu_{10}$ (pre-post in treated)
- Differences-in-Differences estimate of causal effect is: $(\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$

Area	Before	After	Difference (A - B)
Treated	135	100	-35
Control	80	60	-20
T - C	55	40	**-15**

A social epidemiology example

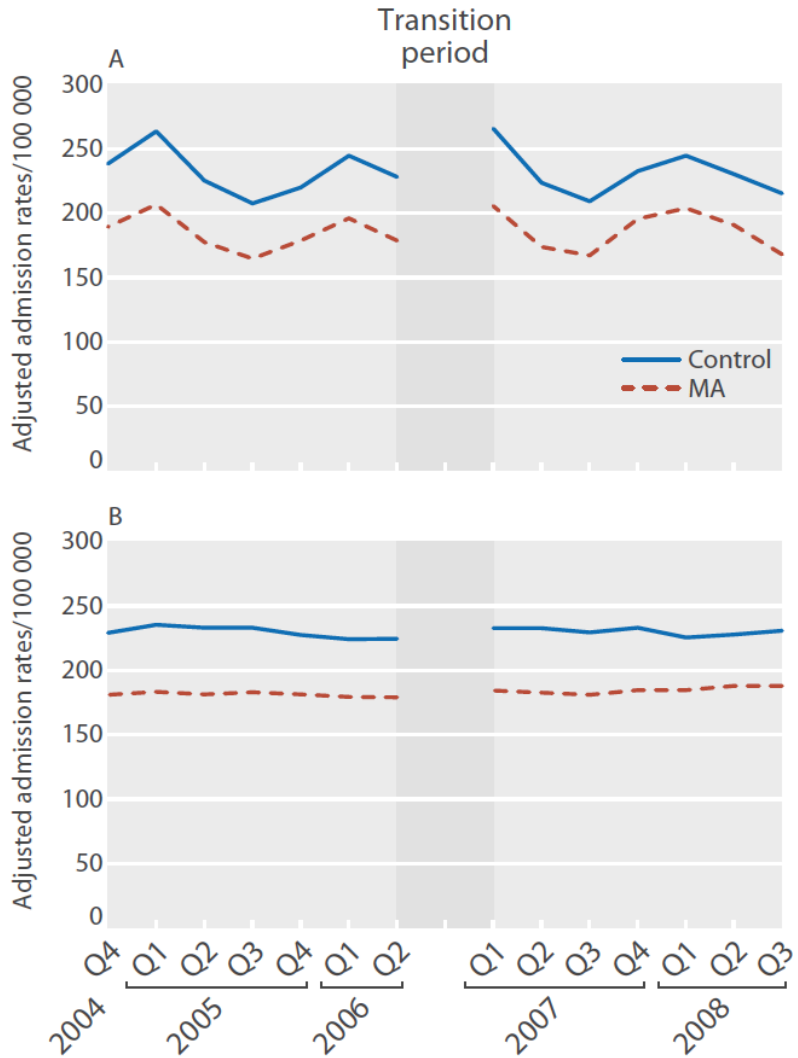
Effect of Massachusetts healthcare reform on racial and ethnic disparities in admissions to hospital for ambulatory care sensitive conditions: retrospective analysis of hospital episode statistics

Danny McCormick,[1] Amresh D Hanchate,[2, 3] Karen E Lasser,[3] Meredith G Manze,[3] Mengyun Lin,[3] Chieh Chu,[3] Nancy R Kressin[2, 3]

- Evaluated impact of MA reform on inequalities in hospital admissions.
- Compared MA to nearby states: NY, NJ, PA.
- Intervention "worked": % uninsured halved (12% to 6%) from 2004-06 to 2008-09.

Evaluating pre-intervention trends

- Adds credibility to assumption that post-intervention trends would have been similar in the absence of the intervention.
- "Null" results help focus on alternative mechanisms linking disadvantage to hospital admissions.



Synthetic Controls

Synthetic control methods

- Inference from comparative case studies is limited if we cannot identify a control to represent the counterfactual scenario.
- Abadie and Gardeazabel (2003) pioneered the synthetic control method to examine the economic impact of terrorism in the Basque country, using other Spanish regions as control groups.
- The synthetic control method uses a data driven approach to compare the trend of an outcome in a treated unit with the trend in a synthetic composite area (the "synthetic control").

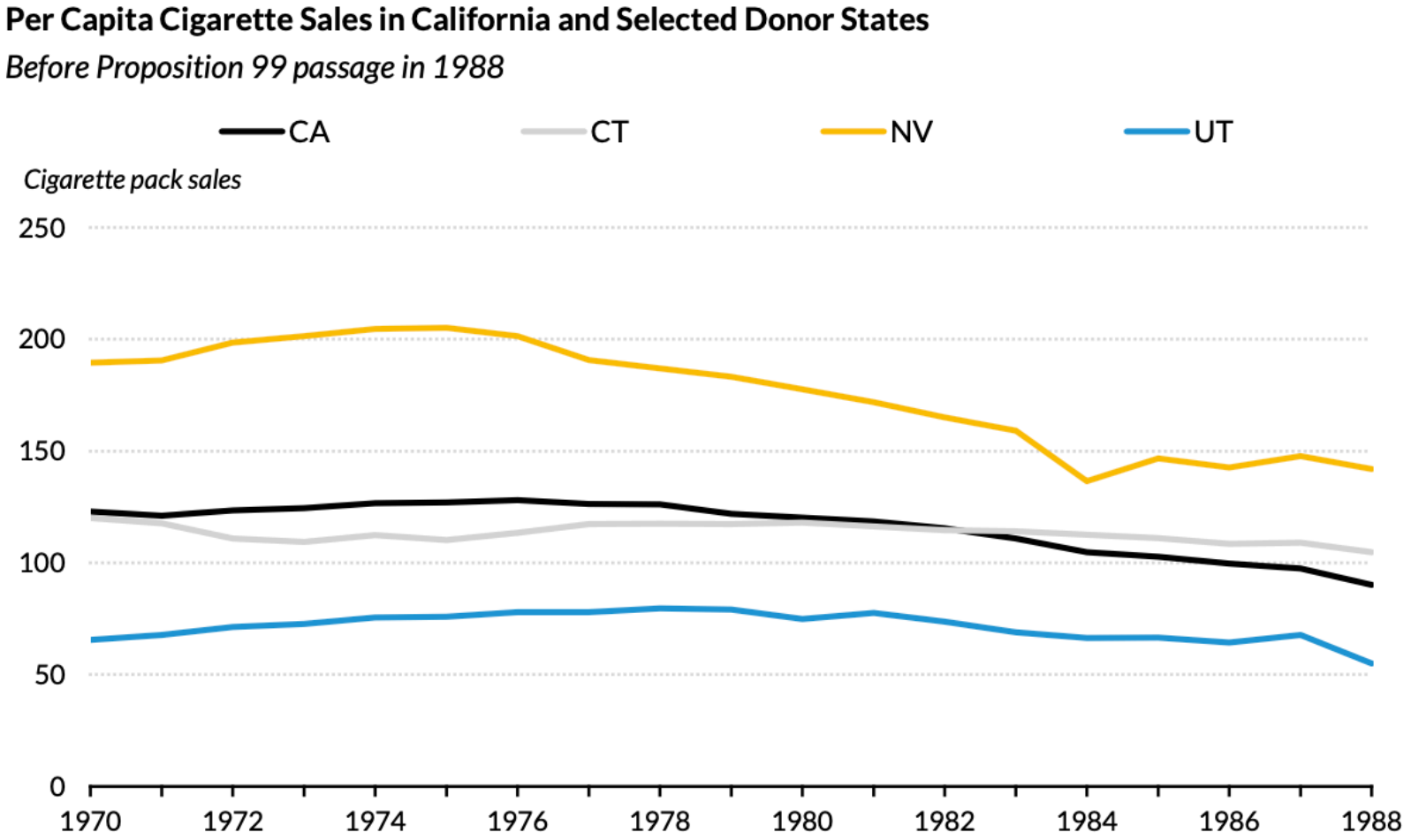
Abadie and Gardeazabel (2003)

What is a synthetic control?

- A synthetic control is a weighted average of available control units that approximates the most relevant characteristics of the treated unit prior to the treatment
- The synthetic control mimics the values of the predictors of the outcome, including pre-intervention values of the outcome, for the treated unit before the intervention occurred
- The synthetic control represents the counterfactual scenario for a treated unit in the absence of the intervention under scrutiny
- Intuition: A weighted combination of comparison units (the “synthetic control”) provides a better comparison for the treated unit than any single comparison unit alone

Example of 1999 cigarette sales tax in California

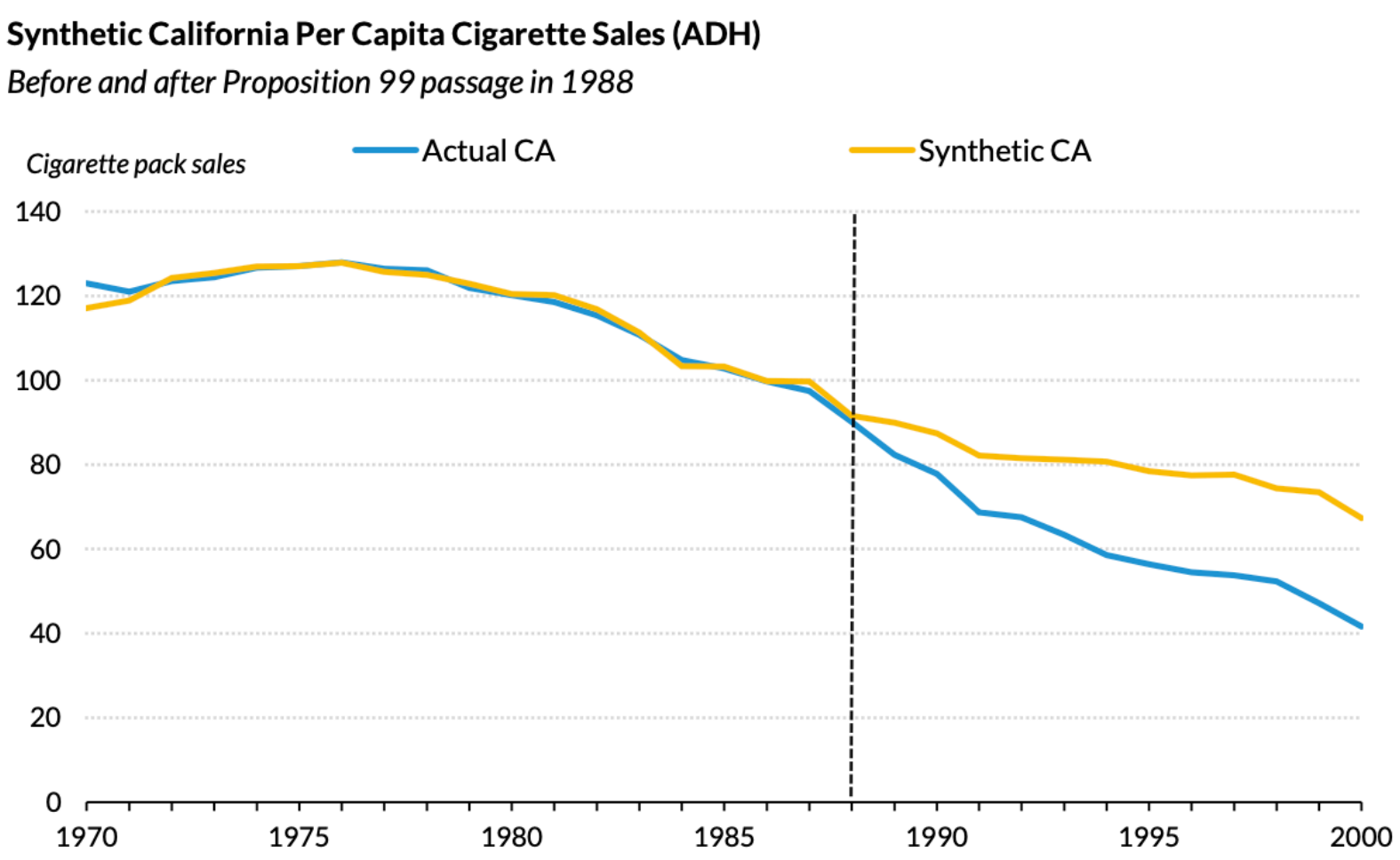
- No control state looks like a good 'match'.
- SC creates a weighted control.



McLelland (2017)

Example of 1999 cigarette sales tax in California

- No control state looks like a good 'match'.
- SC creates a weighted control.



McLelland (2017)

Spruk and Kovac *Swiss Journal of Economics and Statistics* (2020) 156:4
https://doi.org/10.1186/s41937-019-0048-0

Swiss Journal of
Economics and Statistics

ORIGINAL ARTICLE Open Access

Does a ban on trans fats improve public health: synthetic control evidence from Denmark

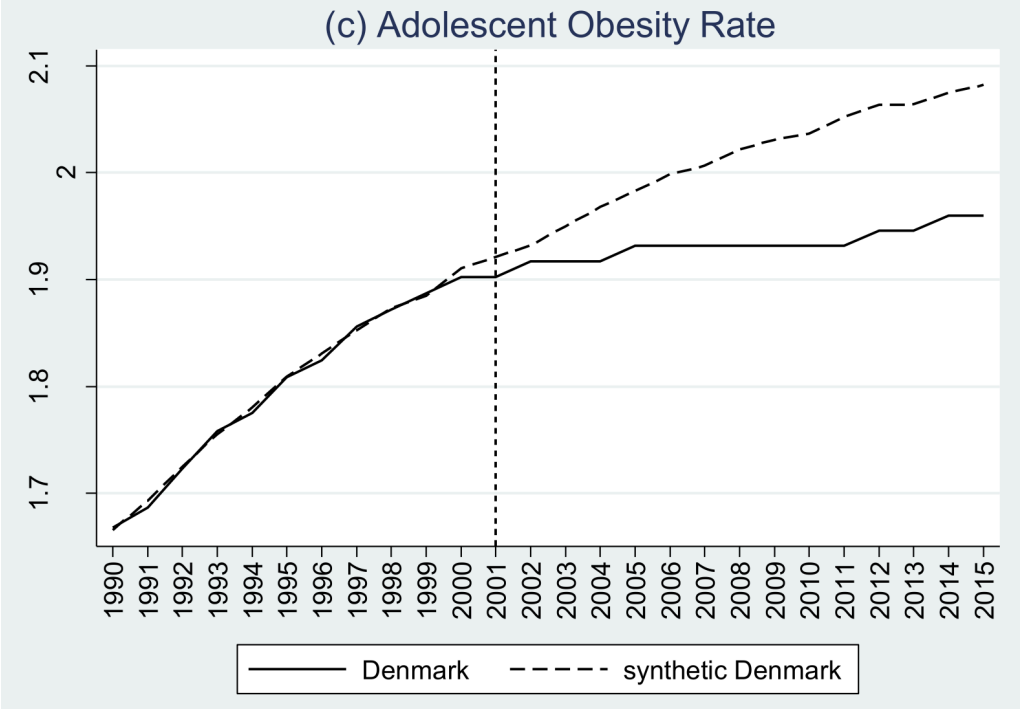
Rok Spruk* and Mitja Kovac

Abstract

We examine the impact of the trans fat ban on a variety of public health outcomes. To this end, we consider a de facto trans fat ban that was introduced by Denmark in 2001. Using the synthetic control method, parallel trends between Denmark and countries in a control group in the years prior to the ban are used to construct a "synthetic Denmark" without any such trans fat ban. Our synthetic control estimates suggest the ban led to substantial improvements in public health. Following the ban, cardiovascular mortality dropped considerably, while the trends of adolescent and child obesity came to a halt and decreased significantly compared to the synthetic control group. Our findings provide new insights into the benefits for public health arising from the banning of trans fats.

- 'Synthetic' DK mostly SWE, ITA, USA, and FIN.
- Also declines in CVD mortality.
- Robustness checks:

> By deliberately assigning TFA policy to wrong dates and other countries, we show the effect of the 2001 TFA policy intervention is specific to Denmark and does not appear to be driven by alternative dates

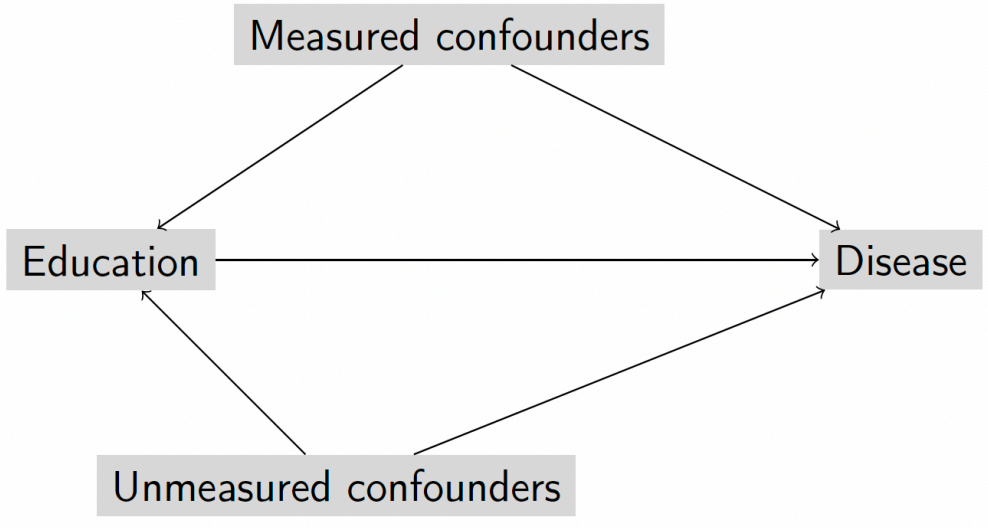


Spruk and Kovac (2020)

Instrumental Variables

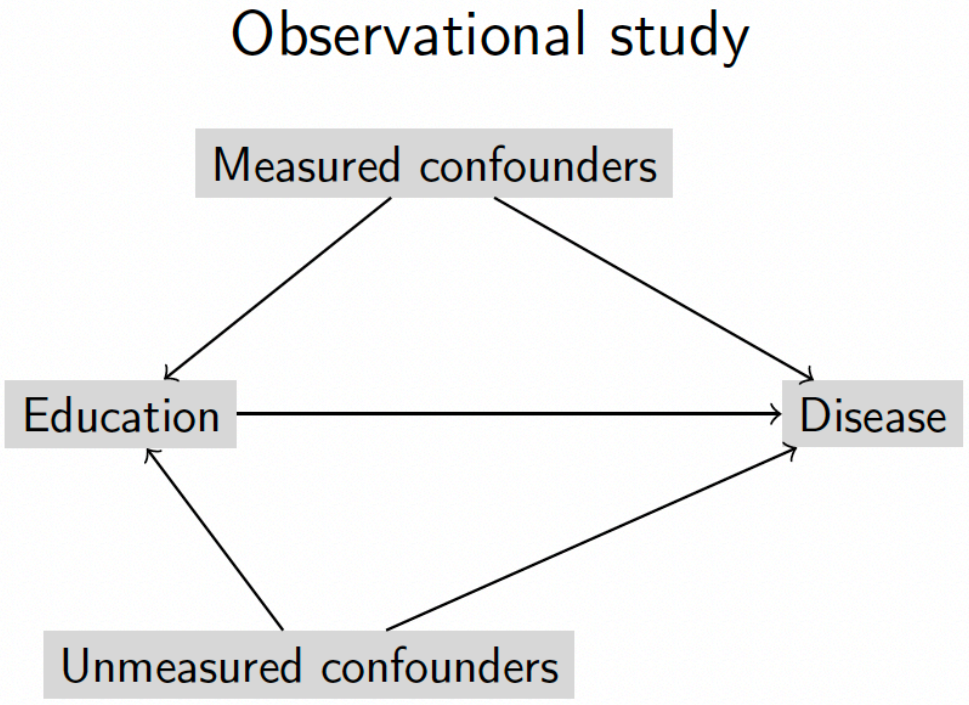
Challenge of conventional observation study (again)

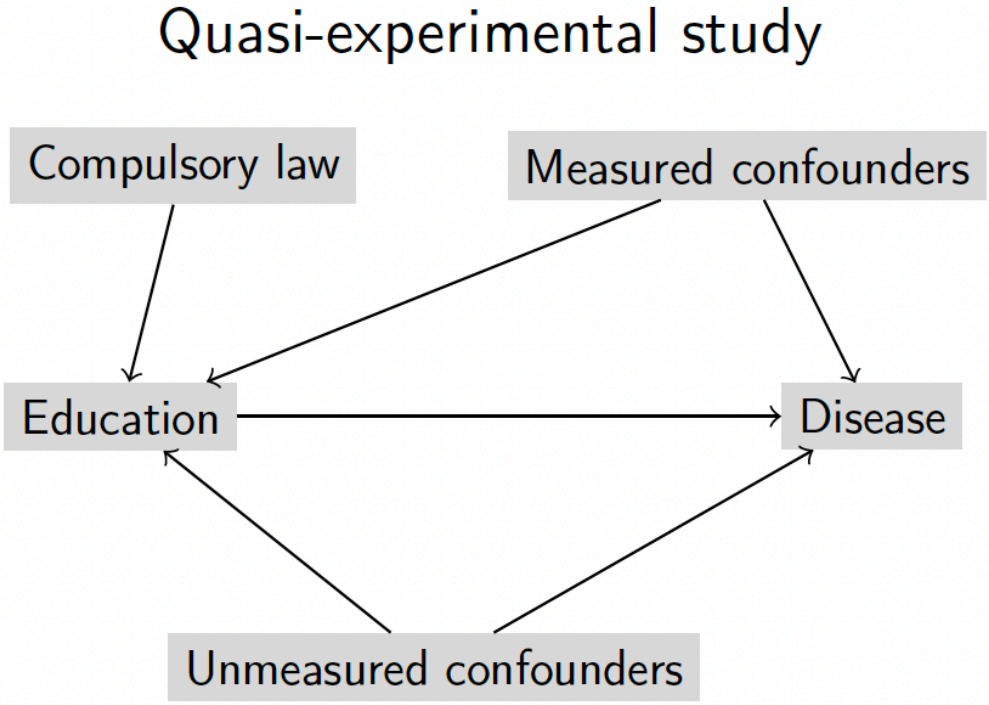
- WHO: "Educational attainment is linked to improved health outcomes."
- But what about unmeasured confounding? Unmeasured factors such as personality traits, cognitive ability, etc. may be predictive of both education and disease.
- Failure to measure such factors will falsely attribute their effects to education.



Possible solution: Quasi-experiment

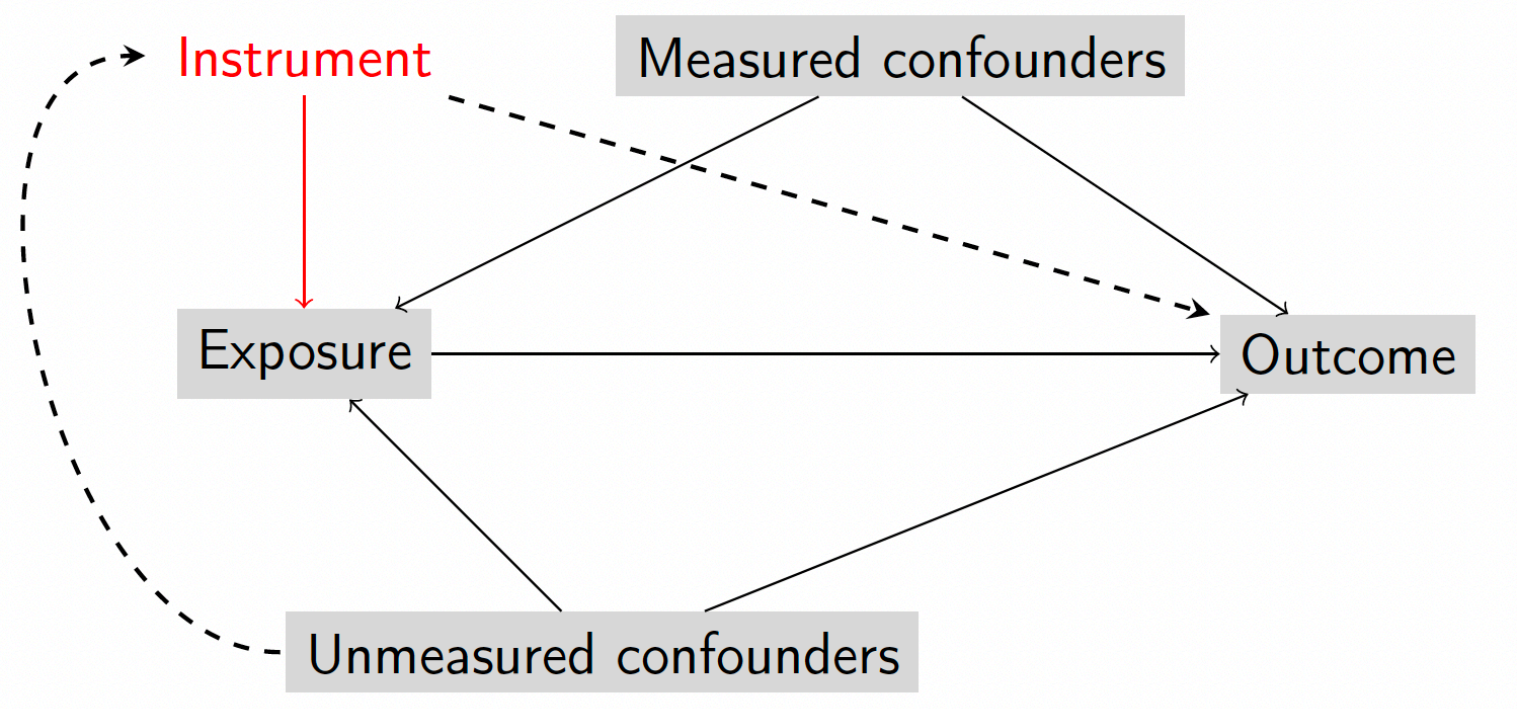
"Instrumental variable": predicts education but not associated with anything else [mimicking random assignment].





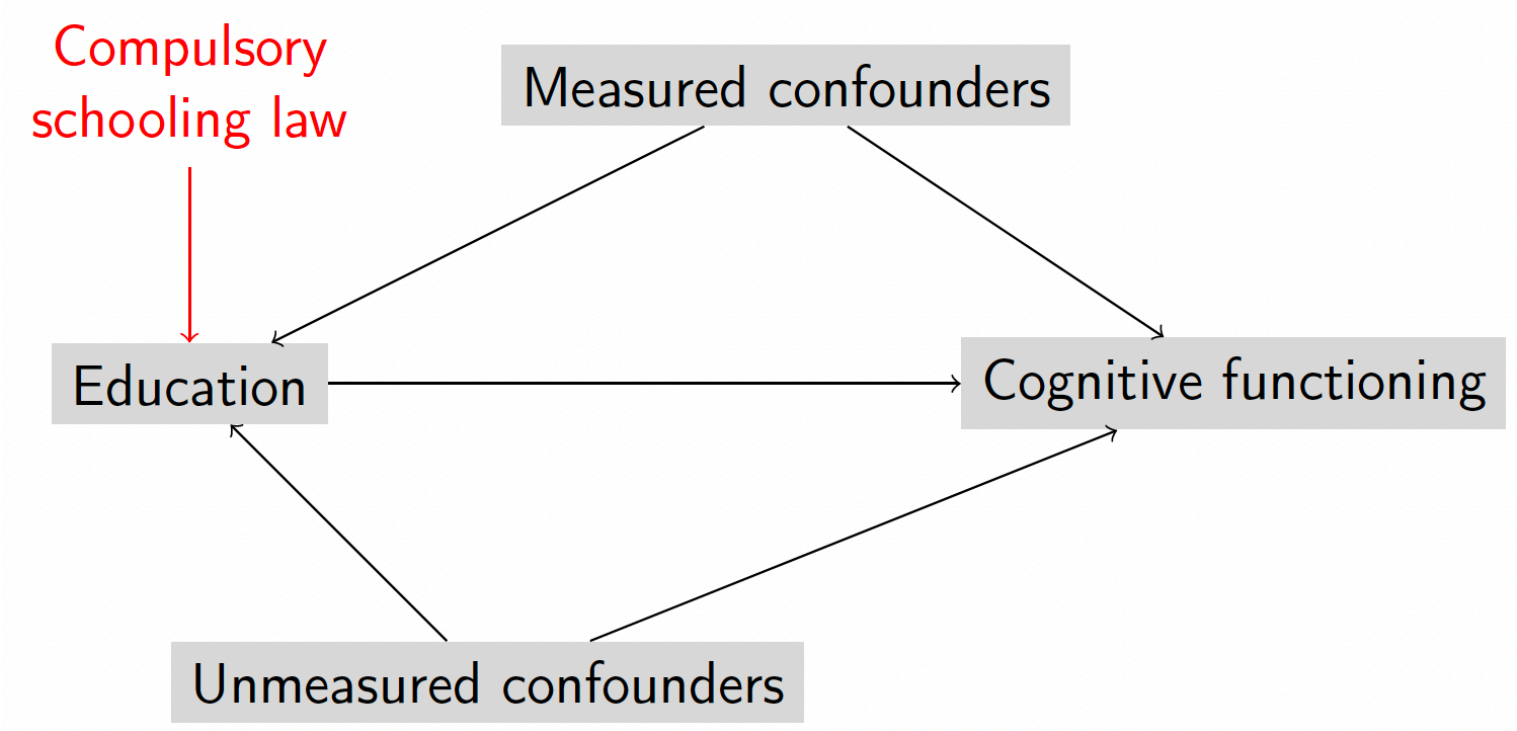
Non-randomized instrument creates additional issues

- In an RCT we know the treatment assignment is not associated directly with the outcome or with other unmeasured common causes.
- This assumption is less credible when our "instrument" is non-randomized.



Non-randomized examples of IV: Policies

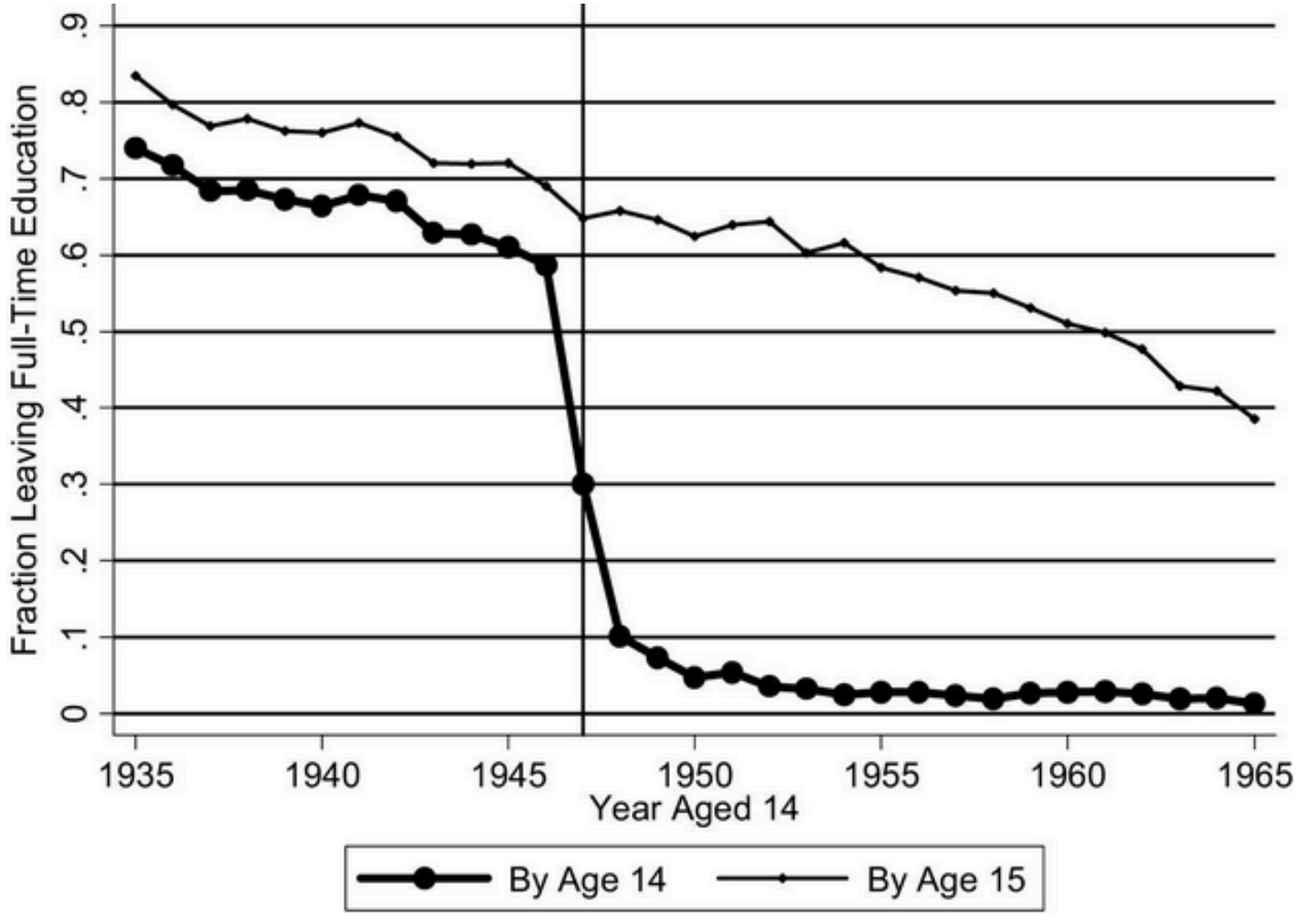
- Does education affect cognitive functioning?
- **Instrument**: changes in compulsory schooling laws [mimicking random assignment].



Glymour et al. (2008)

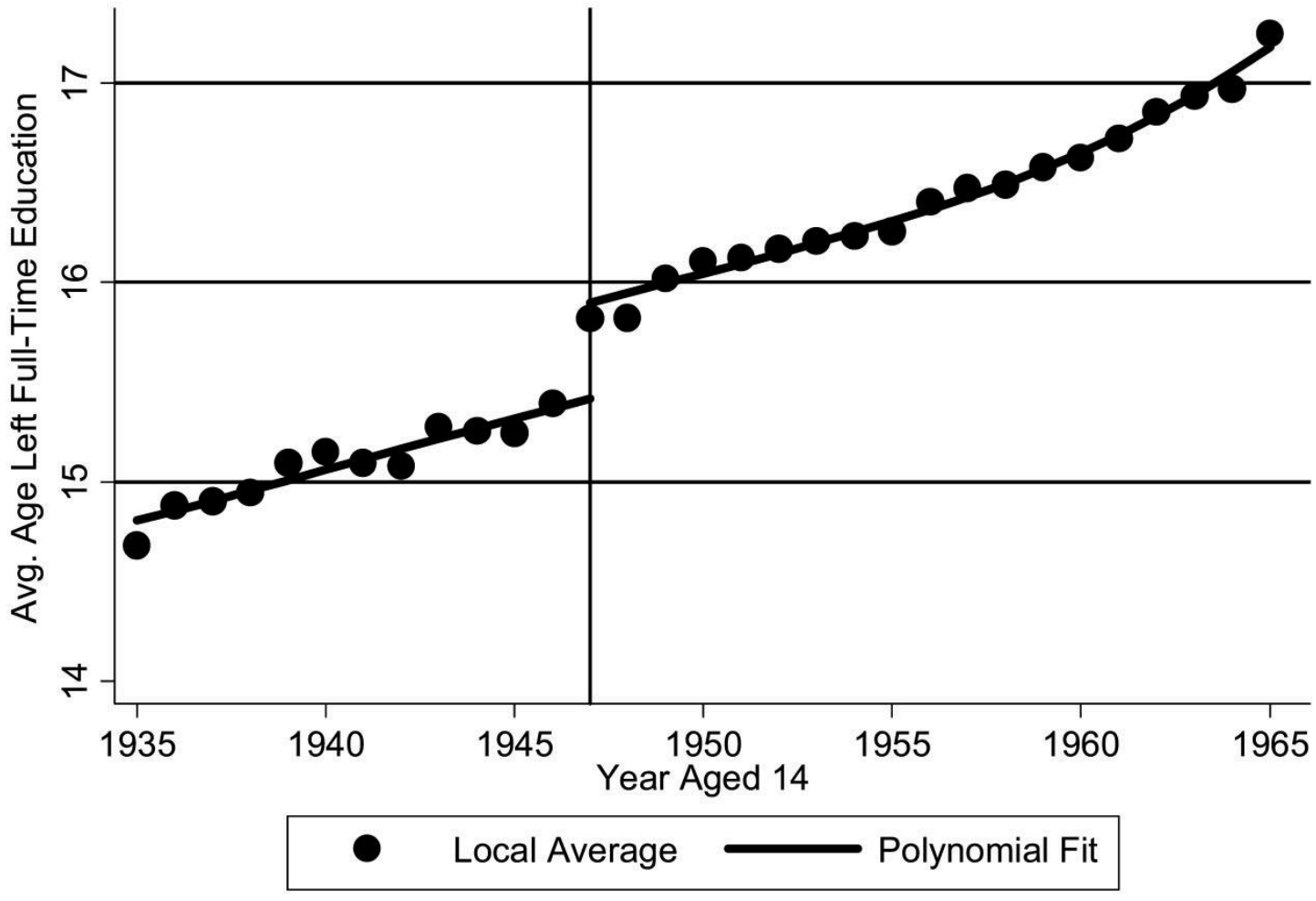
What does a quasi-experiment look like?

Fraction left full-time education by year aged 14 and 15 (Great Britain)



The lower line shows the proportion of British-born adults aged 32 to 64 from the 1983 to 1998 General Household Surveys who report leaving full-time education at or before age 14 from 1935 to 1965. The upper line shows the same, but for age 15. The minimum school-leaving age in Great Britain changed in 1947 from 14 to 15 [Oreopoulos (2006)]

Average schooling increases by exactly half a year between the cohorts that were age 14 in 1946 and in 1948.



Ex: Education and HIV

Length of secondary schooling and risk of HIV infection in Botswana: evidence from a natural experiment

Jan-Walter De Neve, Günther Fink, S V Subramanian, Sikhulile Moyo, Jacob Bor

Summary

Background An estimated 2·1 million individuals are newly infected with HIV every year. Cross-sectional and longitudinal studies have reported conflicting evidence for the association between education and HIV risk, and no randomised trial has identified a causal effect for education on HIV incidence. We aimed to use a policy reform in secondary schooling in Botswana to identify the causal effect of length of schooling on new HIV infection.

Methods Data for HIV biomarkers and demographics were obtained from the nationally representative household 2004 and 2008 Botswana AIDS Impact Surveys (N=7018). In 1996, Botswana reformed the grade structure of secondary school, expanding access to grade ten and increasing educational attainment for affected cohorts. Using exposure to the policy reform as an instrumental variable, we used two-stage least squares to estimate the causal effect of years of schooling on the cumulative probability that an individual contracted HIV up to their age at the time of the survey. We also assessed the cost-effectiveness of secondary schooling as an HIV prevention intervention in comparison to other established interventions.

Findings Each additional year of secondary schooling caused by the policy change led to an absolute reduction in the cumulative risk of HIV infection of 8·1 percentage points (p=0·008), relative to a baseline prevalence of 25·5% in the pre-reform 1980 birth cohort. Effects were particularly large in women (11·6 percentage points, p=0·046). Results were robust to a wide array of sensitivity analyses. Secondary school was cost effective as an HIV prevention intervention by standard metrics (cost per HIV infection averted was US$27 753).

Interpretation Additional years of secondary schooling had a large protective effect against HIV risk in Botswana, particularly for women. Increasing progression through secondary school could be a cost-effective HIV prevention measure in HIV-endemic settings, in addition to yielding other societal benefits.

Funding Takemi Program in International Health at the Harvard T.H.Chan School of Public Health, Belgian American Educational Foundation, Fernand Lazard Foundation, Boston University, National Institutes of Health.

Lancet Glob Health 2015; 3: e470–77

Published Online June 29, 2015 http://dx.doi.org/10.1016/S2214-109X(15)00087-X

See Comment page e428

Department of Global Health and Population (J-W De Neve MD, G Fink PhD), **Department of Social and Behavioral Sciences** (S V Subramanian PhD), **and Center for Population and Development Studies** (G Fink, S V Subramanian), **Harvard T.H. Chan School of Public Health, Boston, MA, USA; Botswana–Harvard AIDS Institute Partnership, Gaborone, Botswana** (S Moyo MSc); **and Department of Global Health and Center for Global Health and Development, Boston University School of Public Health, Boston, MA, USA** (Jacob Bor ScD)

Correspondence to: Jacob Bor, Department of Global Health, Boston University School

Regression Discontinuity

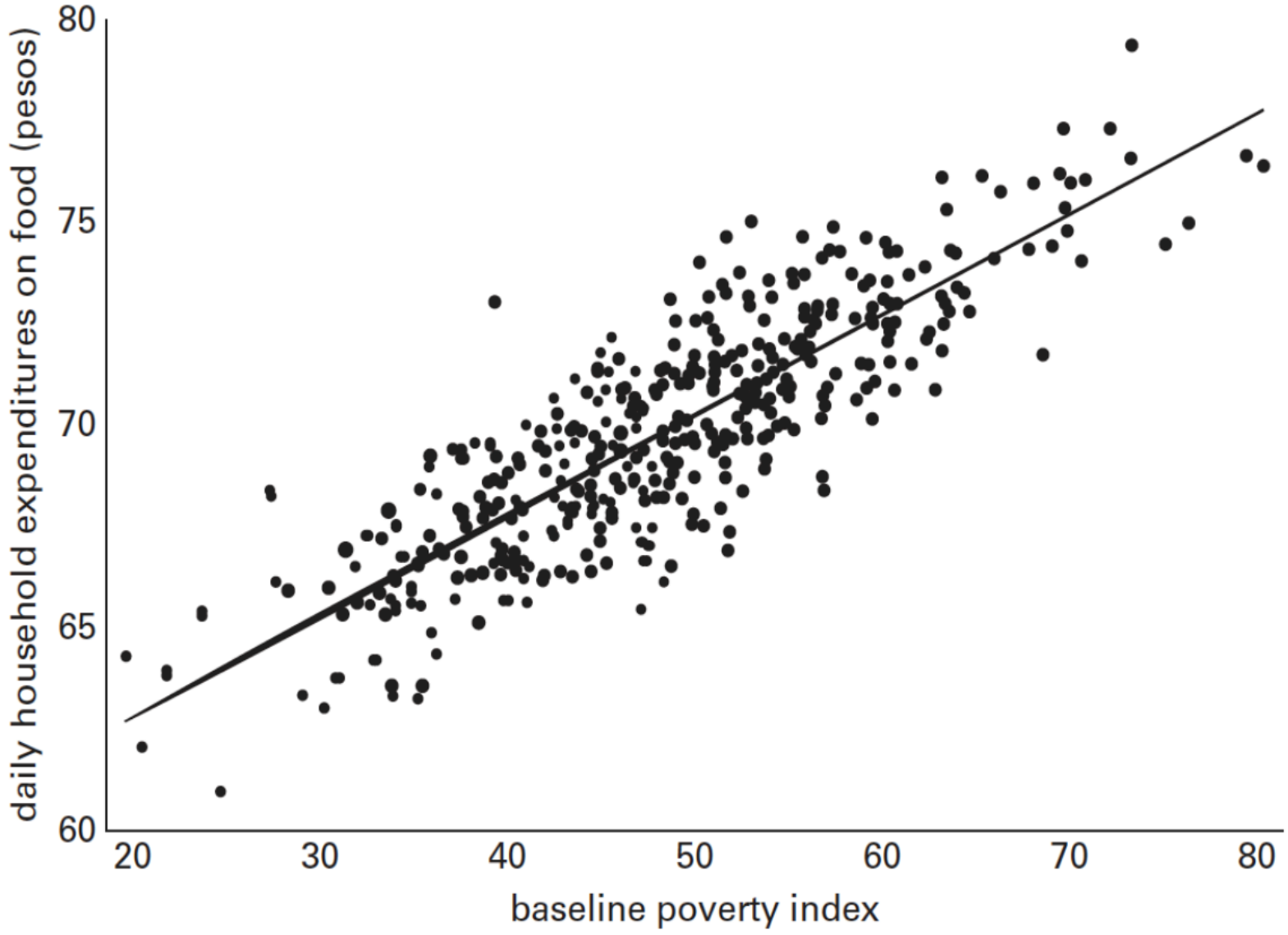
RD: Basic Idea

- Take advantage of arbitrary thresholds that sometimes assign treatment to individuals.
- When an administrative or rule-based cutoff in a continuous variable (present in your data) predicts treatment assignment, being on one side or the other of this cutoff determines, or predicts, treatment received.
- The continuous variable is called the "assignment" or "forcing" variable.
- Groups just on either side are the threshold considered "as good as randomly" assigned to treatment.

RD: Motivating example

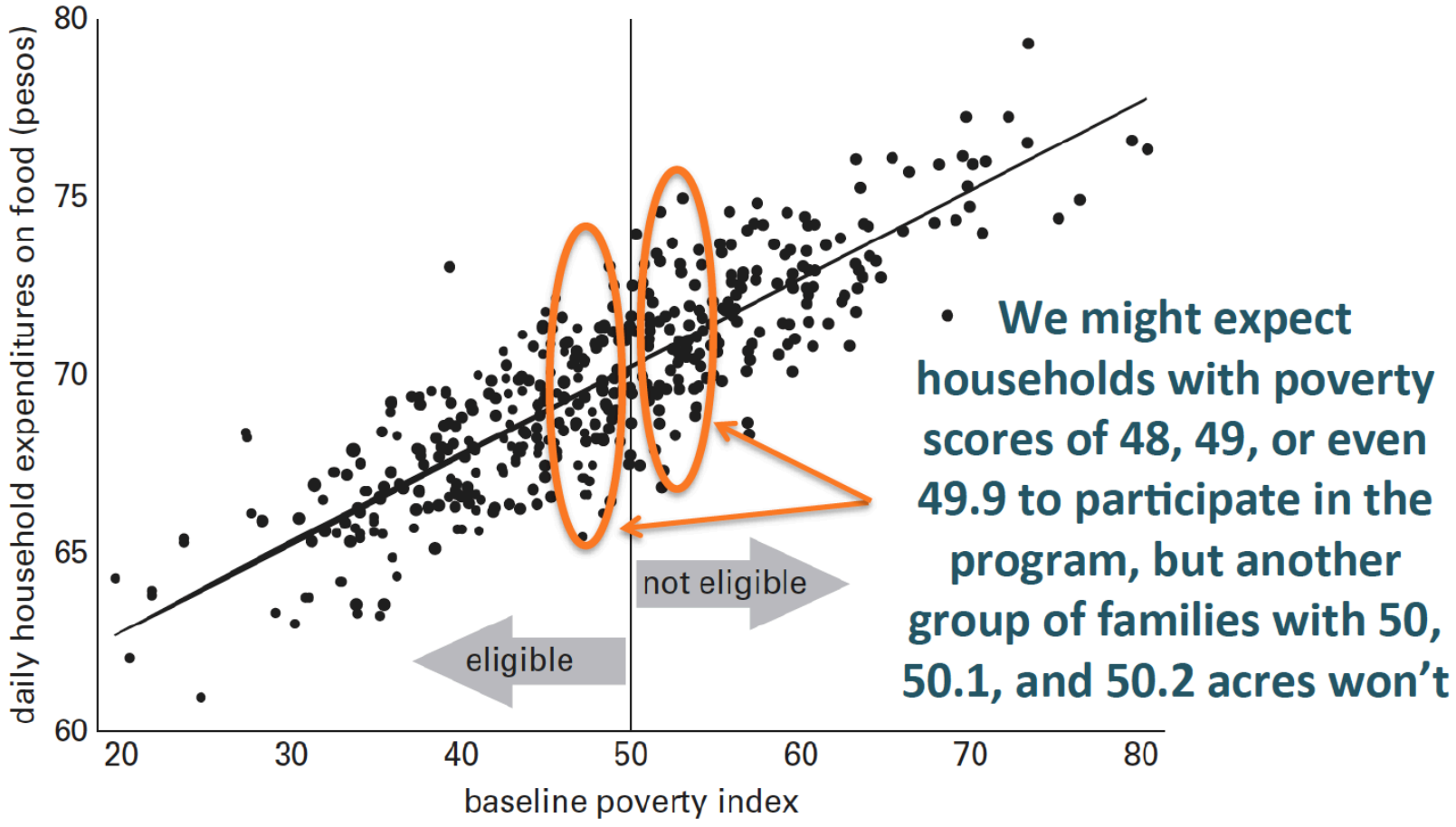
- Suppose we want to estimate the impact of a cash transfer program on daily food expenditure of poor households.
- Poverty is measured by a continuous score between 0 and 100 that is used to rank households from poorest to richest.
- Poverty is the assignment variable, Z, that determines eligibility for the cash transfer program.
- The outcome of interest, daily food expenditure, is denoted by Y.

At baseline, you might expect poorer households to spend less on food, on average, than richer ones, which might look like this



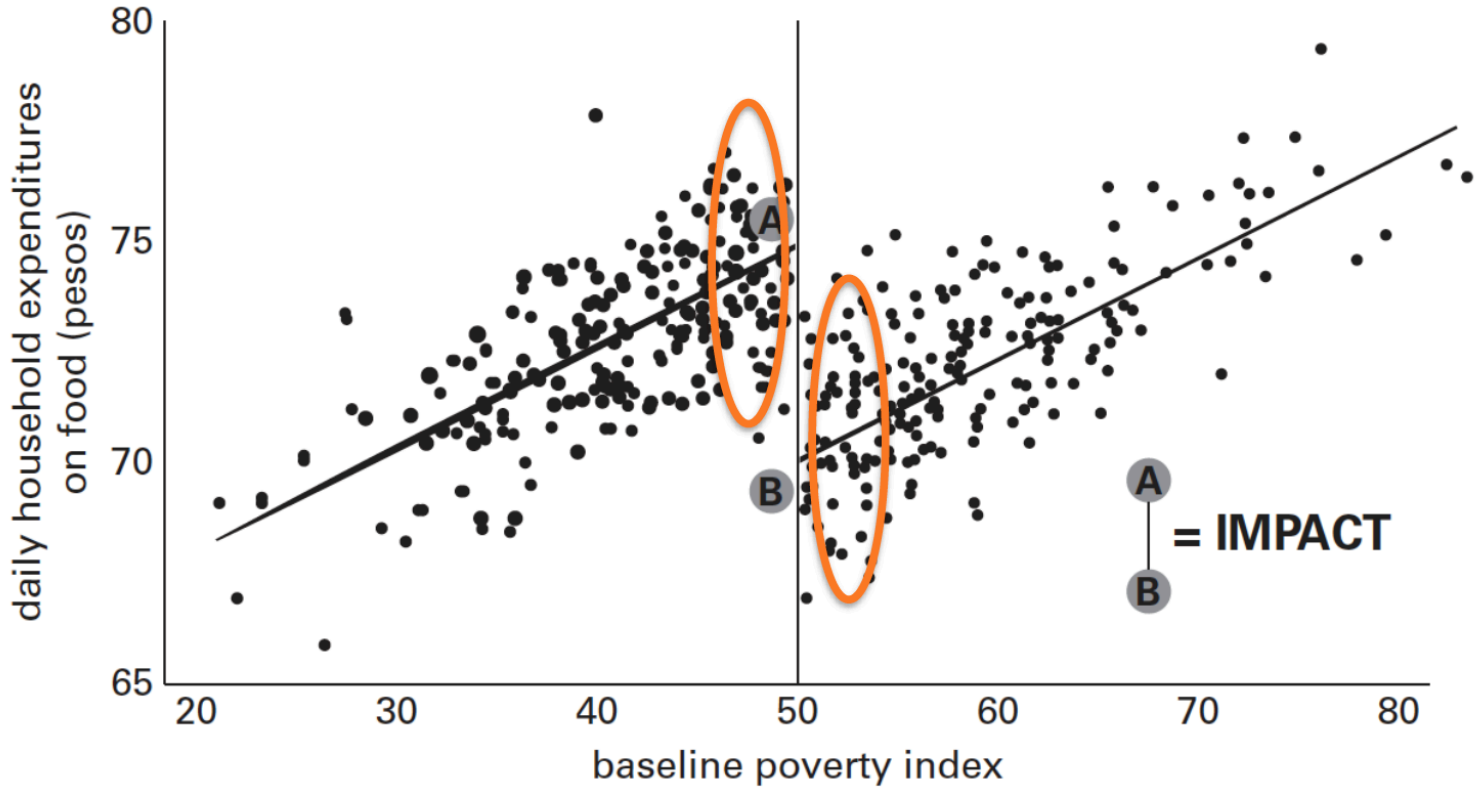
Gertler (2011)

Under the program's rules, only households with a poverty score, Z, below 50 are eligible for the cash payment



Gertler (2011)

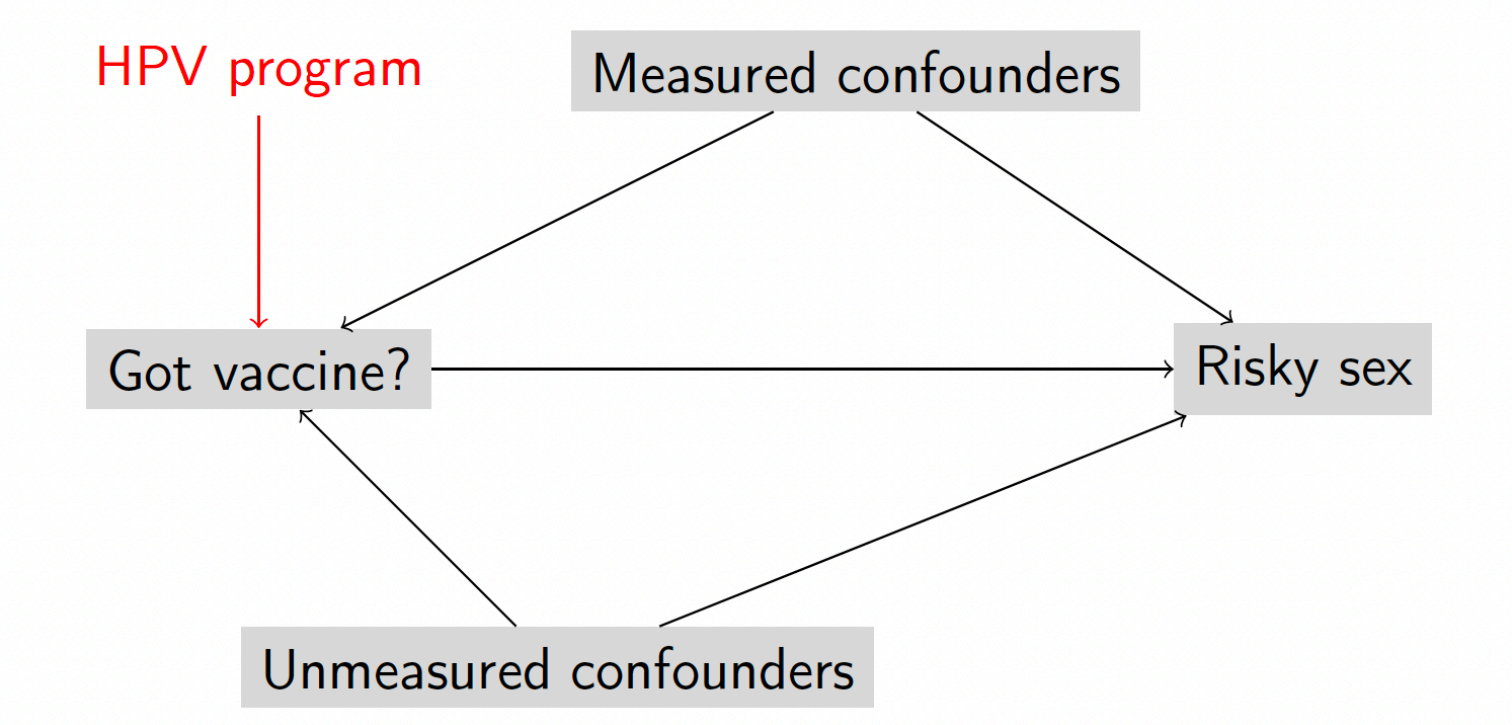
As we approach the cutoff value from above and below, the individuals in both groups become more and more alike, on both measured and unobserved characteristics---in a small area around the threshold, the only difference is in treatment assignment



Gertler (2011)

Applied example: HPV vaccine and sexual behaviors

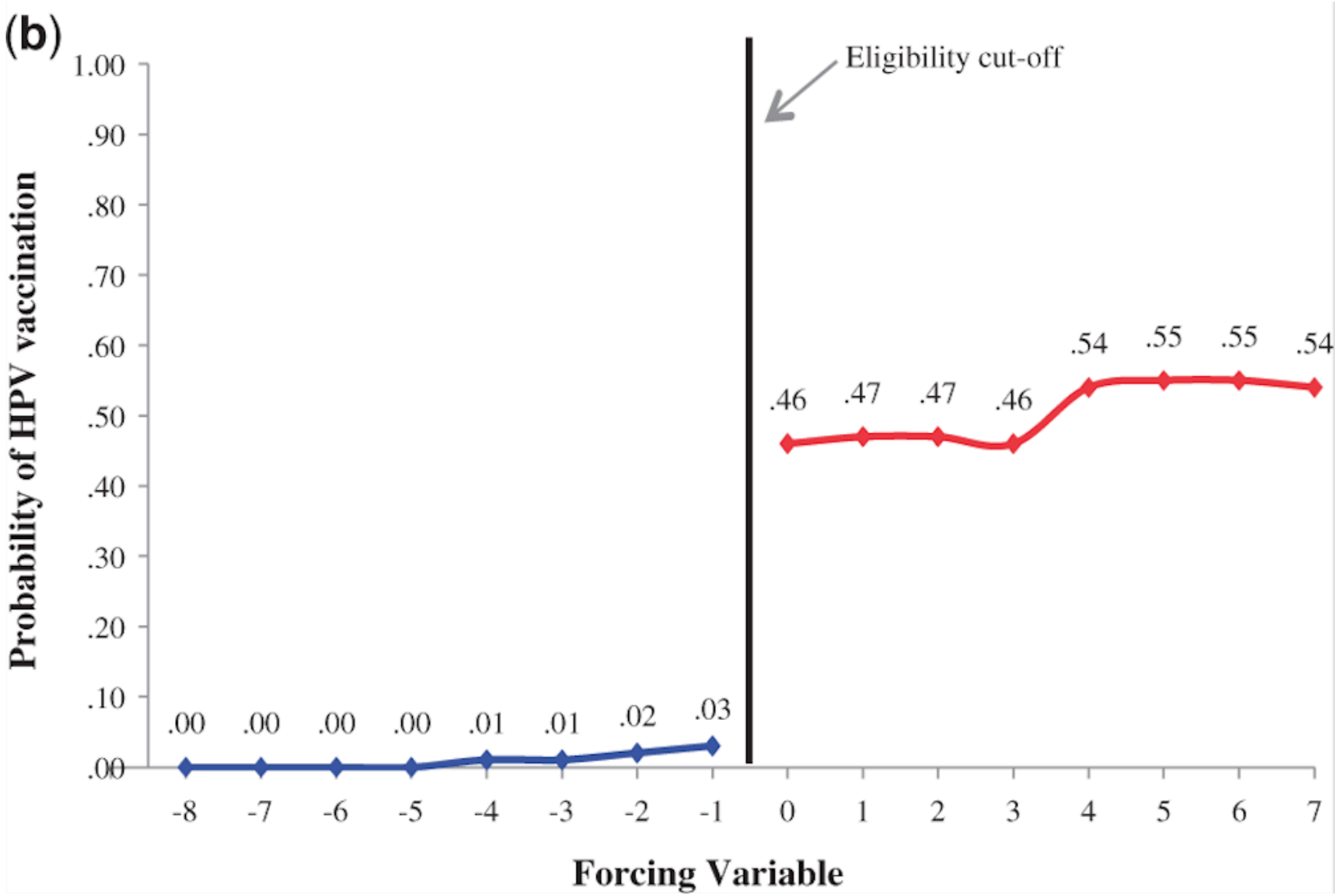
- Does getting the HPV vaccine affect sexual behaviors?
- Vaccine policy: predicts vaccine receipt but (**we assume**) .red[not] associated with anything else [mimicking random assignment].



Smith et al. (2015)

Does the cutoff predict treatment?

- Girls "assigned" to HPV program by quarter of birth.
- Pr(vaccine) jumps discontinuously at cutoff



Smith et al. (2015)

What does a credible natural experiment look like?

Table 1: Baseline characteristics of the eligibility groups in the study cohort

Characteristic	Program eligibility group; % of eligibility group*		Characteristic	Program eligibility group; % of eligibility group*	
	Ineligible (n = 131 781)	Eligible (n = 128 712)		Ineligible (n = 131 781)	Eligible (n = 128 712)
Sociodemographic†			**Health services use**††**		
Age, yr, mean ± SD	13.17 ± 0.28	13.17 ± 0.28	Hospital admission		
Birth quarter			0	98.0	98.2
Jan.–Mar.	24.3	24.2	≥ 1	2.0	1.8
Apr.–June	26.1	26.1	LOS, d, mean ± SD	7.4 ± 15.6	8.0 ± 18.2
July–Sept.	25.7	25.8	Same-day surgery		
Oct.–Dec.	23.9	23.9	0	97.7	97.8
Residency			≥ 1	2.4	2.2
Urban	85.3	85.8	Emergency department visits		
Rural	14.0	13.5	0	70.7	71.1
Missing‡	0.7	0.6	1	18.1	17.8
Income quintile			≥ 2	11.2	11.1
1 (lowest)	16.6	15.0	Outpatient visits		
2	18.4	17.8	0 or 1	22.6	22.8
3	20.6	21.1	2–5	27.4	26.9
4	22.0	23.1	6–12	25.1	24.5
5 (highest)	21.4	22.1	≥ 13	25.0	25.8

Smith et al. (2015)

Note little impact of adjustment

Table 3: Effect of quadrivalent human papillomavirus vaccination on clinical indicators of sexual behaviour*

Outcome	No. of excess cases per 1000 girls (95% CI)	RR (95% CI)	Adjusted† RR (95% CI)
Effect of vaccine			
Composite outcome	–0.61 (–10.71 to 9.49)	0.96 (0.81 to 1.14)	0.98 (0.84 to 1.14)
Pregnancy	0.70 (–7.57 to 8.97)	0.99 (0.79 to 1.23)	1.00 (0.83 to 1.21)
STIs	–4.92 (–11.49 to 1.65)	0.81 (0.62 to 1.05)	0.81 (0.63 to 1.04)
Effect of program			
Composite outcome	–0.25 (–4.35 to 3.85)	0.99 (0.93 to 1.06)	1.00 (0.93 to 1.07)
Pregnancy	0.29 (–3.07 to 3.64)	1.00 (0.92 to 1.09)	1.01 (0.93 to 1.10)
STIs	–2.00 (–4.67 to 0.67)	0.92 (0.83 to 1.03)	0.92 (0.83 to 1.03)

Note: CI = confidence interval, RR = relative risk, STIs = sexually transmitted infections.
*To address the effect of birth timing that we observed, we used the entire bandwidth of data (i.e., all observations in the 1992 to 1995 birth cohorts) and included birth quarter as a covariate in the model. In all analyses, the birth cohorts closest to the cut-off (1993 and 1994) were weighted twice as heavily as those furthest from the cut-off (1992 and 1995).
†In this sensitivity analysis, we adjusted for neighbourhood income quintile, hepatitis B vaccination and history of sexual health–related indictor, as well as for birth quarter.

Smith et al. (2015)

Issues related to generalizability

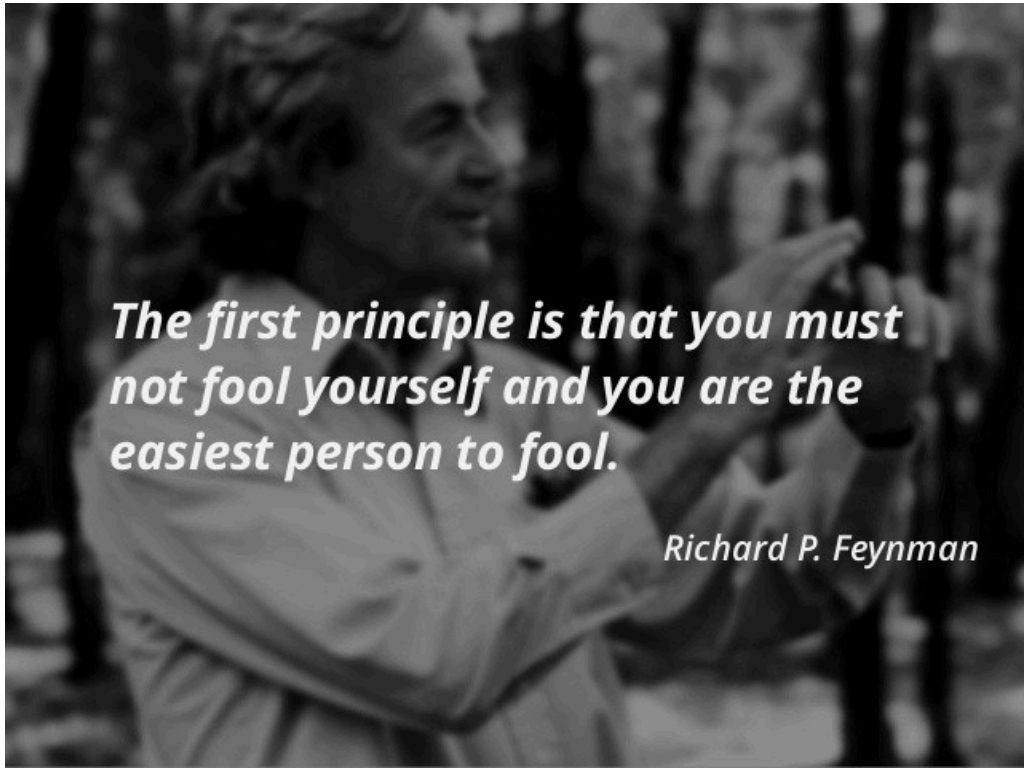
- RD estimates local average impacts around the eligibility cutoff where treated and control units are most similar and results cannot be generalized to units whose scores are further away from the cutoff (unless we assume treatment heterogeneity).
- If the goal is to answer whether the program should exist or not, then RD is likely not the appropriate methodology.
- However, if the question is whether the program should be cut or expanded at the margin, then it produces the local estimate of interest to inform this policy decision

Quasi-Experiments

1. Motivation
2. Randomization and Observation
3. Quasi-Experimental Designs
4. Final Thoughts

Be careful, and skeptical

- Correlations between social factors and health are easy to find.
- They do not necessarily reflect **causal** relationships.
- Need to search hard for alternative explanations.
- Important to consider the strength of evidence in considering interventions.

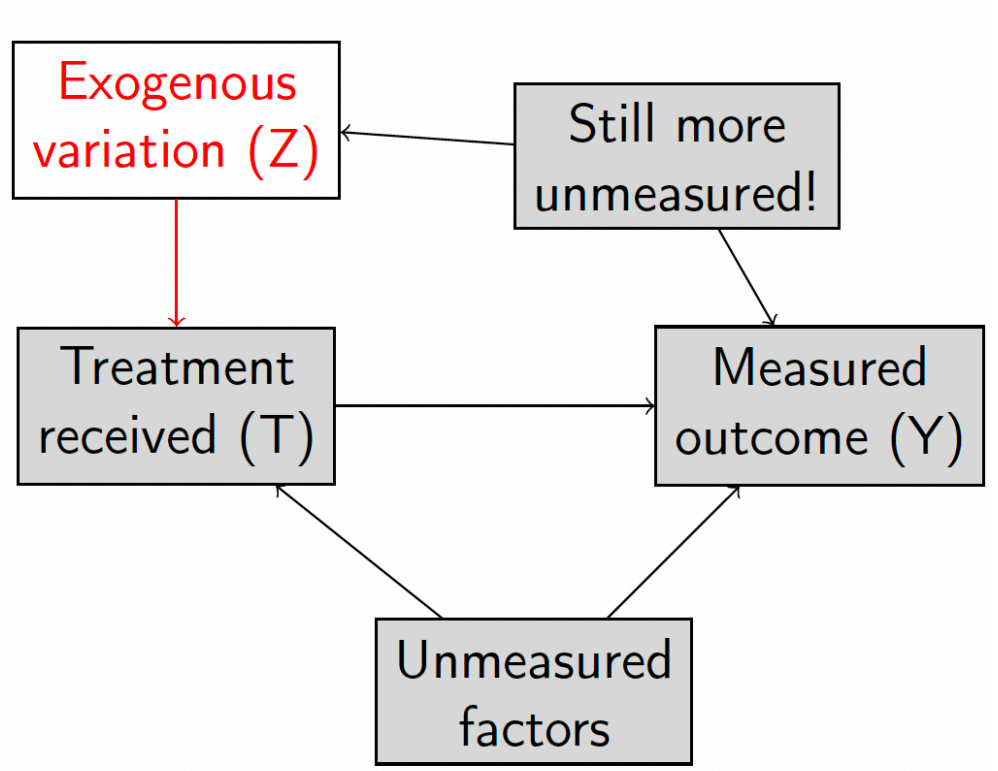


Are natural experiments always more credible?

- Not necessarily, but probably.
- Key is "as-if" randomization of treatment:
  - If this is credible, it is a much stronger **design** than most observational studies.
  - Should eliminate self-selection in to exposure groups.
- Allows for simple, transparent analysis of average differences between groups.
- Allows us to rely on weaker assumptions.

Assumptions still matter!

- Quasi-experimental studies are still observational.
- Most credible if they create unconditional randomized treatment groups (e.g., lottery).
- Credibility is continuous, not binary.
- I worry about the cognitive impact of the "quasi-experimental" label.



Back to basics: assumptions and costs

- Major benefit of randomized evaluations are that few assumptions are needed to estimate a causal effect.
- Necessary assumptions can often be checked.
- Non-randomization means more assumptions, more possibility for assumptions to be violated.
- Should lead us to spend lots of time trying to test the credibility of these assumptions.
    - How good is "as-if random"?
    - Are there compelling non-causal alternative explanations for the observed results?
- Not all non-randomized designs are created equal.