Quasi-experimental designs

Arijit Nandi (arijit.nandi@mcgill.ca)

Departments of Epidemiology & Equity, Ethics, and Policy

Acknowledgements/disclosures

Sam Harper and students of PPHS 617, “Impact Evaluation”

Conflicts of interest: None

What-ifs and counterfactuals

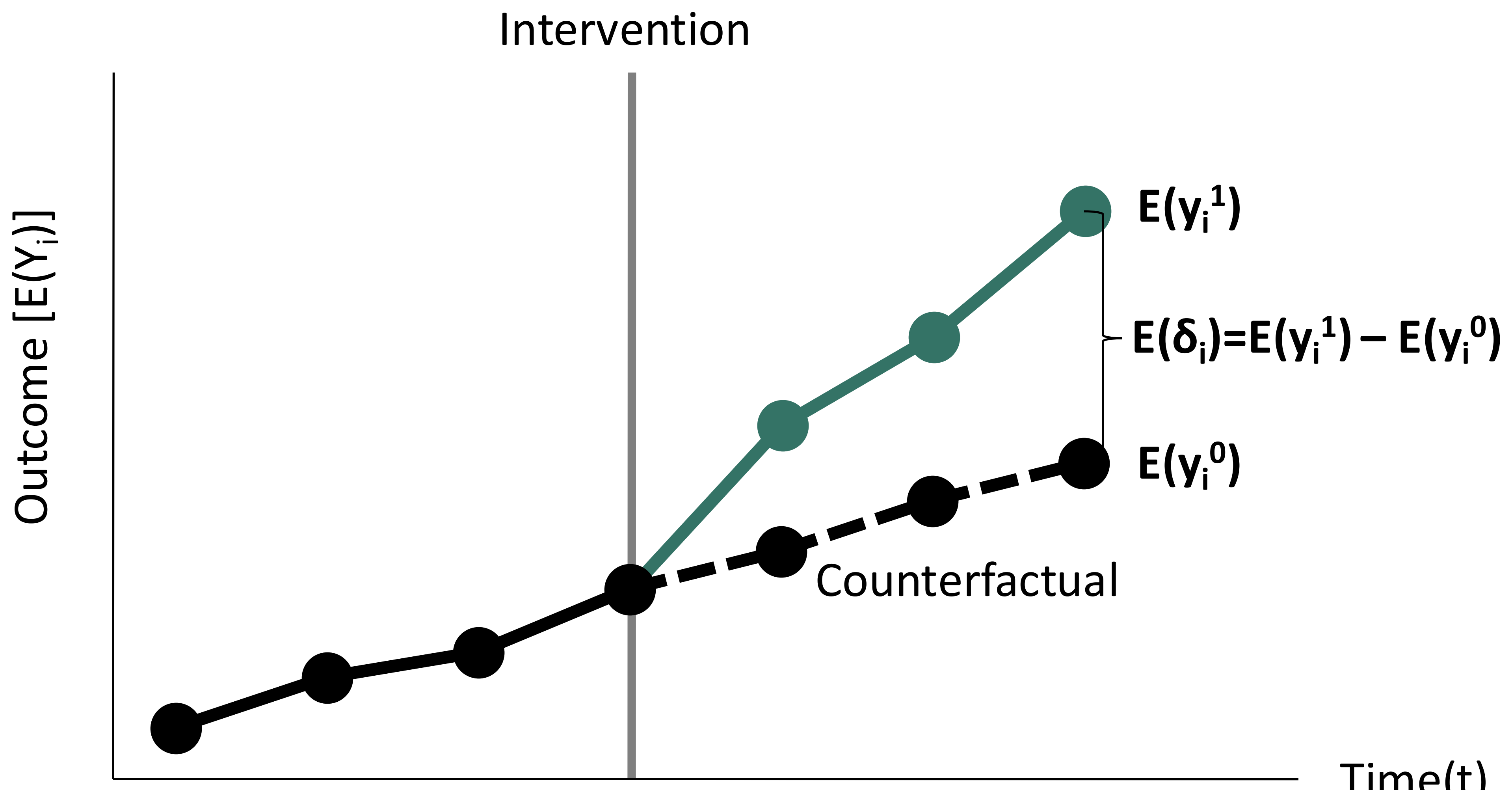
- Questions about the impact of a population-level interventions (e.g., the effect that can be causally attributed to a change in policy) are about **what-ifs.**

- Prospectively, we can think about how the world would be different if we intervened to change the status quo.

- Retrospectively, we can think about what would have been had we not implemented a particular policy or program.

- These alternative causal states are known as **counterfactuals.**

Fundamental challenge of causal inference



The “selection” problem

- Treatment and control groups are rarely exchangeable (i.e., programs/policies are selectively placed in different areas and the decision to participate is often voluntary).

- These differences could affect potential outcomes, creating bias.

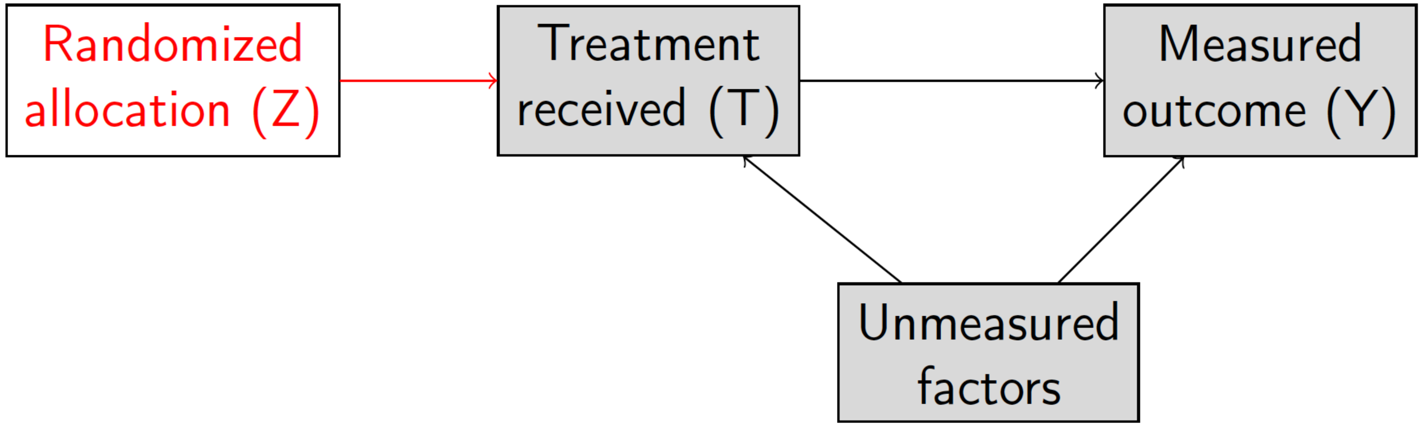
- Economist call this **selection** or **omitted variable** bias.

- In epidemiology, the effect of these pre-existing differences between groups is commonly called **confounding bias.**

RCTs are designed to address selection

- An RCT is characterized by: **(1)** comparison of treated and control groups; **(2)** randomized treatment assignment; and **(3)** investigator control over the randomizing.

- Randomization guarantees exchangeability on measured and unmeasured factors, and we can estimate the causal effect without confounding bias.



Challenges of population-level exposures

- When considering the social determinants of health, many exposures, whether social factors or policies/programs cannot be randomized:
  - Unethical (poverty, parental social class, job loss);
  - Impossible (ethnic background, place of birth);
  - Expensive (neighborhood environments, large-scale poverty policies)

- Moreover, some exposures are hypothesized to have long latency periods (many years before outcomes are observable).

- To measure impact, we need non-randomized alternatives to RCTs.

Consequences of non-randomized assignment

- If we are not controlling treatment assignment, then who is?

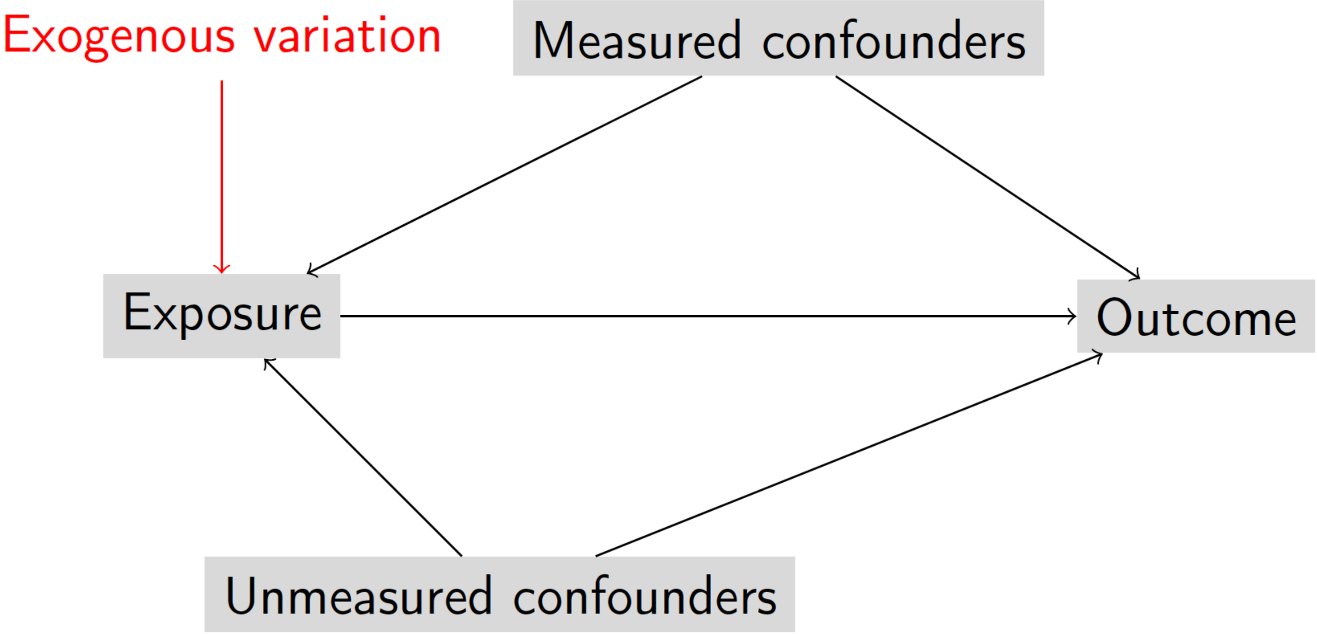
- Policy programs do not typically select people at random:
  - Programs target those that they think are most likely to benefit;
  - Programs implemented non-randomly (e.g., provinces passing drunk driving laws in response to high-profile accidents).

- People do not choose to participate in programs at random—for example:
  - Welfare programs, health screening programs, etc.;
  - People who believe they are likely to benefit from the program.

- **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.

Selection on "observables" and "unobservables"

- Observables: Things you measured or **can** measure.
- Unobservables: Things you **can't** measure (e.g., innate abilities).
- 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;
  - Adjustment;
  - Matching or weighting.

- Quasi-experiments aim to account for unmeasured factors **by design**:
  - Interrupted time series (ITS) and difference-in-differences (DD);
  - Synthetic controls (SC);
  - Instrumental variables (IV) and regression discontinuity (RD).

- In contrast to traditional observational studies, natural and quasi-experimental designs include some strategy for addressing selection on “unobservables”.

Natural experiments vs. quasi-experiments

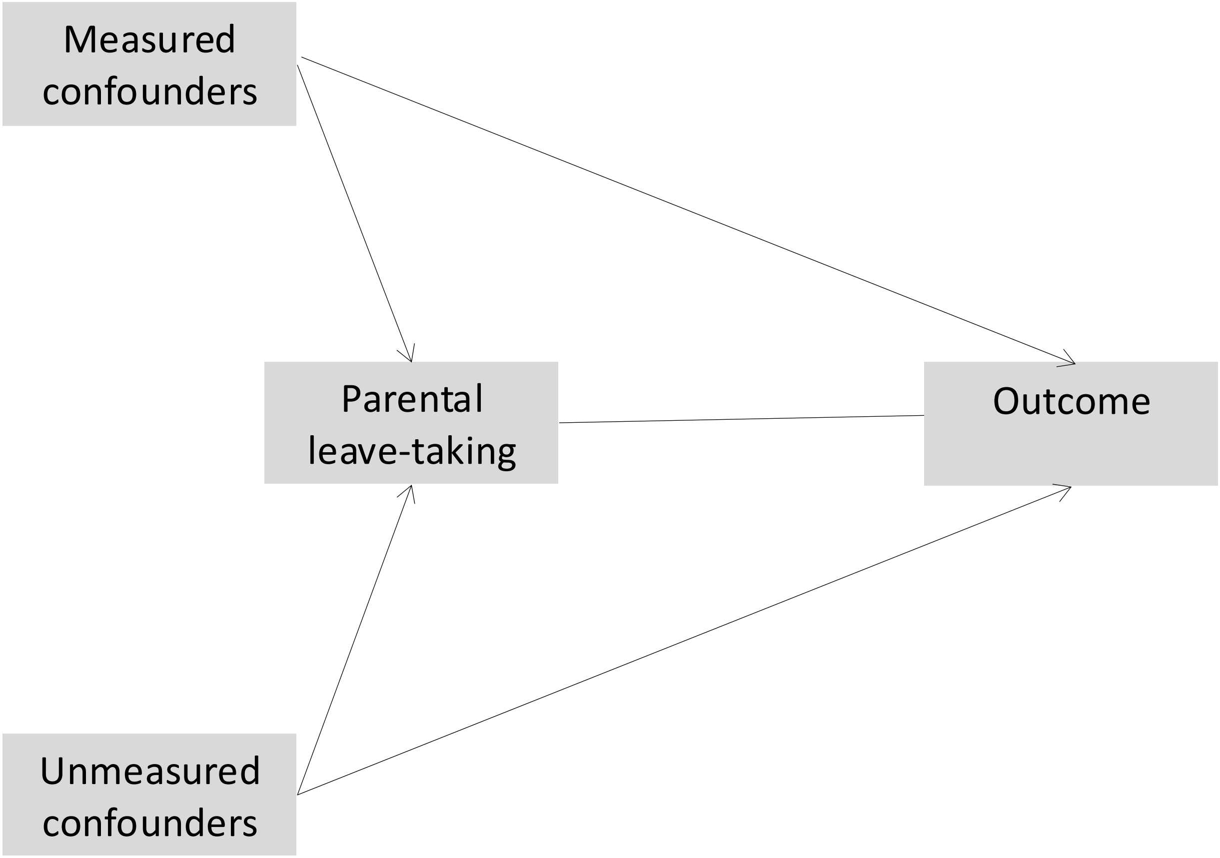
- Natural and quasi-experiments refer to “experiments that have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred.”(Cook, 1979)

- Natural experiments: Treatment groups are random or “as if” randomly assigned, but not by the investigator (e.g., lotteries, arbitrary treatment discontinuities, weather shocks).

- Quasi-experiments:
    - Assignment to treatment groups is not “naturally” random;
    - However, can make a convincing case for “as if” random assignment with added design features, controls, and (of course) assumptions.

Cook (1979); Dunning (2012)

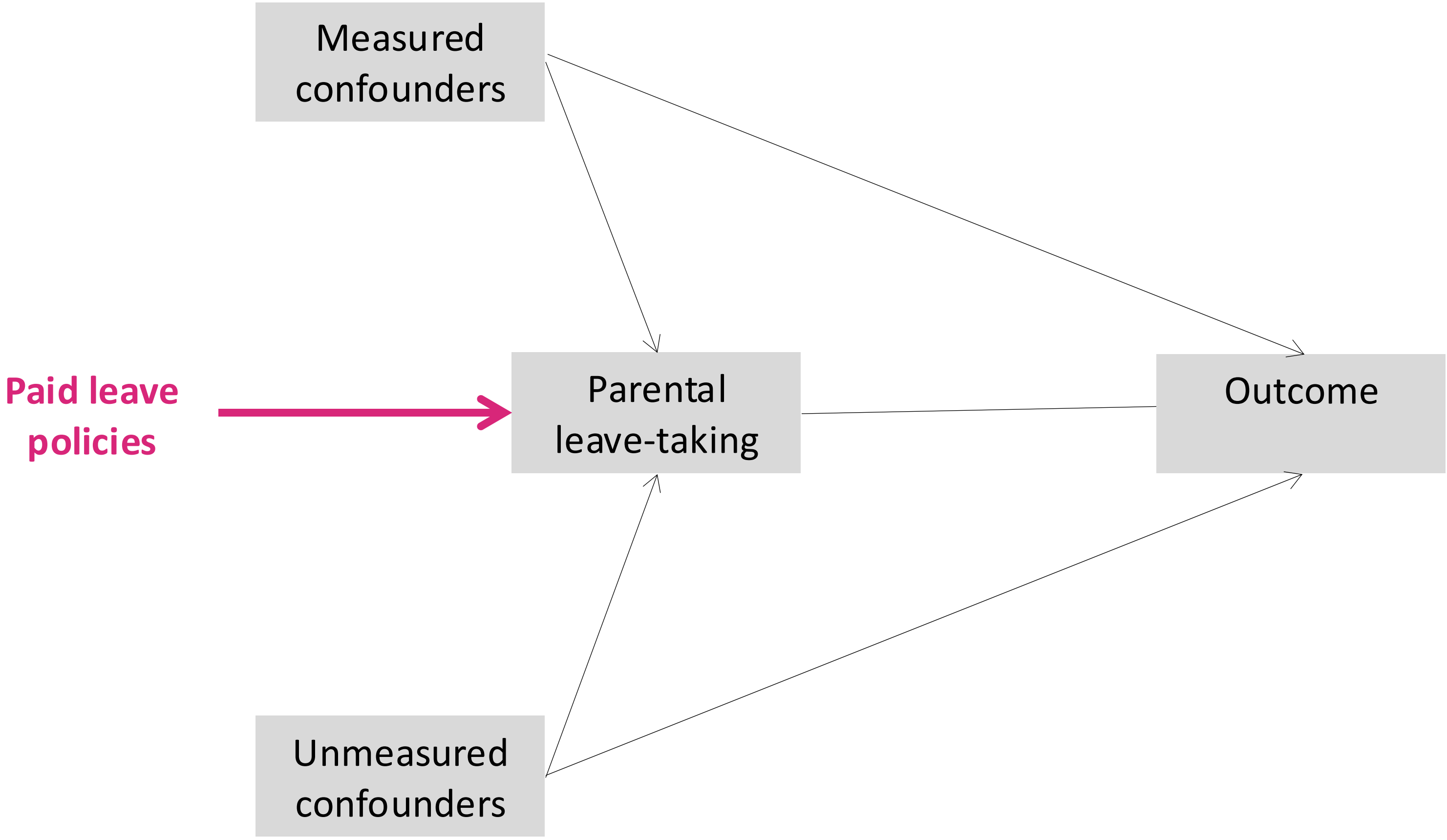
Example: impact of parental leave policies



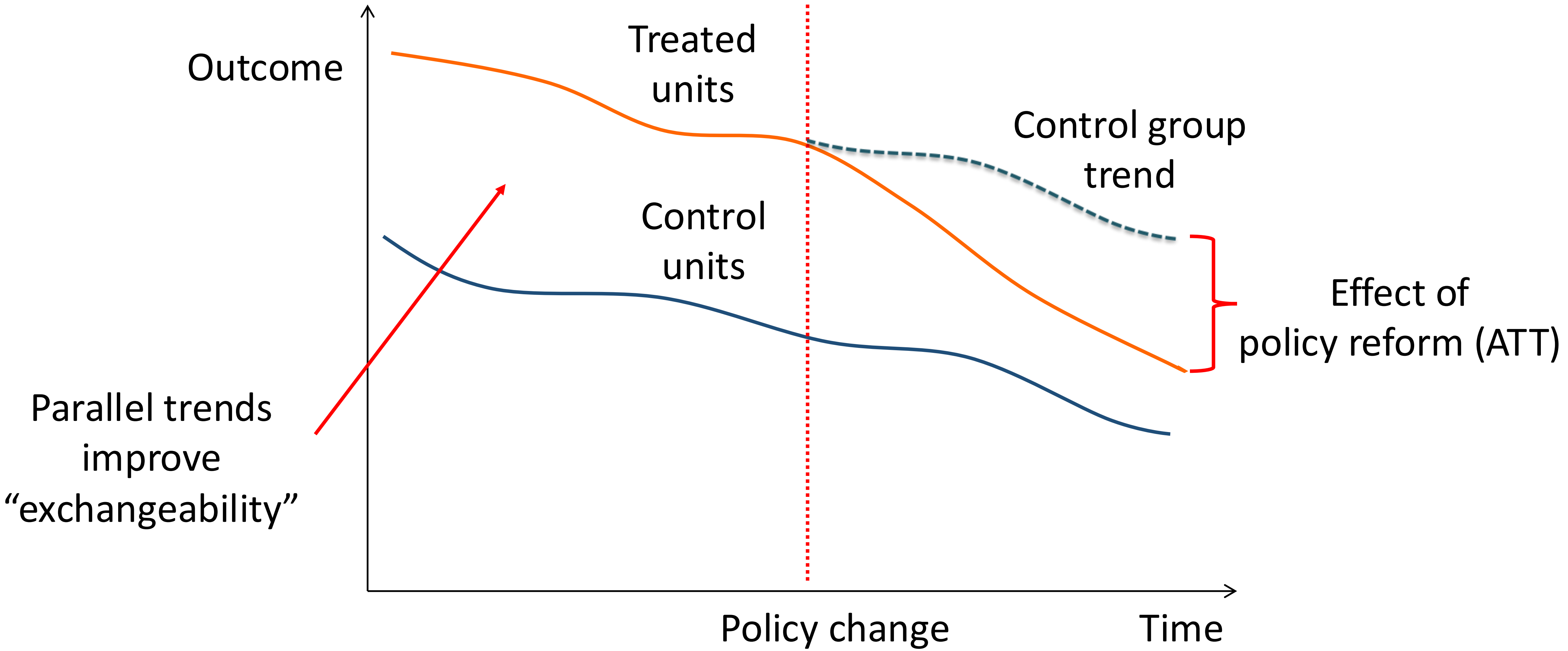
Problem of purely observational approaches

- It is hard (or maybe impossible?) to randomize new parents to different durations of leave after giving birth.
- Hundreds of studies have compared outcomes for parents who took different quantities of leave after the birth of a child.
- These studies rely on the unverifiable assumption that we can adequately measure and properly control for all confounders that explain why people take different quantities of leave and affect the outcome.
- Thus, methods that only address observables, such as regression adjustment or matching, are at high risk of confounding bias.

What if we consider parental leave policies?

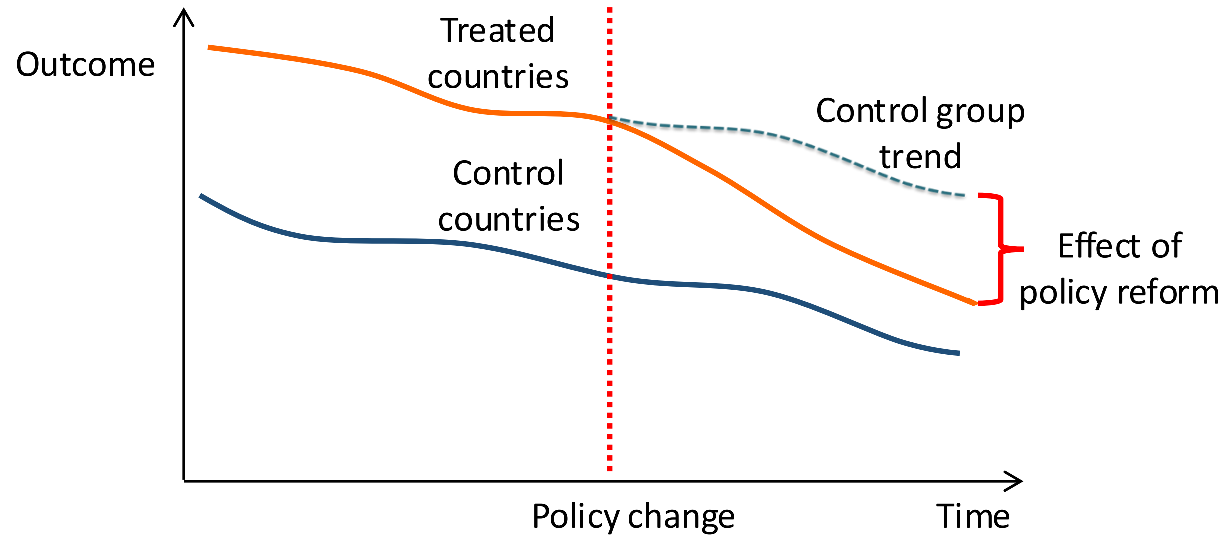


Basic difference-in-differences design (visually)



Source: Adapted from Gertler (2011)

Advantages of fixed effects



$$Y = \beta_0 + \beta_1 * treat + \beta_2 * post + \beta_3 * treat * post$$

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

- Fixed (i.e., non time-varying) differences between those groups;
- Comparisons over time in the treatment group that could be the result of time trends unrelated to the treatment.

DD designs: What's the counterfactual (WTC)?

- Counterfactual: DD designs use a (untreated) control group to substitute for the trend we would have observed in the treated group, **had it been untreated**.

- Core assumptions:
    - Parallel trends: without the intervention, treated and control groups would have displayed similar trends, which is unverifiable but can be explored.
    - No anticipation of treatment

- **Many** extensions:
    - Robustness checks (e.g., triple differences, violations of parallel trends);
    - Dynamic effects (e.g., leads, lags, event study);
    - Staggered treatments; methods that allow for heterogeneous treatment effects.

Paid Family Leave and Mental Health in the U.S.: A Quasi-Experimental Study of State Policies

Amanda M. Irish, DVM, MPH,[1] Justin S. White, PhD,[1,2] Sepideh Modrek, PhD,[3]
Rita Hamad, MD, PhD[2,4]

Introduction: Several U.S. states have implemented paid family leave policies for new parents. Few studies have evaluated the impacts of U.S. paid family leave policies on families' health. This study tests the hypothesis that paid family leave policies in California and New Jersey improved parent and child mental health.

Methods: Using national data from the 1997–2016 waves of the National Health Interview Survey, the study assessed changes in parental psychological distress (measured using the Kessler 6 score, *n*=28,638) and child behavioral problems (measured using the Mental Health Indicator score, *n*=15,987) using difference-in-differences analysis, a quasi-experimental method that compared outcomes before and after the implementation of paid family leave policies in California and New Jersey while accounting for secular trends in states without paid family leave policies. Secondary analyses were conducted to assess differential responses among prespecified subgroups. Data analysis was conducted in 2018–2021.

Results: Exposure to paid family leave policies was associated with decreased psychological distress among parents (−0.49, 95% CI= −0.77, −0.21). There was no association between the paid family leave policies and children's behavioral problems (−0.06, 95% CI= −0.13, 0.012). Associations varied by demographic and socioeconomic characteristics, with some subgroups experiencing benefits, whereas others were negatively impacted.

Synthetic control methods

- What if we can't find a suitable control group for a DD design?

- 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")

- The synthetic control is a weighted combination of comparison units, which arguably provides a better comparison for the treated unit than any single control unit

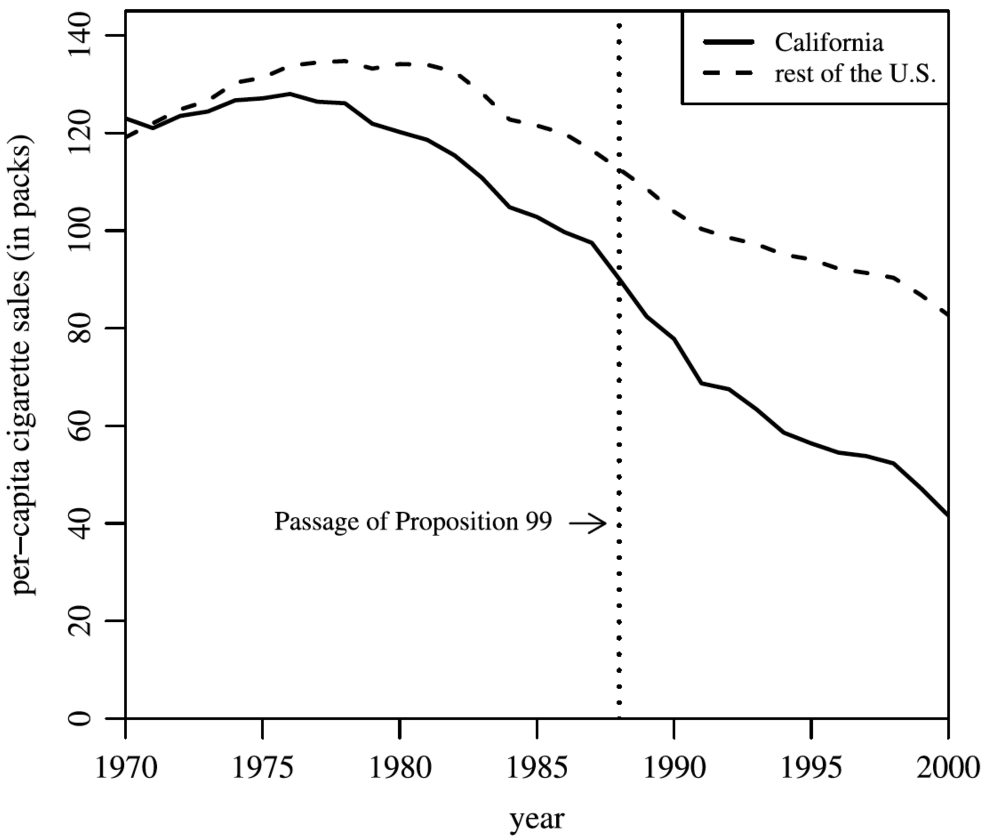


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Source: Abadie et al. (2010)

Synthetic control designs: WTC?

- Counterfactual: The synthetic control represents the counterfactual scenario for a treated unit in the absence of the intervention under scrutiny.

- Core assumptions:
  - Absence of significant shocks that affected the treated unit exclusively;
  - No impact of treatment on control units;
  - Unmeasured confounding? Abadie et al. (2010) argue that effective matching on lagged outcomes and measured covariates controls for time-varying unobserved factors.

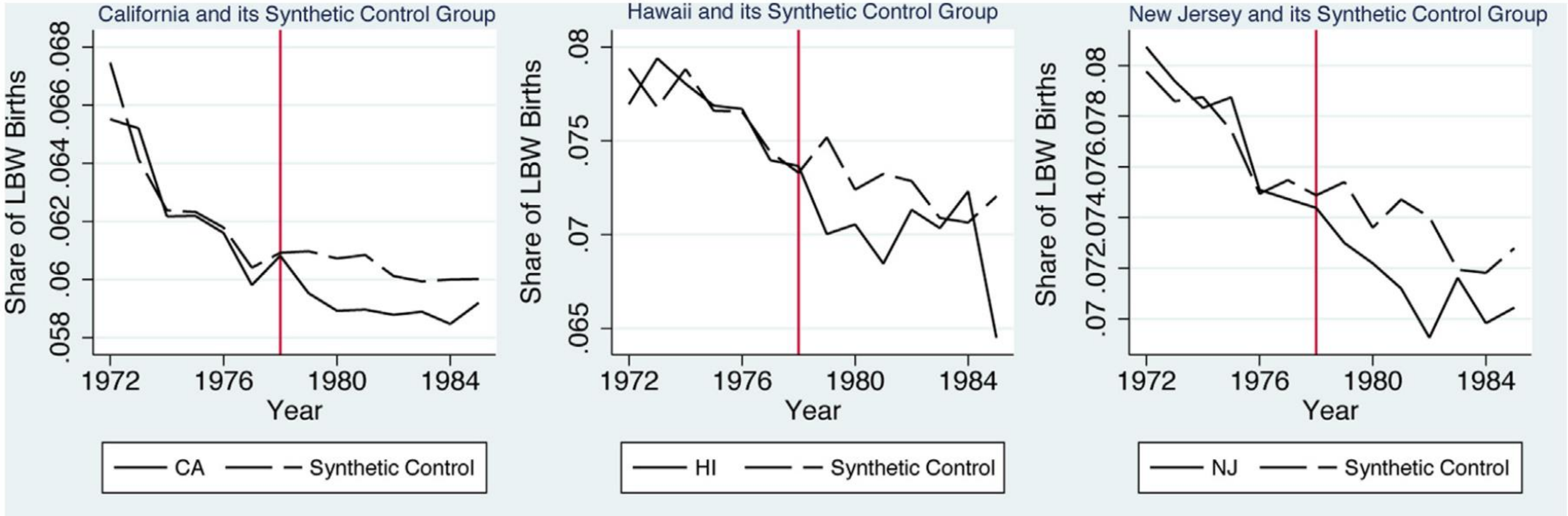
- Extensions:
  - Alternative controls, predictor weights, study periods;
  - Placebo and falsification tests;
  - Augmented synthetic control methods.

The effects of paid maternity leave: Evidence from Temporary Disability Insurance

Jenna Stearns[*,1]

University of California, Santa Barbara, United States

This paper investigates the effects of a large-scale paid maternity leave program on birth outcomes in the United States. In 1978, states with Temporary Disability Insurance (TDI) programs were required to start providing wage replacement benefits to pregnant women, substantially increasing access to antenatal and postnatal paid leave for working mothers. Using natality data, I find that TDI paid maternity leave reduces the share of low birth weight births by 3.2 percent, and the estimated treatment-on-the-treated effect is over 10 percent. It also decreases the likelihood of early term birth by 6.6 percent. Paid maternity leave has particularly large impacts on the children of unmarried and black mothers.



The interrupted time series design

- The **interruption** refers to a population-level intervention that occurs in a known point in time and separates a time series into pre- and post-intervention periods.

- ITS measures the impact of that interruption on the behavior (e.g., level, slope) of the time series and does not necessarily require a control group.

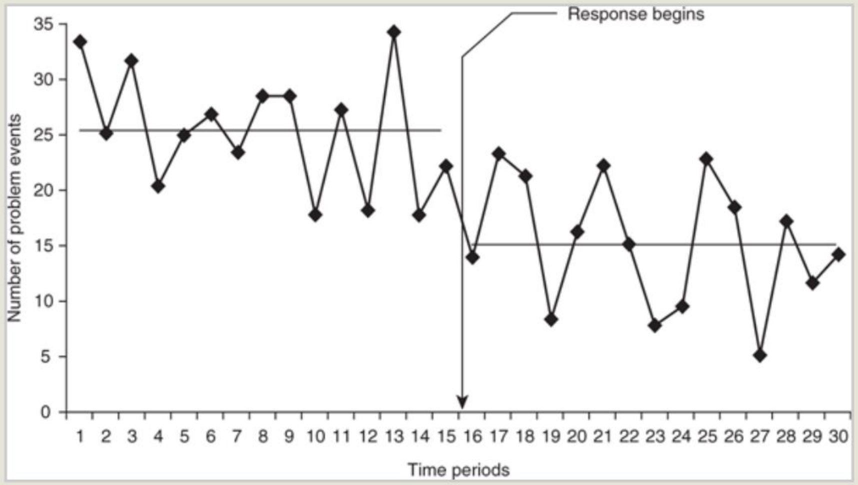


Figure 4.1. Hypothetical Example of a Simple Interrupted Time Series Design.

Interrupted time series designs: WTC?

- Counterfactual: single group ITS compares the trend in an outcome after an intervention against the **extrapolated trend from the pre-intervention time series**.

- Core assumptions:
  - Accurate prediction of how the outcome would have evolved in the absence of the intervention (the “control function”);
  - The absence of co-occurring events that affect the outcome, or confounding by seasonality or other cyclical trends.

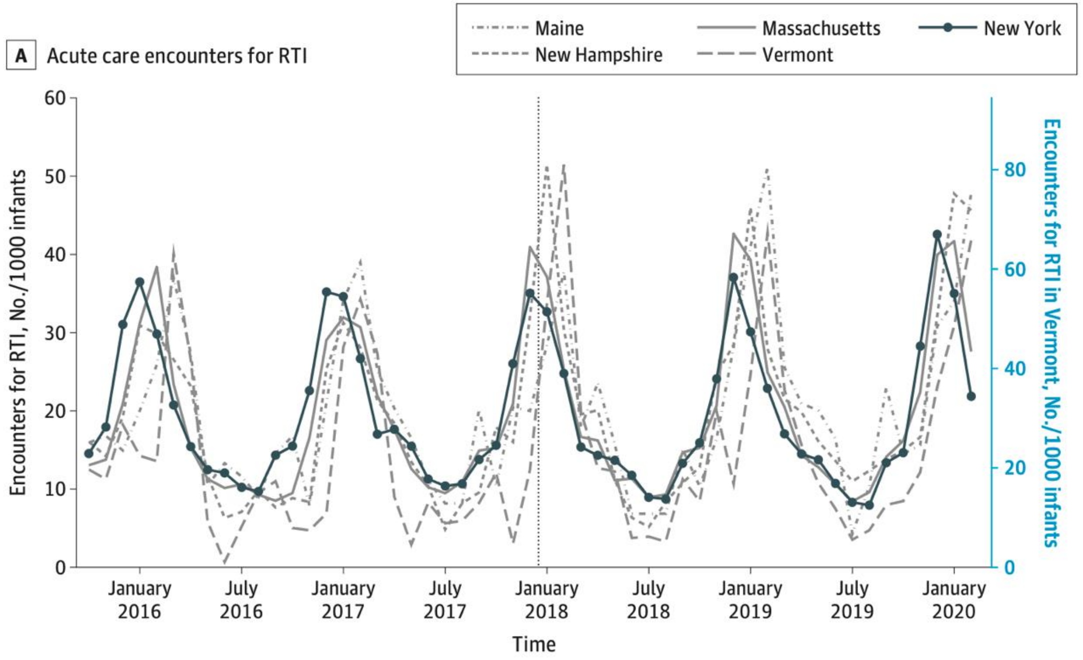
- Extensions:
  - ITS with a control group, known as controlled ITS, which is analogous to DD;
  - Multiple interruptions and allowing for lagged effects;
  - Robustness checks (e.g., testing for lead effects; negative controls) and methods to handle autocorrelation.

Paid Family Leave and Prevention of Acute Respiratory Infections in Young Infants

Katherine A. Ahrens, PhD; Teresa Janevic, PhD; Erin C. Strumpf, PhD; Arijit Nandi, PhD; Justin R. Ortiz, MD; Jennifer A. Hutcheon, PhD

OBJECTIVE To determine if the 2018 introduction of paid family leave in New York state reduced acute care encounters for respiratory tract infections in infants 8 weeks or younger.

DESIGN, SETTING, AND PARTICIPANTS This population-based study of acute care encounters took place in New York state and New England control states (Maine, Massachusetts, New Hampshire, Vermont) from October 2015 through February 2020. Participants included infants aged 8 weeks or younger. Controlled time series analysis using Poisson regression was used to estimate the impact of paid family leave on acute care encounters for respiratory tract infections, comparing observed counts during respiratory virus season (October through March) with those predicted in the absence of the policy. Acute care encounters for respiratory tract infections in 1-year-olds (who would not be expected to benefit as directly from the policy) were modeled as a placebo test.

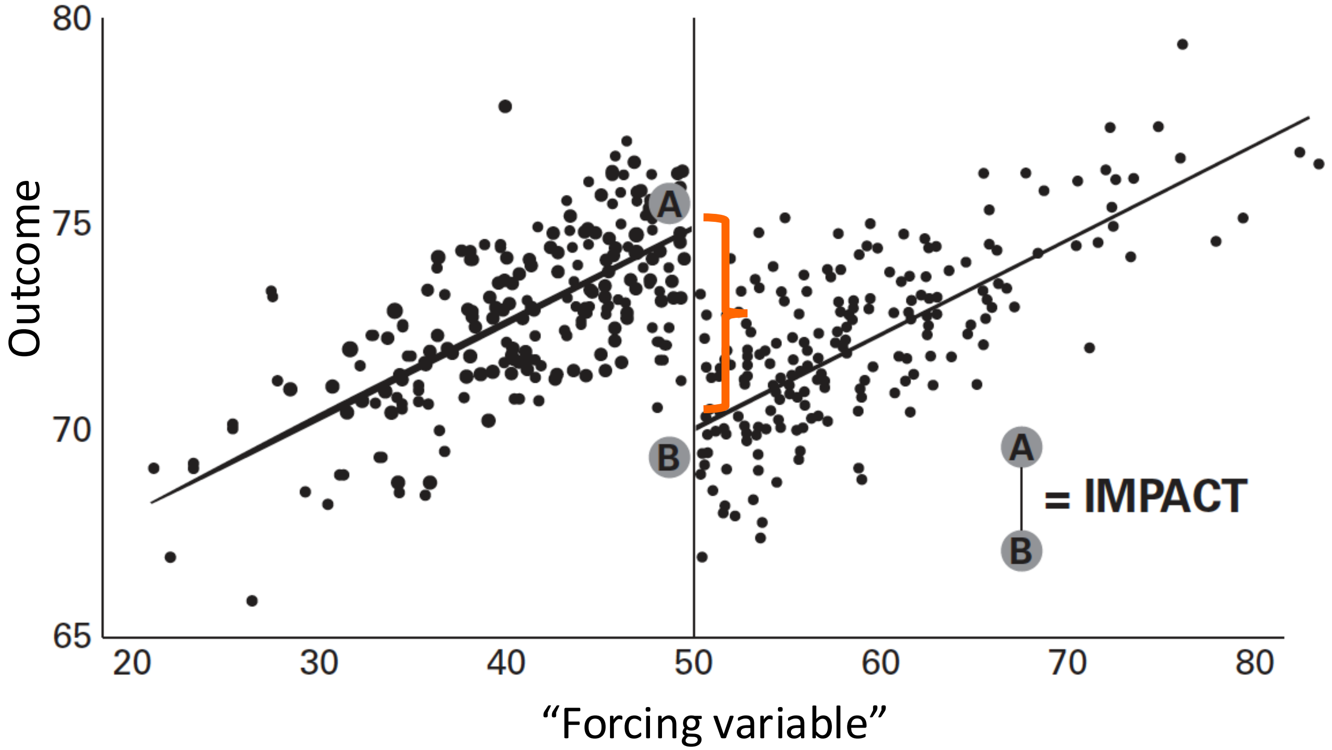


JAMA Pediatr. 2024;178(10):1057-1065.

Regression discontinuity (visually)

- The RD design uses the random variation in treatment assignment created by arbitrary cutoffs as an instrument to evaluate impacts of interventions and other treatments.

- RD measures the difference in post-intervention outcomes between units near the cutoff.



Source: Gertler (2011)

Regression discontinuity: WTC?

- Counterfactual: those who fall just above the cutoff based on some characteristic, called the assignment variable, should be like those who fall just below it on **measured and unmeasured factors** and serve as the counterfactual.

- Assumptions:
  - Continuity of assignment variable near cutoff (no manipulation);
  - In the absence of treatment, no prior discontinuity in the outcome or covariates;
  - Fuzzy RD, which is basically IV analysis, requires standard IV assumptions.

- Extensions:
  - Assessing balance of other covariates;
  - Use different bandwidths, with and without covariates;
  - Alternative parametric and non-parametric modeling strategies.

Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes☆

Astrid Würtz Rasmussen

Department of Economics, Aarhus School of Business, Aarhus University, Hermodsvej 22, 8230 Aabyhoej, Denmark

A B S T R A C T

Investments in children are generally seen as investments in the future economy. In this study I focus on time investments in children as I investigate the long-term educational effects on children of increasing parents' birth-related leave from 14 to 20 weeks using a natural experiment from 1984 in Denmark. The causal effect of the reform is identified using regression discontinuity design to compare a population sample of children born shortly before and shortly after the reform took effect. Results indicate that increasing parents' access to birth-related leave has no measurable effect on children's long-term educational outcomes. Mothers' incomes and career opportunities are slightly positively affected by the reform.

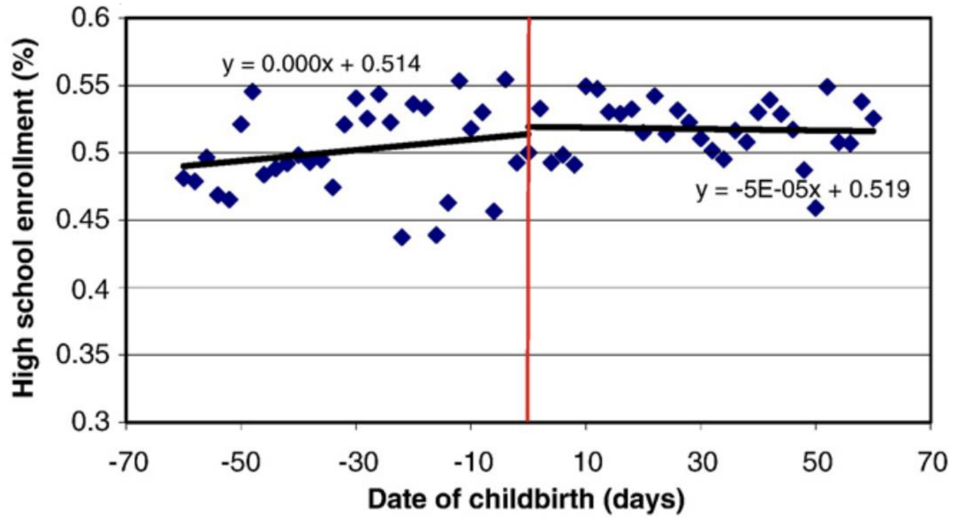


Fig. 5. Probability of high school enrollment in 2005 for children born in 1984 by date of birth. Note: Mean values are calculated using 2-day intervals. The vertical line represents children born at March 26th, 1984.

Concluding remarks

- We often want to estimate the impact of population-level interventions but lack control over treatment assignment.

- Quasi-experimental studies are a family of methods that, **by design**, account for some forms of selection by unobservables (unmeasured confounding).

- However, they are **still observational**—credibility is continuous and results are more credible if we start with unconditional randomized treatment groups.

- We should do our best to examine the robustness of our main findings through carefully designed sensitivity analyses.

Thanks!

SSHRC CRSH

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

CIHR IRSC

Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

For questions, comments, or suggestions: arijit.nandi@mcgill.ca

For information about our project: www.prosperedproject.com

Selected readings

DD

- Benmarhnia, T. et al. (2019). "A rose by any other name still needs to be identified (with plausible assumptions)". Int J Epidemiol 48.6, pp. 2061-2062. DOI: 10.1093/ije/dyz049.
- Caetano, C. et al. (2022). "Difference in Differences with Time-Varying Covariates". arXiv. DOI: 10.48550/ARXIV.2202.02903. URL: https://arxiv.org/abs/2202.02903.
- Callaway, B. et al. (2021). "Difference-in-differences with multiple time periods". Journal of Econometrics 225.2, pp. 200-230.
- Cengiz, D. et al. (2019). "The effect of minimum wages on low-wage jobs". The Quarterly Journal of Economics 134.3, pp. 1405-1454.
- Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing". Journal of Econometrics 225.2, pp. 254-277.
- Lopez Bernal, J. et al. (2019). "Difference in difference, controlled interrupted time series and synthetic controls". Int J Epidemiol 48.6, pp. 2062-2063. DOI: 10.1093/ije/dyz050.

Selected readings

DD (continued)

- Rambachan, Ashesh, and Jonathan Roth. "A more credible approach to parallel trends." *Review of Economic Studies* 90.5 (2023): 2555-2591.
- Roth, Jonathan, et al. "What’s trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235.2 (2023): 2218-2244.
- Sun, L. et al. (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". Journal of Econometrics 225.2, pp. 175-199.

Selected readings

ITS

- Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. International journal of epidemiology. 2017;46(1):348-55.
- Bernal JL, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. International journal of epidemiology. 2018;47(6):2082-93.
- Bernal JL, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. Journal of clinical epidemiology. 2018;103:82-91.
- Jebb AT, Tay L, Wang W, Huang Q. Time series analysis for psychological research: examining and forecasting change. Frontiers in psychology. 2015;6:727.

Selected readings

ITS (continued)

- Linden A. Challenges to validity in single-group interrupted time series analysis. Journal of evaluation in clinical practice. 2017;23(2):413-8.
- McCleary R, McDowall D, Bartos BJ. Design and analysis of time series experiments: Oxford University Press; 2017.
- Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples: Springer; 2017.
- Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. Journal of clinical pharmacy and therapeutics. 2002;27(4):299-309.

Selected readings

SC

- Abadie, Alberto, and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country." American Economic Review 2003:93;113-132.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synth: An r package for synthetic control methods in comparative case studies." Journal of Statistical Software 2011:42;1-17.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program." Journal of the American Statistical Association 2010:105;493-505.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method." American Journal of Political Science 2015:59;495-510.
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "The augmented synthetic control method." Journal of the American Statistical Association 116.536 (2021): 1789-1803.

Selected readings

SC (continued)

- Gerring, John. "What is a case study and what is it good for?." American political science review 2004:98;341-354.
- McLelland R, Gault S. The Synthetic Control Method as a Tool to Understand State Policy. Urban Institute, 2017.
- O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. Health Services Outcomes Research Methods 2016:16;1-21.

Selected readings

RD

- Andalón, Mabel. "Oportunidades to reduce overweight and obesity in Mexico?." Health economics 20.S1 (2011): 1-18.
- Angrist, Joshua D., and Jörn-Steffen Pischke. Mastering'metrics: The path from cause to effect. Princeton University Press, 2014.
- Bor, Jacob, et al. "Regression discontinuity designs in epidemiology: causal inference without randomized trials." Epidemiology 25.5 (2014): 729-737.
- Cattaneo MD, et al. A Practical Introduction to Regression Discontinuity Designs: Parts I & II. Forthcoming in Cambridge Elements: Quantitative and Computational Methods for Social Science.
- Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142.2 (2008): 615-635.
- McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." Journal of Econometrics 142.2 (2008): 698-714.

Selected readings

RD (continued)

- Moscoe, Ellen, Jacob Bor, and Till Bärnighausen. "Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice." Journal of clinical epidemiology 68.2 (2015): 122-133.
- Van der Klaauw, Wilbert. "Regression–discontinuity analysis: a survey of recent developments in economics." Labour 22.2 (2008): 219-245.