Causal Inference II

Mixtape Session

MIXTAPE SESSIONS

Roadmap

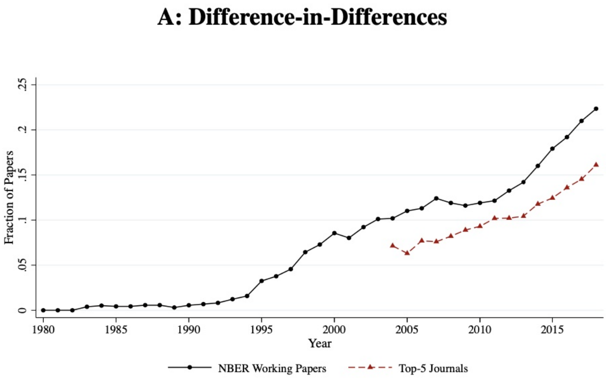
- Difference-in-Differences
  - Origins
- Parallel Trends Violations
  - Results versus Evidence

Introduction

- Scott Cunningham, Ben H. Williams Professor of Economics at Baylor University in Texas USA
- Wanted to walk you through the history of diff-in-diff
- Will show you stories, pictures, and code

Diff-in-diff popularity

Figure: Currie, et al. (2020)



What is difference-in-differences (DiD)

- DiD is when a group of units are assigned some treatment and then compared to a group of units that weren't before and after
- One of the most widely used quasi-experimental methods in economics and increasingly in industry
- Predates the randomized experiment by 80 years, but uses basic experimental ideas about treatment and control groups (just not randomized)
- Uses panel or repeated cross section datasets, binary treatments usually, and often covariates
- We'll do a quick run through the social history of diff-in-diff to set the stage for our workshop this week

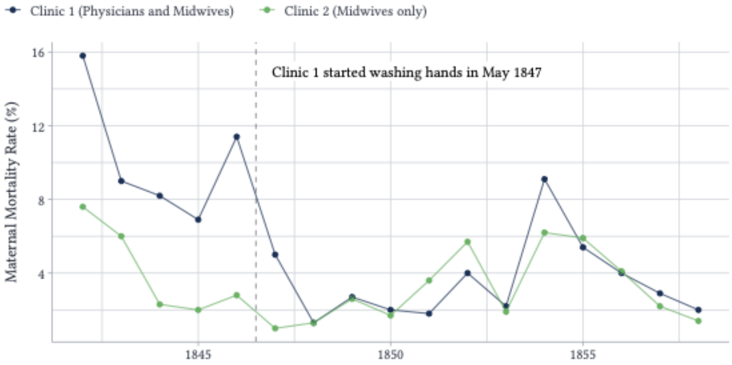
Ignaz Semmelweis and washing hands

- Early 1820s, Vienna passed legislation requiring that if a pregnant women giving birth went to a public hospital (free care), then depending on the day of week and time of day, she would be routed to either the midwife wing or the physician wing (most likely resulting in random assignment)
- But by the 1840s, Ignaz Semmelweis noticed that pregnant women died after delivery in the (male) wing at a rate of 13-18%, but only 3% in the (female) midwife wing – cause was puerperal or "childbed" fever
- Somehow this was also we known – women would give birth in the street rather than go to the physician if they were unlucky enough to have their water break on the wrong day and time

Ignaz Semmelweis and washing hands

- Ignaz Semmelweis conjectures after a lot of observation that the cause is the teaching faculty teaching anatomy using cadavers and then delivering babies *without washing hands*
- New training happens to one but not the other and Semmelweis thinks the mortality is caused by working with cadavers
- Convinced the hospital to have physicians wash their hands in chlorine but not the midwives, creating a type of difference-in-differences design

Semmelweis diff-in-diff evidence



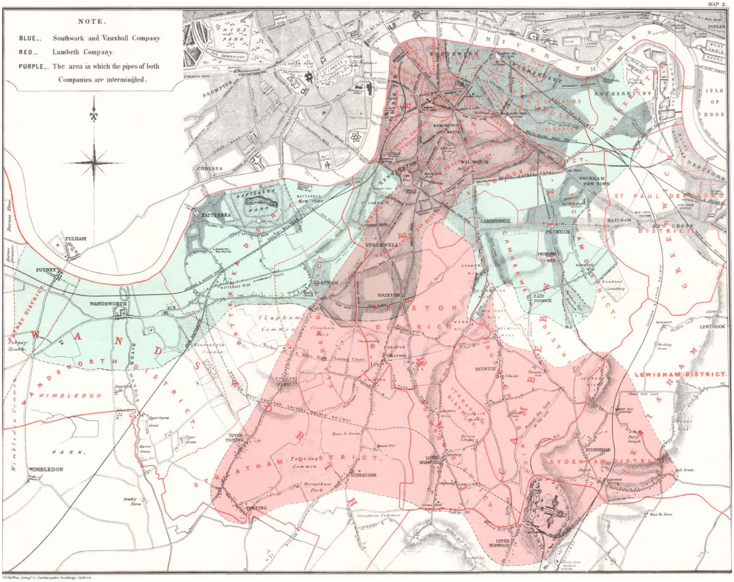
Evidence Rejected

- Diff-in-diff evidence was rejected by Semmelweis' superiors claiming it was the hospital's new ventilation system
- Dominant theory of disease spread was caused by ”odors” or miasma or ”humors”
- Semmelweis began showing signs of irritability, perhaps onset of dementia, became publicly abusive, was committed to a mental hospital and within two weeks died from wounds he received while in residence
- Despite the strength of evidence, difference-in-differences was rejected – a theme we will see continue

John Snow and cholera

- Three major waves of cholera in the early to mid 1800s in London, largely thought to be spread by miasma (“dirty air”)
- John Snow believed cholera was spread through the Thames water supply through an invisible creature that entered the body through food and drink, caused the body to expel water, placing the creature back in the Thames and causing epidemic waves
- London passes ordinance requiring water utility companies to move inlet pipe further up the Thames, above the city center, but not everyone complies
- Natural experiment: Lambeth water company moves its pipe between 1849 and 1854; Southwark and Vauxhall water company delayed

Figure: Two water utility companies in London 1854



Difference-in-differences

Table: Lambeth and Southwark and Vauxhall, 1849 and 1854

Companies	**Time**	**Outcome**	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + L_t + D$	$L_t + D$	
				$D + (L_t - SV_t)$
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + SV_t$	SV_t	

$$\widehat{\delta}_{did} = D + (L_t - SV_t)$$

This method yields an unbiased estimate of D if $L_t = SV_t$, but note that L_t is a counterfactual trend and therefore not known

Two rivers into causal inference

Orley Ashenfelter

↓

Princeton Industrial Relations Section

↓

Quasi-Experimental Design

↓

David Card

↓

Alan Krueger

↓

Don Rubin

↓

Harvard Statistics

↓

Experimental Design

↓

Potential Outcomes

↓

Treatment Effects

↓

Josh Angrist
(Princeton)

Guido Imbens
(Brown)

Harvard Economics

Background I: Harvard Stats and Potential Outcomes

- Don Rubin, former chair of Harvard stats, is the main source of potential outcomes, building on Jerzy Neyman's 1923 work.
- Rubin's influential 1970s papers advocated for causal inference using contrasts of $Y(1)$ and $Y(0)$.
- Neyman's notation, initially in Polish, was translated into English in 1990, likely due to Rubin.
- Rubin expanded Neyman's ideas from experiments to observational studies, leading to developments like propensity score methods.
- Economics was slow to adopt these methods initially.

Background II: Princeton Industrial Relations Section

- Late 1970s and early 1980s: little ”credibility” in empirical labor studies.
- Princeton Industrial Relations Section: older than the economics dept, rigorous, non-partisan focus on US ”manpower”, highly empirical.
- Key faculty are Orley Ashenfelter, David Card, Alan Krueger.
- Key students include Bob Lalonde, Josh Angrist, Steve Pischke, John Dinardo, Janet Currie, Anne Case, and many more.
- Listen to David Card: https://youtu.be/1soLdywFb_Q?si=BCVqYeRz6jYiwHTQ&t=1580

Background II: Princeton Industrial Relations Section

Example of Princeton Paradigm Emerging

- Lalonde (1986) was a groundbreaking study, recently reviewed by Guido Imbens and Yiqing Xu (2024)
- Lalonde, a student of Card and Orley, analyzed an RCT on a job training program, finding an average treatment effect of +$800.
- He then replaced the experimental control group with survey data, reran econometric methods, and couldn’t replicate the results.
- Orley and Card emphasized randomization in their 1985 Restat article, advocating for its exploitation in studies.

Orley Ashenfelter and diff-in-diff

- Diff-in-diff gets rediscovered by Orley Ashenfelter from Princeton
- Leaves academia to work in Washington DC to study job training programs for low skill workers
- Coins the phrase ”difference-in-differences” so as to avoid having to explain regressions to bureaucrats (3:53) `https://youtu.be/WnB3EJ8K7lg?si=uE4clqUIPzvbxm0r&t=2`
- More associated with David Card (Mariel boatlift, minimum wage), but it was earlier that he and Orley worked with the method, and ironically largely, rejected its usefulness for the questions they were working on

Minimum wages

- Card and Krueger (1994) have a famous study estimating causal effect of minimum wages on employment
- New Jersey raises its minimum wage in April 1992 (between February and November) but neighboring Pennsylvania does not
- Using DiD, they do not find a negative effect of the minimum wage on employment leading to complex reactions from economists
- Orley's describes his understanding of people's reaction to the paper. https://youtu.be/MOtbuRX4eyQ?t=1882

Binyamin Appelbaum
@BCAppelbaum

Replying to @BCAppelbaum

The Nobel laureate James Buchanan wrote in the Wall Street Journal that Card and Krueger were undermining the credibility of economics as a discipline. He called them and their allies "a bevy of camp-following whores."

3:49 PM · Mar 18, 2019

179 Reply Share

Read 18 replies

Reaction to the paper

Lots of anecdotes in this interview with Card, but here are just two. First, Card and Krueger received a lot of personal hostility from their peers (1:07 to 1:10)

https://youtu.be/1soLdywFb_Q?si=laAVYf_E2KBZKywG&t=4020

Later Card says Sherwin Rosen accused them of having an agenda. But the worst is what happens to Alan Krueger maybe (1:16 to 1:17)

https://youtu.be/1soLdywFb_Q?si=jsb8h50ZosGDnKrv&t=4556

Card on that study

"I've subsequently stayed away from the minimum wage literature for a number of reasons. First, it cost me a lot of friends. People that I had known for many years, for instance, some of the ones I met at my first job at the University of Chicago, became very angry or disappointed. They thought that in publishing our work we were being traitors to the cause of economics as a whole."

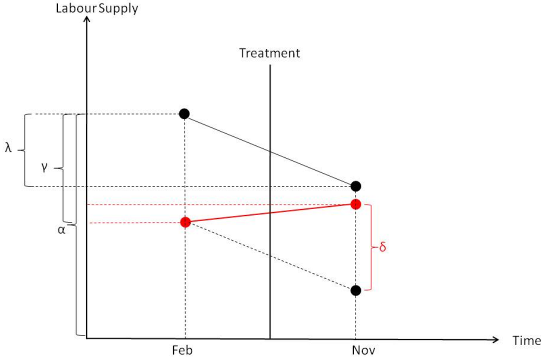
OLS specification of the DiD equation

- The correctly specified OLS regression is an interaction with time and group fixed effects:

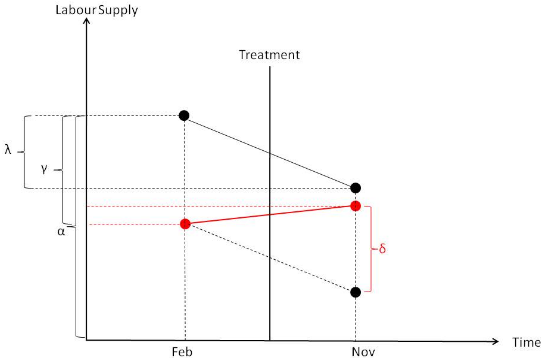
$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

  - → NJ is a dummy equal to 1 if the observation is from NJ
  - → d is a dummy equal to 1 if the observation is from November (the post period)
- This equation takes the following values
  - → PA Pre: α
  - → PA Post: $\alpha + \lambda$
  - → NJ Pre: $\alpha + \gamma$
  - → NJ Post: $\alpha + \gamma + \lambda + \delta$
- DiD equation: (NJ Post - NJ Pre) - (PA Post - PA Pre) $= \delta$

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$



Notice how OLS is "imputing" $E[Y^0|D=1, Post]$ for the treatment group in the post period? It is only "correct", though, if parallel trends is a good approximation

Roadmap

Difference-in-Differences

- Origins

Parallel Trends Violations

- Results versus Evidence

Evidence versus the Main Result

- Causal inference is about *warranted beliefs* – should you or should you not believe the *causal claim*?
- Your DiD *results* are like the claim of guilt, but your DiD results are *not* the smoking gun
  - → Your table of regression coefficients *is not enough* for evidence
  - → You need to do more to provide a justification for parallel trends
- You need to provide evidence for parallel trends against several well known vulnerabilities
- Evidence will be bite, falsifications, mechanisms and event study data visualization
- We will mix the parallel trends violations with the evidence concept before getting into advanced estimators

Court metaphor

- Think of yourself as a prosecutor arguing against a defense attorney to convince a judge and jury of a defendant's guilt
- The claim the defendant is guilty is your table of main results
- But the claim is not the evidence – you have to back up that claim
- Your evidence of guilt is the smoking gun, the fingerprints, the eye witnesses, the footprints in the mud outside the house
- If your claim is supported by weak evidence, then no one *should* convict – it would be borderline corruption if they did

Event study regression

- Alternatively, present estimated coefficients from a dynamic regression specification:

$$Y_{its} = \alpha + \sum_{\tau=-2}^{-q} \mu_\tau (D_s \times \tau_t) + \sum_{\tau=0}^{m} \delta_\tau (D_s \times \tau_t) + \tau_t + D_s + \varepsilon_{ist}$$

  - → With a simple 2x2, you are interacting treatment indicator with calendar year dummies
  - → Includes q leads (dropping the $t-1$ as baseline) and m lags
  - → Since treatment did not happen until $\tau = 0$, then pre-treatment coefficients only capture differential trends
- Estimated $\widehat{\delta}_\tau$ coefficients are estimated ATT for each year under parallel trends but $\widehat{\mu}_\tau$ is your smoking gun evidence
- Just remember that $\mu = 0$ is not the same as parallel trends as parallel trends is untestable.

Reviewing previous slide for emphasis

- Under NA, SUTVA and parallel pre-trends, then mechanically $\widehat{\mu_\tau}$ will be zero as everything cancels out
  - $\rightarrow$ There are still specification and power issues that Jon Roth has written about, but I will skip that
- But also under NA, SUTVA and parallel trends (post trends), then $\widehat{\delta}$ are estimates of the ATT at points in time
- Typically you'll plot the coefficients and 95% CI on all leads and lags

DiD coefficient

$$\delta = \text{ATT} + \text{Non-Parallel Trends Bias}$$

You want to know if trends would've been the same had the policy not happened.

Pre-treatment DiD coefficient

$$\widehat{\delta}_{t-2} = \underbrace{\bigg[E[Y^0_k|t-2] - E[Y^0_k|t-1]\bigg] - \bigg[E[Y^0_U|t-2] - E[Y^0_U|t-1]\bigg]}_{\text{Non-parallel trends bias in 2x2 case}}$$

Under NA, then the $t-1$ period is untreated. But then so are the other pre-periods so the ATT is implicitly zero and the *only* thing that you can be measuring with pre-trend DiD coefficients is differential trends.

Event study coefficients

- Remember that the OLS specification we discuss collapses to ATT plus parallel trends bias
- This is *always* true because it's an identity and holds even in the pre-period as much in the post
- It's just in the pre period, you do not have the missing $E[Y^0|D=1]$ term as no one and nothing is treated in pre-period under NA
- This means pre-period is basically an opportunity to directly verify parallel pre-trends – but it's the past's pre-trends, not the counterfactual pre-trend of the present/future
- And that's how people use the pre-period – they use the pre-period to evaluate whether they think this is a good control group

Event study example

- The notion is really simple: if PT held then, you'll argue that it's reasonable it would've still held
- But this is an assertion, and you need to build the case as we said
- At this point, it's a lot easier to show you what I'm talking about – where the art and the science meet – with a great paper

Medicaid and Affordable Care Act example

Volume 136, Issue 3
August 2021

< Previous Next >

Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data

Get access >

Sarah Miller, Norman Johnson, Laura R Wherry

The Quarterly Journal of Economics, Volume 136, Issue 3, August 2021, Pages 1783–1829,
https://doi.org/10.1093/qje/qjab004
Published: 30 January 2021

Cite Permissions Share ▾

Abstract

We use large-scale federal survey data linked to administrative death records to investigate the relationship between Medicaid enrollment and mortality. Our analysis compares changes in mortality for near-elderly adults in states with and without Affordable Care Act Medicaid expansions. We identify adults most likely to benefit using survey information on socioeconomic status, citizenship status, and public program participation. We find that prior to the ACA expansions, mortality rates across expansion and nonexpansion states trended similarly, but beginning in the first year of the policy, there were significant reductions in mortality in states that opted to expand relative to nonexpanders. Individuals in expansion states experienced a 0.132 percentage point decline in annual mortality, a 9.4% reduction over the sample mean, as a result of the Medicaid expansions. The effect is driven by a reduction in disease-related deaths and grows over time. A variety of alternative specifications, methods of inference, placebo tests, and sample definitions confirm our main result.

JEL: H75 - State and Local Government: Health; Education; Welfare; Public Pensions, I13 - Health Insurance, Public and Private, I18 - Government Policy; Regulation; Public Health

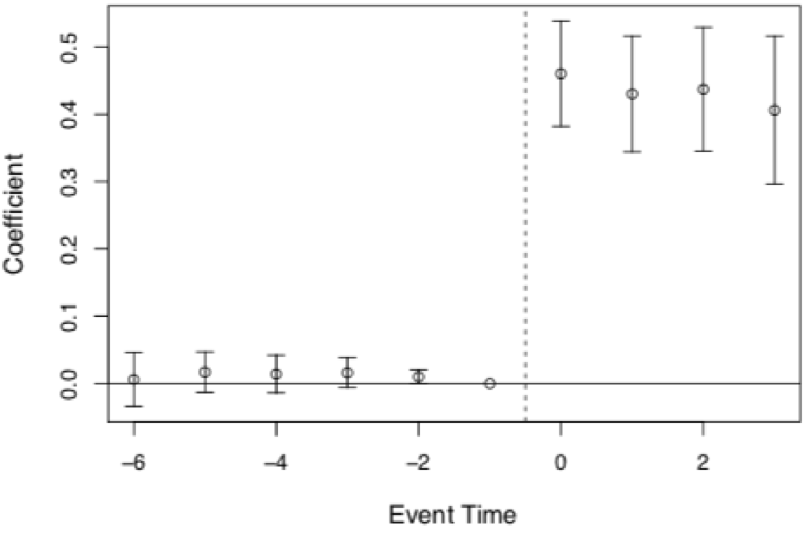
Issue Section: Article

Their Evidence versus Their Result

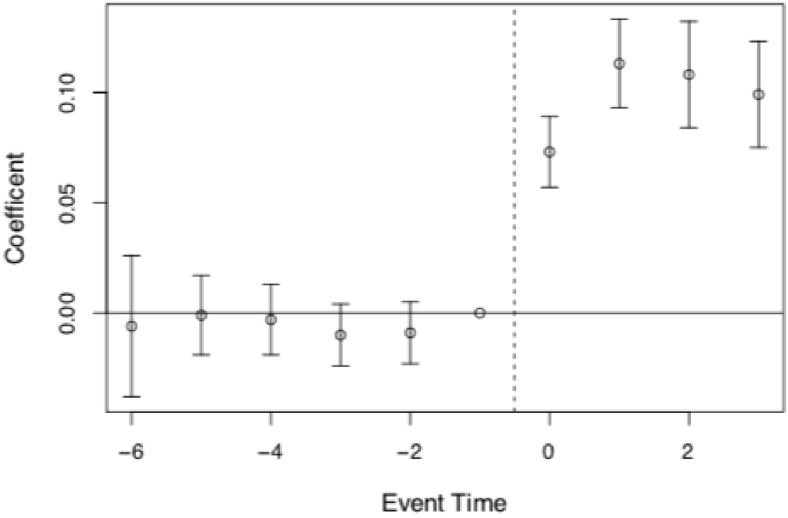
- **Bite** – they will show that the expansion shifted people into Medicaid and out of uninsured status
- **Placebos** – they show that there's no effect of Medicaid on a similar group that didn't enroll
- **Event study** – they will lean hard on those dynamic plots
- **Main results** – with all of this, they will show Medicaid expansion caused near elderly mortality to fall
- **Mechanisms** – they think they can show it's coming from people treating diseases causing mortality declines to compound over time

Bite

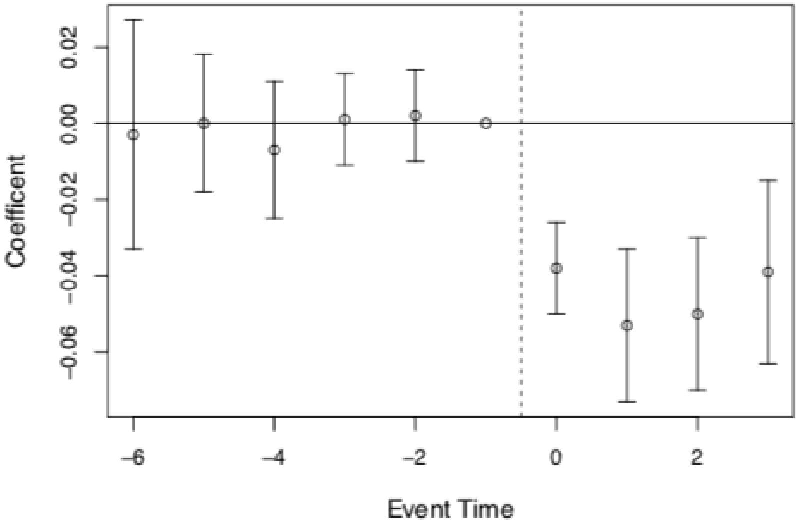
- Bite is a labor economist's phrase, often used with the minimum wage, to say that the minimum wage actually was binding in the first place
- Here it means when US states made Medicaid more generous, people got on Medicaid who would not have been on it otherwise
- And as a bonus, would not have been insured at all without it
- Not the most exciting result, but imagine if the main results on mortality were shown but there was no evidence for bite – is it believable?



(a) Medicaid Eligibility



(b) Medicaid Coverage



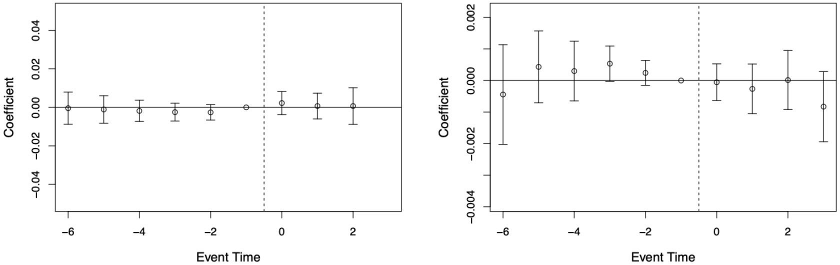
(c) Uninsured

Falsification

- Their study focuses on “near elderly”, which means just under 65
- They choose just under 65 because in the US, 65 and older are eligible for Medicare so more generous Medicaid is irrelevant
- *But* probably the near elderly and the elderly are equally susceptible to unobserved factors correlated with the treatment
- So they painstakingly examine the effects on elderly as a falsification as this will strengthen the parallel trends assumption on the near elderly

Falsifications on elderly

Age 65+ in 2014



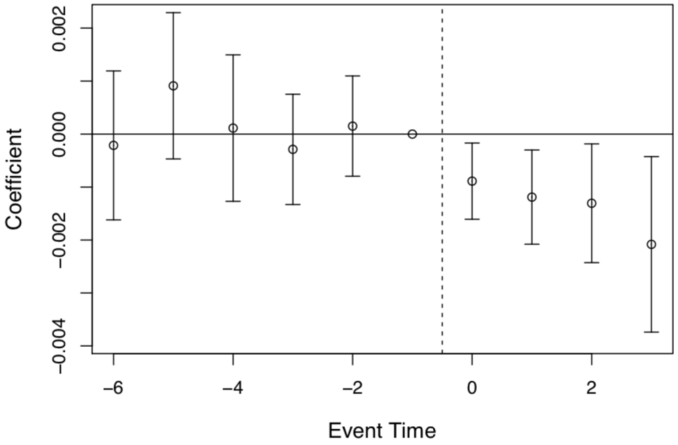
(c) Medicaid Coverage

(d) Annual Mortality

Main result

- Finally they focus on the main result – and there's more in the paper than I'm showing
- Event study plots with same specification as the rest allowing us to look at the pre-trends and the post-treatment coefficients
- If parallel trends holds, then the post-treatment coefficients are interpreted as ATT parameter estimates for each time period
- The result alone isn't nearly as strong the result in combination with the rest, but it could still be wrong as parallel trends is ultimately not verfiable

Near elderly mortality and Medicaid expansion



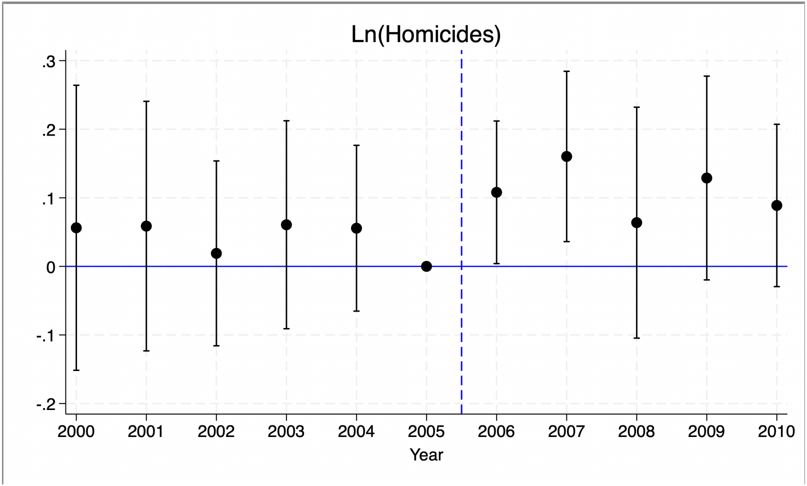
Summarizing evidence and results

- **Bite**: Increases in enrollment and reductions in uninsured support that there is adoption of the treatment
- **Event studies**: Compelling graphics showing similarities between treatment and control
- **Falsifications**: no effect on a similar group who isn't eligible
- Main results: 9.2% reduction in mortality among the near-elderly
- Mechanism: "The effect is driven by a reduction in disease-related deaths and grows over time."

Making event study

- When there is only one treatment group and one comparison group, then you run a regression with an interaction of the treatment group dummy and the calendar year dummies (plus both separately)
- You must drop $t-\tau$ as the baseline (e.g., $t-1$) and it must be Y^0 untreated comparisons (No Anticipation)
- I have included in a do file that will do it for you either manually or using coefplot in `simple_eventstudy.do` at the shared github labs directory

Manually creating the event study

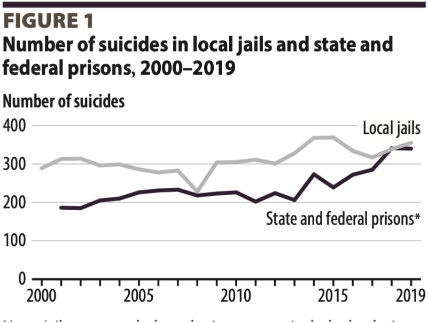


Roadmap

Difference-in-Differences
- Origins

Parallel Trends Violations
- Results versus Evidence

Manually creating the event study



Suicides and Prisons

- Suicide is leading single cause of death in jails
- From 2001 to 2019, number of suicides in state prisons rose 85%
- Texas Dept of Criminal Justice implemented ”self harm prevention office” in 15 prisons
- I’ll walk you through some preliminary analysis