PHS 2000B: Econometrics Interrupted Time Series and Difference in Differences 2023

Maggie McConnell

Department of Global Health and Population

Harvard T. H. Chan School of Public Health

Plan of presentation

1. Recap ITS/Controlled ITS
2. **Difference in Differences model**
3. Example
4. Clustering and standard errors
5. Contrasting difference in differences approach with interrupted time series

Logic of interrupted time series analysis

- Time series data allows us to carefully model the general trends in our outcome of interest

- Most programs/treatments are introduced at a very specific time point: if the program worked, we should observe that the outcome shifts onto a different trajectory after the policy change happened

- The main assumption of this interrupted time series (ITS) approach is that the pre-intervention trend can be continued after the intervention as the counterfactual

ITS and counterfactual

- In a randomized trial the counterfactual is what happens in the control group – we think this is what would have happened without the treatment
- In ITS we fit a trend to the data in the pre intervention period – the counterfactual is that the data would have follows this trend in the post intervention period without the trend
- Assumes we have adequately fitted the trend in the data
- Assumes no other new interventions happening in the post intervention period that could affect the outcome
  - This is a big assumption – not likely to be true when we know the policy landscape changing at the same time

Controlled interrupted time series (CITS)

- Given temporal confounding concerns, one may want to include control regions/areas in the model
  - In these areas the intervention was not implemented
  - If we see a post-intervention effect in these areas → related to common factor affecting both the treated and control areas and <u>not</u> the intervention

- Counterfactual – jump and change in trend in the control group in the post-intervention period would have been the same the same as in the treatment group without treatment

Ideal control group

- Control group may not be identical to treatment group in the pre-intervention period
- But control group must be similar enough that we believe changes from trend in the post intervention period in the control would also have occurred in the intervention group without the intervention
- *Ideal control group:* not affected by the intervention but would react to other things happening around the time of the intervention in a similar way to treatment group
- Sometimes this can be hard to find
  - We will discuss ”synthetic” control methods tomorrow

Difference-in-Differences logic

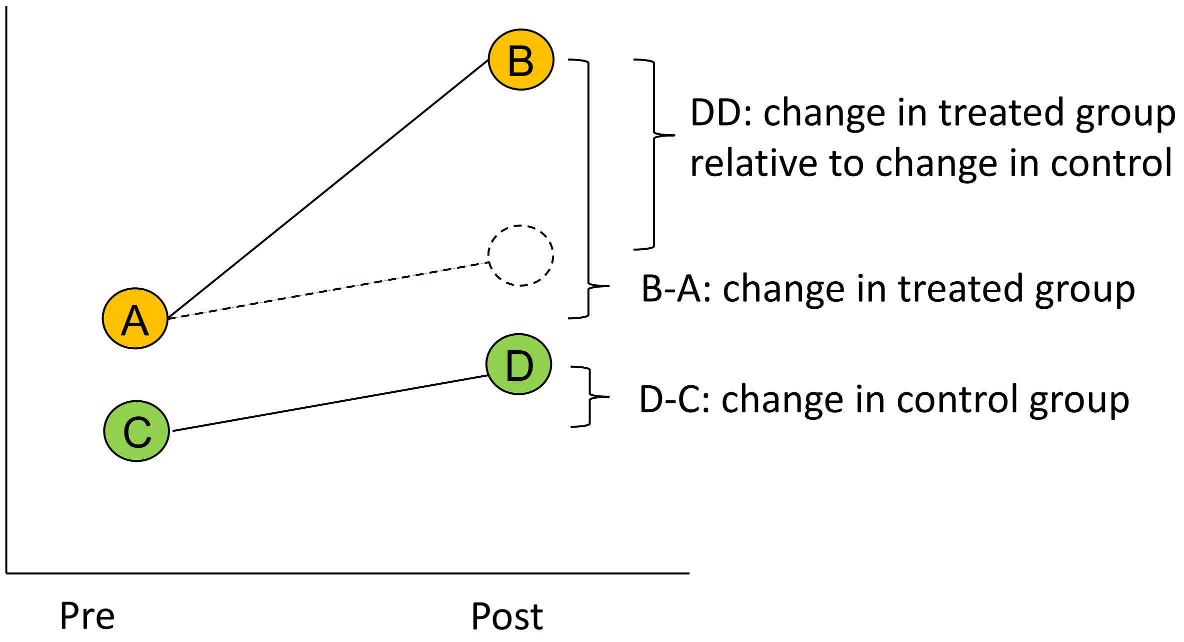
- Suppose we have two areas that are similar at baseline and one gets the intervention (treatment area) and one does not (control area)
- If the areas are similar, we could argue that counterfactual is that the increase over baseline in the treated areas would have been similar to the increase over baseline in the control area without intervention
- Notice this is similar to the logic of controlled ITS

Difference-in-differences estimator:

- Estimate difference in outcome between post- and pre-intervention levels in program areas (“treated”)
- Estimate difference in outcome between post- and pre-intervention levels in non-program areas (“controls”)
- Estimate difference in changes in outcome levels between treated and controls

Difference in difference

	Treatment
Before	A
After	B
Change	B-A



Did the outcomes in the treated areas improve by more than in the control areas?

DD = (change in treatment) – (change in control) = (Treat^{post} - Treat^{pre}) – (Control^{post} - Control^{pre})

Difference-in-Differences estimator: Notation

- Define:
  - Y_{it}: outcome of interest
  - $t = \{0,1\}$ where $t = 0$ represent pre-intervention period while $t = 1$ represents post-intervention period
  - $T = \{0,1\}$ where $T = 1$ indicates being in treatment group while $T = 0$ indicates being in control group

- The basic Difference-in-Differences estimator is given by:

$$[(\overline{Y_1}|T = 1) - (\overline{Y_0}|T = 1)] - [(\overline{Y_1}|T = 0) - (\overline{Y_0}|T = 0)]$$

=Dif (over time) in treatment group – Dif (over time) in control group

Difference-in-Differences logic

- The real causal question is: How would the “treated” have done without the treatment?
- Differences in differences assumes the control group can tell us *what would have* happened in the post intervention period in the treatment group with no intervention

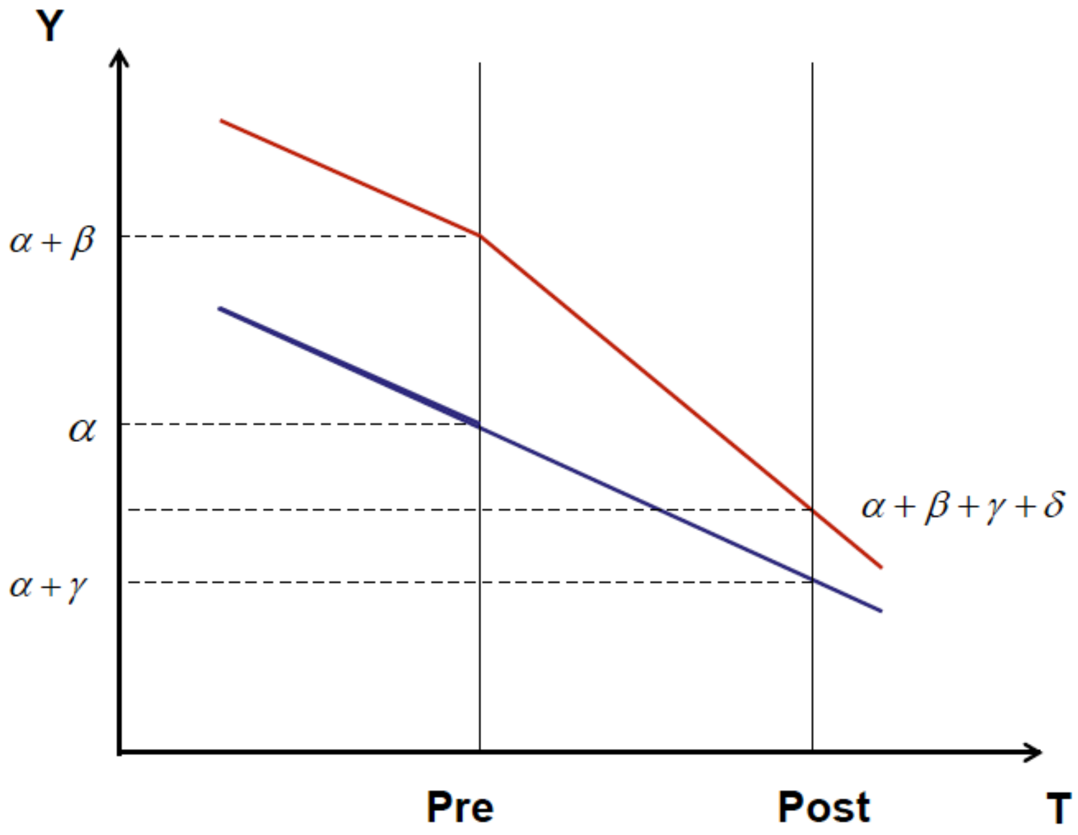
- The counterfactual is that if the treatment group had not got the treatment, the *change* we would have seen is the *change* observed in the control group over the same period

Simplest difference in difference regression

- The simplest difference-in-difference specification is a 2 period (pre, post) and two group (treated, control) specification:

$$Y_{it} = \alpha + \beta Treated_i + \gamma Post_t + \delta(Treated_i * Post_t) + \epsilon_{it}$$

Difference in difference – another example

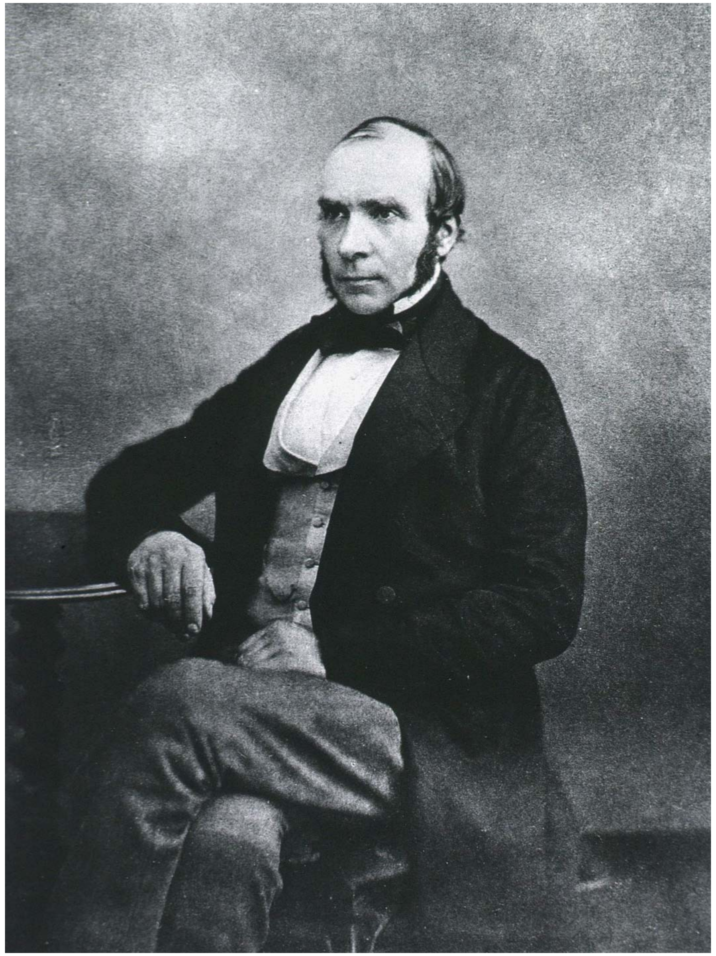


We can compare the change over time in the treated group $(\alpha + \beta + \gamma + \delta)$-$(\alpha + \beta)$ = $\gamma + \delta$ to the change over time in the control group $(\alpha + \gamma - \alpha)$ = γ to get the “treatment effect=δ

What is the sign of γ – the estimate of the change over time in the control group?

What about δ – the estimate of the “treatment” effect?

Difference in differences is a classic strategy for generating public health evidence



John Snow DiD analysis

- London had cholera outbreaks in 1849 and 1853
- Two water companies: Lambeth Waterworks & Southwark and Vauxhall Water Co.
  - Both got their water from the Thames river
  - In 1852 Lambeth moved its water intake to the Thames Ditton river, which is higher upstream

OPPORTUNITY THUS AFFORDED OF GAINING CONCLUSIVE EVIDENCE OF THE EFFECT OF THE WATER SUPPLY ON THE MORTALITY FROM CHOLERA

The experiment, too, was on the grandest scale. No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into two groups without their choice, and, in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and, amongst it, whatever might have come from the cholera patients, the other group having water quite free from such impurity.

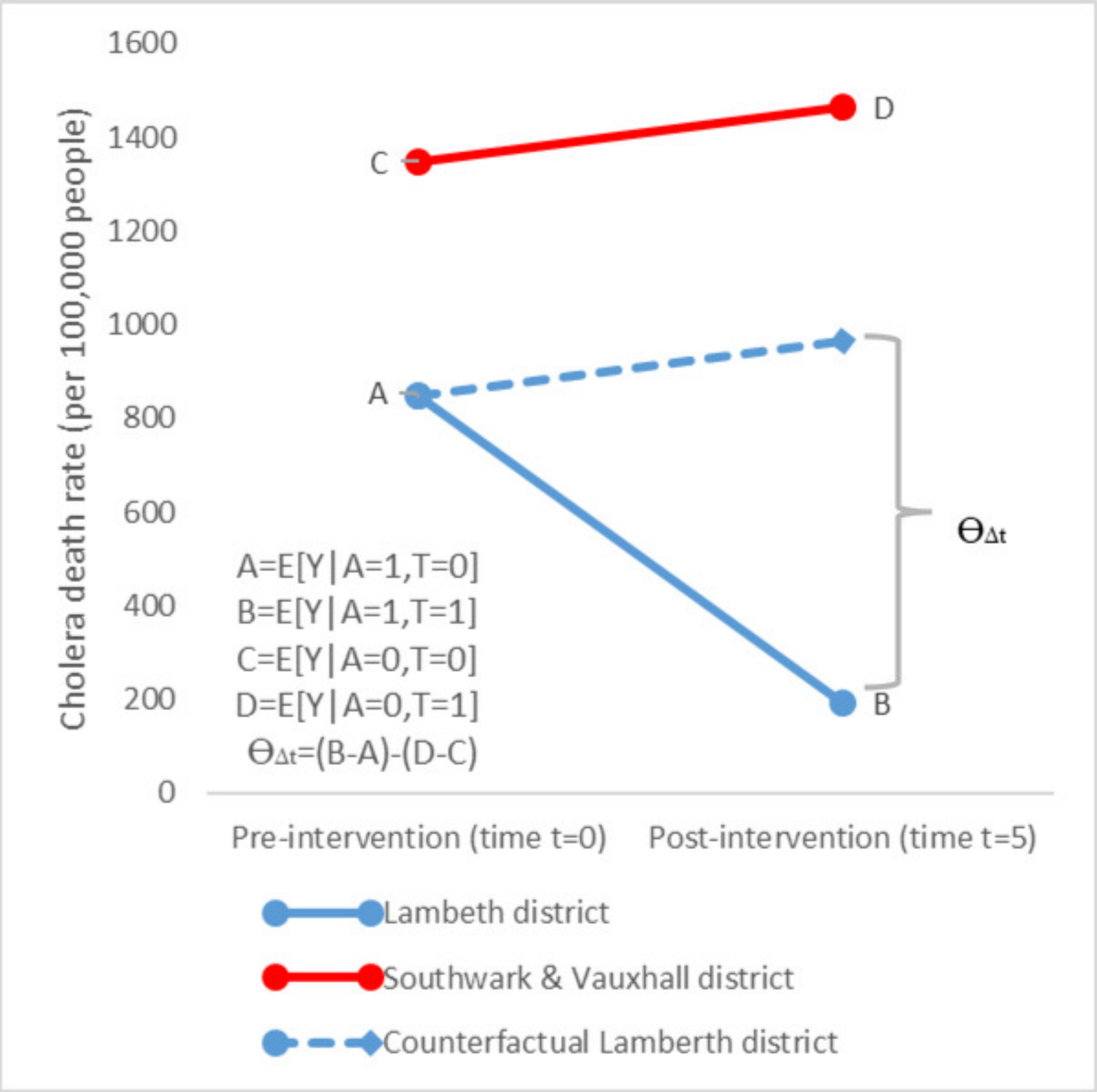
Source: John Snow 1855 "On the Mode of Communication of Cholera"

[Difference-in-Difference in the Time of Cholera: a Gentle Introduction for Epidemiologists]

Variable A indicates different water sources

Variable T indicates measurement takes place before and after the change of water source

Y is a death from cholera



Simplest difference in difference case

- We are comparing 4 data points: control pre, control post, treated pre, treated post

- The difference between control pre and post must be a good counterfactual for what would have happened to the treatment group without treatment

- But what if in the control district there was a hand washing campaign at the same time the treatment group changed its water source
  - Do we find the counterfactual assumption plausible in this case?

Another example – rollout of health insurance

- Assume you want to evaluate the effect of a recently introduced health insurance on health seeking behavior at the individual level
- Further assume that the health insurance was piloted only in 3 out of 25 districts of the respective region of interest
- You have two cross-sectional health surveys: one before the introduction of the insurance scheme, and one after. Both surveys contain information on:
  - Total number of visits at health center
  - Total average health cost per person

Example – rollout of health insurance

- You estimate the following difference in difference model where V is the number of visits:

$$\left[\left(\overline{V_{post}}|T=1\right)-\left(\overline{V_{pre}}|T=1\right)\right]-\left[\left(\overline{V_{post}}|T=0\right)-\left(\overline{V_{pre}}|T=0\right)\right]$$

- Would this provide a correct estimate if the 3 target districts had higher income than the rest of the country?
- What if the government responded to complaints from the areas where the rollout had not yet occurred (control districts) by providing block grants for facility improvement?
  - Which way would the bias go?

Multiple groups and time periods

- We need at least two groups – treatment and control – and two time periods – pre and post – to get a difference-in-differences estimator
  - Essentially at least four observations
  - However -- need more data for inference
  - With more data we can strengthen our belief in the assumptions that make DiD valid
    - Hard for one district to make a convincing control group → averaging across many districts together means the counterfactual assumption is more plausible
    - With just one observation from the pre-period, harder to know that we haven't gotten noise → more pre-period data allows better modeling of trends and changes over time
- DID is a flexible methodology we can extend in many ways
  - For example, we can include two post periods – an initial period of "transition" and a period where policy is fully implemented

Multiple groups and time periods

- Typical DiD specification
  - We have multiple groups ($g = \{1 \ldots M\}$) with some getting the treatment and some not getting it (group fixed effects)
    - ”Treated” binary variable gets subsumed by these group fixed effects
  - We have multiple time periods ($t = \{1 \ldots K\}$) both pre- and post-intervention (time fixed effects)
    - “Post” binary variable gets subsumed by these time fixed effects

$$Y_{igt} = \sum_{g=1}^{M} \alpha_g + \sum_{t=1}^{K} \gamma_t + \beta\left(I_{post} * I_{g\in T}\right) + \epsilon_{igt}$$

  - β tells us the average effect of treatment on the treated
    - The difference in the outcome in the post-intervention period in the treated areas after accounting for state and time level differences
  - Models can be more complex, i.e. modeling time trends, covariates, phased in implementation

Examples

- 50 states in the US with different policies

- States that adopt policy are treatment group, those that do not are control group

- Treatment may start at different time in different groups

- Examples
  - Medicaid expansion under the Affordable Care Act
  - State level rollout of SNAP/benefit programs
  - State-level rollout of COVID-19 lotteries
  - State-level rollout of anti-discrimination laws against LGBT discrimination/state level rollout of legalization of same sex marriage

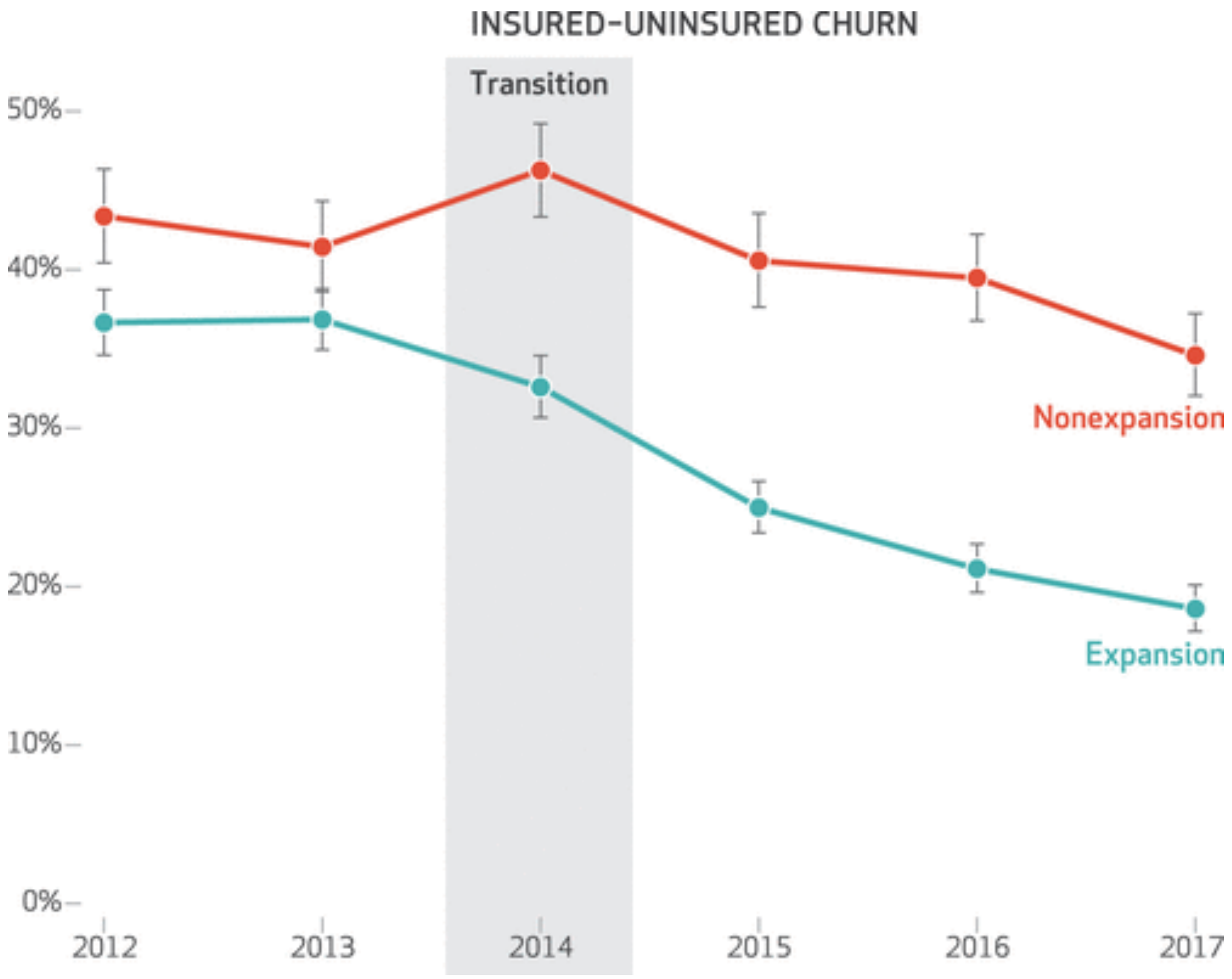
Medicaid Expansion Improved Perinatal Insurance Continuity For Low-Income Women
Daw et al 2020

Outcome: transitions from insured to uninsured during pregnancy/postpartum period

Treated: states that expanded Medicaid

Control: states that did not

How convinced are you that the counterfactual assumption is reasonable here?



Adding covariates

- We could adjust the model covariates – without these the model may suffer from omitted variable bias
- If the intervention is time varying at the group level, the key omitted variables that affect identification of the effect are time varying group level covariates
- Adding individual level covariates may improve fit but is unlikely to effect estimates of program effect
- Adding time varying group level covariates potentially has large effect on the estimates
- *Key question*: are there factors that affect the treatment group but not the control group in the post-period? Or vice versa?

Key assumptions: Counterfactual

- Key assumption in Difference-in-Differences
  - Treatment group would have followed the same trajectory as the control group in the post-intervention period if the intervention had not occurred

- This counterfactual is **not** testable
  - This counterfactual may feel more believable if
    - Control and treatment areas are qualitatively similar before the intervention
    - The composition of control and treatment groups are not changing over time in different ways
    - The control and treatment group are on a similar trajectory before the intervention

Key assumptions: Counterfactual

- We can test if the time trends of the treatment and control groups are the same prior to the intervention
- This is often called the "parallel trends assumption"
  - If we fail the test of similar pre-existing time trends, the assumption of similar trends post-intervention is less defensible
  - However, failing to reject the null hypothesis of divergence in trends does not mean we can be confident that we have parallel trends
    - These tests are often underpowered
    - Can depend on how much pre-trend data you use
    - Can also depend on functional form

Event study option

- One way to demonstrate what is happening with pre-trends is through estimating an “event study"
- We still use basic difference in differences structure – fixed effects for each time period, fixed effects for each state
- Also estimates a time to event variable indexed on when the intervention occurred
- Allows for visualization of whether pre-trends differ in treated areas
- Allows for more flexible specification of the timing of treatment effects

Example

[MEDICAID AND MORTALITY: NEW EVIDENCE FROM LINKED SURVEY AND ADMINISTRATIVE DATA]

Time is indexed to when Medicaid expansion occurred (will be a different year in different states)

Model also includes state, year and survey wave fixed effects

What do you worry about in terms of threats to validity?

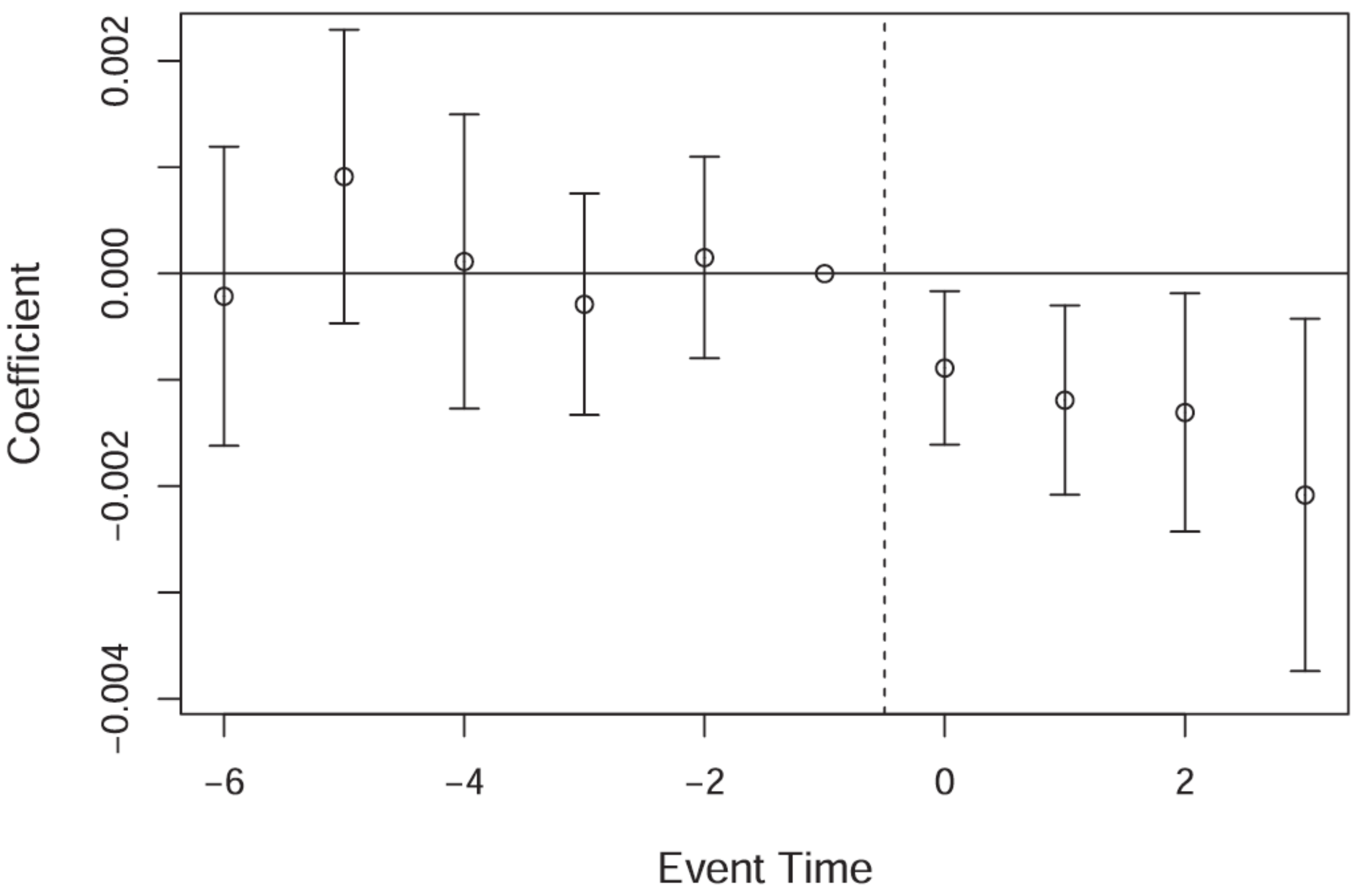


FIGURE II

Effect of the ACA Medicaid Expansions on Annual Mortality

Functional form dependence

Magnitude and sign of the DiD estimator can be sensitive to functional form

DD using level changes

	Treatment	**Control**
Before	15	5
After	25	10
Change	25 – 15 = 10	10 – 5 = 5
DD	10 – 5 = 5	

DD using log changes

	Treatment	**Control**
Before	log(15)	log(5)
After	log(25)	log(10)
Change	log(25) - log(15)	log(10) – log(5)
DD	-0.182	

Adapted from Duflo

Challenge: mean reversion

- Often treatments/programs are targeted based on pre-existing differences in outcomes
  - In fact this may be *why* some regions are chosen to participate in the program

- If outcomes have a natural tendency to mean reversion, comparing outcomes of treated to non-treated regions leads to biased estimates of the treatment effect

Key assumptions: No Contamination

- Stable unit treatment value assumption (SUTVA).
- We require that "the observation on one unit should be unaffected by the particular assignment of treatments to the other units“
- If there is spillover from treatment to control areas due to the intervention this will be violated
- What kind of spillovers would you worry about for a study on the impact of Medicaid expansion?
- Using US states in health studies
  - Contamination due to infectious disease spread
  - Movement of people to get access to policy benefits
  - Movement of health care providers in policy response

Difference in Difference estimation tips

- Show lots of figures
  - Most reviewers want to see evidence of parallel pre-trends
  - Event study can be a nice way to show this
- Demonstrate no major differential compositional changes in population of treatment or control in post period
- Do placebo/falsification tests
  - Fake time period for policy introduction
  - Outcome that shouldn’t be affected by treatment
- Show impact estimates with and without controls
  - Controls only need to be group*time varying (all others will fall out of the regression because of group and time fixed effects)
- Ideal to control for presence of other programs that might have influenced outcome and show that the impact estimate doesn’t change

Valid standard errors: Clustering

- Many difference in difference studies use individual level data to understand group level policy change
  - Challenge is that outcomes are correlated within groups
- One approach is to aggregate data up to the group level and use group level averages to estimate effects like we did in ITS
  - Will substantially restrict our power
- An alternative is to cluster standard errors at the *level of treatment*, to control for correlation of individuals within the same group
- Works well with a large number of clusters

Valid standard errors: Small number of groups

- With many groups, clustering the standard errors at the group level works well to allow for intracluster correlation
- With a small number of groups, this is not a good approach
  - Can actually lead to standard errors that are *too small*
  - Could instead average the data across groups but this will severely limit power
- Can use bootstrap/placebo methods but more complex
  - "Wild" Bootstrap – use boottest in STATA implements Cameron et al 2008
  - Construct placebo intervention data based on random assignment of intervention to groups and construct test statistics based on the distribution of outcomes for placebo interventions
  - Can work with as few as 6 groups

Valid standard errors: Autocorrelation

- With autocorrelation, observations for the same group will be correlated over time
- Observations of multiple time points pre- and post-intervention will not be independent and do not add much additional information
- With a large number of groups, clustering the standard errors across time within the same group can correct for autocorrelation
- We can also estimate the autocorrelation structure, but very biased estimates with group fixed effects and small number of periods

Difference-in-Differences vs. Interrupted Time Series

- *Conceptually similar*: both used to estimate the effect of group level interventions by creating a counterfactual; what would have happened without the intervention

- *Structural differences*
  - ITS uses aggregated data with one treated area time series whereas Difference-in-Differences usually keeps the data at the individual level and adjusts for within-group correlation.
  - ITS has one treatment group (and one control group in CITS) while Difference-in-Differences works well with multiple treated and control groups
  - ITS fits a linear time trend and extrapolates counterfactual whereas Difference-in-Differences uses time fixed effects and uses control group to generate counterfactual
  - ITS will be favored when change is expected to be immediate and large. DiD methods are better for changes that may be more subtle or phase in over time

Tomorrow we will get into

- Recent developments in understanding challenges to differences in differences
  - Heterogeneity
  - Challenges with different groups being treated at different times
- Synthetic control methods
- Other advanced difference in difference topics

- And on Monday we will recap and review all of our econometrics
  - Bring your questions!

Additional Slides

Formal difference in differences assumptions

Laura Hatfield has created a terrific website that provides a lot of formal detail surrounding difference in differences for health policy:

https://diff.healthpolicydatascience.org/

This website includes simulations you can play around with different assumptions and what they mean

Also contains lots of technical details about the different formal assumptions of DiD

Difference in Difference: Recommended readings

- Angrist and Pischke (2009). Mostly Harmless Econometrics, Section 3.2.

- Bertrand, Duflo, and Mullainathan (2004). How much should we trust difference-in-differences estimates? [Link]