# Difference-in-Differences in Action

Centre for Evaluation Seminar Series

Timothy.Powell-Jackson@lshtm.ac.uk

Department of Global Health and Development
London School of Hygiene and Tropical Medicine

# Overview

- Idea behind difference-in-differences
  - Identify when the difference-in-differences method can be used
  - Understand the key limitations of the difference-in-differences method
  - Extensions of the basic approach
- Applications of the approach

# Two naïve approaches

- Before and after
  - Outcome data are available before and after an intervention (same individuals / different individuals)
  - Estimate of impact is the difference in the mean outcome over time
  - Threat from any time varying factor (other than the programme) that influences the outcome
- With and without
  - Outcome data are available for participants and non-participants
  - Estimate of impact is the difference in the mean outcome between the two groups
  - Threat from any unobserved difference between the two groups that influences the outcome

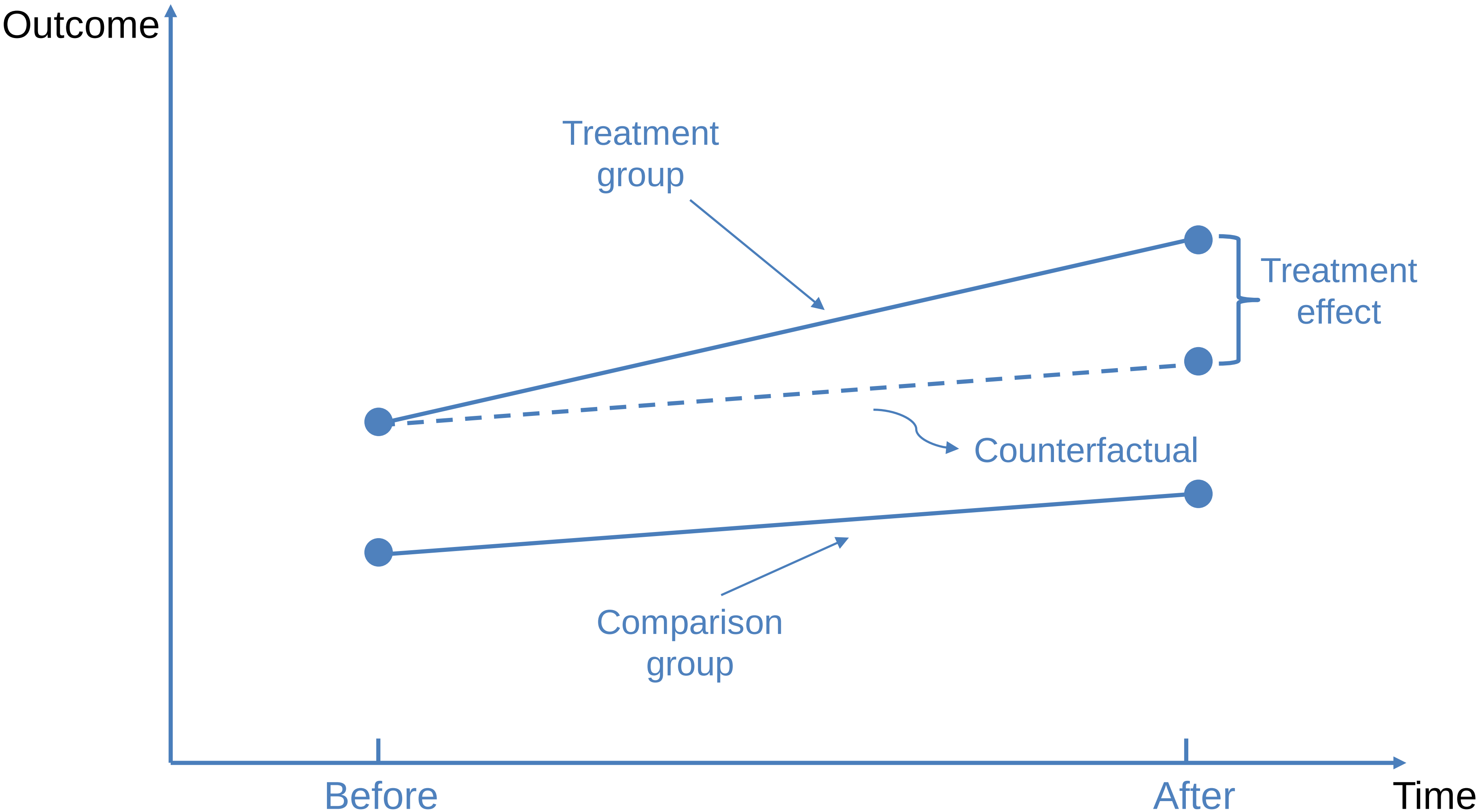
# Differences-in-differences

Idea and concepts

# What is DiD?

- Not a new method – basic idea behind John Snow's research of the cholera epidemic
- Combines the two naïve approaches to generate better estimate of the counterfactual
- Compares before and after changes in intervention group (first difference) with before and after changes in comparison group (second difference)
- Four groups: 1) intervention, before; 2) intervention, after; 3) comparison, before; 4) comparison, after
- Comparison group provides the counterfactual (what would have happened in absence of treatment)

# Basic intuition



# Numerical example

| | Before | After | Difference |
|---|---|---|---|
| Treatment | A | B | B – A |
| Comparison | C | D | D – C |
| Difference | A – C | B – D | DD = (B – A) – (D – C) |

| | Before | After | Difference |
|---|---|---|---|
| Treatment | 0.3 | 0.5 | 0.2 |
| Comparison | 0.2 | 0.3 | 0.1 |
| Difference | 0.1 | 0.2 | DD = 0.2 – 0.1 = 0.1 |

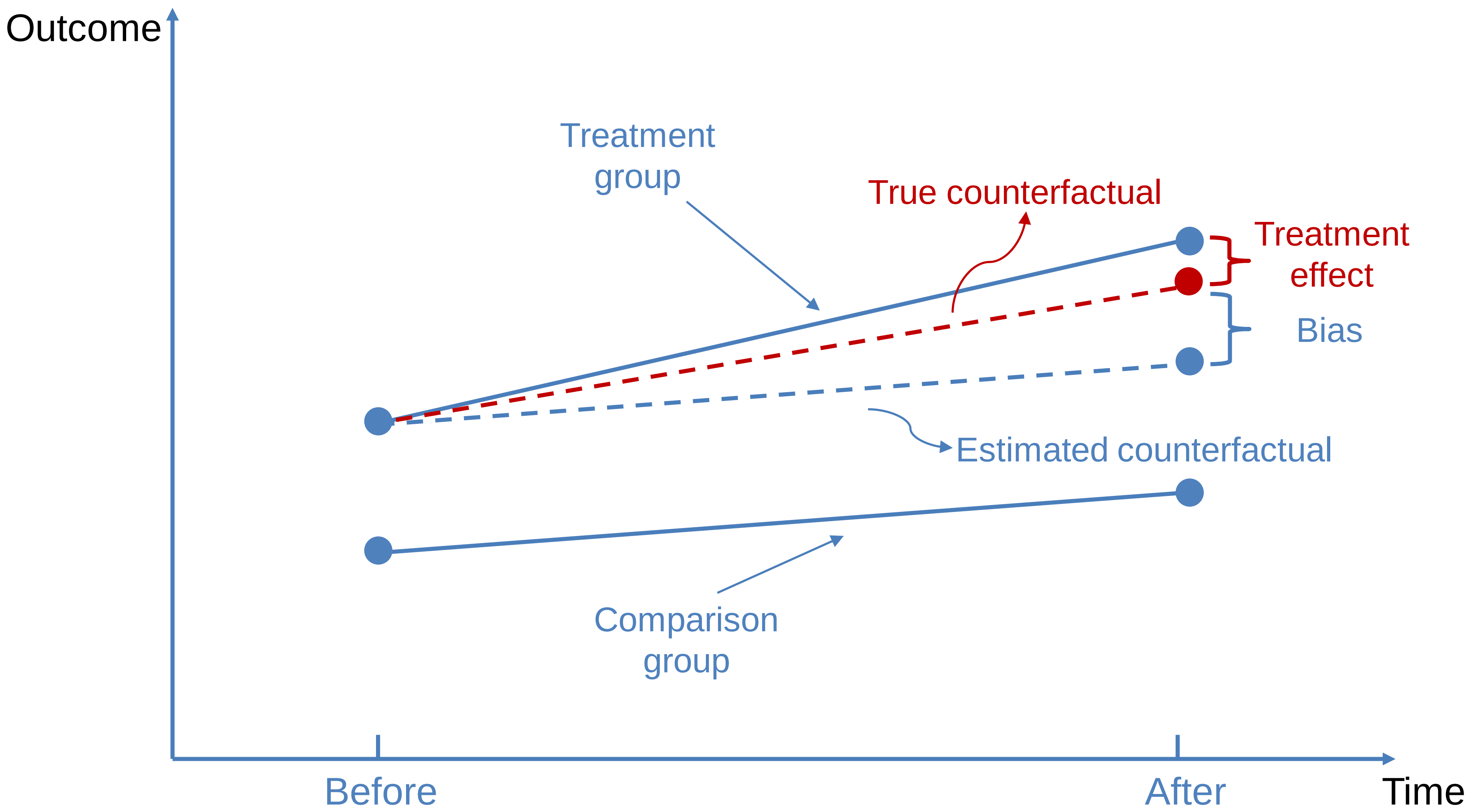
# How is DiD helpful?

- Many characteristics that are linked to outcomes likely to be fixed over time (at least during evaluation period)
  - Personal traits
  - Characteristics of location
- Difference in the outcome before and after the programme sweeps away time-invariant characteristics of the unit or individual
- DiD controls for both *observed* and *unobserved* characteristics that are time invariant

# Key assumption

- DiD cannot eliminate (unobserved) differences between the treatment and comparison groups that change over time
- DiD must assume no such time-varying differences exist between treatment and comparison groups
  - Put another way, in the absence of treatment, outcomes in treatment and comparison groups display equal trends
  - Referred to as the "parallel trends assumption"
- If violated, DiD produces biased estimates
  - Threat comes from any time-varying factor that differentially affects outcomes in treatment and control

# Violation of assumption



# Differences-in-differences

In practice

# Data and analysis

- Time and geographic (location) variation
  - Longitudinal data… on programme region(s) and comparison region before and after introduction of programme
- Groups can be constructed in other ways – e.g. location and poverty level
- More data the better
  - Number of pre-intro time periods, number of units (e.g. districts)
- Beware of changes in how outcomes are measured
- OLS regression typically used to generate DiD estimate (interaction variable)

# Extensions of basic approach

- Multiple time periods and sequential treatment (think stepped wedge trials)
- Triple difference-in-differences
  - Three sources of policy variation
- Difference-in-differences combined with matching
- Study of policies defined in terms of treatment intensity (continuous) variables
- Dealing with serially correlated outcomes
- Allowing for unit specific time trends

# When is DiD best used?

- Placement of policy is "exogenous"
  - Introduced in an arbitrary quasi-random fashion
  - Factors determining placement of policy are known, measured and controlled for
- Policy change is dramatic, introduced with little lead time, and well implemented
- Contamination is minimal
  - Policy does not spillover to comparison sites
  - Migration towards policy locations is difficult

# Social franchising for maternal health in India

(Tougher et al 2018, Lancet Global Health)

# Context

- What should we do about the private sector in India? Engage with it rather than ignore it!
- Does social franchising improve the quality and coverage of health services in the population?
- Evaluation of the Matrika social franchise model to improve maternal health in Uttar Pradesh, India
- Intervention comprised social franchising, training of public and private providers, and demand generation activities

# Implementation of DiD

- Geographical variation in the programme
  - 3 intervention districts and 3 comparison districts
- Time variation in exposure to programme
  - Before introduction programme (pre-September 2013)
  - After introduction of the programme (post-September 2013)
- Difference-in-differences approach using data collected in multiple rounds of a household survey of 6,998 women in 180 villages
- Prospective matching (using Indian Census 2011) generates baseline balance across outcomes and covariates

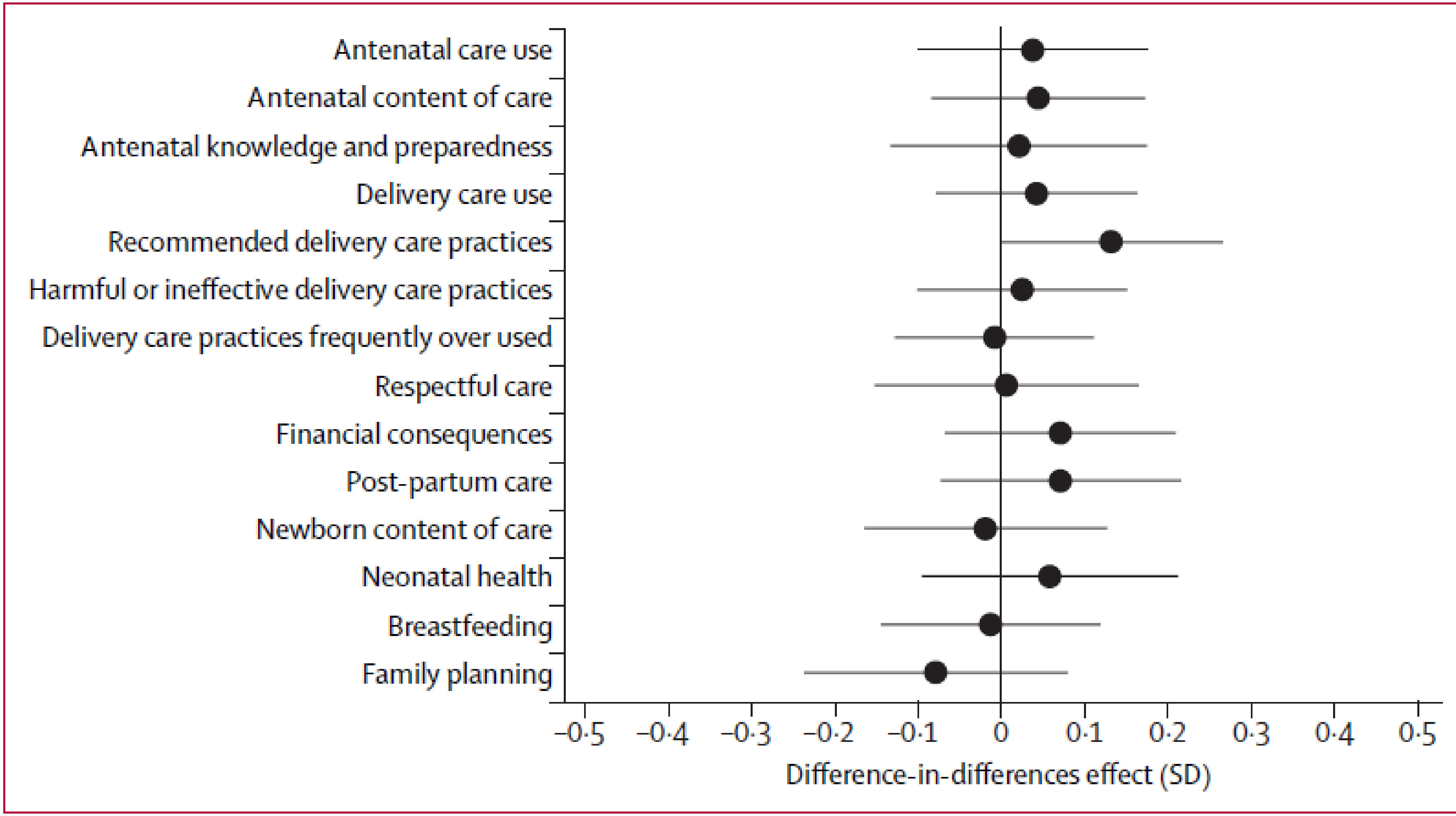



*Figure 3:* **Effect of the Matrika programme on summary measures of outcomes**
Data are from two rounds of a survey of women aged 15–49 years who gave birth in the previous 2 years (round 1) and 18 months (round 2), including those women who had a stillbirth or whose child had died since childbirth. This figure shows standardised treatment effects on indices generated from multiple outcomes within a family. We recoded individual outcomes when necessary so that higher values correspond to better outcomes. Treatment effects are presented in SD units of the comparison group. We did these analyses with all available data.

# Hospital pay-for-performance in England's NHS

(Sutton et al 2012, NEJM)

# Context

- Pilot in northwest region of England in late 2008
- Hospitals required to collect data on 28 clinical quality measures
- Best performing hospitals received bonus payment at end of each year
- Money given internally to clinical teams to be invested in improvements to quality
- Supported by other activities (eg. shared learning events)

# Implementation of DiD

- Patient level data on mortality:
  - Acute myocardial infarction, heart failure, pneumonia incentivised by scheme
  - Six other conditions not incentivised by scheme
- Mortality risk-adjusted, by hospital by quarter year
- Comparison of mortality in northwest hospitals with those in rest of England before and after P4P scheme started
  - 18 months pre- and 18 months post-introduction of P4P
- Mortality trends shown to be similar before the P4P scheme
- Triple difference also performed (time, geography, condition)

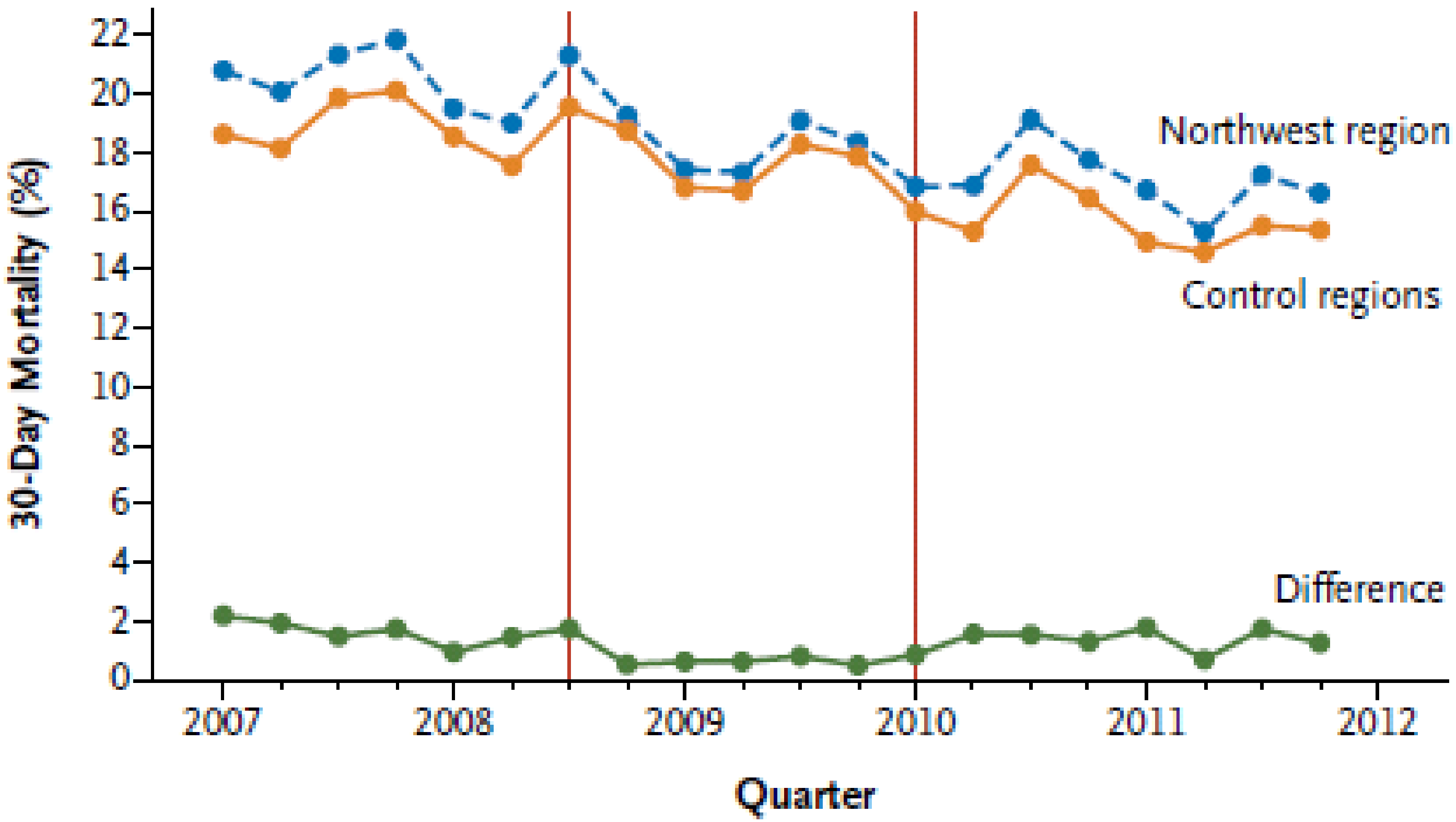



**Figure 2.** **In-Hospital Mortality at 30 Days for Conditions Linked to Incentives.**

The vertical line at the left indicates the start of the short-term period (months 1 through 18 of the program), and the vertical line at the right indicates the start of the long-term period (months 19 through 42 of the program).

Kristensen et al 2014, NEJM

# Newspaper entry and political participation in the US

(Gentzkow et al 2011, AER)

# Background

- Does the entry or exit of local newspapers affect political participation (i.e. voter turnout)?
- Uses county level data over the period 1869 to 2004 in the US
- Treatment variable is number of local newspapers in a county
- Uses a difference-in-differences type approach that includes both leads and lags to detect anticipatory effects (bias) and long-term effects

Panel B. Turnout controlling for demographics

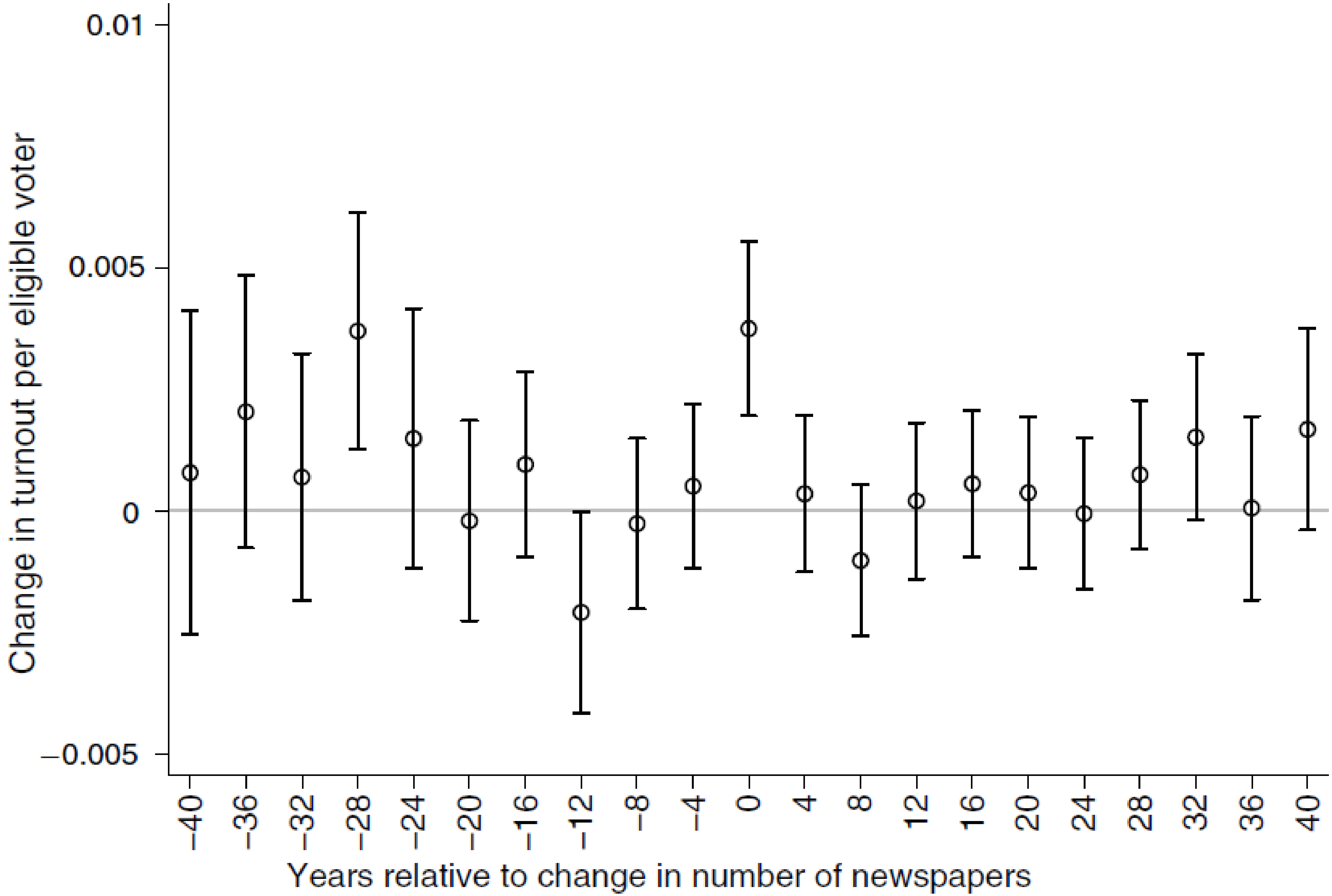



Figure 2. Presidential Turnout and Newspaper Entries/Exits

# Concluding remarks

- Difference-in-differences methods popular in impact evaluation
- Increasingly used in medical journals, sometimes labelled as “difference-in-differences”
- Literature scattered with examples of poor practice in the implementation of the method
- Data requirements can be demanding
- Parallel trends assumption must be defended

# Other useful references


- Babiarz KS, Miller G, Yi H, Zhang L, Rozelle S. New evidence on the impact of China's New Rural Cooperative Medical Scheme and its implications for rural primary healthcare: multivariate difference-in-difference analysis. BMJ. 2010
- Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. Lancet. 2011
- Daw JR, Sommers BD. Association of the Affordable Care Act Dependent Coverage Provision With Prenatal Care Use and Birth Outcomes. JAMA. 2018
- Mohanan M, Babiarz KS, Goldhaber-Fiebert JD, Miller G, Vera-Hernández M.Effect Of A Large-Scale Social Franchising And Telemedicine Program On Childhood Diarrhea And Pneumonia Outcomes In India. Health Aff (Millwood). 2016
- Chetty, R., A. Looney and K. Kroft (2009). "Salience and Taxation: Theory and Evidence." American Economic Review 99(4): 1145-1177.
- Ryan AM, Burgess JF, Dimick JB. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. Health Services Research. 2015
- Athey, S. and G. W. Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation." Journal of Economic Perspectives 31(2): 3-32.

# Assessing validity of DiD

- Validity of parallel trends assumption cannot be formally proved (or disproved)
- Evidence can be shown to support assumption
  1. Pre-programme trends move in tandem
  2. Policy placement uncorrelated with historical levels of outcome
  3. Placebo outcome test

# Descriptive findings

| Health Conditions | Northwest Region | | | Rest of England | | |
|---|---|---|---|---|---|---|
| | Mortality before Introduction | Mortality after Introduction | Difference | Mortality before Introduction | Mortality after Introduction | Difference |
| | | | *percent* | | | |
| Not included in the program | 13.1 | 12.1 | –1.0 | 12.0 | 10.7 | –1.3 |
| Included in the program | 21.9 | 20.1 | –1.8 | 20.2 | 19.3 | –0.9 |
| Acute myocardial infarction | 12.1 | 10.7 | –1.4 | 11.3 | 10.4 | –1.0 |
| Heart failure | 18.8 | 17.5 | –1.3 | 16.9 | 15.8 | –1.1 |
| Pneumonia | 29.4 | 27.0 | –2.4 | 27.1 | 26.3 | –0.7 |

Sutton et al 2012, NEJM

# Impact estimates

- P4P reduced mortality for conditions included in the scheme
- P4P had no effect on mortality for conditions not included in scheme
- Mortality impact driven by effect on pneumonia mortality
- Subsequent paper found mortality impacts did not last!

| Health Conditions | Between-Region Difference in Differences‡ |
|---|---|
| | *percent (95% CI)* |
| Not included in the program | 0.3 (−0.4 to 1.1) |
| Included in the program | −0.9 (−1.4 to −0.4) |
| Acute myocardial infarction | −0.3 (−1.0 to 0.4) |
| Heart failure | −0.3 (−1.2 to 0.6) |
| Pneumonia | −1.6 (−2.4 to −0.8) |

Sutton et al 2012, NEJM