Synthetic Control Methods workshop: Examples and methods

Geraldine Clarke

Autumn 2020

Contents

Background

Difference-in-differences (DiD) methods are widely used to control for confounding.

DID assumes that in the absence of an intervention, outcomes for treated and control units would follow parallel trajectories over time.

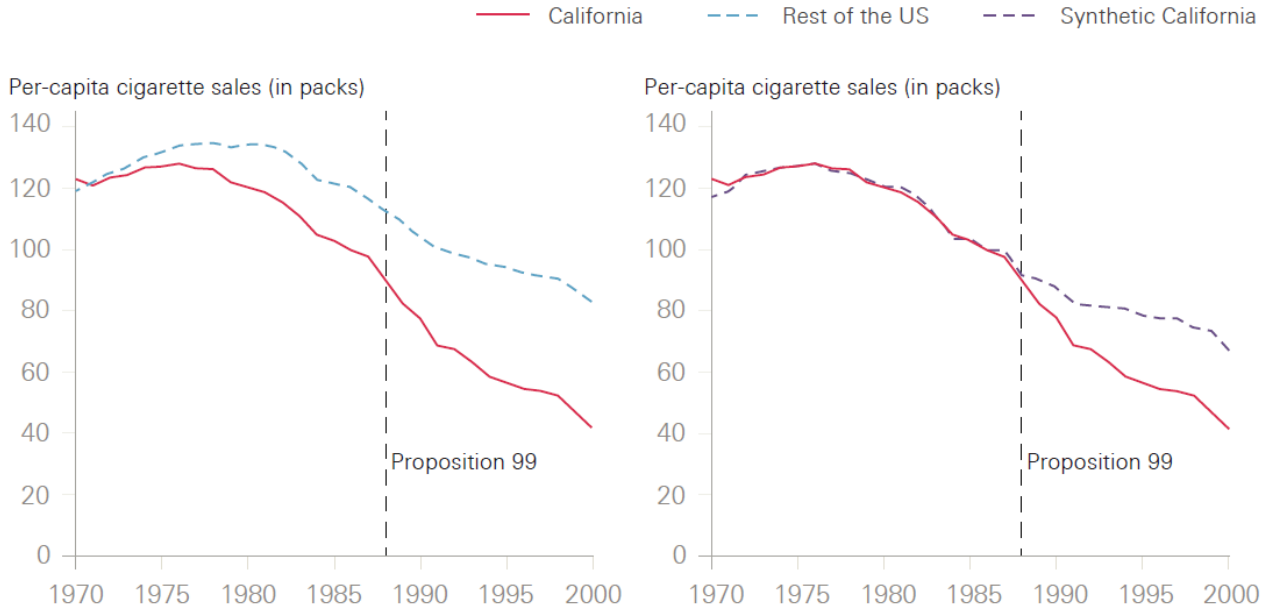
This means no predictors of the outcome that vary over time differently across treated and control groups.

SCM allows predictors to vary over time.

Can provide unbiased estimates as long as pre-intervention period is long enough

Useful for initiatives taking place at an aggregate level with few treated units and multiple controls.

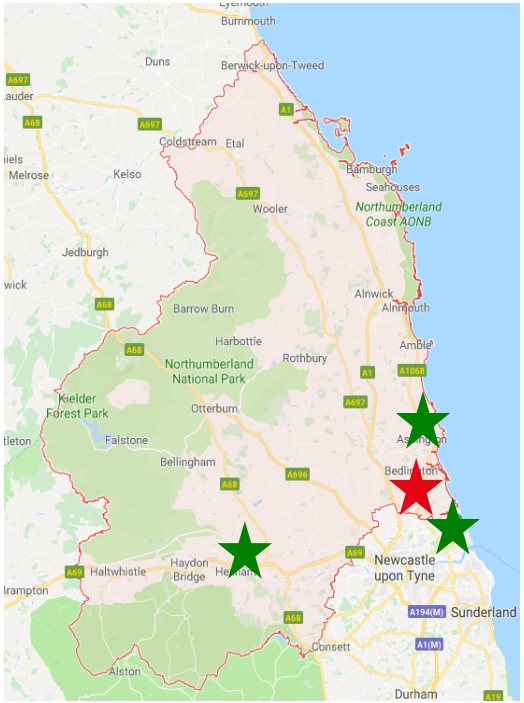
Example 1 | Impact of California 1988 anti-tobacco policies on smoking rates (Abadie et al. 1993)



Synthetic California: 33% Utah, 23% Nevada, 20% Montana, 16% Colorado and 7% Connecticut

Example 2| Impact of redesigning urgent and emergency care in Northumberland (IAU 2017)

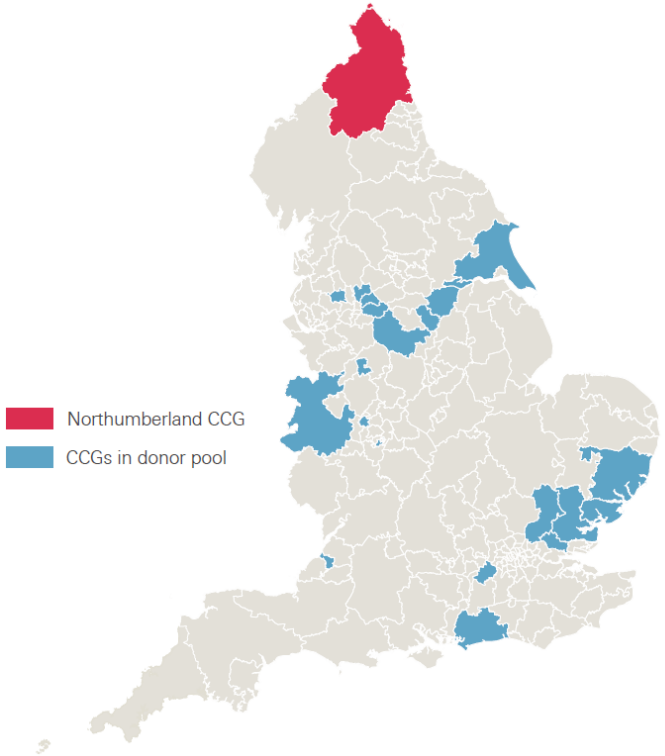
- Northumberland is a Primary and Acute Care Services vanguard serving ~ 320,000.
- In June 2015 a new specialist emergency care hospital was opened in Cramlington, Northumberland and three existing A&Es in North Tyneside, Hexham and Wansbeck were gradually refocused on providing care for minor injuries
- Hospital activity in Northumberland CCG was compared to that of a *synthetic* control area formed by combining data from comparable CCGs in other parts of England.



Example 2: Methods

Selecting a donor pool: 20 CCGs with similar values of

- Age, gender, ethnicity
- Education level
- Population density
- Socioeconomic and health deprivation
- Number of GPs and care home beds
- QOF achievement score
- Disease prevalence and comorbidities
- Hospital utilisation



Example 2| Methods continued

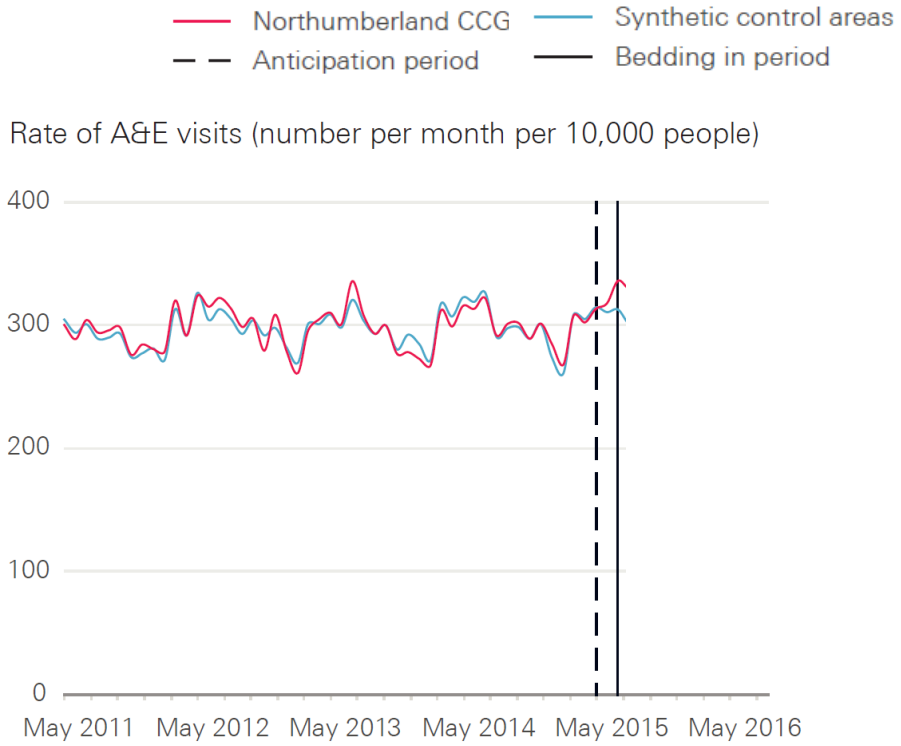
Selecting the synthetic control:

33% Heywood, Middleton and Rochdale

39% Doncaster

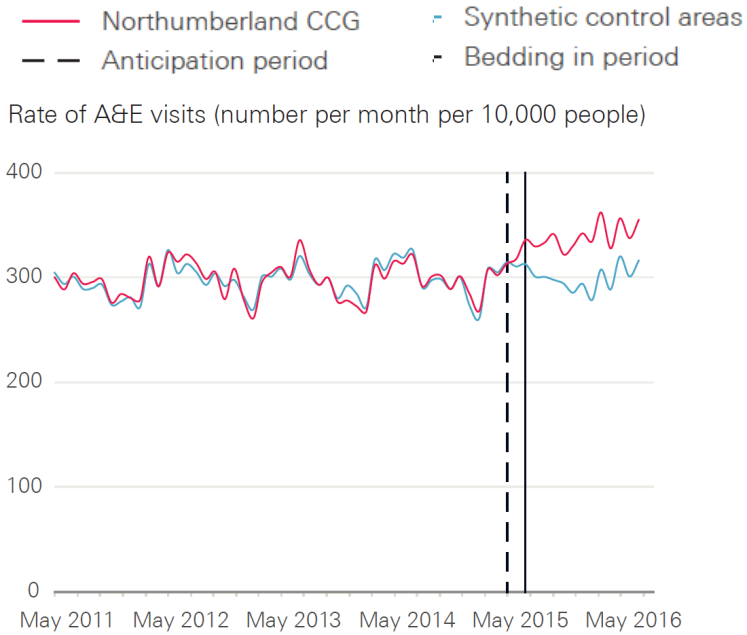
27% North Derbyshire

1% Birmingham South and Central



Example 2| Methods continued

Estimating the impact:



- Differences between red and blue lines in the post-intervention period indicates an increase in A&E attendances in Northumberland CCG compared to the synthetic control after the opening of the hospital in June 2015.

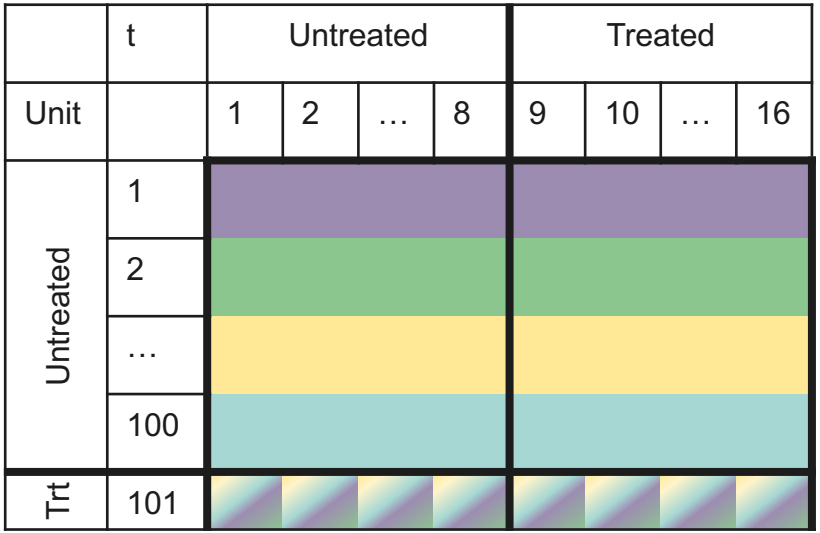
Example 2| Results

	Northumberland CCG (mean, SD)	Relevant synthetic control area (mean, SD)	Difference	Relative difference (%)	Significance score
Rates of hospital use: expressed as activity per month per 10,000 people					
A&E visits	339.6 (12.7)	299.1 (12.1)	40.5	13.6	0%
Admissions	274.6 (9.1)	272.3 (7.1)	2.2	0.8	95%
Emergency admissions	74.2 (1.8)	75.4 (1.4)	-1.2	-1.6	70%
A&E visit length in minutes	143.8 (5.0)	158.1 (6.8)	-14.3	-10.5	0%
Length of admissions in days	2.7 (0.2)	2.3 (0.8)	0.4	12.1	10%
Expressed as a proportion of patients					
A&E visits lasting less than 4 hours	91.8	85.2	6.7	6.9	0%

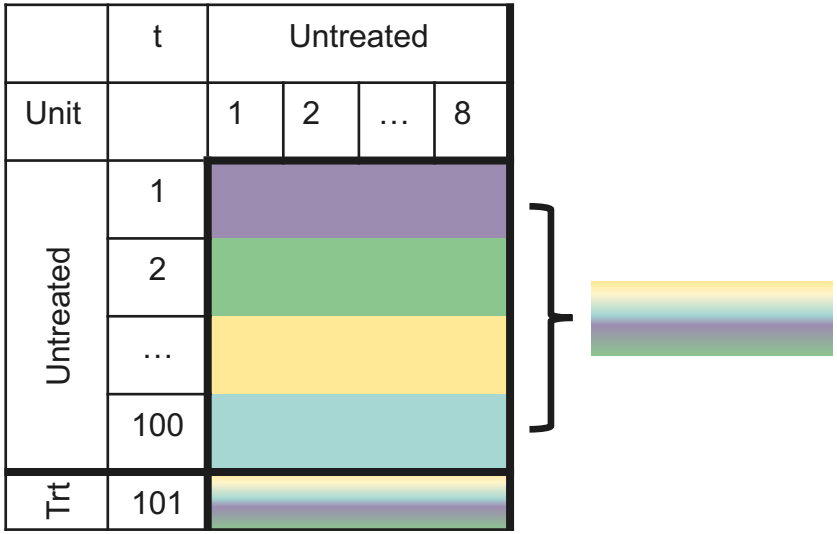
Note: SD=standard deviation. A lower significance score indicates greater confidence that the effect is not due to chance.

- Northumberland residents had a 13.6% increase in rate of A&E visits
- 91.8% of patients in Northumberland were admitted, transferred or discharged within 4 hours of attending A&E, compared with 85.2% in the control area
- On average, the duration of A&E visits were 14.3 minutes shorter than expected

Some maths



Some maths



Suppose we have:

outcomes Y_i and one or more predictors X_i at each time period

Chose 100 weights w_i, w_2,.. 100 such that they

- add up to 1:

$$\sum_{1}^{100} w_i = 1$$

- In each pre-intervention period t0:
  - The weighted average of the control outcomes = the treated outcome

$$\sum_{1}^{100} w_i Y_{it0} = Y_{101,t0}$$

  - The weighted average of the control predictor = the treated predictor

$$\sum_{1}^{100} w_i X_{it0} = X_{101,t0}$$

ATT

ATT at time t1 in post-intervention period

$= ATT_{t1}$

= (weighted outcome in controls) $_{t1}$ − (outcome in treated unit) $_{t1}$

$$= \sum_{1}^{100} w_i Y_{i,t1} - Y_{100,t1}$$

$$ATT = \frac{1}{8}\sum_{9}^{16} ATT_{t1}$$

Now open:

/workshop/SCM-workshop_for_participants.Rmd