

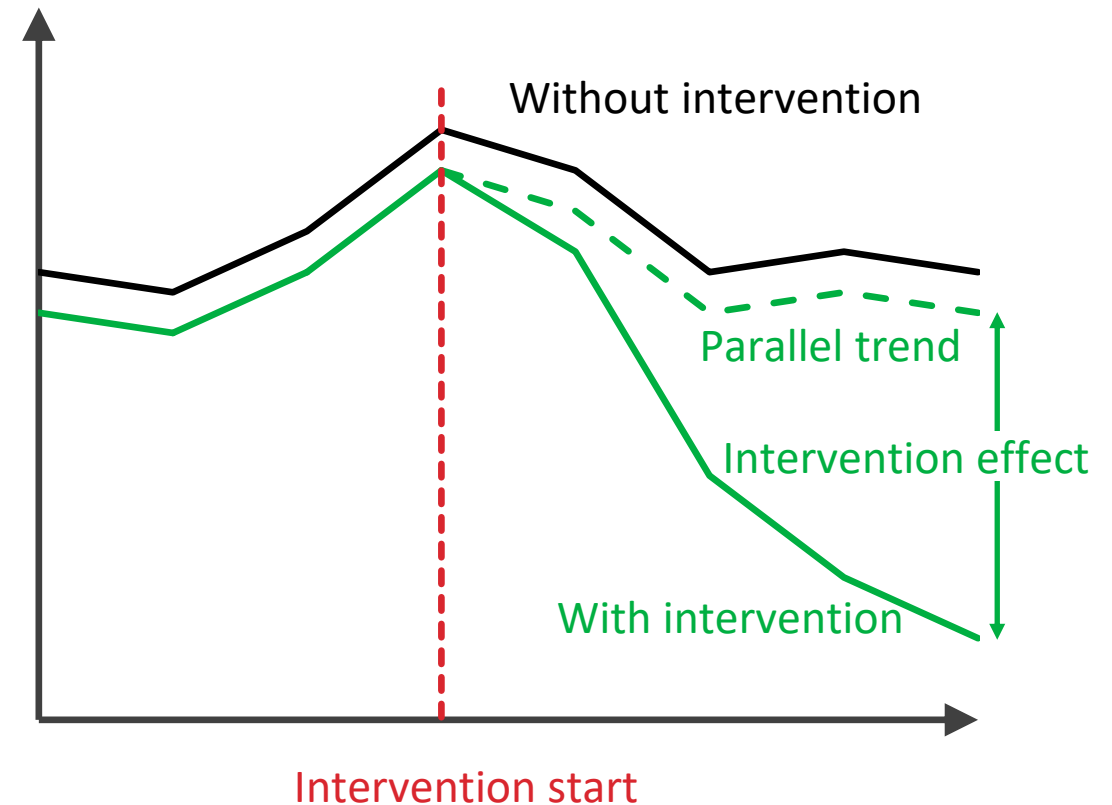
Difference in differences using constraints in Stata

Colin Birch, Dept Epidemiological Sciences, Animal and Plant Health Agency
colin.birch@apha.gov.uk

DOI: [10.1038/s41598-024-54062-4](https://doi.org/10.1038/s41598-024-54062-4)

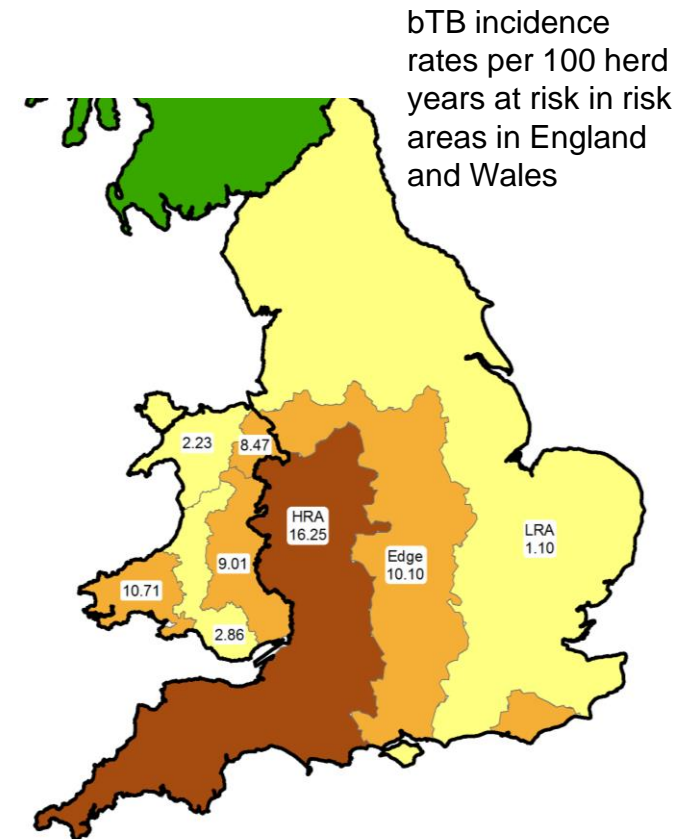
Difference in differences emphasizing the parallel trends assumption

- The parallel trends assumption:
 - Suppose we can estimate an average trend for an observation, excluding the effect of an intervention.
 - Assume the trend is independent of whether or when the intervention is applied.
- We estimate the parallel trend
- The average intervention effect is the divergence from the parallel trend observed after the intervention.
- Parallel trends can't be demonstrated?



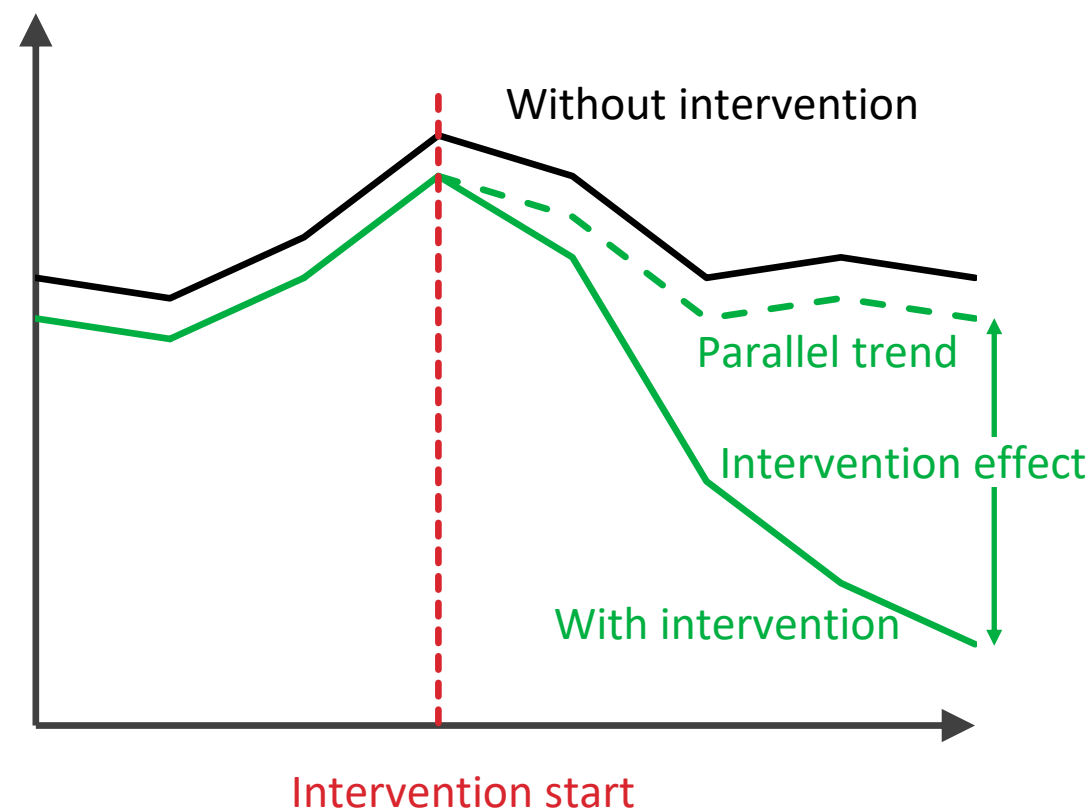
Badger Control Policy (BCP) in England

- An intervention to reduce Tuberculosis (TB) incidence rate in cattle
- Controls in cattle were not reducing TB in large areas
- Cattle and badgers share tuberculosis infection
 - Badger culling licenses reintroduced
- Pilots from 2013, rapid rollout from 2016
- Multiple areas start BCP each year 2016-2020
- Culling licenses issued for 52 areas up to 2020
- Culling maintained for at least 4 years.
- By end of 2022, BCP included > 90% high TB incidence areas in England.



Difference in differences analysis included heterogeneous trends before intervention

- Analysis included 4 years before start of BCP, 2009-2013.
- Trends differed between areas before the start of BCP in each area.
 - TB incidence rate converged. Between area variance in 2013 < 2009
- Fitted descriptive trend in each area to estimate TB incidence rate at start of BCP.
- Trend NOT assumed to continue.
 - Implicit trend after BCP starts = 0
- Constrained mean trend before BCP = 0



Constraints in Stata are very flexible!

* Set a constraint for the sum of heterogeneous observable slopes = 0

```
constraint 1 1.AreaID#c.IntYearX + 2.AreaID#c.IntYearX + 3.AreaID#c.IntYearX +  
4.AreaID#c.IntYearX + 5.AreaID#c.IntYearX + 6.AreaID#c.IntYearX +  
7.AreaID#c.IntYearX + 8.AreaID#c.IntYearX + 9.AreaID#c.IntYearX +  
10.AreaID#c.IntYearX + 11.AreaID#c.IntYearX + 12.AreaID#c.IntYearX +  
13.AreaID#c.IntYearX + 14.AreaID#c.IntYearX + 15.AreaID#c.IntYearX +  
16.AreaID#c.IntYearX + 17.AreaID#c.IntYearX + 18.AreaID#c.IntYearX +  
19.AreaID#c.IntYearX + 20.AreaID#c.IntYearX + 21.AreaID#c.IntYearX +  
22.AreaID#c.IntYearX + 23.AreaID#c.IntYearX + 24.AreaID#c.IntYearX +  
25.AreaID#c.IntYearX + 26.AreaID#c.IntYearX + 27.AreaID#c.IntYearX +  
28.AreaID#c.IntYearX + 29.AreaID#c.IntYearX + 30.AreaID#c.IntYearX +  
31.AreaID#c.IntYearX + 33.AreaID#c.IntYearX + 34.AreaID#c.IntYearX +  
35.AreaID#c.IntYearX + 36.AreaID#c.IntYearX + 37.AreaID#c.IntYearX +  
38.AreaID#c.IntYearX + 39.AreaID#c.IntYearX + 40.AreaID#c.IntYearX +  
41.AreaID#c.IntYearX + 42.AreaID#c.IntYearX + 43.AreaID#c.IntYearX +  
44.AreaID#c.IntYearX + 45.AreaID#c.IntYearX + 46.AreaID#c.IntYearX +  
47.AreaID#c.IntYearX + 48.AreaID#c.IntYearX + 49.AreaID#c.IntYearX +  
50.AreaID#c.IntYearX + 51.AreaID#c.IntYearX + 52.AreaID#c.IntYearX +  
53.AreaID#c.IntYearX = 0
```

Applying the constraint in `cnsreg` (or `glm`)

- * Use `cnsreg` to apply a regression constrained by the
- * constraint above.
- * `TBincidence` = confirmed TB incidence rate
- * `AreaID` and `Year` indicate the fixed effects of location and
- * year.
- * `BCP` = 6 levels of intervention: 0 and 1st - 5+ years after
- * start.
- * `i.AreaID#c.IntYearX` fits heterogeneous slopes up to the start
- * of interventions. (heterogeneous observable trends)
- * Only include up to Area 53, omitting Area 32.

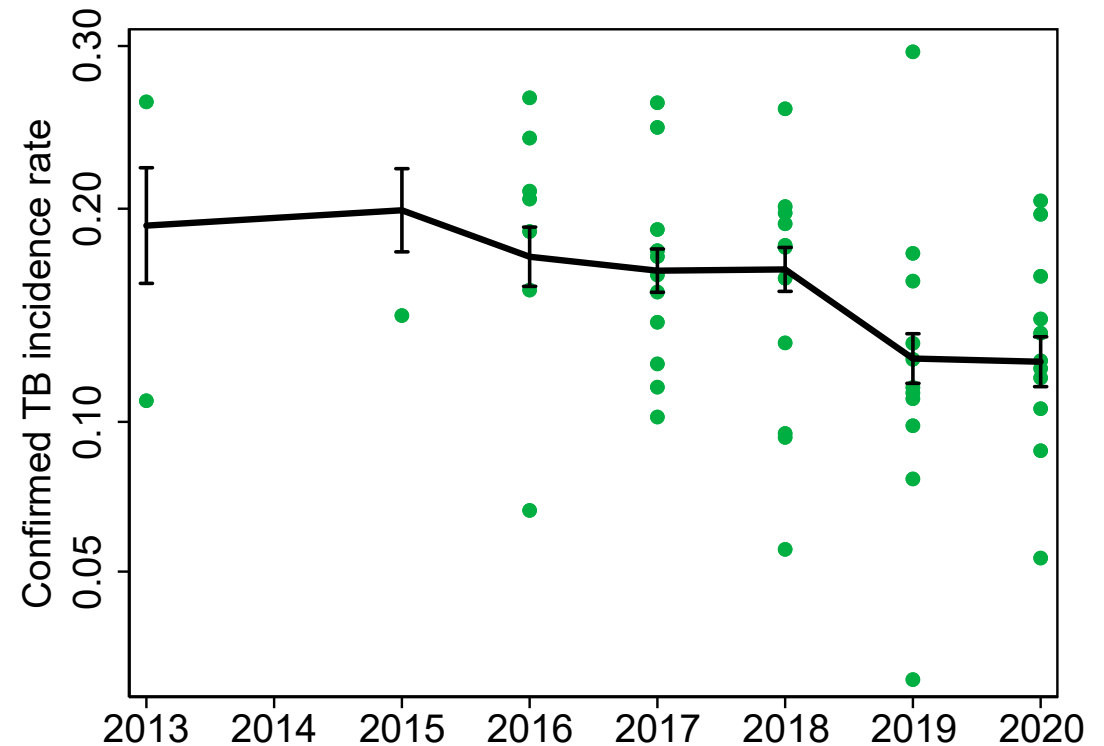
```
cnsreg TBincidence i.AreaID i.Year i.AreaID#c.IntYearX i.BCP if  
AreaID!=32 & AreaID<54, constraints(1) vce(cluster AreaID)
```

Tuberculosis incidence rate just before interventions start, with linear prediction

```
margins if Year == StartYear-1, over(StartYear)
```

```
marginsplot, addplot(scatter TBincidence StartYear ...)
```

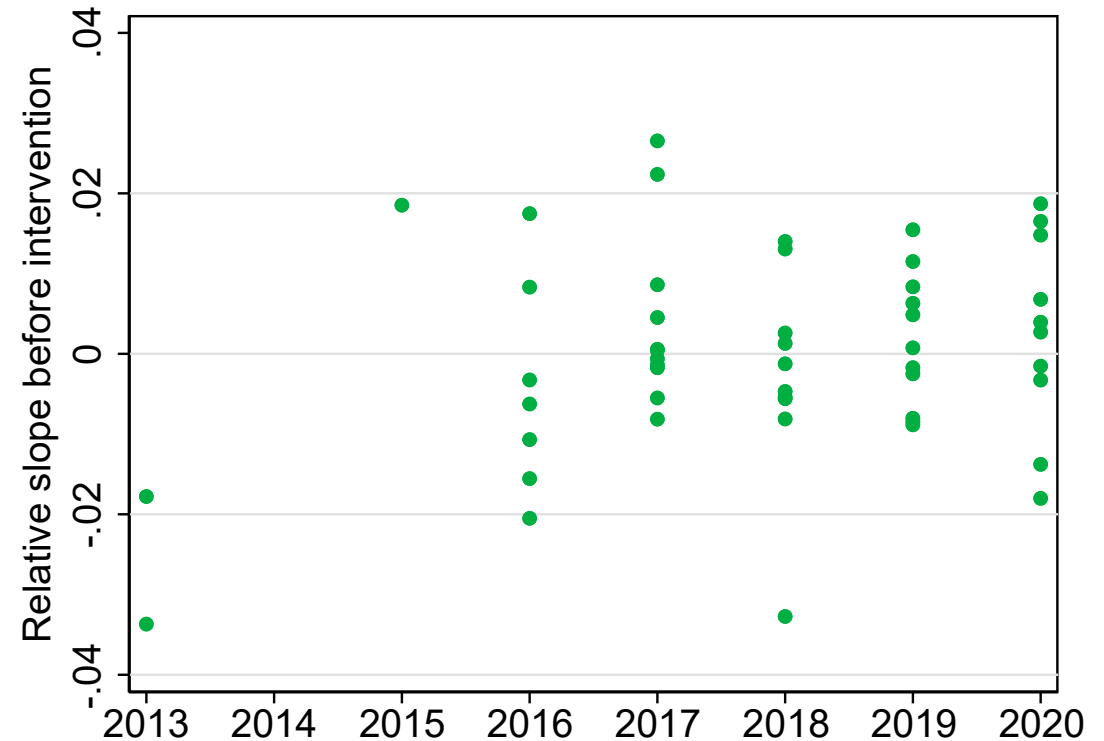
- Selection bias – TB incidence rate tended to be lower in areas starting later
- Reason for using difference in differences



Heterogeneous linear slopes before BCP intervention

* Elegant Stata code not available

- The analysis included distinct linear trends for each area up to the start of interventions.
 - “Heterogeneous observable trends”
- Overall average constrained = 0
- The average deviation of the trends was near zero for 2016 – 2020.
- **Demonstrates the parallel trend assumption?**

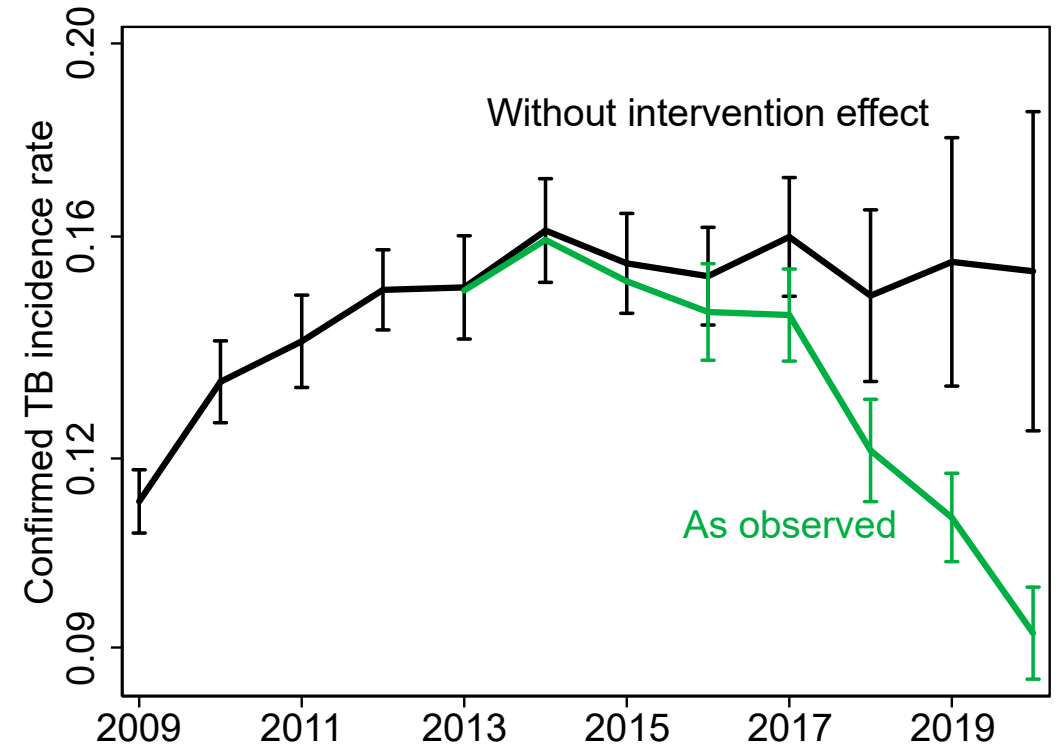


Estimated annual average TB incidence rate without intervention effect vs as observed

margins, over(Year) at (BCP=0)

margins, over(Year)

- The time trend estimated by the analysis shows the parallel trend without the effect of interventions.
- This trend was nearly static from 2014
- Hence the observed decline in TB incidence rate since 2014 roughly matched the cumulative intervention effect.



Estimating the intervention effect

margins BCP

marginsplot...

- The BCP intervention effect increased with time after it started.
- Effect increased at least to the third intervention year
- The estimated decline in the fourth intervention year was over 50%

