Difference in Difference: A tool for policy evaluation

23 June 2022

Heather Brown
h.w.brown@lancaster.ac.uk

Evaluating the effectiveness of policy

Lancaster University

- Try a before or after (pre-post) evaluation
- Compare outcomes before and after the intervention to assess if the intervention has lead to any change
- Subject to bias

What can go wrong?

- There may be unobserved factors that impact on the outcome of interest and may also change with the intervention.

Find a control group

- Is there another area that is similar in terms of key characteristics related to the outcome
- Or another group of people who are similar to the people who were exposed to the intervention.

Some more things to think about for a control group

- Groups are often defined by geography and time period.
- Groups could also be defined by socioeconomic status and time or age/gender and time.
- Differencing needs to be a credible way to deal with non-group equivalence.

Difference in Difference

- If unobserved factors that could bias your pre-post intervention also affect your control group then be preforming double differencing this will remove this type of bias.
- Then you can isolate the treatment effects and estimate the impact of the policy.
- Difference in difference is effectively a variant of a fixed effects model.
- Time varying covariates that may impact on the outcome of interest can be easily added to the model.
- Time constant variables will be differenced out of the model.

What data do you need

- Data is needed on 2 groups: 1) impacted by policy/intervention and 2) not impacted by policy/intervention.
- Data is needed for both groups for at least one period before the policy/intervention and one period after.
- Using data from the same individuals over time means that you can control for time constant unobserved differences between the control and treatment groups that may affect the outcome measure.

UK Data Archive: https://www.data-archive.ac.uk/

- Repeated cross-sections such as the Health Survey of England and Scottish Health Survey
- Longitudinal data such as Understanding Society, English Longitudinal Survey of Aging, Millennium Cohort Survey, National Child Development Survey

Government data sources (some examples)

- https://digital.nhs.uk/data
- https://www.nomisweb.co.uk/
- Linking across data sources

Example 1:https://repec.iza.org/dp14968.pdf

Lancaster University

- Did the implementation of the Welfare Reform Acts 2012 and 2016 increase mental health inequalities between low income single mothers and other low income women
- We used ten waves of data from the UK Understanding Society Survey (USS), covering the period between 2009-2019 (University of Essex, 2020).

The Sample

- Working-age women (aged 16-64) who were on low-income pre-reform (i.e. at wave 1).
- Intervention group➔ Unpartnered mothers who have at least one child (under 16) and remain single for the entire sample period, and live in a low-income household at baseline (wave 1).
- Control group➔ Low income at baseline partnered mothers and low income childless women at baseline

Variables

- **Outcome variable**: mental health measured by GHQ-12 (higher scores mean better mental health) and SF-12
- **Key explanatory variables**: reform implementation dummies (post-April 2013 and post April 2017)
- **Other covariates**: age, age squared, ethnicity, education, number of children, number of adults in the household, age of the youngest child, housing tenure and region

The Model

Lancaster University

$$MHit=\alpha i+\beta 1Xit+\beta 2Comparison\ group+ \beta 3Reform\_2013+\beta 4Reform\_2017+ \beta 5Comparison\ group*Reform\_2013 + \beta 6Comparison\ group*Reform\_2017+\beta 7Year\ \varepsilon it$$

Parallel Trends Assumption

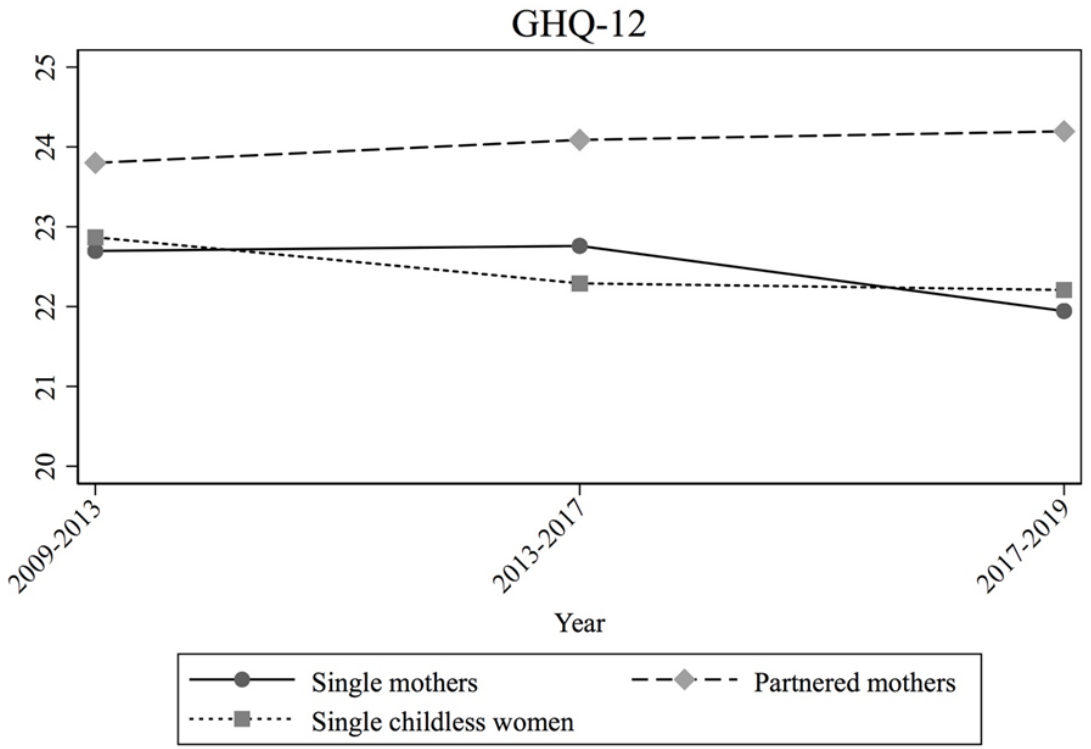
Lancaster University

- Prior to each reform period, the difference between the treatment and control group outcomes is constant over this time period.
- In this study, we test this by an event study analysis by including leads and lags of the first reform (2013).
- The parallel trend assumption holds if the coefficients of the leads were not statistically significant from zero.
- If the effect of the 2013 reforms was a shift in the trend line (rather than the slope), the additional effects of the second round of reforms could simply be estimated using an event study design with data from 2014-2019, with 2016 as the omitted reference year (i.e. an equivalent of the analysis for the first round of reforms described above).

Parallel Trends Assumption Continued

- To check if the effect of the 2013 reforms was a parallel shift (i.e. a change in the intercept only, not the slope), we investigated the equality of the lead coefficients of the years 2014, 2015 and 2016 using an F-test whereby a non-rejection of the null hypothesis means that it can be assumed that the trend shift was parallel.
- As a further test of validity of the difference-in-difference estimates, we also explored if the results were sensitive to alternative reform date specifications, using as alternatives the welfare reform act legislation dates (March 2012 and March 2016 for WRA 2012 and WRWA 2016 respectively).

Trends in mental health



Results

Table 1 Difference-in-difference results, comparison group: partnered mothers

	GHQ-12	SF-12
After April 2013 and before 2017 X Single mother	-0.536	-0.195
	(0.330)	(0.513)
After April 2017 X Single mother	-1.216**	-0.875
	(0.534)	(0.810)
R-squared	0.01	0.02
F-statistic	1.76***	2.80***
Rho	0.63	0.63
Observations	8162	8624

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controlling for age, age squared, number of adults in the household, number of own children, age of the youngest child, housing tenure, year.

Table 1 Difference in difference results, comparison: single childless women

	GHQ-12	SF-12
Single mother X After April 2013 and before 2017	-0.151	1.054*
	(0.341)	(0.564)
Single mother X After April 2017	-0.902	-0.123
	(0.563)	(0.910)
R-squared	0.009	0.021
F-statistic	1.43*	3.88***
Rho	0.62	0.61
Observations	8679	9100

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results of Parallel Trends Assumption

	Partnered mothers		Single childless	
	GHQ-12+	SF-12	GHQ-12	SF-12
Ref. Single mother X 2012				
Single mother X 2009	-0.187**	-0.0809	-0.0942	-0.0742
	(0.0869)	(0.0834)	(0.0853)	(0.0833)
Single mother X 2010	0.00849	-0.0763	-0.0131	-0.0104
	(0.0689)	(0.0679)	(0.0695)	(0.0676)
Single mother X 2011	0.0101	-0.0512	0.0562	0.0191
	(0.0652)	(0.0634)	(0.0651)	(0.0621)
Single mother X 2013	0.00600	-0.00111	-0.0106	0.0683
	(0.0706)	(0.0683)	(0.0710)	(0.0694)
Single mother X 2014	-0.0462	-0.0206	0.000281	0.0723
	(0.0784)	(0.0749)	(0.0755)	(0.0728)
Single mother X 2015	-0.112	-0.0366	-0.0205	0.0696
	(0.0811)	(0.0808)	(0.0821)	(0.0818)
Single mother X 2016	-0.201**	-0.123	-0.0954	0.0595
	(0.0856)	(0.0862)	(0.0838)	(0.0865)
Single mother X 2017	-0.198**	-0.158	-0.0953	-0.0167
	(0.100)	(0.0999)	(0.0920)	(0.0966)
Single mother X 2018	-0.203*	-0.127	-0.0510	0.0519
	(0.108)	(0.107)	(0.0960)	(0.0981)
Single mother X 2019	-0.273*	-0.273*	-0.196	-0.104
	(0.145)	(0.145)	(0.122)	(0.119)
Observations	8167	8683	8629	9104

Table M2 *Testing the equality of coefficients of the pre-2017 reform years (2014, 2015, 2016)*

	GHQ-12		SF-12	
Comparison group	F-test	Result	F-test statistic	Result
Partnered mothers	F(2,2203)	1.06	F(2,2495)	0.03
	Prob > F	0.3451	Prob > F	0.9661
Single childless women	F(2,2552)	0.10	F(2,2823)	0.00
	Prob > F	0.9084	Prob > F	0.9982

More Parallel Trends Results

	Partnered mothers		Single childless	
	GHQ-12[+]	SF-12	GHQ-12	SF-12
Ref. Single mother X 2016				
Single mother X 2014	0.131	0.0578	0.0745	0.00313
	(0.0917)	(0.0882)	(0.0810)	(0.0826)
Single mother X 2015	0.0953	0.0483	0.0693	0.0237
	(0.0758)	(0.0701)	(0.0666)	(0.0644)
Single mother X 2017	0.0133	-0.0441	-0.133	-0.141[*]
	(0.0996)	(0.0900)	(0.0875)	(0.0857)
Single mother X 2018	0.0584	0.0942	-0.0488	0.0244
	(0.116)	(0.103)	(0.102)	(0.100)
Single mother X 2019	-0.0720	-0.113	-0.203	-0.139
	(0.147)	(0.125)	(0.138)	(0.116)
Observations	2842	2830	2811	2815

Additional sensitivity analysis (alternative reform dates)

Table O1 *Difference in difference results, comparison: partnered mothers*

	GHQ-12	SF-12
Single mother X After March 2012 and before 2016	-0.0479	0.0282
	(0.317)	(0.520)
Single mother X After March 2016	-1.119**	-0.902
	(0.479)	(0.741)
Observations	8167	8629

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table O2 *Difference in difference results, comparison: single childless women*

	GHQ-12	SF-12
Single mother X After March 2012 and before 2016	0.405	1.000*
	(0.333)	(0.555)
Single mother X After March 2016	-0.576	0.452
	(0.512)	(0.845)
Observations	8683	9104

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.0$

So what does this all mean?

- Some evidence to suggest that mental health inequalities of single mothers compared to partnered mothers increased.
- However, interpret with caution as some of the underlying modelling assumptions are violated

Some potential policy recommendations

- Additional benefits for single mothers for childcare (targeted payments)
- Additional funding to schools to extend after school childcare provisions

Example 2: https://doi.org/10.1016/j.socscimed.2022.115126

Lancaster University

- Evaluate the impact of planning guidance by Gateshead local authority restricting all new takeaways in the borough
- All data used in the analysis is publicly available and can be freely downloaded from https://data.food.gov.uk/catalog/datasets/
- Data pre-intervention was from June 2012-May 2015 and data post-intervention is from June 2015-December 2019.
- All analysis is undertaken at the lower super output area (LSOA) level

The Sample

Intervention group: All LSOAs in Gateshead

Treatment Group: The control group included LSOAs in the local authorities of Stockton on Tees, Durham, Northumberland, Darlington, and Hartlepool.

The adoption of planning guidance is a complex decision that is based on local area characteristics (Lake et al. 2017; Keeble et al. 2019a; Keeble et al. 2021). To identify a suitable control group, we use propensity score matching method using a logit model (Rosenbaum and Rubin 1983). Each LSOA is assigned only one propensity score. We employ single nearest neighbour matching (i.e., 1:1 matching) within a calliper of 0.01 without replacement. We drop a LSOA if the match is outside the calliper, which removed estimation bias from unmatched covariates.

Characteristics of the control and treatment group

Lancaster University

Table 2. Descriptive statistics on characteristics of treatment and control LSOAs (matched and unmatched).

	Treatment: All		Control: All		Treatment: Matched		Control: Matched		t-tests	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Diff	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No. of fast-food outlets	1.85	2.88	1.60	2.74	1.90	3.01	2.11	4.30	0.22	0.22
Density of fast-food	114.9	179.8	100.8	173.2	119.5	189.1	118.8	206.9	−0.67	0.94
Area Size (km sq)	1.13	2.37	10.12	42.12	1.22	2.53	1.13	1.64	−0.08	0.77
Deprivation Deciles:										
Income	4.29	2.62	4.82	2.97	4.44	2.67	4.61	2.86	0.17	0.66
Employment	3.81	2.55	4.10	2.80	3.87	2.59	4.03	2.72	0.16	0.67
Education	4.73	2.80	5.19	2.96	4.86	2.79	5.06	2.97	0.20	0.61
Health	2.91	1.85	3.63	2.36	3.05	1.89	3.22	1.96	0.17	0.51
Crime	6.61	2.43	6.46	2.74	6.61	2.40	6.61	2.71	0.01	0.98
Housing Services	6.37	2.29	6.52	2.80	6.52	2.33	6.44	2.60	−0.08	0.81
Living Environment	8.48	2.00	8.56	2.26	8.68	1.76	8.56	2.17	−0.12	0.66
IDACI	4.69	2.68	4.99	3.01	4.72	2.69	4.86	2.94	0.14	0.72
No. of LSOAs	126		764		109		109			

Key variables

- **Outcome variables:** Density and proportion of fast food outlets
- **Explanatory Variables:** Population Density

The model

$$FastFood_{it} = \beta\, Treated_i \times Post_t + \delta X_{it} + \alpha_i + \tau_t + \varepsilon_{it}$$

Common Trends Test

Lancaster University

- Follow Bertrand and Mullainathan (2003) and examine the dynamic effect of the intervention on the outcomes

$$Y_{it} = \beta_1\, Treated_i \times Pre_t^{-3} + \beta_2\, Treated_i \times Pre_t^{-2} + \beta_3\, Treated_i \times Pre_t^{-1} + \beta_4\, Treated_i \times Post_t^{+1} + \beta_5\, Treated_i \times Post_t^{+2} + \beta_6\, Treated_i \times Post_t^{+3} + \beta_7\, Treated_i \times Post_t^{+4} + \delta X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (2)$$

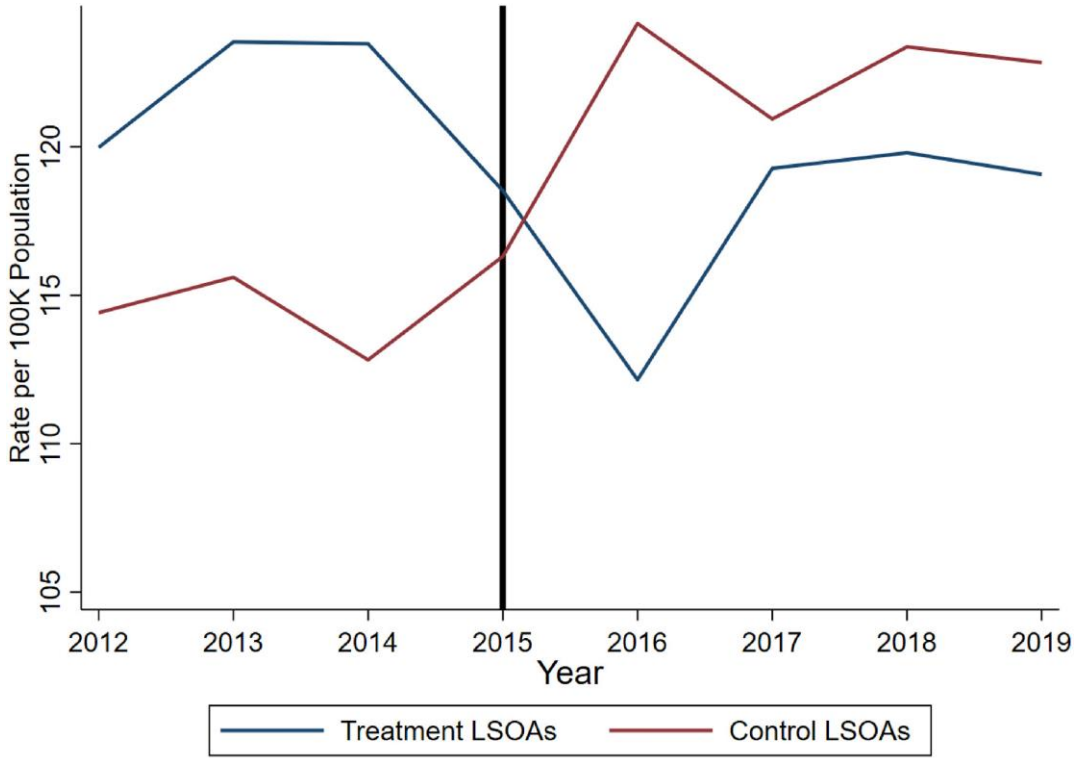
Where,
Pre_t^{-3} is an indicator that is equal to 1 if the year is in 2012, and 0 otherwise;
Pre_t^{-2} is an indicator that is equal to 1 if the year is in 2013, and 0 otherwise;
Pre_t^{-1} is an indicator that is equal to 1 if the year is in 2014, and 0 otherwise;
$Post_t^{+1}$ is an indicator that is equal to 1 if the year is in 2016, and 0 otherwise;
$Post_t^{+2}$ is an indicator that is equal to 1 if the year is in 2017, and 0 otherwise;
$Post_t^{+3}$ is an indicator that is equal to 1 if the year is in 2018, and 0 otherwise;
$Post_t^{+4}$ is an indicator that is equal to 1 if the year is in 2019, and 0 otherwise.

The treated i*Pret0 (i.e., 2015) is the base case. This test examines whether the treatment effects in the given year is significantly different from the 2015. Our main interests in this test are b1, b2, and b3.

Trends over time

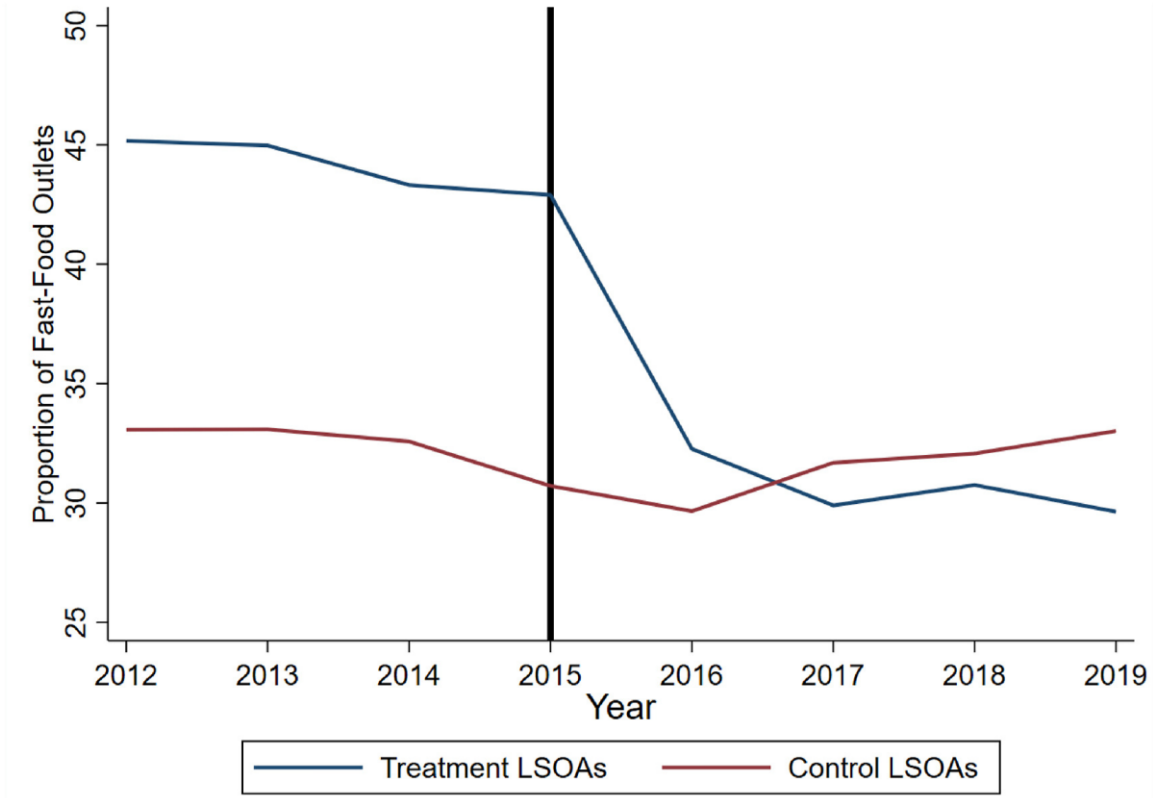
Lancaster University

Density of outlets



Note: 2015 is the year Gateshead adopted the planning guidance.

Proportion of outlets



Note: 2015 is the year Gateshead adopted the planning guidance.

Main Results

Table 4. Estimation results from the PSM-DID model.

Fast-food outlets	Density		Proportion*100%		Number	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated * Post	−11.84**	−12.45***	−13.95***	−13.88***	−0.19*	−0.17
	(4.75)	(4.72)	(2.34)	(2.34)	(0.11)	(0.11)
Population Density		−0.02**		−0.00		0.00
		(0.01)		(0.00)		(0.00)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
LSOA Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1744	1744	1406	1406	1744	1744
R-squared	0.01	0.03	0.17	0.17	0.01	0.03

Note: The proportion has been multiplied by 100. Treated * Post is an indicator that equals 1 if the LSOA is treated and the year is after 2015. Standard errors in parentheses are clustered at the LSOA level. *p < 0.10, **p < 0.05, ***p < 0.01.

Parallel Trends Assumption

Lancaster University

Table A2. Common trend tests

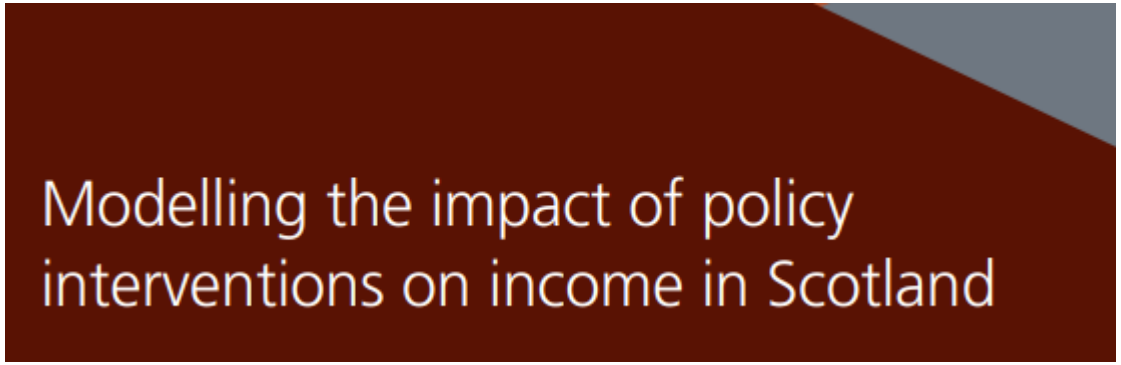
Fast-food outlets	**Density**
	(1)
Treated * Pre^{-3} (2012)	3.824
	(6.105)
Treated * Pre^{-2} (2013)	5.669
	(7.056)
Treated * Pre^{-1} (2014)	8.472
	(6.775)
Treated * Post^{+1} (2016)	-14.291***
	(4.937)
Treated * Post^{+2} (2017)	-4.193
	(6.890)
Treated * Post^{+3} (2018)	-6.491
	(6.671)
Treated * Post^{+4} (2019)	-6.850
	(7.131)
Population Density	-0.028**
	(0.011)
Year Dummies	Yes
LSOA Dummies	Yes
N	1,744
R-squared	0.030

Note: This table presents OLS estimates of the density and proportion of fast -food outlets from 2012 to 2019. The treated is an indicator that is equal to 1 if the LSOA is within Gateshead, and 0 otherwise. Pre^{-3} is an indicator variable set to 1 for the year 2012, and 0 otherwise. Similarly, all the Pres and Posts are indicators for the year in the brackets. Standard errors in parentheses are clustered at the LSOA level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results summary/policy recommendations

- Planning policy that completely restricts new takeaways from opening can significantly change the food environment in the short term (<5 years from introduction.
- Well thought out planning policy involving a multidisciplinary team is a low cost way for local authorities to promote a healthy environment. This can potentially feed into the Levelling Up Agenda.

Another example which may be useful



http://www.healthscotland.scot/media/1698/modelling-the-impact-of-policy-interventions-on-income-in-scotland.pdf

Other extensions to consider