Promoting Normal Birth and Reducing Caesarean Section Rates: An Evaluation of the Rapid Improvement Programme

Prof. Graham Cookson
Dr. Ioannis Laliotis

Faculty of Arts & Social Sciences, University of Surrey
"Delivering Better for Less: Improving Productivity in the Public Services"
The Leverhulme Trust

MSc in Economics & Data Analysis
Seminar Series
University of Patras

Structure

Introduction

- Caesarean section rates have been increasing in England (≈10% in the '80s, ≈25% today)
- Huge variation across hospitals: 9.5%-30% (2000), 18%-35% today
- Extremely expensive procedure
  - Maternities absorb ≈3% of total NHS expenditures
  - Deliveries absorb ≈30% of maternity expenditures
  - C-sections are 3 times more expensive than vaginal deliveries
- Not without maternal/neonatal risks although safer and more convenient due to technological advances

Introduction

Observed caesarean section variation due to:

- Maternal requests (socio-economic status) for convenience and lifestyle reasons
- Physical/psychological benefirs $>$ Adverse outcome risks
- Rising maternal age
- Medical/technological advances (high risk classification)
- Doctor-induced demand: fee differentials and/or more convenient planning
- Defensive medicine (lawsuit threats)
- Maternity workforce skills

Introduction

- Health policy makers concerned about increased C-section rates
- Cuts to public health funding
- Interventions for lower rates (WHO upper threshold estimate ≈10%-15%)
- "Focus on Normal Birth and Reducing Caesarean Section Rates Rapid Improvement Programme" (*circa* mid-2008)
  - 20 hospitals were selected from a wider pool of applicants
  - Organisational and clinical pathways identified from an interdisciplinary expert team
  - Provided a Toolkit to achieve and maintain low rates and promote normal birth

Data sources

- Hospital Episode Statistics, maternity tail
- All NHS admissions on a pseudo-anonymous patient-level administrative dataset
- Individual (mother) level variables (age, ethnic group, socio-economic status, clinical characteristics etc.)
- Hospital (trust) level variables (Strategic Health Authority, teaching hospital, NHS Foundation Trusts)
- Baseline (2008m6-2008m12) and follow-up periods (2009m1-2009m6 & 2009m7-2010m1) for participating and non-participating trusts (treatment and control groups)
- Individual data collapsed to the trust level

Descriptive statistics

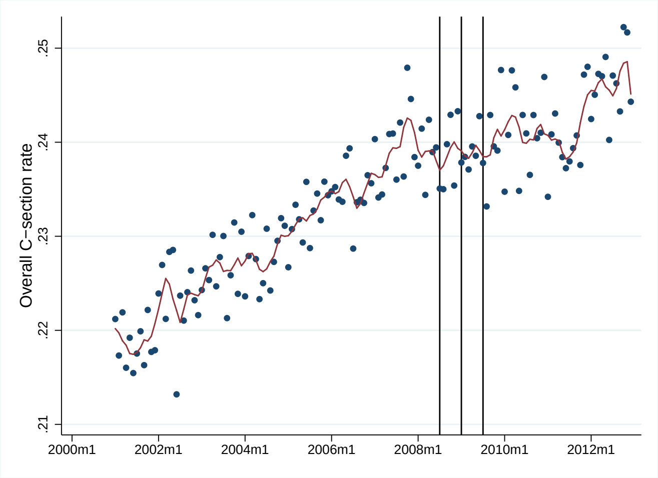
Table 1: Descriptive statistics for main pre-treatment characteristics

Variable name	Total sample Mean	Total sample Std.Dev	Control group Mean	Control group Std.Dev	Treatment group Mean	Treatment group Std.Dev	*t*-test *p*-value	Kolmogorov-Smirnov *p*-value
Caesarean section rate	.240	.077	.232	.045	.292	.169	.001	.033
Planned caesarean section rate	.099	.078	.092	.022	.147	.202	.003	.057
Emergency caesarean section rate	.141	.032	.141	.032	.146	.077	.483	.282
High risk	.448	.091	.443	.082	.480	.134	.097	.100
Healthy mother	.467	.071	.465	.057	.481	.129	.351	.854
Parity	.833	.317	.826	.466	.873	.104	.568	.807
Nulliparous	.481	.114	.484	.103	.456	.171	.308	.887
Urban	.826	.177	.834	.168	.779	.229	.203	.081
Foundation trust	.527	.501	.515	.502	.600	.503	.485	.999
Teaching hospital	.203	.403	.195	.397	.250	.444	.574	.999
Discharged home in 28 days	.417	.077	.419	.070	.401	.114	.328	.896
Readmitted within 28 days	.044	.023	.044	.021	.048	.023	.446	.268
Age: 20-24 years old	.197	.063	.200	.064	.179	.057	.087	.345
Age: 25-29 years old	.270	.038	.273	.032	.252	.064	.019	.073
Age: 30-34 years old	.274	.074	.268	.045	.313	.165	.013	.282
Age: 35-39 years old	.158	.042	.157	.040	.161	.049	.758	.504
Age: >40 years old	.036	.013	.036	.013	.037	.014	.929	.290
Mixed ethnic origin	.014	.018	.014	.019	.012	.008	.551	.261
Asian ethnic origin	.095	.108	.099	.112	.075	.073	.357	.721
Afro-Caribbean ethnic origin	.049	.077	.053	.082	.028	.034	.181	.733
Other/Unknown ethnic origin	.086	.085	.085	.088	.093	.059	.706	.241
IMD II	.220	.091	.224	.091	.196	.088	.208	.228
IMD III	.188	.076	.186	.073	.197	.092	.559	.484
IMD IV	.171	.095	.168	.096	.184	.092	.496	.743
IMD V	.162	.158	.150	.145	.234	.216	.027	.021
Hospital load	12.763	5.427	12.786	5.381	12.618	5.856	.898	.887
Observations	148		128		20		148	148

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).

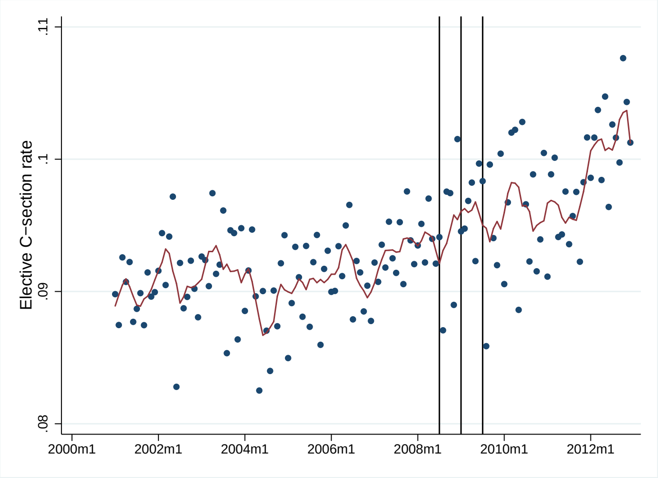
Descriptives

Figure 1: Overall Caesarean section rate, monthly time series



Descriptives

Figure 2: Planned (elective) Caesarean section rate, monthly time series



Descriptives

Figure 3: Emergency (non-elective) Caesarean section rate, monthly time series

Econometric methodology

Programme evaluation with observational data

- Estimation of the average effect of a programme on outcome(s)
- Comparison of outcomes between treated and control units
- Standard regression-based approach:

$$y = \alpha + \beta X + \gamma D + \epsilon$$
$$y = \alpha + \beta X + \gamma D + \delta T + \phi(D \times T) + \epsilon$$

- Usual exogeneity assumption: $\epsilon \perp\!\!\!\perp D, X$
- Constant treatment effects across units
- Programme participation not random
- Treated $\neq$ controls in many (un)observable aspects

Econometric methodology

Propensity score matching

- Matching treated & control units is “easy” in a low-dimensional world
- Given reality, matching problem can be reduced to a single dimension: the propensity score

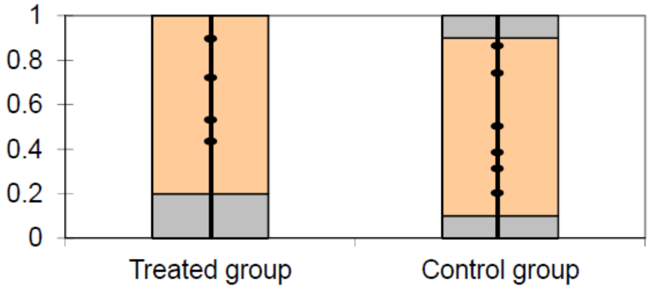
$$p(x) = prob(D = 1|X) = E(D|X)$$

- Treatment status as a function of (pre-treatment) variables:
  - Relevant theory
  - Previous research
  - Administrative selection criteria / programme eligibility
  - Same data source
- Estimated probability of programme participation (logit/probit)

Econometric methodology

Matching methods

- Match each i-th treated to a j-th control with similar characteristics (in propensity score terms):
    - Without replacement: each j is matched to an i just once
    - With replacement: each j is matched to several i's
- $\hat{p}(X)$ for treated and control units:

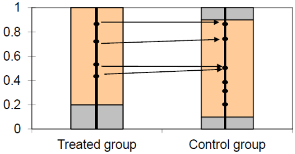


Matching methods

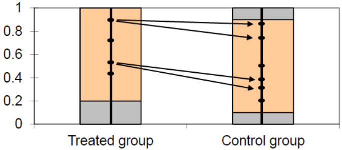
Nearest 1 : m neighbour(s) matching

- For each i select a j with the closest $\hat{p}(X)$

$$min \parallel p_i - p_j \parallel$$



$m = 1$



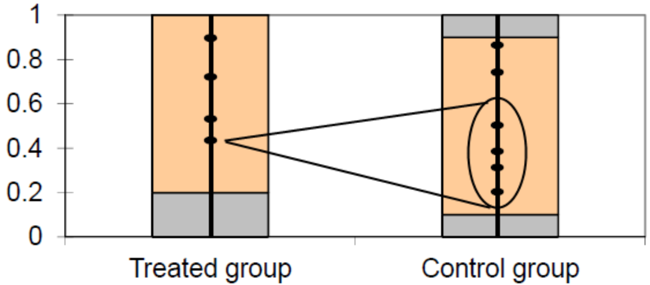
$m > 1$

Matching methods

Radius (caliper) matching

- Each i is matched to a number of j's falling within a specified radius (or caliper)

$$\| p_i - p_j \| < r$$

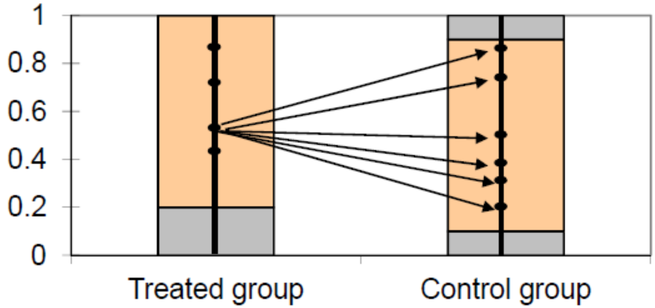


Matching methods

Kernel & Local Linear Regression matching

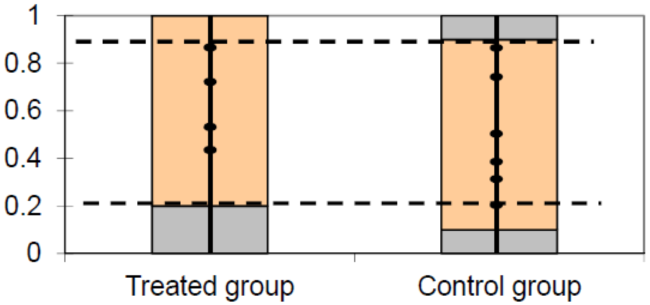
- Each i is matched to all j's with their weights being inversely proportional to their $\hat{p}(X)$ distance

$$g(p_i, p_j) = \frac{K[\frac{(p_i - p_j)}{h}]}{\sum_{k \in C} K[\frac{(p_i - p_k)}{h}]}$$



Matching methods

- The choice of a matching algorithm involves a trade-off between variance and bias
- Important to restrict matching in the common propensity score range of treated and control units



Econometric methodology

Programme evaluation with propensity score matching

- After matching, the Average Treatment Effect (ATE) can be calculated: $\Delta = y_1 - y_0$

$$ATE = E(\Delta) = E(y_1|X, D = 1) - E(y_0|X, D = 0)$$

- ATE is fine for randomised experiments
- Often biased in observational studies

Econometric methodology

Programme evaluation with propensity score matching

- More often, the Average Treatment Effect on the Treated (ATT) is estimated:

$$ATT = E(\Delta|D=1) = E(y_1|X, D=1) - \underbrace{E(y_0|X, D=1)}_{counterfactual}$$

- The counterfactual is proxied by $(E(y-0|X, D=0))$ and matching is performed using $p(X)$

$$\begin{aligned} ATT &= E(\Delta|p(X), D=1) \\ &= E(y_1|p(X), D=1) - E(y_0|p(X), D=0) \\ &= \frac{1}{n_1} \sum_{\substack{i=1 \\ D_i=1}}^{n_1} [y_{1i}(X_i) - \hat{E}(y_{0i}|p(X_i), D_i=0)] \\ \Delta &= ATT + SB \end{aligned}$$

- where $SB = E(y_0|D=1) - E(y_0|D=0)$

Programme evaluation with propensity score matching

The Difference-in-differences case

- Data availability on outcomes before (t) and after (t') programme participation
- Time-invariant unobserved heterogeneity
- "Selection on observables" assumption is relaxed

$$
\begin{aligned}
ATT &= E(\Delta_t - \Delta_{t'} | D = 1) \\
&= E[(y_{1t} - y_{0t}) - (y_{1t'} - y_{0t'}) | X, D = 1] \\
&= E(y_{1t} - y_{1t'} | X, D = 1) - E(y_{0t} - y_{0t'} | X, D = 1) \\
&= \frac{1}{n_{1t}} \sum_{\substack{i=1 \\ D_i=1}}^{n_{1t}} [y_{1ti}(X_i) - \hat{E}(y_{0ti} | p(X_i), D_i = 0)] \\
&- \frac{1}{n_{1t'}} \sum_{\substack{j=1 \\ D_j=1}}^{n_{1t}} [y_{0t'j}(X_j) - \hat{E}(y_{0t'j} | p(X_j), D_j = 0)]
\end{aligned}
$$

Programme evaluation with propensity score matching

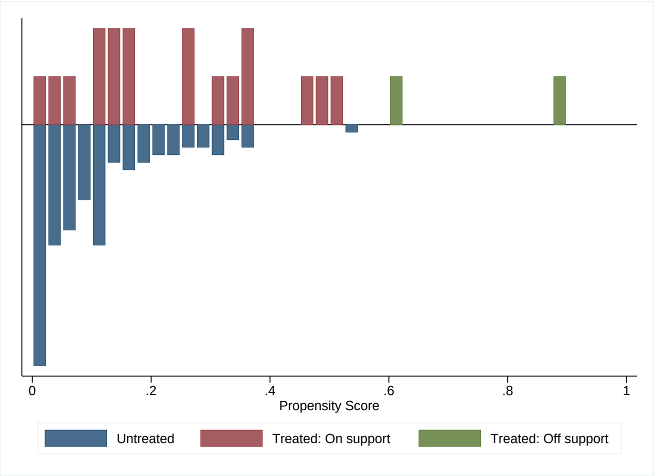
Required assumptions for the propensity score matching validity

- No general equilibrium effects (treatment only affects the treated)
- Conditional Independence Assumption (CIA): $(y_0, y_1) \perp D|X$
- Control group outcome and programme participation are conditionally independent (weaker than CIA): $y_0 \perp D|X$
- Overlapping assumption (common support): $0 < p(D = 1|X) < 1$
- Balancing condition $D \perp X|p(X)$

Results

Check of common support region

Figure 4: Histograms of propensity scores for treated and control units



Results

Check of balancing properties

- Paired t-test between treated and matched comparison units
- Calculate standardised difference (bias) for each X variable:

$$Bias(X) = \frac{100 \frac{1}{N} \sum_{i \in T} [X_i - \sum_{j \in C} g(p_i, p_j) X_j]}{\sqrt{\frac{Var_{i \in T}(X) + Var_{j \in C}(X)}{2}}}$$

- Regression-based F-test:

$$X = \alpha_0 + \sum_{k=1}^{m} \alpha_k \hat{p}(X)^k + \sum_{k=1}^{m} \beta_k D \hat{p}(X)^k + \varepsilon$$

$$H_0 : \beta_1 = 0; \beta_2 = 0, \cdots \beta_k = 0$$

Results

Check of balancing properties

Table 2: Balancing tests based on Gaussian kernel matching and regression functions

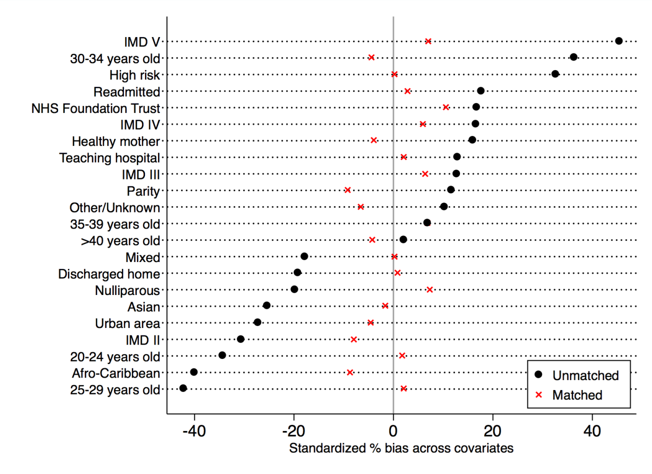
Variable	Mean Treated	Mean Control	% Bias	% Bias reduction	*t*-test *t*-stat	Regression-based *F*-stat
High risk	0.4521	0.4519	0.3	99.1	0.01 (0.989)	0.15 (0.965)
Healthy mother	0.4569	0.4608	-3.9	75.4	-0.22 (0.826)	0.21 (0.933)
Parity	0.7442	0.7805	-.9.1	22.4	-0.39 (0.700)	1.51 (0.203)
Nulliparous	0.5060	0.4955	7.4	62.7	0.35 (0.729)	1.92 (0.111)
Urban	0.8164	0.8254	-4.5	83.5	-0.19 (0.853)	0.24 (0.914)
Foundation trust	0.6111	0.5583	10.5	37.4	0.31 (0.756)	1.38 (0.243)
Teaching hospital	0.2222	0.2134	2.1	83.8	0.06 (0.950)	0.20 (0.937)
Discharged home in 28 days	0.4197	0.4189	0.9	95.1	0.04 (0.970)	0.19 (0.944)
Readmitted within 28 days	0.0504	0.0498	2.9	83.6	0.08 (0.937)	0.40 (0.810)
Age: 20-24 years old	0.1873	0.1863	1.8	94.9	0.06 (0.954)	0.50 (0.738)
Age: 25-29 years old	0.2640	0.2630	2.1	95.0	0.10 (0.920)	0.62 (0.646)
Age: 30-34 years old	0.2768	0.2821	-4.3	88.1	-0.36 (0.722)	0.72 (0.577)
Age: 35-39 years old	0.1701	0.1689	6.9	-1.2	0.24 (0.813)	0.58 (0.681)
Age: >40 years old	0.0388	0.0394	-4.1	-98.4	-0.13 (0.900)	1.26 (0.289)
Mixed ethnic origin	0.0128	0.0127	0.3	98.3	0.01 (0.992)	0.06 (0.994)
Asian ethnic origin	0.0786	0.0801	-1.6	93.6	-0.05 (0.957)	0.66 (0.618)
Afro-Caribbean ethnic origin	0.0294	0.0348	-8.6	78.4	-0.33 (0.740)	1.69 (0.156)
Other/Unknown ethnic origin	0.0985	0.1034	-6.5	36.5	-0.19 (0.854)	0.43 (0.789)
IMD II	0.2060	0.2130	-7.8	74.4	-0.26 (0.800)	0.14 (0.969)
IMD III	0.2112	0.2059	6.4	49.8	0.20 (0.840)	0.55 (0.698)
IMD IV	0.1975	0.1920	5.9	64.3	0.20 (0.845)	0.35 (0.843)
IMD V	0.1986	0.1857	7.0	84.6	0.28 (0.781)	0.16 (0.960)
Hospital load	12.974	12.508	8.3	-179.2	0.27 (0.788)	0.52 (0.724)

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).
Notes: *p*-values in parentheses. For the Gaussian kernel matching the bandwidth parameter is set to 0.06.

Results

Check of balancing properties

Figure 5: Standardised bias in matched and unmatched samples



Results

Table 3: Difference-in-differences estimates

	Follow-up period 1			Follow-up period 2		
	Total	Planned	Emergency	Total	Planned	Emergency
Panel A: D-i-D without covariates						
Baseline difference ($\hat{\gamma}$)	.060	.055	.005	.060	.055	.005
	(.018)[a]	(.019)[a]	(.008)	(.018)[a]	(.019)[a]	(.008)
Follow-up difference ($\hat{\gamma} + \hat{\phi}$)	.012	.008	.004	.021	.003	.018
	(.018)	(.013)	(.014)	(.018)	(.019)	(.008)[b]
Difference-in-differences ($\hat{\phi}$)	-.048	-.047	-.002	-.039	-.052	.013
	(.026)[c]	(.019)[b]	(.020)	(.026)	(.026)[b]	(.011)
Panel B: D-i-D with covariates						
Baseline difference ($\hat{\gamma}$)	.024	.014	.010	.026	.012	.014
	(.009)[b]	(.006)[b]	(.008)	(.008)[a]	(.006)[b]	(.006)[b]
Follow-up difference ($\hat{\gamma} + \hat{\phi}$)	.011	-.003	.014	.026	.008	.018
	(.009)	(.006)	(.008)[c]	(.008)[a]	(.006)	(.006)[a]
Difference-in-differences ($\hat{\phi}$)	-.013	-.017	.004	.000	-.004	.004
	(.013)	(.008)[b]	(.011)	(.011)	(.008)	(.008)

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).

Results

Table 4: The impact of programme participation on the overall caesarean section rate: Difference-in-differences propensity score matching estimates

	Follow-up period 1		Follow-up period 2	
Matching algorithm	Estimate	Standard error	Estimate	Standard error
Unmatched sample	-.04195	.01554[a]	-.03536	.01511[b]
Nearest *1:1* neighbour w/o replacement	-.00979	.00641	-.00419	.00596
Nearest *1:1* neighbour	-.01623	.00688[a]	-.00352	.00614
Nearest *1:m* neighbour	-.01072	.00552[c]	.00243	.00604
Radius	-.01113	.00632[c]	.00193	.00554
Gaussian kernel	-.01085	.00600[c]	.00139	.00548
Epanechnikov kernel	-.01098	.00634[c]	.00184	.00555
Local linear regression	-.01276	.00654[b]	.00218	.00557

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).
Notes: In the case of the *1:m* nearest neighbour matching, *m* is equal to 5. The bandwidth for the Gaussian kernel and the Local Linear Regression matching is set to 0.06. The caliper in the case of Radius matching is set to 0.06. [a], [b] and [c] denote statistical significance at the 1%, 5% and 10% levels, respectively. Estimates were performed using the user-written Stata command `psmatch2` (Leuven and Sianesi, 2003).

Results

Table 5: The impact of programme participation on the planned (elective) caesarean section rate: Difference-in-differences propensity score matching estimates

	Follow-up period 1		Follow-up period 2	
Matching algorithm	Estimate	Standard error	Estimate	Standard error
Unmatched sample	-.04711	.01790[a]	-.04546	.01789[a]
Nearest *1:1* neighbour w/o replacement	-.00946	.00399[a]	-.00332	.00352
Nearest *1:1* neighbour	-.01252	.00458[a]	-.00469	.00352
Nearest *1:m* neighbour	-.00740	.00336[b]	-.00290	.00328
Radius	-.00530	.00351	-.00269	.00313
Gaussian kernel	-.00580	.00329[c]	-.00263	.00310
Epanechnikov kernel	-.00554	.00351	-.00279	.00352
Local linear regression	-.00683	.00353[b]	-.00277	.00315

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).
Notes: In the case of the *1:m* nearest neighbour matching, *m* is equal to 5. The bandwidth for the Gaussian kernel and the Local Linear Regression matching is set to 0.06. The caliper in the case of Radius matching is set to 0.06. [a], [b] and [c] denote statistical significance at the 1%, 5% and 10% levels, respectively. Estimates were performed using the user-written Stata command `psmatch2` (Leuven and Sianesi, 2003).

Results

Table 6: The impact of programme participation on the emergency (non-elective) caesarean section rate: Difference-in-differences propensity score matching estimates

	Follow-up period 1		Follow-up period 2	
Matching algorithm	Estimate	Standard error	Estimate	Standard error
Unmatched sample	.00505	.00589	.00999	.00529[c]
Nearest *1:1* neighbour w/o replacement	-.00045	.00554	-.00094	.00617
Nearest *1:1* neighbour	-.00390	.00610	.00113	.00625
Nearest *1:m* neighbour	-.00356	.00459	.00529	.00595
Radius	-.00598	.00507	.00459	.00562
Gaussian kernel	-.00522	.00478	.00399	.00557
Epanechnikov kernel	-.00561	.00508	.00463	.00563
Local linear regression	-.00612	.00516	.00492	.00564

Source: Hospital Episode Statistics (HES) database (maternity tail). Health & Social Care Information Centre (HSCIC).
Notes: In the case of the *1:m* nearest neighbour matching, *m* is equal to 5. The bandwidth for the Gaussian kernel and the Local Linear Regression matching is set to 0.06. The caliper in the case of Radius matching is set to 0.06. [a], [b] and [c] denote statistical significance at the 1%, 5% and 10% levels, respectively. Estimates were performed using the user-written Stata command `psmatch2` (Leuven and Sianesi, 2003).

Conclusions

- Need of initiatives/policies to reduce caesarean section rates
- Expensive and often not medically justified
- Rapid Improvement Programme had had a small impact on that direction
- Still work to be done:
  - Alternative specifications of programme participation equation
  - Alternative baseline and follow-up periods
  - Control for programme anticipation
  - More on programme's specifics (awaiting)
  - Fine tuning (bandwidth, caliper, bootstrapping etc)
  - Quantile-specific ATT effects
  - Kernel ridge regression propensity score matching
  - Combine individual-level analysis

This work is part of the *"Delivering Better for Less: Improving Productivity in the Public Services"* research program funded by The Leverhulme Trust and based at the University of Surrey

The Leverhulme Trust

Better for Less

Improving Productivity in the Public Services

www.leverhulme.ac.uk

www.deliveringbetter.com

We are offering 2 PhD studenships to extend our research:

Delivering Better for Less in Education

Delivering Better for Less in the Local Government

Spread the word!

Thank you!