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

Graham Cookson*[a] and Ioannis Laliotis[a]

[a]University of Surrey

January 28, 2016

Abstract

This paper seeks to evaluate the impact of the 2008 Rapid Improvement Programme that aimed in promoting normal birth and reducing caesarean section rates in the English NHS. Using Hospital Episode Statistics data and combining difference-in-differences with propensity score matching estimators, the average treatment effect on the treated is found to be small, significant and short-lived. The uncovered reduction concerned the overall and the planned trust-level caesarean section rates and it occured only shortly after the programme implementation. No effect was found in the case of the emergency caesarean section rate.

Keywords: Caesarean sections; English NHS; Difference-in-differences; Propensity score matching

JEL Classification: I10; I11; C21

*Address for correspondence: Faculty of Arts and Social Sciences, University of Surrey, UK. E-mail: g.cookson@surrey.ac.uk. The authors acknowledge financial support from The Leverhulme Trust (RL-2012-681). The views expressed here are those of the authors and do not necessarily reflect those of the funding organisation. The usual disclaimer applies.

1 Introduction

Ceasarean section deliveries have been rising across English NHS hospitals (Declercq *et al.*, 2011). According to Hospital Episode Statistics data, the overall caesarean section rate increased from 21.8% in 2000 to over 25% in 2013, and was as low as 9% and 12% in 1980 and 1990, respectively (Bragg *et al.*, 2010; Francome and Savage, 1993; Lancet, 2000). This increase and the often unjustified routine use on very healthy mothers has concerned policy makers and medical professionals across the world. Moreover, there is evidence of considerable variation in caesarean rates within the UK. In England, the variation between providers ranged from 9.53% to 30.01% in 2000 and from 17.67% to 34.31% in 2013 (Hospital Episode Statistics). To date, there is no consensus on the ideal caesarean section rate and the observed variations between countries, regions and providers may indicate clinical uncertainty and practice style heterogeneity (Baicker *et al.*, 2006; Francome and Savage, 1993; Paranjothy *et al.*, 2005). The World Health Organization recommended rate for caesarean sections in 1985 was 10-15% (WHO, 1985) with this upper threshold theoretical estimate being supported by later studies (Althabe *et al.*, 2006; Villar *et al.*, 2006). In their 2009 Handbook, WHO acknowledged the existence of a growing body of research showing the negative impact of a high caesarean section rate; that both very high and very low rates were dangerous but that the optimum rate was unknown. They identified a lack of empirical evidence for an optimum percentage or range of percentages for caesarean sections (WHO, 2009).

Despite the lack of a consensus, there are concerns about whether the high incidence of caesarean sections is justified given that it is not a riskless procedure (Shorten, 2007). It is also extremely expensive. Over £3 billion of the £97 billion gross expenditure was spent on maternity services in 2010, of which over £1 billion was spent on deliveries. Caesarean delivery is reimbursed at approximately 3 times the rate of normal, vaginal deliveries. Besides the perverse financial incentive to perform more caesareans, planned (elective) procedures offer predictability and convenience, shorter procedure timing, advanced staff planning, weekday working hours for staff (scheduling births by time of day, day of week and non-holidays, which is also cheaper when outsourcing staff), quick turnover of delivery rooms and higher fees (Sakala and Corry, 2008).

The reasons for practising such a procedure have been under scrutiny, with numerous studies analysing their risks and benefits (Menacker and Hamilton, 2010). During caesarean deliveries major abdominal surgeries take place and some serious risks are involved (Bragg *et al.*, 2010; Menacker and Hamilton, 2010; Shorten, 2007). Women may experience surgical complications, they are more likely to be rehospitalized and they face increased probabilities for complications in subsequent pregnancies (Bragg *et al.*, 2010;

Deneux-Tharaux *et al.*, 2006; Landon *et al.*, 2004; Lavender *et al.*, 2012; Shearer, 1993; Yang *et al.*, 2007; Villar *et al.*, 2006). Although in a less frequent basis, serious neonatal complications requiring intensive care may occur (DiMatteo *et al.*, 1996; Lavender *et al.*, 2012; Shorten, 2007). At the same time, hospital costs for caesarean section deliveries are significantly higher as compared to those of a normal (vaginal) delivery (Menacker and Hamilton, 2010; Shearer, 1993). On the other hand, some of the benefits linked with planned caesarean deliveries include greater safety for the mother and the baby due to technological advances in the procedure, avoidance of labour pain and convenience (Bragg *et al.*, 2010; Lavender *et al.*, 2012; Shearer, 1993).

Several factors seem to explain the observed variation in caesarean section deliveries. Perhaps the most popular one is the increased maternal requests in cases where medical or obstetrical indications are small or absent, mostly for lifestyle reasons (Alves and Sheikh, 2005).[1] In these cases, women seek to plan a caesarean delivery because the physical or psychological benefits outweigh the risk of an adverse outcome (Fenwick *et al.*, 2010; Lavender *et al.*, 2012). The ibserved upward trends have also been atttributed to the rising maternal age, improvements in medical and technological equipment which have made the procedure safer and the growing portion of women who had previous caesarean sections in the past (Bragg *et al.*, 2010; Lancet, 2000). Some authors also mention that malpractice claims risks faced by hospitals and physicians may also lead to defensive medicine since the threat of lawsuits wary clinicians about the childbirth risks (Dubay *et al.*, 1999; Localio *et al.*, 1993; Yang *et al.*, 2009).[2] Grant (2009) and Gruber *et al.* (1999) have examined the role of financial incentives showing that caesarean rates increase with the fee differentials between caesarean and vaginal childbirth. Moreover, based on an induced-demand model, Gruber and Owings (1996) demonstrated that declines in state-level fertility rates have led obstetricians and gynaecologists to substitute vaginal deliveries with more highly reimbursed alternatives. However, according to recent evidence using individual level data for the US, the convenience-driven physician-induced demand is small and the decision takes place in the ward rather than being planned in advance (Lefèvre, 2014). The role of maternity staffing levels has also been examined. Increased, better trained and experienced maternity workforce may contribute in lowering the caesarean section rates, especially the emergency ones (Alves and Sheikh, 2005; Lancet, 2000).[3]

With respect to the effectiveness of interventions aiming to lower the caesarean section rates, several studies have been conducted (Marshall *et al.*, 2015). In England, the 'Focus on Normal Birth and

[1]However, Kalström *et al.* (2011) reported that the rising caesarean section rates seem to be related to factors other than preferences, after analysing a Swedish regional cohort of women.

[2]Dubay *et al.* (1999) also found that the defensive response of physicians varies with the mother's socio-economic status, with the effect being stronger for those women with the lowest socio-economic status.

[3]Roberts and Nippita (2015) argue that the required skill for a justified medical decision between caesarean and vaginal delivery can be greater that the skills actually needed to perform the procedure alone.

Reducing Caesarean Section Rates' was part of the *Spread and Adopt Rapid Improvement Programme* that was implemented in July 2008. It was mainly influenced by previous work performed during the 1990s by a Working Group in Ontario which examined how specific hospitals were able to maintain low caesarean section rates (Baldwin *et al.*, 2010; Marshall *et al.*, 2015). They found that cultural aspects, such as willingness to keep low rates, normal birth culture, teamwork, leadership, quality-improvement activities and the ability to manage change were the driving factors of their success. In the same spirit, the 'Focus in Normal Birth and Reducing Caesarean Section Rates Rapid Improvement Programme' targeted in promoting vaginal deliveries and reducing the caesarean section rates. The programme was implemented in 20 NHS Trusts selected from a wider pool of applicants. Participating trusts were offered a Toolkit containing four self-improvement pathways; one with respect to the characteristics of each organisation and three clinical pathways in order to keep first pregnancy and labour normal, promote vaginal birth after caesarean section and plan caesarean sections (Marshall *et al.*, 2015). Furthermore, various other tools were offered in order to support service improvements identified by hospital teams (NHS Institute for Innovation and Improvement, 2007). Marshall *et al.* (2015) have recently performed a first attempt to evaluate the effectiveness of this programme by using a mixed-methods study, e.g. by collecting data, sending questionnaires and interviewing key individuals from the participating trusts.

In this study, we attempt a more formal evaluation of the Rapid Improvement Programme by using Hospital Episode Statistics data and adopting a non-parametric difference-in-difference propensity score matching estimator. Hence, we seek to identify the causal effect of programme participation at the trust level adopting a flexible estimation technique that does not rely on strong exclusion restrictions (Blundell and Costa Dias (2000). According to the results, there was a small reduction in the overall and the planned caesarean section rates. However, it was detected only in the period shortly after the programme participation.

The rest of the paper is organized as follows: Section 2 presents the data sources and some preliminary descriptive analysis. Section 3 outlines the adopted empirical strategy the results of which are discussed in Section 4. Section 5 concludes.

2 Data

We draw data from the Hospital Episode Statistics (HES hereafter) database. HES is a pseudo-anonymous patient level administrative database containing details of all admissions, outpatient appointments and Accident & Emergency attendances at all NHS trusts in England, including acute hospitals, primary care trusts and mental health trusts.[4] Each HES record contains details of a single consultant episode: a period of patient care overseen by a consultant or other suitably qualified healthcare professional (e.g. a registered midwife). It is more common to work with spells or admissions, which is a continuous period of time spent as a patient within a trust. This may include more than one episode. The anonymous, unique patient identifiers in the HES records help to append or derive relevant information from previous delivery and spells (e.g. parity - the number of live births that a woman has had). This allowed for a more complete picture of a woman's obstetric history to be compiled. Primary care trusts, mental health trusts and private providers were excluded from the dataset. This was done mostly to avoid any confounding errors. For example, primary trusts provide a great deal of community based midwifery care (e.g. antenatal care and home deliveries), which will distort the representation somewhat.

Attached to a mother's delivery episode is the 'maternity tail', i.e. records for up to nine babies. Each baby has its own HES birth record, but this is not linked to the mother's delivery record. Following Marshall *et al.* (2015) we extracted individual delivery records from the HES database for two periods: a baseline period (before the program implementation) from July 2008 to December 2008 and two follow-up periods (after the program implementation). The first follow-up period is from January 2009 to June 2009 and the second follow-up period is from July 2009 to January 2010. For the purposes of the analysis, the sample was restricted to singleton births and all the variables were collapsed to the trust level. Table 1 presents descriptive statistics regarding the outcomes (overall, planned and emergency caesarean rates) and some basic trust-level variables. These include the proportions of women by age category, ethnic group, urban and socio-economic status. The latter is based on the socio-economic quintile of their residence area and it is measured using the 2007 Index of Multiple Deprivation (IMD hereafter) at the super output area (DCLG, 2011).[5] Moreover, there are variables measuring the mean parity, the percentages of nulliparous women, healthy mothers, those who were discharged to their home within 28 days, those who were readmitted within 28 days and those classified as 'high risk' women. [6] Also, there

[4]These were stored in an SQL database on a secure, private network. Full details on data storage, data management and information governance procedures are available upon request. The University of Surrey is compliant with the research and Information Governance frameworks for health and social care in the United Kingdom and is compliant with the University's best practice standards. It adheres to all of the conditions imposed by the NHS and HSCIC under the HES and Electronic Staff Record (ESR) data sharing agreements.

[5]The index is constructed from 38 indicators across seven weighted domains measuring an area's income, deprivation, employment deprivation, health deprivation and disability, education, skills and training, barriers to housing and service, crime and the local environment. The index is produced periodically for the Department of Communities and Local Government by researchers at the University of Oxford. The raw scores are meaningless, and it is the relative deprivation that is relevant. Here we categorize the raw scores into deprivation quintiles.

[6]In this paper we adopted the innovative method developed in Sandall *et al.* (2014) to exploit the rich clinical history available in HES records to identify women with "higher risk" pregnancies because of pre-existing medical conditions, a complicated previous obstetric history or conditions that develop during pregnancy. These women and their babies may have different outcomes from women regarded as at "lower risk". They used the National Institute for Health and Care Excellence (NICE hereafter) intrapartum care guideline (NICE, 2007) and matched the conditions listed in the guideline to relevant four-alphanumeric digit ICD-10 codes. For certain conditions, other types of codes were matched, such as OPCS-4

is a hospital load variable indicating the average daily number of deliveries within each trust, regional dummies indicating the Strategic Health Authority each trust belongs to and binary variables indicating the teaching or the foundation trust status of each hospital.[7]

The descriptive statistics have been calculated for the total sample of trusts as well as by programme participation status. Furthermore, the results from some standard tests examining whether their means and distributions are statistically different are displayed. For most of the variables, the differences in means and distributions are not statistically different. However, there are some notable differences regarding the overall and the planned caesarean section rates, the high risk, age, socio-economic and urban profiles.

[Table 1 about here]

3 Econometric methodology

The objective of this study is to examine whether the participation in the Rapid Improvement Programme had had a causal effect on the trust-level caesarean section rate. In a conventional linear regression framework the problem would be summarized as follows:

$$cs_i = \alpha + \beta T_i + \gamma RIP_i + \delta(T_i \times RIP_i) + \sum_{j=1}^{m} \phi_j X_{ji} + u_i \tag{1}$$

where cs_i is the caesarean section rate for the i-th trust, T is a binary variable indicating the baseline (pre-treatment) and follow-up (post-treatment) periods, RIP indicates programme participation and X is a set of control variables. Hence, $\hat{\gamma}$ is the baseline difference between the treated and non-treated trusts, $\hat{\gamma} + \hat{\delta}$ is their follow-up difference and $\hat{\delta}$ is the difference-in-differences parameter (D-i-D) of main interest. A problem with the estimation of such an impact is the possible endogeneity of the trusts participated in the programme. They are likely not to be random and their characteristics may be systematically different from those of non-participated trusts. Therefore, estimates on outcome variables would be biased if such non-randomness is left unaccounted for. Selection bias issues are often being dealt with by the use of an instrumental variables approach. However, appropriate instruments are hard to be found in this context since variables affecting program participation are also likely to affect caesarean section rates. An alternative method is the propensity score matching combined with the difference-in-difference

or HES Data Dictionary data items, for example to identify breech presentation or multiple pregnancy.

[7] The relationship between the teaching status of the hospital and the caesarean probability has been documented in older studies (e.g. Oleske *et al.*, 1991).

estimator (Blundell and Costa Dias, 2000).

The objective of the matching procedure is to find a group of non-participating trusts that have similar characteristics with the participating ones. Let $RIP \in \{T, C\}$ denote a programme participation indicator equal to T for the participating trusts (treatment group) and equal to C for the trusts not selected into the programme (control group). Moreover, let cs^T denote the caesarean section rate for the i-th treated trust. Likewise, cs^C denotes the caesarean section rate of the i-th trust that would have been observed had that trust not been selected to participate the programme. Apparently, none of the trusts can be observed simultaneously in two different states, given that either cs^C or cs^T is missing for each trust. This is the fundamental problem of causal inference and it is often being described as the evaluation problem of missing data. The microeconometric programme evaluation literature (Heckman *et al.*, 1997, 1998; Dehejia and Wahba, 2002) has shown that under certain assumptions the average treatment effect on the treated (ATT) trusts can be identified as:

$$ATT = E(\Delta|RIP = T) = E(cs^T|RIP = T) - E(cs^C|RIP = T) \tag{2}$$

where $\Delta = cs^T - cs^C$. In this case causal inference relies upon the construction of the counterfactual outcome contained in the last term of Equation (1). Since it cannot be observed, it can be approximated by the average caesarean section rate of the trusts that did not participate in the programme, i.e. $E(cs^C|RIP = C)$. However, this is subject to the assumption that there are no contemporaneous effects correlated with the programme participation status, otherwise the analysis is biased due to endogeneity and simultaneity. Hence, a valid control group must be selected in order to construct the counterfactual outcome and various matching algorithms can be employed for that purpose. The objective is to pair each participating trust with a non-participating one, based on some observed characteristics. In this way, an estimation bias is less likely to be induced as the trusts consisting the comparison group do not have substantially different characteristics given that they are not randomly chosen.

The matching process requires the comparison of participants and non-participants across a range of observable pre-programme characteristics. Rosenbaum and Rubin (1983) demonstrated the advantages of performing the matching based on a single index capturing all the information contained in a vector of observables affecting programme participation. Using standard parametric procedures (logit or probit), this single index is calculated as the predicted probability of programme participation conditional on some characteristics observed prior programme participation, i.e. $p(X) = Pr(RIP = T|X_{it-1})$. The choice of covariates is guided by the relevant theory, previous research on factors influencing self and institutional selection as well as any explicit criteria governing programme eligibility (Baum, 2013; Smith and Todd, 2005; Sianesi, 2004). Hence, the matching problem is reduced to a one-dimensional non-parametric estimation problem.

The identifying assumption behind the matching procedure is that, conditional on a set of observables, treatment status is orthogonal to treatment outcome, i.e. $(cs^T, cs^C) \perp RIP|X$. This is the Conditional Independence Assumption (CIA), also referred to as the "unconfoundedness" or the "selection on observables" assumption, and it implies that conditional on a set of observed characteristics, treatment status is purely random (Rosenbaum and Rubin, 1983; Lechner, 1999). Moreover, for the causal identification of the treatment effect, the balancing property of the estimated propensity score, $\hat{p}(X)$, must also be satisfied. This implies that trusts with the same propensity score must have the same distribution of characteristics, independently of the programme participation status, i.e. $RIP \perp X|p(X)$. In this way, treated and control observations with the same propensity score are observationally identical, on average, hence exposure to the treatment is a random process (Rosenbaum and Rubin, 1983). Finally, the common support assumption implies that the probability of programme participation for each value of the X vector lies strictly within the unit interval, as the probability of not receiving treatment does, i.e. $0 < P(RIP = T|X) < 1$. This ensures that there is sufficient overlap in the characteristics of participating and non-participating trusts to find adequate matches.

Once the propensity score has been estimated, participants are matched to untreated observations through a matching algorithm. Here we report results based on several algorithms, i.e. 1:m nearest neighbour(s) matching with and without replacement, radius matching, kernel and local linear regression matching (Dehejia and Wahba, 2002; Heckman *et al.*, 1997, 1998; Rosenbaum and Rubin, 1985; Rubin, 1973). Once the control group of trusts is obtained, the causal impact of programme participation can be evaluated by combining the propensity score matching with the difference-in-difference estimator. If t and t' denote a pre-programme benchmark and a follow-up period, respectively, the ATT can be identified as:

$$\begin{aligned} ATT_{DiD} &= E(\Delta_t - \Delta_{t'}|RIP = T) = E[(cs_t^T - cs_t^C) - (cs_{t'}^T - cs_{t'}^C)|RIP = T] \\ &= E[(cs_t^T - cs_{t'}^T)|RIP = T] - [(cs_t^C - cs_{t'}^C)|RIP = T] \end{aligned} \tag{3}$$

In this case, by comparing changes in the caesarean section rates of treated and control trusts before and after the programme implementation, we can estimate δ, which is the difference-in-differences parameter of interest from Equation 1. One main advantage over standard cross-sectional matching estimators is that any time-invariant unobserved differences between participating and non-participating trusts are eliminated, hence the "selection on observables" assumption can be relaxed.

4 Results

4.1 Balancing tests

The impact from programme participation estimated using the propensity score matching method will be reliable and robust only if we are strongly ignorable about the treatment status. Under the CIA assumption, pre-programme participation variables should be balanced between the treated and control groups otherwise the propensity score estimation will be misspecified (Rosenbaum and Rubin, 1983). A series of suggested balancing tests are performed in this subsection in order to ensure that the balancing condition is adequately satisfied by the data (Dehejia and Wahba, 2002; Smith and Todd, 2005).

[Figure 1 about here]

First, for each variable entering the propensity score model specification, a formal paired t-test is conducted. This will reveal whether the means of those variables are statistically different between treated and matched comparison trusts. The second balancing test examines the standardised difference (bias) in the common support region (Figure 1) for every explanatory variable.[8] For each variable, the bias is calculated as the standardised mean difference between participating and appropriately matched non-participating trusts, scaled by the average variances of that variable in both groups (Girma and Görg, 2007; Smith and Todd, 2005):

$$Bias(X) = \frac{100\frac{1}{N}\sum_{i\in T}[X_i - \sum_{j\in C} g(p_i, p_j)X_j]}{\sqrt{\frac{Var_{i\in T}(X)+Var_{j\in C}(X)}{2}}} \tag{4}$$

where N is the number of participating trusts and $g(\cdot)$ is a function assigning weights to the j-th control trust while constructing the counterfactual for the i-th treated trust.[9] The lower the standardised difference, the more similar are treated and control groups with respect to the variable under consideration. Smith and Todd (2005) have also proposed a regression-based balancing test. If $\hat{p}(X)$ is the

[8] As seen in Figure 1, two trusts lie outside the common support region and therefore they are excluded from the analysis.

[9] In this case it is the Gaussian kernel function that is defined as $g(p_i, p_j) = \frac{K[\frac{(p_i - p_j)}{h}]}{\sum_{k\in C} K[\frac{(p_i - p_k)}{h}]}$, where $K(\mu) \propto exp(\frac{-\mu^2}{2})$ is the Gaussian normal function, h is the bandwidth parameter and C is the group of non-treated trusts.

estimated propensity score of programme participation and RIP is a treatment status indicator, then the following regression is estimated for each variable entering the propensity score model:

$$X = \alpha_0 + \sum_{k=1}^{m} \alpha_k \hat{p}(X)^k + \sum_{k=1}^{m} \beta_k RIP \hat{p}(X)^k + \varepsilon \tag{5}$$

where $m = 4$ and the joint significance of the βs is tested. If the balancing property is sufficiently satisfied by the propensity score, then the βs should not be jointly statistically significant (Smith and Todd, 2005). In this way, the ability of D to provide any additional information for each covariate, conditional on various levels and non-linearities of $\hat{p}(X)$ is tested (Lee, 2013).

The results regarding the alternative balancing tests are displayed in Table 2. They are based on a Gaussian kernel matching procedure with replacement and a bandwidth parameter set equal to 0.06.[10] In the matched sample, the standardised differences between participating and non-participating trusts are all less than 10.5%. In most of the cases, adopting a matching method leads to a substantial bias reduction. The results obtained from the regression based tests also support the relevance of the propensity score matching method. For every variable entering the treatment participation equation, the null hypothesis about the βs being jointly significant is rejected. Moreover, in order to test whether the differences in the explanatory variables are jointly insignificant we also run a Hotteling's T-squared generalised means test. According to the results, the differences were not jointly significant (F-stat$= 0.945$), hence the balancing conditions seem to be satisfied. These results provide evidence that the propensity score model has been adequately specified regarding the factors influencing programme participation (Table A1 in the Appendix).

[Table 2 about here]

4.2 Difference-in-difference matching estimates

Based on Equation 1, we begin by presenting some plain DiD estimates with and without the use of covariates. These include the variables listed in Tables 1 and 2 plus a set for Strategic Health Authority (SHA) indicators to control for regional effects that are fixed across trusts (East Midlands is the base region). With respect to the age, ethnic group and socio-economic status sets of variables, the reference categories are women younger than 20 years old, women with white ethnic background and women living in areas classified in the first IMD quintile, respectively. All models control for the full set of explanatory

[10] We have also examined the balancing properties based on other matching algorithms. The results are fairly similar and available upon request from the authors. Moreover, in the regression-based balancing test, we have also set $m = 3$ with the results suggesting that the balancing property is adequately satisfied.

variables. We report results regarding the total, planned (elective) and emergency (non-elective) caesarean section rates. As seen in Table 3, there are statistically significant differences between the treated and control groups in the cases of the total and the planned caesarean delivery rates. These differences remain significant after controlling for a large number of confounding variables and regional dummies. According to the results, caesarean section rates were higher in treated trusts during the baseline period (2008m7-2008m12); .024 and .014 on average for the overall and the planned caesareans, respectively. Moreover, those differences vanished during the first follow-up period (2009m1-2009m6), while in the case of the total caesarean section rate, the difference in the second follow-up period (2009m7-2010m1) remained the same as in the baseline period. Regarding the emergency caesarean section rates, their follow-up differences seem to have been increased. Yet, within the DiD framework adopted here, what interests is the deviations of these post-participation differences from the baseline ones. In the case of the fully specified model (Panel B) it seems that only the planned caesarean section rate has been decreased (-0.017) and only in the first post-participation period.

[Table 3 about here]

However, these relationships cannot be assumed to be causal ones, given the selectivity into programme participation and the unobservability of the counterfactual outcomes of the participating trusts. According to the preceding analysis, the treated and control groups of trusts are comparable with respect to their observable pre-treatment attributes, conditional on the estimated propensity score. In order to identify the causal effects of programme participation on the trust-level caesarean section rates, we proceed by combining difference-in-differences with propensity score matching estimators. Throughout the estimations, the common support condition is imposed in order to compare non-participating trusts that fall within the propensity score distribution of the participating trusts. We also report results obtained after using various matching algorithms and the two alternative post-participation periods. Table 4 displays the results regarding the total caesarean section rate. With respect to the unmatched sample, the estimates are relatively close to the DiD ones when no covariates were used (Panel A). According to the matching estimates displayed next, the impact of programme participation ranges between -0.010 and -0.016, depending on the matching algorithm. In the case of the 1:1 nearest neighbour matching without replacement, each treated trust is paired to its closest comparison trust in terms of the estimated propensity score, with the latter being only once considered. If replacement is allowed, a comparison trust is used more than once. This reduces the number of distinct non-participating trusts used to construct the counterfactual outcome, improves the matching quality and reduces the bias (Caliendo and Kopeinig, 2008; Smith and Todd, 2005). In the oversampling case, where more than 1 nearest neighbours are used ($m = 5$), the result remains significant, however, a trade-off between variance and bias is involved because the matches can be poorer as more information is used to construct the counterfactual caesarean rate.[11]

[Table 4 about here]

In order to avoid the risk of poor matching given that some close neighbours may be "far away" in terms of propensity score, we can impose a caliper tolerance level, i.e. a maximum propensity score distance (Caliendo and Kopeinig, 2008). Dehejia and Wahba (2002) have proposed a radius matching variant that uses all the control observations lying within the specified caliper (set equal to 0.06) and avoids the use of bad matches. The result remains significant and close to that obtained in the case of *1:m* nearest neighbour matching. In the case of kernel matching, weighted averages of all the control trusts are used in order to construct the counterfactual and the variance is even lower (at least when the Gaussian kernel is used). However, including all the observations may incorporate some bad matches, therefore it is important to impose the common support condition. In the kernel matching case, a weighted regression of the counterfactual caesarean section rate on an intercept is estimated, whereas the weights increase with the propensity score distance between treated and control trusts. In the case of local linear matching, a linear term is also included in the regression, in order to account for any asymmetries or gaps in the estimated propensity score distributions for treated and untreated units, e.g. as those in Figure 1 (Caliendo and Kopeinig, 2008).[12] As seen in Table 4, in that case the estimated ATT effect is marginally higher (-0.013) and more significant. A final point regarding the kernel and local linear matching, is the selection of the bandwidth parameter (here it is set to 0.06) that again involves a trade-off between bias and variance.[13]

The same set of estimation was also performed for planned and emergency caesarean section rates, separately. Tables 5 and 6 display the obtained results. With respect to the planned caesarean rate, the estimated ATT effect ranges between -0.006 and -0.013 depending on the matching method, however, a statistically significant difference is observed only in the first follow-up period. Hence, the Rapid Improvement Programme for Promoting Normal Birth in English NHS hospitals, seem to had had a small impact on reducing the increasing caesarean section rates.

[Table 5 about here]

[11] This is more evident when the number of matching partners is allowed to vary. The results are available upon request.

[12] The Gaussian kernel has been used in the case of the local linear matching, however, experimenting with other kernels led to similar results that are available upon request.

[13] The bandwidth parameter was allowed to vary between 0.010 and 0.095 and this trade-off was evident in the case of kernel matching. Regardig the local linear matching, the results were not so sensitive to the bandwidth selection.

[Table 6 about here]

5 Conclusions

Caesarean section rates have been increasing during the last years and efforts have been made to moderate this trend. One such attempt was the 'Rapid Imrovement Programme' implemented in the English NHS circa mid-2008. This programme aimed in promoting vaginal birth over caesarean section by providing the participating trusts a Toolkit to reduce and keep their caesarean section rates low. Twenty trusts selected from a wider pool of applicants participated into the programme. This provides a setting for an evaluation of the impact the programme had had on the trusts that participated. In the empirical analysis, we used data from the maternity tail of the Hospital Episode Statistics for the period before and after the programme participation. Moreover, we adopted an estimation technique combining difference-in-differences with propensity score matching estimators in order to evaluate whether programme participation had had a causal effect on the caesarean section rates of English NHS hospitals. According to the results, participating trusts exhibited a decrease in their overall and their planned caesarean section rates. However, the uncovered reduction was small, observed only shortly after the programme participation and disappeared after that. No effect was found in the case of the emergency caesarean section rate.

References

[1] Althabe, F., Sosa, C., Belizán, J.M., Gibbons, L., Jacquerioz, F. and Bergel, E. (2006), "Cesarean Section Rates and Maternal and Neonatal Mortality in Low-, Medium-, and High-Income Countries: An Ecological Study", *Birth*, 33(4): 270–277

[2] Alves, B. and Sheikh, A. (2005), "Investigating the Relationship Between Affluence and Elective Caesarean Sections", *British Journal of Obstetrics and Gynaecology*, 112: 994-996

[3] Baicker, K., Buckles, K.S. and Chandra, A. (2006), "Geographic Variation in the Appropriate Use of Cesarean Delivery", *Health Affairs*, 25(5): w355–w367

[4] Baldwin, J., Brodrick, A., Cowley, S. and Mason, N. (2010), "An innovative kit to support normal birth", *Journal of Midwifery & Women's Health*, 55(3): 270–272

[5] Baum, C.F. (2013), "Propensity score matching, regression discontinuity, limited dependent variables", EC 823: Applied Econometrics, Boston College, Spring 2013

[6] Blundell, R. and Costa Dias, M. (2000), "Evaluation methods for non-experimental data", *Fiscal Studies*, 21(4): 427–468

[7] Bragg, F., Cromwell, D.A., Edozien, L., Gurol-Urganci, I., Mahmood, T.A., Templeton, A. and van der Meulen, J.H. (2010), "Variation in Rates of Caesarean Section Among English NHS Trusts After Accounting for Maternal and Clinical Risk: Cross Sectional Study", *British Medical Journal*, 341: c5065

[8] Caliendo, M. and Kopeinig, S. (2008), "Some practical guidance for the implementation of propensity score matching", *Journal of Economic Surveys*, 22(1): 31–72

[9] DCLG (2011), "The English Indices of Deprivation 2010", Neighbourhood Statistics Release, URL: `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf`, accessed on 10 December 2014

[10] Declercq, E., Young, R., Cabral, H. and Ecker, J. (2011), "Is a Rising Caesarean Delivery Rate Inevitable? Trends in Industrialized Countries, 1987 to 2007", *Birth*, 38(2): 99–104

[11] Dehejia, R.H. and Wahba, S. (2002), "Propensity score-matching for nonexperimental causal studies", *Review of Economics and Statistics*, 84(1): 151–161

[12] Deneux-Tharaux, C., Carmona, E., Bouvier-Colle, M.H. and Bréart, G. (2006), "Postpartum Maternal Mortality and Cesarean Delivery", *Obstetrics & Gynecology*, 108(3, Part 1): 541–548

[13] DiMatteo, M.R., Morton, S.C., Lepper, H.S., Damush, T.M., Carney, M.F., Pearson, M. and Kahn, K.L. (1996), "Cesarean Childbirth and Psychosocial Outcomes: A Meta-Analysis", *Health Psychology*, 15(4): 303–314

[14] Dubay, L., Kaestner, R. and Waidmann, T. (1999), "The Impact of Malpractice Fears on Cesarean Section Rates", *Journal of Health Economics*, 18(4): 491–522

[15] Fenwick, J., Staff, L., Gamble, J., Creedy, D.K. and Bayes, S. (2010), "Why do Women Request Caesarean Section in a Normal, Healthy First Pregnancy?", *Midwifery*, 26(4): 394-400

[16] Francome, C. and Savage, W. (1993), "Caesarean Section in Britain and the United States 12% or 24%: Is Either the Right Rate?", *Social Science & Medicine*, 37(10): 1199–1218

[17] Grant, D. (2009), "Physician Financial Incentives and Cesarean Delivery: New Conclusions From the Healthcare Cost and Utilization Project", 'emphJournal of Health Economics, 2(1): 244–250

[18] Gruber, J., Kim, J. and Mayzlin, D. (1999), "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery", *Journal of Health Economics*, 18(4): 473–490

[19] Gruber, J. and Owings, M. (1996), "Physician Financial Incentives and Cesarean Section Delivery", *RAND Journal of Economics*, 27(1): 99–123

[20] Heckman, J.J., Ichimura, H. and Todd, P.E. (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme", *Review of Economic Studies*, 64(4): 605–654

[21] Heckman, J.J., Ichimura, H. and Todd, P.E. (1998), "Matching as an econometric evaluation estimator", *Review of Economic Studies*, 65(2): 261–294

[22] Imbens, G.W. and Wooldridge, J.M. (2009), "Recent developments in the Econometrics of program evaluation", *Journal of Economic Literature*, 47(1): 5–86

[23] Kalström, A., Nystedt, A., Johansson, M. and Hildingsson, I. (2011), "Behind the Myth: Few Women Prefer Caesarean Section in the Absence of Medical or Obstetrical Factors", *Midwifery*, 27(5): 620–627

[24] Lancet [editorial] (2000), "Caesarean Section on the Rise", *Lancet*, 356(9243): 1697

[25] Landon, M.B., Hauth, J.C., Leveno, K.J., Spong, C.Y., Leindecker, S., *et al.* (2004), "Maternal and Perinatal Outcomes Associated with a Trial of Labor After Prior Cesarean Delivery", *New England Journal of Medicine*, 351: 2581–2589

[26] Lavender, T., Hofmeyr, G.J., Neilson, J.P., Kingdon, C., Gyte, G.M.L. (2012), "Caesarean Section for Non-Medical Reasons at Term", *Cochrane Database of Systematic Reviews*, (3): CD004660

[27] Lechner, M. (1999) "Earnings and employment effects of continuous off-the-job training in East Germany after unification", *Journal of Business and Economics Statistics*, 17(1): 74–90

[28] Lee, W.S. (2013), "Propensity score matching and variations on the balancing test", *Empirical Economics*, 44(1): 47–80

[29] Lefèvre, M. (2014), "Physician Induced Demand for C-Sections: Does the Convenience Incentive Matter?", WP 14/08, Health Econometrics and Data Group (HEDG), University of York

[30] Leuven, E. and Sianesi, B. (2003), "`psmatch2`: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing", `http://ideas.repec.org/c/boc/bocode/s432001.html`, version 4.0.11

[31] Localio, A.R, Lawthers, A.G., Bengtson, J.M., Hebert, L.E., Weaver, S.L., Brennan, T.A. and Landis, J.R. (1999), "Relationship Between Malpractice Claims and Cesarean Delivery", *Journal of the American Medical Association*, 269(3): 366–373

[32] Marshall, J.L, Spiby, H. and McCormick, F. (2015), "Evaluating the 'Focus on Normal Birth and Reducing Caesarean Section Rates Rapid Improvement Programme': A mixed method study", *Midwifery*, 31: 332–340

[33] Menacker, F. and Hamilton, B.E. (2010), "Recent Trends in Cesarean Delivery in the United States", *National Center for Health Statistics Data Brief*, 35: 1–8

[34] National Institute for Health and Care Excellence (2007), "Intrapartum Care: Care of Healthy Women and Their Babies During Childbirth", NICE Clinical Guideline 55, URL: `www.nice.org.uk/guidance/cg55/resources/guidance-intrapartum-carepdf`, accessed on 10 September 2014

[35] Oleske, D.M., Glandon, G.L., Giacomelli, G.J. and Hohmann, S.F. (1991), "The Cesarean Birth Rate: Influence of Hospital Teaching Status", *Health Services Research*, 26(3): 325–337

[36] Paranjothy, S., Frost, C. and Thomas, J. (2005), "How Much Variation in CS Rates Can Be Explained by Case Mix Differences?", *British Journal of Obstetrics and Gynaecology*, 112(5): 658–666

[37] Roberts, C.L. and Nippita, T.A. (2015), "International Caesarean Section Rates: The Rising Tide", *Lancet Global Health*, 3(5): e241–e242

[38] Rosenbaum, P.R. and Rubin, D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70(1): 41–55

[39] Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a control group using multivariate matched sampling methods that incoroporate the propensity score", *American Statistician*, 39(1): 33–38

[40] Rubin, D.B. (1973), "Matching to remove bias in observational studies", *Biometrics*, 29(1): 159–183

[41] Sakala, C. and Corry, M.P. (2008), "Evidence-Based Maternity Care: What It Is and What It Can Achieve", New York: Milbank Memorial Fund 2008, URL: `http://www.childbirthconnection.com/pdfs/evidence-based-maternity-care.pdf`, accessed on 5 June 2015

[42] Sandall, J., Murrells, T., Dodwell, M., Gibson, R., Bewley, S., Coxon, K., Bick, D., Cookson, G., Warwick, C. and Hamilton-Fairley, D. (2014), "The Efficient Use of the Maternity Workforce and the Implications for Safety and Quality in Maternity Care: A Population-Based, Cross-Sectional Study", *Health Services and Delivery Research*, 2(38)

[43] Shearer, E.L. (1993), "Cesarean Section: Medical Benefits and Costs", *Social Science & Medicine*, 37(10): 1223–1231

[44] Shorten, A. (2007), "Maternal and Neonatal Effects of Caesarean Section", *British Medical Journal*, 335: 1003–1004

[45] Sianesi, B. (2004), "An evaluation of the active labour market programmes in Sweden", *Review of Economics and Statistics*, 86(1): 133–155

[46] Smith, J. and Todd, P. (2005), "Rejoinder", *Journal of Econometrics*, 125(1-2): 365–375

[47] Villa, J.M. (2012), "Simplifying the estimation of difference-in-differences treatment effects with Stata", MPRA Working Paper, 43943, November

[48] Villar, J., Valladares, E., Wojdyla, D., Zalaveta, N., Carroli, G., Velazco, A., *et al.* (2006), "Caesarean Delivery Rates and Pregnancy Outcomes: The 2005 WHO Global Survey on Maternal and Perinatal Health in Latin America", *Lancet*, 367: 1819–1829

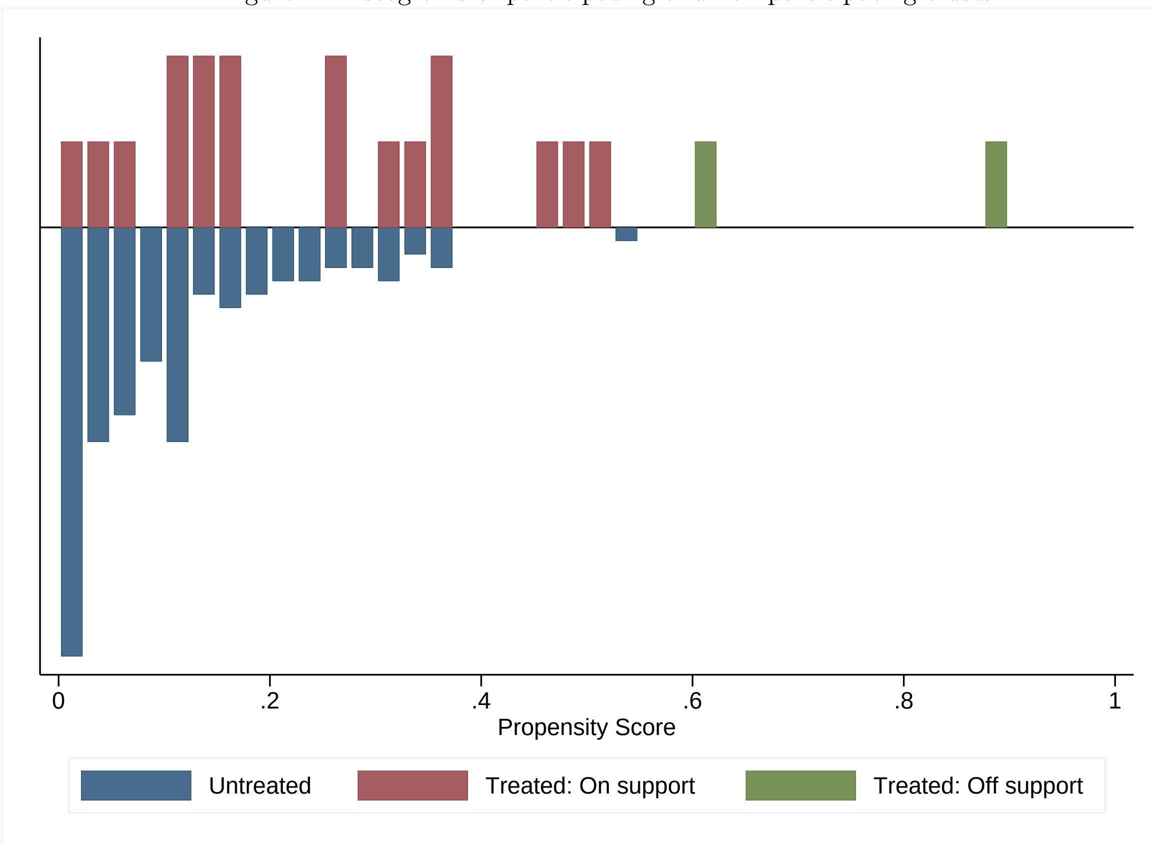
[49] World Health Organization, UNFPA, UNICEF and AMDD, “Monitoring Emergency Obstetric Care: a Handbook”, Geneva: WHO; 2009

[50] Yang, Y.T., Mello, M.M., Subramanian, S.V. and Studdert, D.M. (2009), “Relationship Between Malpractice Litigation Pressure and Rates of Cesarean Section and Vaginal Birth After Cesarean Section” *Medical Care*, 47(2): 234–242

[51] Yang, Q., Wen S.W., Oppenheimer L., Chen X.K., Black D., Gao J. and Walker, M.C. (2007), “Association of Caesarean Delivery for First Birth with Placenta Praevia and Placental Abruption in Second Pregnancy”, *British Journal of Obstetrics and Gynaecology*, 114(5): 609–613

Figures & Tables

Figure 1: Histograms of participating and non-participating trusts



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

Table 1: Descriptive statistics for main pre-treatment characteristics

	Total sample		Control group		Treatment group		t-test	Kolmogorov-
Variable name	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	p-value	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 years old	.064	.023	.065	.023	.059	.023	.339	.624
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
White ethnic origin	.755	.183	.749	.191	.792	.121	.323	.189
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 I	.260	.017	.271	.019	.189	.035	.095	.328
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, Health & Social Care Information Centre (HSCIC).

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, Health & Social Care Information Centre (HSCIC).
Notes: p-values in parentheses. For the Gaussian kernel matching the bandwidth parameter is set to 0.06.

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 ($\hat{\gamma} + \hat{\delta}$)	.012	.008	.004	.021	.003	.018
	(.018)	(.013)	(.014)	(.018)	(.019)	(.008)^b
Difference-in-differences ($\hat{\delta}$)	-.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 ($\hat{\gamma} + \hat{\delta}$)	.011	-.003	.014	.026	.008	.018
	(.009)	(.006)	(.008)^c	(.008)^a	(.006)	(.006)^a
Difference-in-differences ($\hat{\delta}$)	-.013	-.017	.004	.000	-.004	.004
	(.013)	(.008)^b	(.011)	(.011)	(.008)	(.008)

Source: Hospital Episode Statistics (HES) database, Health & Social Care Information Centre (HSCIC).
Notes: Estimates were performed using the user-written Stata command `diff` (Villa, 2012.)

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, 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).

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, 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).

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, 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).

Appendix

Table A1: Programme participation equation: First-step logit results

Pre-treatment variable name	Parameter estimates: Coefficient	Parameter estimates: Standard error	Marginal effects: Coefficient	Marginal effects: Standard error
High risk	0.849	3.349	0.083	0.328
Healthy mother	6.672	16.249	0.650	1.603
Parity	-2.653	2.807	-0.258	0.266
Nulliparous	-9.420	8.793	-0.917	0.828
Urban	3.545	3.981	0.345	0.384
Foundation trust	1.059	0.830	0.103	0.078
Teaching hospital	0.393	0.732	0.038	0.072
Discharged home in 28 days	2.725	15.487	0.265	1.518
Readmitted within 28 days	34.002	18.079[b]	3.311	1.707[b]
Age: 20-24 years old	-19.197	39.112	-1.869	3.881
Age: 25-29 years old	-35.038	35.980	-3.412	3.590
Age: 30-34 years old	-19.684	32.720	-1.917	3.258
Age: 35-39 years old	-39.997	30.963	-3.895	3.108
Age: >40 years old	54.929	75.739	5.354	7.243
Mixed ethnic origin	1.772	12.935	0.173	1.257
Asian ethnic origin	4.216	5.229	0.411	0.517
Afro-Caribbean ethnic origin	-4.984	7.392	-0.485	0.728
Other/Unknown ethnic origin	6.898	4.101[c]	0.672	0.392[c]
IMD II	5.958	5.441	0.580	0.531
IMD III	13.009	9.387	1.267	0.879
IMD IV	2.583	7.803	0.252	0.766
IMD V	10.468	6.218[c]	1.019	0.594[c]
Hospital load	0.046	0.064	0.004	0.006
Regional dummies	Yes		Yes	
Pseudo-R^2	0.191			
Number of trusts	148		148	

Source: Hospital Episode Statistics (HES) database, Health & Social Care Information Centre (HSCIC).
Notes: For the parameter estimates, robust standard errors have been calculated. For the marginal effects, the standard errors have been calculated using the Delta method. [a], [b] and [c] denote statistical significance at the 1%, 5% and 10% levels, respectively.