Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 195 (2015) 93 – 102

Procedia Social and Behavioral Sciences

World Conference on Technology, Innovation and Entrepreneurship

Evaluation of a Pay for Performance Scheme in Maternity Care: The Commissioning Quality and Innovation Payment Framework in England

Merve Ertok[a]*

[a]*University of Brsitol Department of Economics, Bristol, BS8 1TH, United Kingdom*

Abstract

Pay for performance schemes (P4P) have been considered by many governments and implemented (for example in the UK and the US) for many years. The English health care system is an example of this: it has recently experienced a wide range of policy interventions which pay providers of health care for quality with the aim of improving the quality of care delivered to the population. Commissioning Quality and Innovation Payment Framework started in 2009/10 with the aim to improve quality of care provided by NHS Trusts. This study examines the impact of the introduction of the Commissioning Quality and Innovation Payment framework (CQUIN) on elective and emergency c-section rates in England for the 2010/11 CQUIN taxonomy. I show that the CQUIN scheme does not have any statistically significant impact on c-section rates.

Peer-review under responsibility of Istanbul Univeristy.

Keywords: "performanc; hospital; pay for performance; maternity care"

1. Introduction

Pay for performance schemes (P4P) have been considered by many governments and implemented (for example in the UK and the US) for many years. The English health care system is an example of this: it has recently experienced a wide range of policy interventions which pay providers of health care for quality with the aim of

* Corresponding author: Tel.:00905078351964.
E-mail address: merve.ertok@bristol.ac.uk

Peer-review under responsibility of Istanbul Univeristy.
doi:10.1016/j.sbspro.2015.06.333

improving the quality of care delivered to the population. The majority of the UK health care services are tax free at the point of use and provided by the National Health Service (NHS). The NHS employs more than 1 million people and delivers care to a population of 53 million people only in England (NHS website, 2015). Therefore the performance of NHS health care workers and patient satisfaction have been a focus of the English government's health care interventions. The motivation behind pay for performance schemes in health care is to provide financial rewards to health care providers for the level of their performance and therefore reward the quality of their services by making the link between quality and reimbursement more explicit. Here I focus on England. During the past two decades English NHS has adopted a bundle of P4P schemes varying in terms of design and the level at which financial rewards are given. The first (and well-known) performance scheme is the Quality and Outcomes Framework (QOF), introduced at national level in 2004 to improve quality of primary care in the UK. This was followed by the introduction of Advancing Quality (AQ) in Northwest England (2008), which targeted hospitals (secondary and tertiary care providers). In 2009 the AQ scheme became a part of Commissioning Quality and Innovation Framework (Meacock et al. 2014). The CQUIN scheme is a national framework intended to improve quality indicators determined locally as part of negotiation between hospitals and their NHS funders (Department of Health CQUIN summary guide, 2010).

In this paper, I examine a part of the CQUIN scheme. I focus on an important indicator of the quality of maternity care; the rate of caesarean sections (c-sections). C-section rates were included in the CQUIN contract as a (negative) target when it was first introduced in 2009/10 financial year in England, with the intention to reduce costs and promote normal births in maternity care. I group c-sections into elective and emergency c-sections depending on the type of admission and analyse them separately. I start with a simple difference in differences (DiD) design to investigate whether the introduction of the CQUIN scheme reduced c-section rates in 2010/11. The 2010/11 Hospital Episode Statistics maternity data are used since this set of data was the only available (officially released) data for the post-policy analysis at the time of the analysis. The initial year of the CQUIN scheme (2009/10) was a transition period, therefore, the 2009/10 taxonomy is excluded from the study. Sutton at el. (2013) suggest that providers considered the first year of the scheme more as a test of proof of concept and they did not really know much about how the scheme was supposed to work. The communication between local providers and commissioners was not efficient due to the tight schedule of the policy, thus the indicators agreed in the taxonomies were not clearly understood by providers (Sutton et al. 2013). By design, the CQUIN scheme allows me to use naturally occurring control and treatment groups within a DiD design. In addition to the baseline DiD, I control for patient characteristics and time invariant hospital factors which might potentially be correlated with whether a certain hospital is in the treatment group but independent from the main impact of the policy on c-section rates. I examine whether the policy was endogenous to any key observables such as patient characteristics, previous years' c-sections and overall performance of NHS trusts before the main analysis. Most concerns over endogeneity arises from the selection of goals at local level by commissioners and providers as the CQUIN indicators/topics could be selected based on hospitals' previous performances (Sutton et al. 2013). For instance, if a hospital did not perform well for a particular indicator/topic in previous years, the indicator/topic could be included in its CQUIN scheme in the following year. In addition to the selection of goals, patient characteristics might have considerable effects on the selection of hospitals. High risk patients with informed knowledge on hospital quality can select hospitals with high performances and service quality.

2. Literature Review on Pay for Performance schemes

There is an increasing number of developed countries who have adopted pay for performance schemes with a drive to improve provision of health care services (including the UK, US, Denmark, Canada, Germany, Sweden, Norway and the Netherlands). A large body of literature has sought to provide evidence on the impact of pay for performance schemes on quality, especially from the US and the UK. The main inference from the literature is that the impact of such schemes on quality is ambiguous and limited in the long term (Werner et al. 2011; Meacock et al. 2014; Epstein 2007). A set of studies from the UK and the US suggests that the introduction of P4P schemes have had a positive impact on quality (Gillam et al. 2012; Werner et al. 2011; Steel et al. 2007). For example, a systematic review by Gilliam et al. (2012) examines 575 papers from the MEDLINE, EMBASE, and PsycINFO

databases which focus on the Quality and Outcomes Framework (QOF) in the UK. The review suggests that the introduction of the QOF was associated by modest improvements on incentivised quality indicators for chronic diseases compared to the pre-incentivised period. A study by Rosenthal and Frank (2006) finds that most studies for the US health care market suggest that the effectiveness of P4P schemes is weak.

The classical principal-agent theory assumes that a principal who is responsible for a multi-dimensional task might exert his/her effort to verifiable/incentivised dimensions of quality at the cost of a reduction in effort towards unverifiable/unincentivised dimensions (i.e. hospitals are multi-dimensional organizations). Within the health economics literature, doctors are assumed to be altruistic up to a certain level (Jack 2005; Kaarboe and Siciliani 2011). This suggests that they care about not only their patients' health and welfare but also the payment they receive by achieving the overall verifiable/incentivised dimensions of quality. Eggleston (2005) suggests that P4P schemes might improve the overall welfare and health of patient's measured by indicators covered by the scheme at the cost of a reduction in patient's benefit measured by indicators of quality which remains unverifiable (cited in Kaarboe and Siciliani 2011). Despite the implications of principle agent theory for the impact of P4P schemes on unincentivesed/unveriable indicators of quality, from the empirical literature point of view there is a lack of clear evidence of the impact of such schemes on unincentivised domains/indicators in health economics. A number of studies suggest that the quality of health domains/indicators which are not covered by P4P schemes remains stable over time whereas others suggest that such initiatives lead to comprehensive improvements in the entire health care system (Cookson and Fleetcroft 2006; Doran et al. 2011; Steel et al. 2007; Ganz et al. 2007; Asch et al. 2004; Campbell et al. 2007). Eijkenaar et al. (2013) suggests that P4P schemes may crowd out health care providers' intrinsic motivation and may deteriorate the non-financial motivation and lead to the diversion of efforts, therefore, a large number of outcome measures may be necessary in designing such schemes. Incentives on public reporting via patient choice and feedback reports on peer performance may help attenuate the diversion of efforts from non-incentivised activities to incentivised activities (Eijkenaar 2013). Mehrotra et al. (2009) suggest that public reporting contributes to improvements in non-incentivised activities. However, a systematic review on the impact of P4P schemes on clinical quality by Schatz (2008) finds weak evidence for the impact of public reporting (PR) and feedback on the achievement of P4P schemes. In addition to the impact of P4P schemes on non-incentivised activities, pay for performance schemes also affect the cost effectiveness of the delivery of health care services. Cost effectiveness is to obtain a higher level of quality with lower costs or the same level of quality with either equal or a lower level of costs (Eijkenaar et al. 2013). Gillam et al. (2012) and Smith et al. (2011) suggest that the introduction of the QOF (Quality and Outcomes framework) scheme led to a modest reduction in cost effectiveness of mortality and hospital admissions for some health care domains in the UK. In the UK, the demand for secondary care has been rising as the population is ageing (National Health Service, 2013). Therefore to the extent that demand of people with chronic conditions is increasing and the cost of preventive care is high, the NHS struggles under the pressure of financial stability (National Health Service, 2013; Smith et al. 2011). The QOF scheme provides a direct link between secondary care costs and the quality of primary care. Smith et al. (2011) suggests that rewarding GP practices with financial incentives for ensuring preventive quality in primary care leads to a reduction in avoidable admissions, therefore in secondary care costs. In contrast to the impact of P4P schemes on cost efficiency in the UK, a survey report from the US discussing the impact of the Physician Group Practice Demonstration scheme on hospital efficiency suggests that this P4P scheme had a positive impact on hospital costs and savings (Centers for Medicare & Medicaid Services, 2011). With respect to health care inequalities, the fundamental aim of P4P schemes is generally to raise quality rather than reduce disparities across different patient groups. Therefore we may not initially expect an improvement in equity. However, Gillam et al. (2012) and Doran et al. (2011) provide modest evidence that QOF in the UK has reduced inequalities across patient characteristics such as age, gender, ethnicity and socio economic deprivation. They suggest that the variation in the levels of performance was reduced in less deprived areas compared to deprived regions.One major concern for policy makers is how to design P4P schemes. Such schemes should be organized in a way leading to an overall improvement in population health and should not lead to any type of deterioration in patient satisfaction and reliability of data reporting. Such initiatives create opportunities for providers subject to the scheme to misreport data and/or to concentrate on patient selection rather than their own performance. A study by Smith (1995) examines the unintended consequences of publishing

performance data in public sector with a wide range of examples from the UK public sector. He suggests that to minimize public sector managers gaming (referred as "ratchet" effect in the study) P4P schemes, governments should use a set of performance indicators and develop reference points for indicators independent from the organization's former behaviours (Smith 1995).

3. The CQUIN scheme

The Commissioning for Quality and Innovation payment framework (known as CQUIN) commenced in April 2009 in England. The main motivation of the scheme is to link a proportion of NHS provider's income (provider refers to an NHS trust (hospital) in the CQUIN concept) to the achievement of locally defined quality targets. The scheme is based on "local discretion and a discussion between commissioner and provider" (Using the CQUIN payment framework, 2008, p.13). One of the main goals of the scheme is to provide local improvements in quality and efficiency in health care within an NHS payment system (Department of Health, summary guide 2010). It is not intended to "provide a mechanism for investment in large-scale change. Instead, it supports a shift towards a culture where innovation is part of what the NHS does and recognizes locally" (Using the CQUIN payment framework 2008). Moreover, the mission of the CQUIN scheme is set parallel to the previous payment schemes such as Quality and Outcomes Framework which started in 2004 with the intention that there is no conflict between those payment frameworks all aiming quality improvement, equity, safety and efficiency in health care services. The main focus of my study is on NHS acute trusts (NHS secondary and tertiary care providers) and Primary Care Trusts (PCTs) who are the commissioners of secondary care services. I examine the 2010/11 CQUIN scheme and exclude the 2009/10 taxonomy which was the transition period for providers and commissioners. The CQUIN scheme is not a recurrent scheme (the content of the scheme might differ for each financial year). The areas/domains on which a provider trust focuses on, and the rewards gained for improvements in those health areas and indicators, are determined by local negotiations between commissioners and trusts. In the 2010/2011 CQUIN framework (but not the 2009/10 scheme), there are two national goals in schemes for acute trusts. One of them is to reduce avoidable death, disability and responsiveness to the personal needs of patients (Using the Commissioning for CQUIN framework, 2010). The other goals are defined at either local or regional level. In 2010/11 is the financial value of the CQUIN scheme was increased to 1.5% of a provider contract in 2010/11 from 0.5% in the 2009/10 (although 0.3 % of the 1.5% was set aside for the national goals). In the CQUIN concept, a topic is defined according to a specified goal with specific indicators. A topic refers to a particular area/domain such as maternity, health & safety, data collection or diabetes where an improvement in performance is judged to be required by the commissioner and provider. An indicator is a measure of attainment of locally agreed quality targets. For some topics, more than one indicator might be defined depending on the number of targets agreed to be improved. In the English NHS health care setting, Primary Care Trusts (PCTs) are the primary organizations responsible for purchasing NHS services for their local area. Strategic Health Authorities (higher level purchasers) are responsible for monitoring PCTs and providers of health care who have not met Foundation Status requirements and ensuring that these PCTs and providers are implementing national policies (Fischera et al. 2012). The selection of topics, indicators and baseline performance, at the regional or trust level, is non-random. Topics and indicators might be regionally mandated (known as a Strategic Health Authority (SHA) wide topic/indicator), locally negotiated (known as fully flexible) or SHA suggested. An SHA wide indicator/topic or SHA mandated topic/indicator applies to all trusts in an SHA while locally flexible ones are negotiated between trusts and their commissioners. Therefore, it is up to both to determine the indicator and/or the topic. More flexibility is associated with a locally negotiated topic or indicator in terms of selection of both topic and indicator. Unfortunately, there is no information available for researchers on how topics and indicators are selected. Therefore, the CQUIN topic/indicator could be partly endogenous. I examine the effect of the scheme on trusts which have SHA mandated, SHA suggested or fully flexible c-section rates as an indicator. Table A.1 (Appendix) shows that for the 2010/11 taxonomy, almost 12% of acute trusts (whose content of the CQUIN scheme is known) have fully flexible c-section rates as an indicator. Almost 10% of the trusts have SHA wide c-section rates while 3% of those have SHA suggested c-section rates in their contracts.

4. Methodology

4.1. Research Goal

I investigate the impact of the CQUIN scheme on two sets of outcomes: elective and emergency caesarean section rates. I analyse these two sets of outcomes separately due to the differences in the nature of these deliveries. The study investigates two main questions. First, given that topics (maternity)/indicators (c-section rate) are chosen by trusts in consultation with the SHA (or purchasers) do trusts who choose to adopt a topic or an indicator differ from those who do not ("Is the adoption of the policy endogenous?"), and second, did the scheme affect rates of c-sections? Trusts are grouped into 4 depending on the content of their local CQUIN scheme by the 2010/11 taxonomy. All statistical analyses are carried out at the NHS trust level for elective and emergency caesarean section rates and mother's and baby's characteristics are aggregated to the same level. Trusts whose contents of the CQUIN scheme are unknown are excluded from the analysis (these are referred to as Group A). For the study population, 5% of all trusts are in group A (Table A.1, Appendix). Trusts are grouped as:

Group A: Excluded from the analysis because their CQUIN content is unknown.
Group B: Maternity is not a topic in CQUIN scheme.
Group C: Topic is maternity but indicator is not c-section rate.
Group D: Topic is maternity and indicator is c-section rate. Therefore the treated group is group D.

If the indicator is not the c-section rate then the trust is not rewarded for hitting the target (be this a general or a specific reduction). Therefore the treated group, in terms of getting a reward, is group D. But group C also has a focus on maternity care, but reductions in c-sections are not rewarded. So in the analyses I examine groups C and D separately compared to group B. I first examine the effect of patient type and second, the previous year's outcome on the probability of adopting the scheme. I use a "dprobit model" to see the effect of patient profile (i.e. mother's age, weeks of gestation, birth weight, number of babies born at the end of a single pregnancy, and gender of baby) and the previous year's c-section rates on the probability of adopting the scheme. A separate analysis is undertaken for each covariate separately. Secondly, I provide a test on whether there are any regional differences in the adoption of the policy across NHS hospitals. Hospitals are classified by NHS regions (south, middle and north) for the 2010/11 taxonomy. I also examine whether there is any effect of being a good performer to see whether "high quality" trusts are more likely to adopt the policy. As a measure of good performance, I use performance scores defined by the NHS Performance Ratings published by the NHS Quality regulator (The Care Quality Commission). NHS trusts are assessed on various aspects of their performance including performance on clinical outcomes, waiting times, patient access, cost effectiveness of health services, health inequalities, population health etc. (NHS Performance Rating 2008/2009, 2009). The rating system is based on 4 point scale, excellent, good, fair and weak respectively (NHS Performance Rating 2008/2009, 2009). I exploit "Overall quality of services score" where trusts are assessed depending on the performance against core standards, existing commitments and national priorities scores (NHS Performance Rating 2008/2009 2009). Trusts with excellent and good overall quality scores are treated as "trusts with good performance" and trusts with fair and weak scores are considered as "trusts with bad performance". The classification of trusts is based on whether or not these three standards are met. Any treatment effect is investigated by a difference in differences method. As I examine only the 2010/11 CQUIN scheme, 2010/11 data are considered as after and 2007/08 to 2008/09 data as before and the 2009/10 data is dropped from the whole analysis. In general, I have three different sets of comparison groups. With the first comparison group, which is group 1, I compare (B vs. C or D). Here, I am testing whether inclusion of maternity as a topic leads to a reduction in c-section rates compared to trusts without maternity as a topic. With comparison group 2, I compare (B or C vs. D). Therefore, I am testing whether being rewarded for the reduction in c-section rates makes a difference in the outcome of interest compared to trusts whose topics are either maternity or another topic and where the indicator is not c-section rate. With comparison group 3, where I look at only those trusts where maternity is a topic and I compare (C vs. D), I am investigating whether inclusion of indicator reduces c-section rates when maternity as a topic. In a seperate analysis, the effect of having more than one maternity indicator is examined. I investigate

whether trusts' performances in terms of c-sections change less or more depending on whether they have one maternity indicator which is c-section rate only, or have other maternity indicators as well as c-section rate. Therefore, the focus is on variation within group D (all trusts in this group have c-section rates as an indicator in their local CQUIN contract). The test is based on the idea that having only one maternity might reduce c-section rate more than the ones with more maternity indicators as well as c-sections since those trusts might allocate more time on c-sections. In a final analysis, I investigate whether there is any effect of mandation of the indicator on c-section rates. Trusts in group D are classified into two sets according to whether they have SHA mandated (SHA wide) or not (SHA suggested/fully flexible) c-sections rates as the indicator in their local contract. I use a difference in differences design which allows me to use different sets of control and treatment groups. First, I estimate an OLS model with and without controls. Second, I provide controls for unobserved time invariant hospital factors which could be correlated with treatment status using a fixed effect model. This accounts for some proportion of omitted variables bias. The fixed effect model is:

$$Y_{it} = \alpha + \beta D_i + \mu(t) + X'\theta_i + \lambda D_i t + \phi_i + e_{it}$$

where Y_{it} is the outcome for trust i, D_{it} is the group dummy indicating whether or not the topic is maternity or c-section rate is an indicator (depending on the comparison we are undertaking), t is the time dummy, X_{it} is average patient characteristics at trust level, $D_i t$ is the interaction between group and time dummy that indicates whether there is an effect of the policy on the outcome, ϕ_i is the trust fixed effect, and e_{it} is the error term.

4.2. Data

The study uses anonymised maternity data from the Hospital Episode Statistics (HES) database. Ethical approval over and above that required for access to anonymised HES records was not sought. HES database contains information on all inpatient, outpatient and A&E admissions for English NHS trusts (Health and Social Care Information website, 2014.[1] The taxonomy data are from Department of Health. The 2010/11 taxonomy is linked to 2007/08, 2008/09 and 2010/11 HES data respectively. The post-policy period only consists of the 2010/11 HES data as later data were not available at the time of the analysis. The HES maternity data provide information from two main resources; (a) babies' records and (b) mothers' records. The main analysis uses information from both records (but mainly from birth records). Outcome variables (rates of elective and emergency c-sections), mother's age, birth weight (kg), number of babies born at the end of a single pregnancy, ethnicity, gender of the baby, mode of delivery and onset of labour are aggregated to the trust level for each financial year from individual patient records. Performance data are used to group trusts as "trusts with good" and "trusts with bad performances". 2007/08, 2008/09 and 2010/11 HES data and the 2010/11 taxonomy data are linked to the overall performance scores for relevant years. Therefore, the classification of trusts is based on their lagged overall performances (*t*-2). Performance data are extracted from the Care Quality Commission database.

4.3. Analyses and Results

Table 1.1. presents descriptive statistics on key variables used in the evaluation of the CQUIN scheme for the 2010/11 taxonomy. The mean proportion of emergency c-sections is higher than the mean proportion of elective c-sections as expected. Table 1.1 also shows a relatively high amount of within trust variance in key variables. This provides support for identification of CQUIN effect by exploiting within trust variation.

I also examine the distribution of regionally mandated (SHA wide), locally flexible or SHA suggested c-sections only or with multiple maternity indicators including c-sections. These are presented in Table A.1 in Appendix. With

---

[1] Hospital Episode Statistics collect data and information on all the inpatient and outpatient NHS hospital admissions in England. (all financial years for Hospital Episode Statistics start on the 1st of April end on the 31st of March) For more details see: http://www.hscic.gov.uk/hes.

regards to the endogeneity of the policy, there is no association between patient characteristics such as number of previous pregnancies, mother's age, weeks of gestation (days), birth weight (kg), number of babies born at the end of a single pregnancy and baby's gender (male) and the probability of adopting the CQUIN scheme (results are available from the author). In addition to patient characteristics, my results also suggest that there are regional differences in adoption of the policy. Hospitals are classified by NHS regions (south, middle and north). The reference group is being located at the south part of England. I find that trusts located in the northern and middle parts of England are more likely to be treated in groups C or D compared to the reference group. I also find that there is no statistically significant association between previous' years outcomes and the treatment status. However, there is some indication for an association between always being a good performer and the treatment status. I find that hospitals who are always good performers are more likely to adopt the policy. This could due to the fact that these hospitals may want to do better on every aspects of maternity care. Tables 1.2. and 1.3. present results for the impact of the CQUIN scheme on elective and emergency c-sections respectively. OLS estimates in Table 1.2, Columns (1) and (2) show no indication of an impact of the policy on rates of elective c-sections. This is consistent across different DiD specifications. These results are robust to the inclusion of trust fixed effects (Columns (3) and (4)). The inclusion of patient characteristics along with other controls does not add much to the precision of estimates (Columns (2) and (4)). Table 1.3 suggests that for 2010/11, after the inclusion of patient characteristics, there is some indication that having maternity as a topic and c-section rate as an indicator decreases emergency c-section rates shown by the negative and significant coefficient in Column (4) for the first comparison group (B vs. C/D, 10% significance level). I test for whether trusts with only one maternity indicator (which is the c-section rate) tend to have reductions in c-sections. In 2010/11, for elective and emergency c-section rates, there is no effect of having only c-section rate as an indicator in the CQUIN scheme. In other words, the number of maternity indicators is not associated with any significant change in c-section rates. I further investigated whether the scheme being SHA mandated (i.e. applying to all trusts in an SHA) has a larger effect than schemes negotiated between trusts and their local commissioners. I therefore split those trusts with c-section rate as an indicator into two groups: those with SHA mandated indicator vs. not SHA mandated (i.e. SHA suggested and full flexible). I find no evidence for such an effect on c-section rates (results are available from the author). Overall, my results suggest that the 2010/11 CQUIN scheme does not lead to any particular improvements in c-section rates. OLS and fixed effects models provide almost similar results; therefore the inclusion of average patient characteristics does not improve the precision of the estimated coefficients for the impact of the CQUIN as the results remain fairly the same

5. Conclusion

This is the first study examining the effect of the 2010/11 CQUIN schemes on caesarean c-section rates using a difference in differences methodology. The 2010/11 CQUIN taxonomy is used along with 2007/08, 2008/09 and 2010/11 HES data and average patient characteristics are included to account for possible observable confounders. Allowing for unobserved heterogeneity at hospital level and average patient characteristics, my results weakly indicate that hospitals tend to reduce the amount of emergency c-sections after the introduction of the CQUIN scheme but not change elective c-sections. Therefore, there is no strong indication for hospitals controlling over elective c-sections. Emergency c-sections are deliveries where mothers have to give birth immediately to reduce risks associated with their babies's as well as their own health status. Therefore, this weak evidence is consistent with the fact that hospitals have less control over the occurrence of this type of deliveries. In the second year of the policy (2010/11), a higher proportion of NHS provider's income was linked to the achievement of locally defined quality targets (0.5% in 2009/10 to 1.5 in 2010/11, Department of Health summary guide 2010). From the commissioners' point of view, this increased the expectation of the achievement of the locally defined CQUIN targets. However, Sutton et al. (2013) provides further information on the fact that concerns over the achievement of goals and targets and the communication between providers and commissioners persisted in the second year. Therefore, this provides a consistent explanation why the 2010/11 CQUIN scheme did not lead to any significant improvements in c-section rates.

Table 1.1. Descriptive statistics (2010/2011 CQUIN taxonomy)

Variable observations	Mean	SD	Between trusts SD	Within trusts SD	Number of observation

2010/11 CQUIN taxonomy					
Elective c-section rate	0.090	0.044	0.033	0.029	291
Emergency c-section rate	0.150	0.132	0.087	0.099	291
Normal birth rate	0.580	0.196	0.137	0.141	291
Number of previous pregnancies	1.058	0.681	0.491	0.473	291
Mother's age	28.97	2.102	1.538	1.439	291
Weeks of gestation (days)	38.77	2.108	1.618	1.358	291
Birth weight (kg)	3.29	0.265	0.176	0.199	291
Number of babies born at the end of a single pregnancy	1.03	0.015	0.011	0.009	291
Male babies (%)	0.509	0.120	0.070	0.097	291
Index of multiple deprivation	25.382	8.722	8.647	1.346	291

The estimation sample consists of 291 NHS acute trusts (97 groups) for 2007/08, 2008/09 and 2010/11 financial years. Group A (N = 15) is excluded. SD indexes standard deviation.

Table 1.2. The impact of the CQUIN on elective c-section rate, 2010/2011 taxonomy

	OLS		**Fixed effect model**	
Groups	w/o controls	w controls	w/o controls	w controls
	(1)	(2)	(3)	(4)
Maternity is the topic: B vs. C|D				
Policy = on	0.008	0.010*	0.008	0.011**
	(0.006)	(0.005)	(0.006)	(0.005)
Trusts = *C|D*	-0.011	-0.003		
	(0.008)	(0.007)		
Policy on among *C|D*	0.012	0.003	0.012	-0.002
	(0.011)	(0.009)	(0.011)	(0.008)
N	291	291	291	291
R^2	0.029	0.462	0.054	0.488
C-section is the indicator: B|C vs. D				
Policy = on	0.014**	0.012**	0.014**	0.013**
	(0.006)	(0.005)	(0.006)	(0.005)
Trusts = *D*	-0.012	-0.009		
	(0.009)	(0.006)		
Policy on among *D*	-0.002	-0.004	-0.002	-0.010
	(0.010)	(0.010)	(0.010)	(0.010)
N	291	291	291	291
R^2	0.035	0.476	0.045	0.493
C-section is the indicator given maternity is the topic: C vs. D				
Policy = on	0.031*	0.015	0.031*	0.028**
	(0.017)	(0.013)	(0.017)	(0.011)
Trusts = *D*	-0.007	-0.012		
	(0.012)	(0.009)		
Policy on among *D*	-0.019	-0.014	-0.019	-0.029
	(0.019)	(0.016)	(0.019)	(0.018)
N	123	123	123	123
R^2	0.082	0.523	0.109	0.658

Table 1.3. The impact of the CQUIN on emergency c-section rate, 2010/2011 taxonomy

	OLS		**Fixed effect model**	
Groups & Years	w/o controls	w controls	w/o controls	w controls
	(1)	(2)	(3)	(4)
Maternity is the topic: B vs. C|D				
Policy = on	-0.033*	-0.017	-0.033*	-0.018
	(0.018)	(0.016)	(0.018)	(0.015)
Trusts = *C|D*	-0.033	-0.041**		
	(0.024)	(0.020)		
Policy on among *C|D*	0.021	0.028	0.021	0.036*
	(0.023)	(0.023)	(0.023)	(0.022)
N	291	291	291	291

R^2	0.018	0.40	0.016	0.607
C-section is the indicator: B|C vs. D				
Policy = on	-0.028*	-0.009	-0.028*	-0.005
	(0.014)	(0.013)	(0.014)	(0.012)
Trusts = *D*	-0.029	-0.046**		
	(0.027)	(0.022)		
Policy on among *D*	0.017	0.018	0.017	0.009
	(0.027)	(0.022)	(0.027)	(0.019)
N	291	291	291	291
R^2	0.014	0.40	0.014	0.601
C-section is the indicator given maternity is the topic: C vs. D				
Policy = on	-0.012	0.002	-0.012	0.007
	(0.020)	(0.018)	(0.020)	(0.013)
Trusts = *D*	-0.008	-0.019		
	(0.032)	(0.023)		
Policy on among *D*	0.001	0.005	0.001	-0.027
	(0.030)	(0.028)	(0.030)	(0.020)
N	123	123	123	123
R^2	0.003	0.660	0.003	0.882

Acknowledgements

I thank my supervisors Carol Propper and Frank Windmeijer for their support and continuous guidance. Matt Sutton, Eleonora Fichera and Simon Burgess deserve my warmest thanks for their comments on my paper.

Appendix

A.1. Distribution of trusts with SHA mandated, SHA suggested and fully flexible c-section rate

Table A.1: Distribution of trusts with SHA mandated, SHA suggested and fully flexible c-section rate given c-section rate is the only indicator/there are multiple maternity indicators as well as c-sections.

Panel A					**Panel B**			
	Groups				**Type of indicator**			
	A	B	C	D		Fully flexible indicator	SHA wide indicator	SHA suggested indicator
2010 taxonomy for 2010/11 HES maternity data								
N	5	56	18	23	*Group 1*	8	0	0
					Group 2	3	9	3

The only group who have c-section rate as an indicator is group D. The distribution is given for the set of trust in the study population for 2010/11 HES maternity data. Groups 1 and 2 are subsets therefore of Group D. Group 1 are trusts with only c-sections as an indicator. Group 2 are trusts with multiple maternity indicators as well as c-section rate. N = number of trusts in each group.

References

Asch, S.M., McGlynn, E.A., Hogan, M.M., Hayward, R.A., Shekelle, P., Rubenstein, L., Keesey, J., Adams, J., & Kerr, E.A. (2004). Comparison of quality of care for patients in the veterans health administration and patients in a national sample. *Annals of Internal Medicine*, *141(12),* 938-W213.

Campbell, S.M., Reeves, D., Kontopantelis, E., Sibbald, B., & Roland, M. (2007). Effects of pay for performance on the quality of primary care in England. *The New England Journal of Medicine*, *357(2),* 181-190.

Care Quality Commission. (2009). NHS performance ratings, an overview of the performance of NHS trusts in England. [Accessed 22nd October 2014]. Available from: http://webarchive.nationalarchives.gov.uk/20100303202142/http://www.cqc.org.uk/_db/_documents/0809_NHS_ratings_overview_document_161009_200910164847.pdf.

Centers for Medicare and Medicaid Services. (2011). Press release: Physician group practice demonstration succeds in improving quality and reducing costs. [Accessed 22nd October 2014]. Available from: http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2011-Press-releases-items/2011-08-08.html.

Cookson, C. & Fleetcroft, R. (2006). Do the incentive payments in the new NHS contract for primary care reflect likely population health?. *Journal of Health Services Research & Policy*, *11(1),* 27–31.

Department of Health (2008) Using the commissioning for quality and innovation (CQUIN) payment framework. London: [Accessed 22nd November 2012]. Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091435.pdf.

Department of Health (2010). Using the commissioning for quality and innovation (CQUIN) payment framework- A summary guide. [Accessed 22nd October 2014]. Available. from:http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/@ps/documents/digitalasset/dh_123008.pdf.

Department of Health. (2010). Using the commissioning for quality and innovation (CQUIN) payment framework – an addendum to the 2008 policy guidance for 2010/11. [Accessed 22nd October 2014]. Available from: http://www.londonsexualhealth.org/uploads/CQUIN%20addendum%20201011.pdf.

Eijkenaar, F., Emmertb, M., Scheppachb, M., & Schöffskib, O. (2012). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy, 110(2-3),* 115-30.

Eggleston K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics*, *24(1),* 211–223.

Epstein, A. M., Rosenthal, M.B., Landon, B. E., Howitt, K., & Songand, H.R. (2007). Climbing up the pay-for-performance learning curve: where are the early adopters now?. *Health Affairs, 26(6),* 1674-1682.

Fichera, E., Gravelle, H., Pezzino, M., & Sutton, M., (2012). Specification of financial incentives for quality in health care contracts the school of economics discussion paper series 1218. *Economics*, The University of Manchester. Available at SSRN: http://ssrn.com/abstract=2196363 or http://dx.doi.org/10.2139/ssrn.2196363.

Doran, T., Kontopantelis, E., Valderas, J.M., Campbell, S., Roland, M., Salisbury, C., & Reeves, D. (2011). Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ. 342.

Ganz, D.A. , Wenger, N.S., Roth, C.P., Kamberg, C.J., Chang, C.T., MacLean, C.H.,Young, R.T, Solomon, D.H., Higashi, T., Min, L., Reuben, B.D., & Shekelle, P.G. (2007). The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences?. *Medical Care, 45(1),* 8-18.

Gillam, S., Siriwardena, A., & Steel, N. (2012). Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework-a systematic review. *Annals of Family Medicine*, *10(5),* 461-468.

Jack, W. (2005). Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, *24(1),* 73–93.

Meacock, R., Kristensen, S., & Sutton, M. (2014). Paying for improvements in quality: recent experience in the NHS in England. *Nordic. Journal of Health Economics. 2(1),* 239-255.

Mehrotra A, Damberg C .L., Sorbero M. E. S., & Telek¸i S. S. (2009). Pay for performance in the hospital setting: What is the state of the evidence?. *American Journal of Medical Quality*, *24(1),* 19-28.

National Health Service England. (2013). The NHS belongs to the people: a call to action. [Accessed 22nd October 2014]. Available from: http://www.england.nhs.uk/wp-content/uploads/2013/07/nhs_belongs.pdf.

Rosenthal, M.B., & Frank, R.G. (2006). What is the empirical basis for paying for quality in health care?. *Medical Care Research and Review*, *63(2),* 135-57.

Kaarboe, O., &, Sicilian, L. (2011). Multi-tasking, quality and pay for performance. *Health Economics*, *20*, 225–238.

Schatz, M. (2008). Does pay-for-performance influence the quality of care?. *Current Opinion in Allergy and Clinical Immunology*, *8(3),* 213–21.

Smith, P. (1995). On the unintended consequence of publishing performance data in the public sector. *International Journal of Public Administration*, *18(2-3),* 277-310.

Smith, P.C., Martin, S., Dusheiko, M., Gravelle, H., &, Rice, N. (2011). Do quality improvements in primary care reduce secondary care costs? The Health Foundation. [Accessed 25th November 2014]. Available from:http://www.health.org.uk/public/cms/75/76/313/2194/Do_quality_improvements_in_primary_care_reduce_secondary_care_costs.pdf?realName=gdrj7N.pdf.

Steel, N., Maisey, S., Clark, A., Fleetcroft, R., & Howe, A. (2007). Quality of clinical primary care and targeted incentive payments: an observational study. *British Journal of General Practice*, *57,* 449-45.

Sutton, M., McDonald, R., Zaidi, S., Todd, S., Konteh, F., Hussein, K., Brown, S., & Kristensen, S. (2013). Evaluation of the commissioning for quality and innovation framework, final report. [Accessed 24th November 2014]. Available from: http://hrep.lshtm.ac.uk/publications/CQUIN_Evaluation_Final_Feb2013-1.pdf.

The National Health Service website. (2015). About the National Health Service [Online]. [Accessed 24th November 2014]. Available from:http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx.

Werner, R., Kolstad, J., Stuart E., & Polsky, D. (2011). The effect of pay for performance in hospitals: lessons for quality improvement. *Health Affairs*, *30(4),* 690-698.