Better for Less
DISCUSSION PAPER SERIES

DP No. 11/2018

Labour Productivity and Skill Mix in Maternity Services: Evidence from the English NHS

Ioannis Laliotis
University of Surrey
Graham Cookson
University of Surrey

School of Economics, Faculty of Arts and Social Sciences,
University of Surrey, Elizabeth Fry Building 04 AD 00,
Guildford, GU2 7XH, Surrey, UK. T: +44(0)01483686623,
E-mail: deliveringbetter@surrey.ac.uk,
Web: www.deliveringbetter.com.

Labour productivity and skill mix in maternity services: Evidence from the English NHS

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

[a]University of Surrey

Abstract

This paper analyses the role of medical and non-medical staff in the production of maternity services in the English NHS. Using hospital panel data (2004-2012) and estimating flexible production functions using system GMM estimators, we explore the output contribution of maternity services labour inputs. The results suggest that consultants and doctors have the highest marginal productivities while the productivity of support workers is insignificantly different from zero. Moreover, there is evidence for some degree of complementarity between midwives, support workers and consultants. Moreover, midwives could replace doctors and doctors could replace consultants in the production of maternity services.

Keywords: Production function; Health; Labour markets

JEL Classification: D24; I10; J24

---

*Address for correspondence: School of Economics, University of Surrey, Guildford, GU2 7XH, UK. E-mail: i.laliotis@surrey.ac.uk. The authors are thankful to Jo Blanden, Heather Gage, Jake Jordan and to participants of the 17th Annual Conference of the International Network for Economic Research for their useful comments and suggestions. The authors acknowledge financial support from The Leverhulme Trust (RL-2012-681). Earlier stages of this work have been also partially funded by the National Institute of Health and Care Excellence (NICE). The views expressed here are those of the authors and do not necessarily reflect those of the funding organisations. The usual disclaimer applies.

1 Introduction

Medical and non-medical staff play a crucial role in the delivery of healthcare services and they are considered essential factors in the production process of health outcomes. So far, several studies have examined the relationship between the healthcare workforce and outcomes, however, the evidence they have provided arises from within acute secondary care settings. A number of reports have identified staffing as a critical component of safe and effective care. The estimated contribution of staffing levels on all reported safety incidents across the English National Health Service (NHS) is around 3.5 percent (National Patient Safety Agency, 2009). The research so far points towards a positive relationship between higher levels of registered nurse staffing and higher quality care (i.e. Jarman *et al.*, 1999; Rafferty *et al.*, 2007; Schubert *et al.*, 2008; Shuldham *et al.*, 2009; Van den Heede *et al.*, 2009). However, the association between higher midwife staffing and reduction in length of stay is not widely supported (Sandall *et al.*, 2014). For example, a meta-analysis by Kane *et al.* (2007), supported such a relationship in the case of surgical patients but not for medical ones.

Literature on the medical workforce is more limited compared to that on nurses and it comes from both the US and the UK settings (i.e Jarman *et al.*, 1999; Pronovost *et al.*, 2002). However, there are significant concerns regarding any causal interpretation of the effects of workforce on outcomes regarding the UK maternity services (Sandal *et al.*, 2014; Cookson *et al.*, 2015). Evidence regarding the optimal levels of staffing for doctors come primarily from the US setting (i.e. Harris *et al.*, 2004; Sucov *et al.*, 2009). The issues of complementarity and substitutability of nurses, midwives and doctors are even less documented, even in large scale studies. Yet, this is important since healthcare outcomes may be sensitive to ratios between nurses and medical staff. For example, Jones *et al.* (2011) showed that higher levels of clinically qualified staff (i.e. doctors and nurses) per bed, and higher ratios of doctors to nurses, were associated with lower mortality-based failure to rescue.

Regarding the maternity care setting, healthcare professionals (Smith and Dixon, 2008) have demonstrated that low staffing levels are widely believed to exert a direct impact on safety of maternity services. The respondents of their survey believed that higher staffing levels would improve a series of quality indicators. However, the empirical evidence is rather scarce since only a few studies have investigated the links between obstetric and midwifery staffing and outcomes. A study by Joyce *et al.* (2004) relied on cross-sectional data from 65 maternity units in a specific UK area for the period 1994-1996 and showed that higher levels of consultant obstetric staff were systematically associated with lower stillbirth rates. A more recent study by Gerova *et al.* (2010) provided supporting evidence of a negative relationship between higher levels of full-time equivalent midwifery staffing and the probability of readmission within 28 days. However, these observational studies have limited capacity, if any, to identify and establish a causal relationship between maternity staffing and outcomes. Hence, there is still much to learn in order to identify the optimum skill mix for quality and productivity. As pointed out in Sandall *et al.* (2014) there is little and inconclusive evidence regarding the effects of changing the workforce skill mix or role substitution on outcomes and costs in maternity care or other settings (i.e. Goryakin *et al.*, 2011). The authors mention that this is particularly important for some worker types for which development programmes are under way, e.g. for support workers.

As mentioned above, the majority of empirical evidence regarding the links between workforce and outcomes in the healthcare sector arises from within the acute care setting. Within the maternity services, the accumulated work to date favours higher levels of registered midwives for better outcomes, and a richer skill mix for both better outcomes and cost effectiveness. Economic evaluations have also suggested that the substitution of nurses for doctors can be cost effective or lead to a net cost reduction. Since medical and non-medical staff are scarce and expensive resources, their optimal use will depend upon whether these resources are complements or substitutes. In the production economics literature there have been two fundamental approaches to answer this question: p-complementarity and q-complementarity (Hicks, 1970; Thurston and Libby, 2002). The evaluation of p-complementarity in healthcare studies is problematic because it requires the estimation of a cost function (i.e. Uzawa, 1962). However, data regarding the input prices of medical and non-medical staff are either unobservable or not often available for all labour inputs (Jensen and Morrisey, 1986a; 1986b; Sandall *et al.*, 2014).[1] On the other hand, q-complementarity can be investigated via the estimation of a production

[1] Thurston and Libby (2002) discuss in brief some alternatives for the estimation of cost functions.

function, where productivity is measured by the volume of treated cases or the length of stay. Thurston and Libby (2002) note that since the study of Reinhardt (1972), almost every paper that has examined the issue of productivity or efficiency in the healthcare sector has adopted the production function approach. Analysing the production of primary care visits, they showed that nurses are q-complements for physicians, while technicians and unqualified nurses are q-substitutes for nurses.

From an economic perspective, despite the crucial role of skill mix and the changing composition of the workforce over the last years, little is known with respect to the complementarity or substitutability of staff groups within the English NHS. Sandall *et al.* (2014) mention the lack of examples regarding these issues from within any acute care setting. They also point out that according to the available evidence, the existence of a general relationship between skill mix and productivity which can be generalized across different care settings is unlikely to hold. Focusing on the NHS maternity setting for the year 2010-11 and adopting a production function analysis, they found that registered midwives are q-complements with doctors and consultants but q-substitutes with support workers in the total number of deliveries. Also, consultants were found to be complements with support workers but substitutes with doctors. However, they advise caution given that, like many other studies, they rely on cross-sectional data that deem any attempt of causal inference problematic.

This paper revisits the relationship between workforce and maternity outcomes in the English NHS in an attempt to contribute knowledge to an important policy question for which there has been a paucity of research.[2] The main objective is to try to address some of the drawbacks of previous studies by using richer sources of data and adopting an estimation strategy in order to tackle the issue of endogeneity. More specifically, we estimate generalized linear production functions in the spirit of Diewert (1971), in order to examine the relationship of both staffing levels and skill mix with the total number of maternities. However, unlike the hitherto presented cross-sectional studies, we utilize a panel dataset at the hospital level covering the period between 2004 and 2012 and employ the dynamic panel estimation framework proposed by Arellano and Bond (1991) and Blundell and Bond (1998). Based on the results obtained from

[2] An early attempt to work with panel data was recently made by Cookson *et al.* (2015).

system GMM regressions we find that the estimates often reported in cross-sectional studies can be misleading. According to our results, consultants and doctors have the highest marginal productivities, while the productivity of support workers has a negative sign and it is not statistically significant. Moreover, we present evidence for some degree of complementarity between registered midwives and support workers and consultants. According to our results, midwives can replace doctors and doctors can replace consultants in the production of maternity services.

The remainder of the paper is organised as follows: Section 2 presents the data sources and some preliminary descriptive analysis. Section 3 outlines the adopted econometric methodology the results of which are discussed in Section 4. Section 5 concludes.

2 Data

Several data sources were linked together for the analysis, i.e. the Hospital Episodes Statistics (HES) database for the years 2003-2013, the Office for National Statistics (ONS) Birth Registration Records for the years 2004-2012 and the Medical and Non-Medical Workforce Censuses for the years 2004-2013 provided by the Health and Social Care Information Centre (HSCIC). HES is a pseudo-anonymous patient level administrative database containing details of all admissions, outpatient appointments and Accident and Emergency attendances across the NHS. Each HES record contains details of a single consultant episode: a period of patient care overseen by a suitably qualified healthcare professional. The unique patient identifiers in the HES records help to append relevant information from previous deliveries and compile a more complete picture of a woman's obstetric history, e.g. the number of live births that a woman has had (parity). Other patient-level variables extracted from HES are maternal age, ethinicity, socioeconomic deprivation as measured by the Index of Multiple Deprivation (DCLG, 2011) and clinical risk at the end of the pregnancy as measured by the National Institute for Health and Care Excellence guideline for intrapartum care (NICE, 2007). We adopted the method developed by Sandall *et al.* (2014) and exploited the rich clinical history available in HES records to identify women with "higher risk" pregnancies due to pre-existing medical conditions, complicated previous obstetric history or conditions that develop during the pregnancy.

These women may have different maternity outcomes from women classified as at "low risk". The extracted HES variables are considered to partially explain the variation in the outcomes between mothers and for the purposes of the analysis they have been aggregated to the trust level. As the composition of mothers (i.e. the case-mix) varies across trusts, their omission could lead to confounding variation between the service user population and the service itself. For example, since clinical risk is an important predictor of maternity outcomes, excluding it from the analysis would appear to show trusts with higher shares of high risk women to have worse outcomes. Moreover, as these case-mix variables can be correlated with the staffing variables, their omission could induce an endogeneity bias. Other important explanatory variables such as smoking status, drug or alcohol consumption and maternal obesity are not available. However, as they are correlated with a number of co-morbidities and conditions included in the clinical risk variable and because they are unlikely to be correlated with the staffing variables, their omission is not expected to bias the results. These are the same variables as used in Sandall *et al.* (2014) with the exception of the service configuration variable (e.g. midwifery-led unit). However, service configuration was not found to be a statistically significant predictor of outcomes and the panel structure of the dataset will account for the effects of time-invariant characteristics.

The trust level dataset was assembled from three distinct sources. The HSCIC provided staffing data for English trusts under a Data Sharing Agreement. The staffing data were Full Time Equivalent (FTE) members by maternity worker type. Data provided for 2004 to 2013 are taken from the Non-Medical Workforce Census as at 30 September in each specified year. NHS Hospital and Community Health Service (HCHS) medical staff in Obstetrics and Gynaecology by organisation and grade are taken from the Medical Workforce Census as at 30 September in each specified year. Output is measured by the number of maternities contained in the Birth Registration Records data provided by the Office for National Statistics (ONS). Alternatively, the total delivery count per trust each year as extracted from the HES data could be used as the output variable. However, the correlation coefficient between these two variables was found to be remarkably high (0.97) while the results of our analysis remain the same using either of the two.

Throughout the analysis the decision making unit is the hospital trust. Merging all datasets resulted in a trust-level panel covering the period between financial years 2004/05 and 2012/13. Merging the data resulted in an unbalanced panel dataset of 352 distinct providers for 10 years, with 228 of them were observed in every year. However, after keeping in the sample only secondary care NHS trusts we were left with 161 providers, with 133 of them being observed every year.[3] Table 1 presents some basic descriptive statistics for the variables used in the analysis, by year and for the total period. Regarding the output variable, the sample mean is 4,376 maternities and the standard deviation is 2,092 indicating a large degree of variation across providers, while the demand for the service has been steadily increasing throughout the period.

[Table 1 about here]

With respect to the staffing data, the main focus is on the following four categories: registered midwives (RD), support workers (SW), consultants (C) and all other doctors (D). The last two categories are considered separately in order to examine their substitutability and complementarity with the rest of the labour input types. Registered midwives are clearly the largest group with a mean FTE of 129.04, followed by doctors (22.99), consultants (10.43) and support workers (5.36). The mean FTE of support workers may seem small, however, their use has been following a steadily upward trend, from a mean FTE of 3.62 in 2004/05 to a mean FTE of 7.57 in 2012/13. The evolution in the use of doctors and consultants has increased more modestly while the mean FTE of registered midwives has been increased from 117.26 in 2004/05 to 147.65 in 2012/13. Moreover, Table 2 displays how the profile of the population has evolved over the period. The increase in the proportion of mothers classified as "high risk" ones (from 41% in 2004 to 53% in 2012) could be partly explained by the improvement in the level of clinical coding of particular conditions or procedures that would render a woman at higher risk of experiencing a difficult delivery. The age profile has increased rather slightly while the mean parity at the trust level has increased from 0.83 in 2004 to 0.99 in 2012.

[3] Primary care trusts, mental health trusts and private providers were excluded from the dataset in order to eliminate confounding errors (primary care trusts often provide community-based midwifery care which could distort the representation somewhat).

3 Econometric methodology

In healthcare studies,

Skill mix is an important topic, specifically regarding the extent to what staff groups and professions are substitutes or complements. Understanding the relationships between staff groups is important for optimising the healthcare workforce to maximise the amount of work that can be done. Changes in healthcare staffing in recent years have implicitly assumed that staff groups are substitutes, at least for certain tasks. Production economics can be used to test whether this assumption is correct and provide important insights regarding the optimal skill mix for maternity services.

In economics, a production function describes the mechanism for converting a vector of inputs (e.g. midwives and doctors) into output which, in this study, is the annual number deliveries within a hospital. Estimating the parameters of an appropriately chosen functional form allows the output elasticities to be calculated and the returns to scale to be found. The output elasticity measures the responsiveness of the annual number of deliveries to a change in the amount of input (e.g. maternity staff). Given the absence of data on input prices, a production (i.e. quantity) function approach has been adopted. Many healthcare studies using production functions (as opposed to cost functions) have adopted Reinhardt's (1972) specification of the production function, which was the first to include multiple labour inputs (registered midwives, technicians, administrative staff and doctors). However, as outlined by Thurston and Libby (2002), given the total absence of any cross-products, this function assumes all inputs to be substitutes and discounts the possibility that different staff groups could be complements. The advance in production function analysis of the 1970s gave rise to two flexible econometric specifications that allow researchers to relax this overly strict assumption. Berndt and Christensen (1973) introduced the transcendental-logarithmic (translog) production function and Diewert (1971) introduced the generalized linear production function (also known as the Allen, McFadden and Samuelson production function). Early research assumed that hospital production can be approximated by the more traditional Cobb-Douglas technology (i.e. Pauly, 1980). However, later studies showed that assuming a Cobb-Douglas technology resulted in biased estimates, hence favouring the use of more flexible functional forms such as the translog (i.e.

Jensen and Morrisey, 1986a; 1986b).

Using either of these functions would have allowed us to estimate the relationship between the labour inputs because the regression coefficient on the cross-products (interaction effects) can be simply used to calculate the Hicks (1970) elasticity of complementarity (Sato and Koizumi, 1973; Syrquin and Hollender 1982). However, an advantage of the Diewert (1971) specification is that it allows zero quantities for some inputs. This is a more realistic assumption, especially when labour inputs are disaggregated as in this study. This modelling enabled us to examine the output contribution of the different staff inputs (output elasticities) and their influence upon the productivity of other staff inputs (i.e. whether they are complements or substitutes). With these results available, we were able to investigate the input substitution possibilities available to hospitals under different scenarios.

The adopted generalized linear production function is of the following form:

$$Q = F(X) = F(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{K}\sum_{j=1}^{K} \alpha_{ij}\sqrt{X_i}\sqrt{X_j} \tag{1}$$

where $K = 4$ (i.e. $X = C, D, RM, SW$) and Q corresponds to the total annual number of deliveries within a given NHS trust. The marginal product of each labour input is simply the partial derivative of the specified production function with respect to that specific input. To examine the issue of q-complementarity and therefore to answer the question relating to skill mix, we calculated the Hicks (1970) elasticity of complementarity, η^H, defined for any two staffing inputs i, j, $(i \neq j)$:

$$\eta_{ij}^{H} = \frac{f f_{ij}}{f_i f_j} \quad \forall \quad i \neq j \tag{2}$$

where

$$f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \tag{3}$$

The marginal productivities for each labour type as well as the elasticities of complementarity are evaluated at the sample means. The total annual number of maternities within a hospital as the output measure as well as the Diewert (1971) generalized linear production function have also been recently adopted by Sandall *et al.* (2014) while modelling the maternity services output in the English NHS. However, instead of using a single cross-section, we constructed a panel dataset at the trust level in order to control for time effects as well as for unobserved heterogeneity at the trust level. Moreover, the panel data structure may alleviate some sources of endogeneity for the key variables of interest, i.e. the staffing groups variables and their interactions. However, given the absence of any strictly exogenous explanatory variables that could serve as instruments while establishing causal relationships between the regressors of interest and the outcome variable, an alternative method in order to exploit the time dimension of the dataset was chosen instead. Hence, the identification strategy will rely upon limited serial correlation in the error term of the model by adopting a Generalized Method of Moments (GMM hereafter) estimator which is widely used in applications with panel data (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). Moreover, we experimented with a dynamic econometric specification of the generalized linear production function. The insertion of a lagged dependent variable into the model accounts for the past behaviour of hospital trusts with respect to the total number of maternities. As discussed in Bond (2002), controlling for dynamics can lead to more consistent estimates of the other parameters that may interest, even if the dynamics themselves are not the primary focus of the analysis. Indeed, controlling for inertia in the production of health services is not something common as most studies adopt static expressions, mostly due to data limitations. Yet, the previous year's actual level of output in the maternity services should be accounted for as it may be correlated with both the current level of output and the current staffing levels. Moreover, the total number of deliveries is most likely to evolve gradually over time in response to shocks, e.g. increased pressure due to immigration. Therefore, its omission can lead to biased estimates of the production function parameters. Including a lagged dependent variable as an additional regressor makes the estimated model take the following form:

$$Q_{it} = \alpha_0 + \alpha_1 Q_{it-1} + \alpha_2 X_{it} + \alpha_3 C_{it} + \delta_i + \lambda_t + \epsilon_{it} \tag{4}$$

where Q_{it} is the total number of maternities in hospital the i-th hospital during the t-th year and X_{it} is a vector containing the staffing level variables and their cross-products as those are derived from equation (1). Obtaining consistent estimates of their parameters will help us in examining the issues of substitutability and complementarity between various combinations of labour inputs. C_{it} is a vector containing time-varying case-mix controls which can be correlated with the total number of deliveries in each hospital and λ_t stands for common time effects across hospitals. Last, δ_i and ϵ_{it} denote the hospital-specific and the idiosyncratic components of the error term which are assumed to be independently distributed across hospitals with $E(\delta_i) = 0$, $E(\epsilon_{it}) = 0$, $E(\delta_i \epsilon_{it}) = 0$ and $E(\epsilon_{it}\epsilon_{is}) = 0$ for every $t \neq s$.

Estimating a dynamic version of Equation (4) using Ordinary Least Squares (OLS) as many related studies do, will lead to rather biased estimates. This is because the staffing variables, the effects of which we are interested in estimating, are very likely to be correlated with the error term in the maternities production function ($E(L_{it}\epsilon_{it}) \neq 0$).[4] Applying an appropriate transformation using the Within (or Least Squares Dummy Variable, LSDV) estimator would address the problem arising from the correlation between the staffing level variables (and their cross-products) and the hospital-specific time invariant component of the disturbance term. However, given the short time dimension in our panel, this will lead to a considerable dynamic panel bias of order $1/T$ (Nickell, 1981).[5] Moreover, the within estimator will treat the staff-related variables as strictly exogenous, while in fact, a plausible assumption is that they are correlated with present or past shocks in the maternities equation ($E(L_{it}\epsilon_{it}) \neq 0$). Nonetheless, both the OLS and the Within estimators are useful starting points for our empirical analysis, despite the fact that the main concern is to address the endogeneity issues which are present in other cross sectional studies. More specifically, these two estimators can provide the value range where a good estimator of the lagged dependent variable should lie within or be near to it (Bond, 2002).

It is common in the empirical literature to employ difference and system GMM estimators to estimate equations like (4) using small *T*, large *N* panel data. For example, Blundell and Bond (2000) applied GMM estimators to Cobb-Douglas production functions using panel data

[4]Earlier studies (i.e. Jensen and Morrisey, 1986a;1986b) have discussed the potential problems arising from a possible correlation between the inputs and the disturbance term. Moreover, they mention a number of studies which reported negligible biases, if any, after applying OLS estimators using cross-sectional data. Jensen and Morrisey (1986b) describe a model according to which inputs and the disturbance term are independent, justifying the use of the OLS estimator.

[5]Another strategy is to instrument the autoregressive term using its own past realizations. However, these are also correlated with the error term, while the *T* dimension will get even smaller leading to considerably biased estimates.

covering a small number of time periods for a large sample of manufacturing companies. These estimators can address several important modelling issues like fixed effects, the endogeneity of multiple regressors, while avoiding dynamic panel bias, and they can be easily implemented in unbalanced panels (Roodman, 2009a). However, the difference GMM estimator suffers from potentially large finite sample bias in the cases where some persistency is present in the series and when the variance of the time invariant group-specific effect increases with respect to the overall variance of the error term. (Alonso-Borrego and Arellano, 1995; Blundell and Bond, 2000; Soto, 2009). More specifically, if the instruments employed in the first-differences equations are weak, then the difference GMM estimator will be biased towards the Within estimator (Blundell & Bond, 2000).

Blundell and Bond (1998) showed that such biases can be reduced if more informative moment conditions are incorporated. This is possible by using lagged first-differences as instruments for the levels equation in addition to the lagged levels as instruments for equations in first-differences. Therefore, the system GMM estimator, as introduced by Arellano and Bover (1995) and Blundell and Bond (1998), is often used to circumvent the finite sample bias, given some reasonable stationarity assumptions. However, any efficiency gains stemming from the additional orthogonality conditions imposed by the system GMM estimator, come with the cost of instrument proliferation: The number of instruments increases exponentially with the number of time periods which leads to finite sample bias and increases the likelihood of false positive results as well as suspiciously high pass rates of specification tests like the Hansen J-test (Roodman, 2009a). Hence, we follow the suggestions of Roodman (2009a; 2009b) and apart from performing various sensitivity tests, we also collapse the instrument matrix. This ensures that the instrument count grows linearly in T. We report estimates on both the one- and two-step variants of the GMM estimator. The two-step method produces coefficient estimates that are asymptotically more efficient, however the estimated standard errors tend to be downwards biased (Arellano and Bond, 1991; Blundell and Bond, 1998). Hence, we apply the Windmeijer (2005) small sample correction regarding the standard errors in the two-step GMM estimators.

4 Results

The presentation of our results begins with Table 3 which reports some basic OLS estimates of the specified production function. Apart from the various labour inputs, namely doctors, consultants, registered midwives and support workers as well as their cross-products, the vector of explanatory variables is gradually augmented with average maternal characteristics calculated by trust and year (i.e. average maternal age, mean parity, proportion of high risk women, proportion of women with British background, proportion of women from the most deprived areas), year and regional fixed effects (as captured by the nine SHA binary indicators with North East being the reference Authority) in order to assess the sensitivity of the results to different model specifications. These year and regional fixed effects help in controlling for factors which are common across trusts for each year (e.g. a common technological shock) and for each SHA region, respectively. Finally, a lagged dependent variable is also inserted into the model in order to account for the past behaviour of hospital trusts with respect to the total number of maternities. The relevant literature has ignored the importance of dynamics since most of the studies rely on cross-sectional evidence, however, controlling for dynamics can remove some bias from the estimated coefficients of the other explanatory variables. For example, the number of deliveries in previous periods may affect the decision of a trust regarding the current staffing levels in its maternity services, therefore the inclusion of dynamics is deemed necessary. Throughout the estimations the standard errors have been corrected for clustering at the trust level in order to account for any unobserved factors which cannot be attributed to the explanatory variables. Hence, we avoid inflating the precision of the estimated parameters by controlling for common error components within trusts.

According to the results, the models have a quite high explanatory power ranging between .82 and .90. The inclusion of the lagged dependent variable brings about a great improvement in the results, given that its inclusion seems to deflate considerably almost all the estimated coefficients of staffing level variables and their interactions. Moreover, the results indicate a significant degree of persistency regarding the annual number of deliveries at the trust level (0.552). However, these OLS estimates may be biased due to time-invariant unobserved heterogeneity at the hospital level. Hence, Table 4 reports the results after applying the Within estimator to the same set of empirical specifications. Regarding the staffing level variables, the results remain the same except for the FTE support worker variable which now always enters with a negative sign. The same happens to the cross-products between midwives and consultants, support workers and doctors and consultants and doctors. Moreover, the coefficient of the lagged dependent variable is now much lower (0.126) than the one estimated using OLS. However, these two values provide the bounds within which a credible estimate should lie. Moreover, the fact that the within and the overall R-squared values are quite far from each other, indicates that individual (i.e. hospital) effects have to be controlled for while modelling the outcome of maternity services. Similar to the results presented by Sandall *et al.* (2014) and Cookson *et al.* (2015), despite the high explanatory power of all models, very few parameters are found to be statistically significant. Sandall *et al.* (2014) argued that one possible cause for that is the relatively low degrees of freedom since they had only 141 observations (i.e. trusts) available to fit 15 parameters. Given the fact that here we report standard errors which have been adjusted for 154 clusters, our degrees of freedom are not much more.[6] However, even if the degrees of freedom are increased by choosing another clustering scheme (e.g. at the regional level) the results do not change with respect to the statistical significance of the estimated parameters. The major reason is that given the functional form of the Diewert (1971) production function, the model suffers from multicollinearity due to the inclusion of the cross-product terms. Indeed, for most of the staffing level variables and the interaction terms, the variance inflation factor (VIF) was considerably larger than 10 but in any case not as large (in thousands) as in Sandall *et al.* (2014).[7] However, given the theoretical importance of the variables, omitting some of them or attempting a transformation of any kind is ruled out for the rest of the empirical analysis.

Despite the fact that all models appear to have a high explanatory power and the inclusion of a lagged dependent variable has a significant contribution, the estimated regression coefficients are rather unhelpful in examining the impact of staffing level and skill mix on the total output. Instead, the elasticities of substitution and complementarity can be more informative.

[6]However, gathering more data than those of the single cross-sectional study of Sandall *et al.* (2014) seems to have contributed in getting smaller standard errors.

[7]The results are available upon request from the authors.

The marginal productivities can be calculated using the estimated regression coefficients and the sample means from the estimation sample and they indicate the number of additional deliveries that would be expected, on average, if the FTE of a particular staffing group is marginally increased, *ceteris paribus*. More specifically, given the generalized linear production function we employed, the following formula was used in order to obtain the estimated marginal productivities for the i-th medical or non-medical type of labour:

$$\text{Marginal productivity}_i = \alpha_i + \frac{1}{2}\sum_{j=2}^{K} \alpha_{ij}\sqrt{\frac{X_j}{X_i}} \tag{5}$$

where α_i and α_{ij} are the estimated coefficients regarding the i-th type of labour and its interaction with the j-th type of labour respectively, while X_i and X_j are the respective estimation sample means for the i-th and the j-th types of labour. Based on the estimated results for the production function and the estimation sample, we can also calculate the Hicks (1970) elasticities of complementarity between the different staffing groups in the production function of deliveries within a given hospital trust each year. A positive elasticity indicates that the two labour inputs are complements (i.e. can be used together) while a negative elasticity indicates that the two staffing groups are substitutes (i.e. one can be used in the place of another). These elasticities can be derived using the formula described by equation (2), where the partial derivative f_{ij} is given by:

$$f_{ij} = \frac{\alpha_{ij}}{4\sqrt{X_i}\sqrt{X_j}} \tag{6}$$

However, both the OLS and the Within estimator treat the variables regarding the staffing levels and the skill mix as strictly exogenous. This can be a rather unrealistic assumption since the staffing levels of a hospital can be correlated with present and past shocks in the maternities production function. Therefore, any attempt to calculate elasticities of substitution and complementarity would rely on biased estimates. The difference and the system GMM estimates displayed in Table 5 are an attempt to deal with this endogeneity issue in the absence of any external instruments. In all models, the lagged dependent variable is treated as predetermined and the case-mix controls are treated as strictly exogenous. All the models use lagged levels dated $t-2$ and back as instruments while the instrument matrix is collapsed in order to avoid instrument proliferation and the suspiciously high pass rates in the diagnostic tests (Roodman, 2009a). Regarding the difference GMM estimates (columns [1] and [2] of Table 5), the coefficient of the lagged dependent variable is biased towards the direction of the Within estimator and, in fact, falls outside the credible range mentioned before. This may be due to a weak instruments problem in the sense that lagged levels of the explanatory variables do not serve well as instruments for their own differences. On the other hand, the system GMM estimates (columns [3] and [4] of Table 5) seem to be more reasonable. The coefficient of the lagged dependent variable now falls within the credible range bounded by the OLS and the Within estimators. Regarding the coefficients of the staffing levels and skill mix, there are some notable differences as compared to the estimates assuming strict exogeneity, especially the ones for FTE doctors, FTE consultants and the interaction term between FTE midwives and FTE consultants, which now has a negative sign. The diagnostic tests tend to support the model specification, especially for the two-step estimates where the null hypothesis on second order autocorrelation is rejected.

Tables 6 and 7 report the results on some robustness checks regarding the system GMM estimates. More specifically, the only difference is that in Table 6 we use lagged levels dated $t-3$ and $t-4$ and earlier as instruments. The coefficients for FTE consultants and their cross products with FTE doctors have the same sign as before, however, they are significantly lower. Moreover, the model specification is again supported from the diagnostic tests, although there is marginal evidence of second order autocorrelation in the error term regarding the one-step estimators. In Table 7, we relax the assumption regarding the exogeneity of the case-mix controls and we treat them as endogenous while we use lagged levels dated $t-2$ and $t-3$ as instruments. The results do not alter substantially except from the fact that support workers now enter with a negative sign. The rows for the Hansen J-test report the p-values regarding the null hypothesis on the validity of the overidentifying conditions which is not rejected in any model. The rows for the Difference-in-Hansen test display the p-values regarding the validity of the additional moment restrictions which are necessary in the cases of the system GMM estimations. Expect from the case where the case mix controls are treated as endogenous and lagged levels dated $t-3$ and back are used as instruments (columns [3] and [4] of Table 7; although the p-value is relatively high), the additional moment conditions appear to be valid. Hence, in most of the cases a proper specification is indicated by the test statistics.

Table 8 reports the marginal products and the elasticities of complementarity between the various staffing groups in the production of maternities. Marginal products indicate the change in output that would occur from a marginal change in each labour input, *ceteris paribus*. Regarding the Hicks elasticities, a positive elasticity indicates that the two staffing groups are complements and hence can be used together in the production of deliveries, while a negative elasticity indicates that the two staffing groups are substitutes in the sense that one can be used in the place of another in the production process. We report marginal products and elasticities of complementarity based on the regression parameters of the production function obtained from various estimation techniques. The reported standard errors have been obtained via the delta method. More specifically, the results Panel A of Table 7 are based on the OLS estimates (column [4] of Table 3), those of Panel B are based on the Within estimates (column [4] of Table 4), those of Panel C have been derived from the GMM regression coefficients displayed in column [4] of Table 5 and those reported in Panel D are obtained using the parameter estimates from column [2] of Table 7.[8]

Regarding the results based on OLS estimates, all the marginal productivities are positive indicating that increasing the staffing level of any medical or non-medical group would lead to an increase in the total number of maternities that the provider could handle. For example, adding an additional FTE registered midwife would allow a hospital to produce an additional 9.5 deliveries on average each year. Regarding the rest staffing groups, doctors have the largest marginal productivity (17 additional deliveries), followed by consultants (12.2), registered midwives and support workers (1.4 additional delivery). However, the marginal productivities are not statistically significant for consultants and support workers while they are significant at the 1% level for doctors and registered midwives. These figures are well below those reported in Sandall *et al.* (2004). As shown by Cookson *et al.* (2015) this comes mostly from the inclusion of the autoregressive term into the model. Of the six possible combination of maternity

[8]We have also calculate the marginal products and elasticities of complementarity based on several other model specification reported in earlier tables. The results are available upon request from the authors.

labour inputs, three of them are complements (i.e. consultants and midwives, consultants and support workers, doctors and support workers) and three of them are substitutes (i.e. midwives and support workers, midwives and doctors, consultants and doctors). This makes clear the advantage of the flexible form of the Diewert (1971) production function which allows some of the staffing groups to be complements rather than posing any *a priori* restrictions forcing all of them to be substitutes. Regarding the results based on the OLS parameter estimates, consultants are quantity substitutes with doctors, with the magnitude of the elasticity of complementarity being close enough to those reported in Sandall *et al.* (2014) and Cookson *et al.* (2015). This makes sense since their work is most likely to overlap. Likewise, registered midwives are also found to be quantity substitutes with support workers and doctors. This was also the case in Cookson *et al.* (2015) but not in Sandall *et al.* who estimated a positive elasticity for doctors and midwives. For the rest of the staffing combinations, there is strong evidence that they can be used together in the production of deliveries.

Regarding the Within estimates (panel B of Table 8), a striking difference with the results presented so far in this study and elsewhere, is that the marginal productivity for support workers enters with a negative sign, indicating that increasing their staffing level within a provider would result in a smaller number of deliveries on average per annum; however this result is not systematically different from zero. This could be important given that developing plans for this labour input are under way across the UK. Another difference is that now consultants appear to have the largest marginal productivity (53 additional deliveries per year) which now is statistically significant at the 1% level, and they are followed by doctors (14.8 additional deliveries) and registered midwives (4.3 additional deliveries). Regarding the results on the elasticities of complementarity, they are not in total accordance with those of Panel A. More specifically, support workers are found to be quantity complements with midwives and doctors while in every other combination there is evidence for substitutability of some degree.

The negative sign for the marginal productivity of support workers remains when the results are derived using the parameter estimates obtained from the system GMM regressions (panel C of Table 8), however it is still insignificantly different from zero. Regarding the productivities of the other staffing groups, the pattern remains the similar to that of panel B, i.e. consultants have the highest marginal productivity (66.1 additional deliveries), followed by doctors and midwives (48.5 and 6.5 additional deliveries, respectively) and they are all statistically significant at conventional levels. Doctors are again found to be quantity substitutes with registered midwives and consultants. Moreover, the Hicks elasticity between doctors and consultants is now found to be negative and statistically significant at the 10% level. Midwives seem to be quantity complements with consultants and support workers, addressing some concerns raised by Sandall *et al.* (2014) about the quality of care (conditional on the groups of women involved or the care setting) if the latter could substitute for registered midwives. Strikingly, as in the case of the Within estimates, consultants are substitutes with support workers. In Panel D where the case-mix controls have been treated as endogenous (column [2] of Table 7), the results are similar to those of Panel C, i.e. support workers have a negative and non-significant marginal product where those for the other types of maternity workers are positive and highly significant. Regarding the Hicks elasticities, there is further evidence that doctors and consultants are quantity substitutes in the delivery of maternity services.

5 Conclusions

This paper attempted an evaluation of the relationship between outcomes and workforce in the English NHS maternity services. The contribution of labour inputs in the production of healthcare visits has been recognized by many studies, although the evidence so far regarding this relationship was coming either from papers that generalized across care settings or from maternity-specific studies which used cross-sectional datasets. Hence, there is still lot to learn regarding the optimum staffing levels and skill mix for productivity.

In this paper we attempted to overcome some of the limitations often spotted in the relevant literature by developing a panel dataset at the trust level covering the period 2004-2012. More specifically, we matched information by linking several data sources, i.e. the Hospital Episode Statistics dataset, the ONS Birth Registration records, and the Health and Social Care Information Centre Medical and Non-Medical Workforce Census. Adopting a generalized linear production function specification (Diewert, 1971) and exploiting the time dimension of our dataset, we tried to go further than the available studies by estimating dynamic panel data models which would address some of the endogeneity issues from which other studies suffer. Our analysis, indicated that accounting for dynamics seems to be a very important issue since it contributed in moderating the estimated coefficients of the explanatory variables indicating both the staffing levels and the skill mix. Moreover, based on the results obtained from system GMM regressions which deal with the endogeneity issue by using appropriate lagged values of the explanatory variables in levels and in differences, we found that the estimates often reported in cross-sectional studies can be misleading. This may cause the derived marginal productivities and elasticities of complementarity and substitution to be severely biased. Based on the system GMM results of a preferred specification, we found that consultants and doctors have the highest marginal productivities, while the productivity of support workers is not statistically different from zero. Moreover, we presented evidence for some degree of complementarity between registered midwives and support workers and consultants. According to our results, midwives could replace doctors and doctors could replace consultants in the production of maternity services.

Future research should be taken in order to improve the knowledge obtained so far. Better data collection of ward-level staffing data that could be linked to patients and their outcomes is of great importance. The existing research so far is severely limited by the use of aggregated staffing data. Designing the implementation of a Maternity Safe Staffing guideline (NICE, 2015) or other staffing intervention in such a way as to create a quasi or natural experiment would be invaluable. At present any causal links between staffing and outcomes should be drawn with caution and the best available evidence does not support a strong relationship between these variables. An experimental design would enable researchers to answer the question conclusively by randomising the numerous omitted variables.

References

[1] Alonso-Borego, C. and Arellano, M. (1999), "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", *Journal of Business & Economic Statistics*, 17(1): pp. 36–49.

[2] Arellano, M. and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58(2): pp. 277–297.

[3] Arellano, M. and Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Component Models", *Journal of Econometrics*, 68(1): pp. 29–51.

[4] Berndt, E. and Christensen, L. (1973), "The Translog Function and the Substitution of Equipment, Structures and Labour in U.S. Manufacturing", *Journal of Econometrics*, 1(1): 81–114.

[5] Blundell, R. and Bond, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87(1): pp. 115–143.

[6] Blundell, R. and Bond, S. (2000), "GMM Estimation With Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, 19(3): 321–340.

[7] Bond, S. (2002), "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice", *Portuguese Economic Journal*, 1(2): 141–162.

[8] Cookson, G., Jones, S., van Vlymen, J. and Laliotis, I. (2015), "The Cost-Effectiveness of Midwifery Staffing and Skill Mix on Maternity Outcomes", A Report for the National Institute for Health and Care Excellence.

[9] Diewert, W.E. (1971), "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function", *Journal of Political Economy*, 79(3): pp. 481–507.

[10] 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.

[11] Gerova, V., Griffiths, P., Jones, S. and Bick, D. (2010), "The Association Between Midwifery Staffing and Outcomes in Maternity Units in England: Observational Study Using Routinely Collected Data", Preliminary Report and Feasibility Assessment, London: National Nursing Research Unit.

[12] Harris, D.A., Vernon, J.P., Sidman, R.D., Sucov, A., Aarbelaez, C. and Monks, A. (2004), "Is There a Doctor in the House? Comparing Attendings, Residents and Midlevel Providers to Increase Clinician Staffing in an Academic Emergency Department: A Cost-Efficient Analysis", *Annals of Emergency Medicine*, 44(4, suppl.): pp. S18–S19.

[13] Hicks, J.R. (1970), "Elasticities of Substitution Again: Substitutes and Complements", *Oxford Economic Papers*, 22(3): pp. 289–296.

[14] Jarman, B., Gault, S., Alves, B., Hider, A., Dolan, S., Cook, A., Hurwitz, B. and Iezzoni, L.I. (1999), "Explaining Differences in English Hospital Death Rates Using Routinely Collected Data", *British Medical Journal*, 318: pp. 1515–1520.

[15] Jensen, G.A. and Morrisey, M.A. (1986a), "Medical Staff Specialty Mix and Hospital Production", *Journal of Health Economics*, 5(3): pp. 253–276.

[16] Jensen, G.A. and Morrisey, M.A. (1986b), "The Role of Physicians in Hospital Production", *Review of Economics and Statistics*, 68(3): pp. 432–442.

[17] Jones, S., Bottle, A. and Griffiths, P. (2011), "An Assessment of "Failure to Rescue" Derived from Routine NHS Data as a Nursing Sensitive Patient Safety Indicator for Surgical Inpatient Care", National Nursing Research Unit, King's College London.

[18] Joyce, R., Webb, R. and Peacock, J.L. (2004), "Associations Between Perinatal Interventions and Hospital Stillbirth Rates and Neonatal Mortality", *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 84: pp. F51–F56.

[19] Kane, R., Shamliyan, T., Mueller, C., Duval, S. and Wilt, T. (2007), "The Association of Registered Nurse Staffing Levels and Patient Outcomes: Systematic Review and Meta-Analysis", *Medical Care*, 45(12): pp. 1195–1204.

[20] 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.

[21] National Institute for Health and Care Excellence (2015), "Safe Midwifery Staffing for Maternity Settings Guideline", NICE Guidelines [NG4], URL: `https://www.nice.org.uk/guidance/ng4/resources/safe-midwifery-staffing-for-maternity-settings-51040125637`, accessed on 28 May 2015.

[22] National Patient Safety Agency (2009), Quarterly Data Summary Issue 13: Learning from Reporting-Staffing, URL: `www.nrls.npsa.nhs.uk/resources/collections/quarterly-datasummaries/?entryid45=62052`, accessed on 5 September 2014.

[23] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49(6): pp. 1417–1426.

[24] Pauly, M.V. (1980), "Doctors and their Workshops", Chicago: University of Chicago Press.

[25] Pronovost, P.J., Angus, D.C., Dorman, T., Robinson, K.A., Dremsizov, T.T. and Young, T.L. (2002), "Physician Staffing Patterns and Clinical Outcomes in Critically Ill Patients", *Journal of the American Medical Association*, 288(17): pp. 2151–2162.

[26] Rafferty, A.M., Clarke, S.P., Coles, J., Ball, J., James, P., McKee, M. and Aiken, L.H. (2007), "Outcomes of Variation in Hospital Nurse Staffing in English Hospitals: Cross-Sectional Analysis of Survey Data and Discharge Records", *International Journal of Nursing Studies*, 44(2): 175–182.

[27] Reinhardt, U. (1972), "A Production Function for Physician Services", *Review of Economics and Statisics*, 54(1): pp. 54–66.

[28] Roodman, D. (2009a), "A Note on the Theme of Too Many Instruments", *Oxford Bulletin of Economics and Statistics*, 71(1): pp. 135–158.

[29] Roodman, D. (2009b), "How to Do xtabond2: An Introduction to Difference and System GMM in Stata", *Stata Journal*, 9(1): pp. 86–136.

[30] 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).

[31] Sato, R. and Koizumi, T. (1973), "On the Elasticities of Substitution and Complementarity", *Oxford Economic Papers*, 25(1): pp. 44–56.

[32] Schubert, M., Glass, T.R., Clarke, S.P., Aiken, L.H., Schaffert-Witvliet, B., Sloane, D.M. and De Geest, S. (2008), "Rationing of Nursing Care and its Relationship to Patient Outcomes: The Swiss Extension of the International Hospital Outcomes Study", *International Journal for Quality in Health Care*, 20(4): pp. 227–237.

[33] Shuldham, C., Parkin, C., Firouzi, A., Roughton, M. and Lau-Walker, M. (2009), "The Relationship Between Nurse Staffing and Patient Outcomes: A Case Study", *International Journal of Nursing Studies*, 46(7): pp. 986–992.

[34] Smith, A. and Dixon, A. (2008), "Healthcare Professionals' Views about Safety in Maternity Services", London: The King's Fund.

[35] Soto, M. (2009), "System GMM Estimation With a Small Sample", *Barcelona Economics Working Paper Series*, Working Paper No. 395.

[36] Sucov, A., Sidman, R. and Valente, J. (2009), "A Cost-Efficiency Analysis to Increase Clinician Staffing in an Academic Emergency Department", *Academic Medicine*, 84(9): pp. 1211–1216.

[37] Syrquin, M. and Hollender, G. (1982), “Elasticities of Substitution and Complementarity: The General Case”, *Oxford Economic Papers*, 34(3): pp.515–519.

[38] Thurston, N.K. and Libby, A.M. (2002), “A Production Function for Physician Services Revisited”, *Review of Economics and Statistics*, 84(1): 184–191.

[39] Uzawa, H. (1962), “Production Functions with Constant Elasticity of Substitution”, *Review of Economics and Statistics*, 29(4): pp. 291–299.

[40] Van den Heede, K., Sermeus, W., Diya, L., Clarke, P.K., Lesaffre, E., Vleugels, A. and Aiken, L.H. (2009), “Nurse Staffing and Patient Outcomes in Belgian Acute Hospitals: Cross-Sectional Analysis of Administrative Data”, *International Journal of Nursing Studies*, 46(7): pp. 928–939.

[41] Windmeijer, F. (2005), “A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators”, *Journal of Econometrics*, 126(1): pp. 25–51.

Tables

Table 1: Summary statistics for key variables, overall and by year.

Variable name	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total
Mean maternal age[a]	28.95	28.96	29.05	29.06	28.97	28.95	29.02	29.03	29.16	29.02
	(1.14)	(1.12)	(1.39)	(1.32)	(1.26)	(1.16)	(1.24)	(1.25)	(1.16)	(1.23)
Mean parity[a]	.825	.841	.822	.761	.820	.892	.910	.947	.989	.866
	(.422)	(.405)	(.357)	(.290)	(.296)	(.317)	(.293)	(.315)	(.340)	(.347)
% “high risk”[b]	.405	.417	.435	.447	.454	.468	.492	.513	.530	.461
	(.089)	(.085)	(.109)	(.100)	(.094)	(.075)	(.088)	(.098)	(.083)	(.099)
% British[a]	.676	.679	.687	.674	.687	.687	.690	.682	.679	.682
	(.217)	(.217)	(.218)	(.220)	(.217)	(.218)	(.221)	(.232)	(.230)	(.220)
% most deprived[a]	.256	.258	.261	.264	.262	.257	.265	.257	.257	.260
	(.209)	(.210)	(.217)	(.210)	(.211)	(.199)	(.201)	(.191)	(.190)	(.204)
Doctors (FTE)[d]	19.37	20.95	21.48	21.80	22.64	24.72	25.04	25.28	26.13	22.99
	(8.92)	(9.87)	(10.36)	(11.36)	(13.04)	(13.29)	(13.24)	(12.85)	(14.15)	(12.15)
Consultants (FTE)[d]	8.49	8.93	9.42	9.55	9.96	10.90	11.80	12.16	12.96	10.43
	(4.07)	(4.34)	(4.67)	(4.67)	(5.11)	(5.53)	(5.61)	(5.93)	(6.69)	(5.41)
Registered midwives (FTE)[e]	117.26	118.55	120.41	123.77	125.99	132.50	137.01	140.37	147.65	129.04
	(53.22)	(53.65)	(53.67)	(55.01)	(56.02)	(57.78)	(60.82)	(62.74)	(71.24)	(59.05)
Support workers (FTE)[e]	3.62	4.11	4.31	4.34	5.17	5.93	6.52	6.98	7.57	5.36
	(6.61)	(8.13)	(8.60)	(8.55)	(9.61)	(10.78)	(10.82)	(11.29)	(11.43)	(9.69)
Maternities[c]	3825.4	3888.3	4397.6	4370.8	4374.3	4481.9	4627.0	4661.9	4829.1	4376.4
	(1607.7)	(1670.7)	(2509.4)	(1956.2)	(2000.9)	(2133.8)	(2081.4)	(2136.4)	(2426.2)	(2091.7)
Observations	151	151	148	145	147	144	143	143	139	1,311

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: Standard deviations in parentheses.

Table 2: OLS parameter estimates for the Generalized Linear Production Function.

Variable name	[1]	[2]	[3]	[4]
$\text{Maternities}^{c}_{t-1}$	-	-	-	0.552(8.19)
RM^{e}	39.327(3.98)	40.448(4.28)	41.803(4.74)	14.105(2.12)
SW^{e}	10.063(0.84)	6.901(0.56)	2.509(0.20)	-0.364(-0.06)
C^{d}	124.156(1.58)	180.962(2.21)	174.131(1.97)	-1.786(-0.02)
D^{d}	121.894(2.43)	120.648(2.55)	88.017(1.91)	35.268(1.43)
$\text{RM}^{1/2}\times\text{SW}^{1/2}$	-29.884(-1.25)	-31.783(-1.42)	-33.255(-1.51)	-22.422(-1.88)
$\text{RM}^{1/2}\times\text{C}^{1/2}$	-41.979(-0.89)	-53.885(-1.13)	-69.381(-1.45)	8.238(0.20)
$\text{RM}^{1/2}\times\text{D}^{1/2}$	-48.794(-1.24)	-40.918(-1.08)	-34.050(-0.94)	-16.320(-0.82)
$\text{SW}^{1/2}\times\text{C}^{1/2}$	49.358(0.67)	40.972(0.59)	59.287(0.87)	36.686(0.96)
$\text{SW}^{1/2}\times\text{D}^{1/2}$	28.905(0.66)	43.284(0.98)	40.457(0.91)	29.963(1.28)
$\text{C}^{1/2}\times\text{D}^{1/2}$	-67.194(-0.63)	-113.661(-1.13)	-60.152(-0.64)	-18.678(-0.32)
Maternal effects[a,b]	No	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes
SHA fixed effects[c]	No	No	Yes	Yes
R^2	.821	.834	.848	.899
Observations	1287	1287	1285	1133
Providers	156	156	155	154

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: The total number of deliveries in each hospital per year is used as the dependent variable. *t*-statistics computed using standard errors which have been corrected for clustering by trust in parentheses.

Table 3: Within-trusts (fixed effects) parameter estimates for the Generalized Linear Production Function.

Variable name	[1]	[2]	[3]	[4]
Maternities$^{c}_{t-1}$	-	-	-	0.126(2.36)
RM^{e}	20.491(1.49)	20.211(1.43)	20.798(1.49)	15.466(1.20)
SW^{e}	-5.545(-0.55)	-7.602(-0.76)	-11.242(-1.15)	-12.770(-1.10)
C^{d}	106.504(0.96)	91.559(0.76)	77.474(0.57)	55.615(0.41)
D^{d}	47.432(2.34)	40.029(2.03)	28.728(1.54)	25.524(1.37)
RM$^{1/2}\times$SW$^{1/2}$	-12.402(-0.82)	-9.818(-0.64)	-9.497(-0.64)	-8.810(-0.54)
RM$^{1/2}\times$C$^{1/2}$	-44.572(-0.67)	-48.512(-0.70)	-55.966(-0.76)	-35.306(-0.49)
RM$^{1/2}\times$D$^{1/2}$	-32.777(-1.28)	-32.177(-1.22)	-30.488(-1.04)	-24.511(-0.86)
SW$^{1/2}\times$C$^{1/2}$	60.033(1.35)	59.165(1.32)	64.179(1.40)	49.799(1.11)
SW$^{1/2}\times$D$^{1/2}$	-4.349(-0.20)	-8.809(-0.39)	-9.744(-0.42)	-1.654(-0.06)
C$^{1/2}\times$D$^{1/2}$	37.608(0.42)	54.097(0.60)	71.856(0.77)	55.291(0.60)
Maternal effects^{a,b}	No	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes
SHA fixed effects^{c}	-	-	-	-
Within R^2	.209	.221	.272	.229
Between R^2	.868	.834	.776	.859
Overall R^2	.789	.748	.687	.783
Observations	1287	1287	1285	1133
Providers	156	156	155	154

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: The total number of deliveries in each hospital per year is used as the dependent variable. *t*-statistics computed using standard errors which have been corrected for clustering by trust in parentheses.

Table 4: GMM parameter estimates for the Generalized Linear Production Function.

Variable name	Diff-1 GMM $t-2$ [1]	Diff-2 GMM $t-2$ [2]	Sys-1 GMM $t-2$ [3]	Sys-2 GMM $t-2$ [4]
Maternities$^{c}_{t-1}$	0.024(0.29)	0.015(0.16)	0.252(3.59)	0.276(3.63)
RM^{e}	-15.608(-0.60)	-14.028(-0.61)	14.698(0.73)	10.423(0.55)
SW^{e}	-53.749(-1.65)	-47.224(-1.33)	-7.929(-0.42)	-19.541(-1.06)
C^{d}	270.889(1.27)	306.609(1.28)	247.111(1.53)	236.026(1.48)
D^{d}	101.318(1.23)	100.418(1.08)	197.662(3.43)	197.676(3.44)
RM$^{1/2}\times$SW$^{1/2}$	16.987(0.39)	17.961(1.08)	-10.805(-0.28)	-6.743(-0.18)
RM$^{1/2}\times$C$^{1/2}$	-33.029(-0.27)	-39.985(-0.30)	17.104(0.17)	31.716(0.30)
RM$^{1/2}\times$D$^{1/2}$	88.933(1.17)	89.102(1.16)	-41.520(-0.90)	-36.457(-1.00)
SW$^{1/2}\times$C$^{1/2}$	168.277(1.32)	147.943(1.30)	30.969(0.29)	33.256(0.31)
SW$^{1/2}\times$D$^{1/2}$	-94.679(-1.52)	-92.829(-1.21)	10.413(0.17)	8.817(0.14)
C$^{1/2}\times$D$^{1/2}$	-422.709(-2.44)	-425.911(-2.12)	-306.582(-2.06)	-320.757(-1.89)
Controls	Yes	Yes	Yes	Yes
Hansen J-test	[0.353]	[0.353]	[0.310]	[0.310]
Diff-in-Hansen test	-	-	[0.903]	[0.903]
AR(1)	[0.002]	[0.009]	[0.000]	[0.003]
AR(2)	[0.912]	[0.890]	[0.040]	[0.124]
Observations	983	983	1133	1133

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: The total number of deliveries in each hospital per year is used as the dependent variable. All models include controls for year, Strategic Health Authority and case-mix. All regressions treat the lagged dependent variable as predetermined and the case-mix controls as strictly exogenous. For the two-step GMM estimates, *t*-statistics computed using finite sample corrected standard errors (Windmeijer, 2005) in parentheses. For the Hansen J-test, the Diff-in-Hansen test and the AR(1) and AR(2) tests, the p-values are reported in brackets.

Table 5: GMM parameter estimates for the Generalized Linear Production Function (robustness checks).

Variable name	Sys-1 GMM $t-3$ [1]	Sys-2 GMM $t-3$ [2]	Sys-1 GMM $t-4$ [3]	Sys-2 GMM $t-4$ [4]
Maternities$^c_{t-1}$	0.238(3.40)	0.255(3.14)	0.244(3.81)	0.294(4.25)
RM^e	27.213(1.00)	16.376(0.69)	43.861(1.68)	37.147(1.20)
SW^e	-1.962(-0.09)	-12.963(-0.52)	12.617(0.55)	6.004(0.26)
C^d	275.250(1.31)	220.904(1.18)	239.551(1.02)	206.052(1.01)
D^d	193.577(2.33)	197.534(2.29)	246.018(2.35)	257.744(2.47)
RM$^{1/2}\times$SW$^{1/2}$	-66.589(-1.50)	-39.626(-0.82)	-84.347(-1.35)	-56.452(-1.05)
RM$^{1/2}\times$C$^{1/2}$	-37.906(-0.31)	7.414(0.07)	-57.733(-0.40)	-23.593(-0.16)
RM$^{1/2}\times$D$^{1/2}$	-60.812(-0.94)	-51.540(-0.93)	-116.550(-1.29)	-119.464(-1.41)
SW$^{1/2}\times$C$^{1/2}$	176.237(1.64)	147.934(1.46)	205.957(1.37)	129.121(0.96)
SW$^{1/2}\times$D$^{1/2}$	38.821(0.53)	5.456(0.07)	33.941(0.38)	28.467(0.36)
C$^{1/2}\times$D$^{1/2}$	-253.934(-1.65)	-274.052(-1.77)	-196.411(-1.07)	-206.097(-1.04)
Controls	Yes	Yes	Yes	Yes
Hansen J-test	[0.397]	[0.397]	[0.362]	[0.362]
Diff-in-Hansen test	[0.783]	[0.783]	[0.322]	[0.322]
AR(1)	[0.000]	[0.003]	[0.000]	[0.001]
AR(2)	[0.080]	[0.120]	[0.094]	[0.116]
Observations	1133	1133	1133	1133

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: The total number of deliveries in each hospital per year is used as the dependent variable. All models include controls for year, Strategic Health Authority and case-mix. All regressions treat the lagged dependent variable as predetermined and the case-mix controls as strictly exogenous. For the two-step GMM estimates, *t*-statistics computed using finite sample corrected standard errors (Windmeijer, 2005) in parentheses. The Hansen J-test row displays the p-values for the null hypothesis of instrument validity. The Diff-in-Hansen row displays the p-values for the validity of the additional moment restrictions imposed for the system GMM estimations. The AR(1) and AR(2) rows display the p-values for first and second order autocorrelation test in the error term of the first differences equations.

Table 6: GMM parameter estimates for the Generalized Linear Production Function (robustness checks).

Variable name	Sys-1 GMM $t-2$ [1]	Sys-2 GMM $t-2$ [2]	Sys-1 GMM $t-3$ [3]	Sys-2 GMM $t-3$ [4]
$\text{Maternities}^{c}_{t-1}$	0.264(3.71)	0.270(3.49)	0.244(3.68)	0.255(3.61)
RM^{e}	14.486(0.78)	10.524(0.62)	26.234(1.16)	24.455(1.02)
SW^{e}	-14.009(-0.64)	-11.673(-0.57)	-1.859(-0.09)	-10.108(-0.42)
C^{d}	123.749(0.77)	141.202(0.93)	96.343(0.56)	102.439(0.58)
D^{d}	148.313(2.70)	144.067(2.33)	95.342(1.28)	75.982(0.86)
$\text{RM}^{1/2}\times\text{SW}^{1/2}$	-27.880(-0.74)	-25.975(-0.79)	-67.798(-1.82)	-56.213(-1.46)
$\text{RM}^{1/2}\times\text{C}^{1/2}$	30.722(0.32)	32.537(0.34)	-25.945(-0.23)	-36.179(-0.31)
$\text{RM}^{1/2}\times\text{D}^{1/2}$	-36.091(-0.84)	-21.680(-0.57)	-47.401(-0.85)	-34.360(-0.54)
$\text{SW}^{1/2}\times\text{C}^{1/2}$	119.306(1.07)	118.698(1.16)	215.006(2.04)	204.630(1.94)
$\text{SW}^{1/2}\times\text{D}^{1/2}$	0.002(0.00)	-8.911(-0.13)	17.227(0.29)	0.930(0.02)
$\text{C}^{1/2}\times\text{D}^{1/2}$	-196.732(-1.44)	-226.871(-1.48)	-43.226(-0.30)	-25.715(-0.15)
Controls	Yes	Yes	Yes	Yes
Hansen J-test	[0.267]	[0.267]	[0.331]	[0.331]
Diff-in-Hansen test	[0.442]	[0.442]	[0.090]	[0.090]
AR(1)	[0.000]	[0.004]	[0.000]	[0.003]
AR(2)	[0.020]	[0.110]	[0.020]	[0.080]
Observations	1133	1133	1133	1133

Source: [a] Hospital Episode Statistics (HES) with categories defined in the Data Dictionary (NHS HSCIC, 2010).
[b] Derived from NICE Clinical Guideline 55 for intrapartum care (NICE, 2007) following the methodology outlined in Sandall *et al.* (2014) using HES data.
[c] ONS Birth Registration Records.
[d] Health and Social Care Information Centre (2003-2013) Medical Workforce Census.
[e] Health and Social Care Information Centre (2003-2013) Non-Medical Workforce Census.
Notes: The total number of deliveries in each hospital per year is used as the dependent variable. All models include controls for year, Strategic Health Authority and case-mix. All regressions treat the lagged dependent variable as predetermined. For the two-step GMM estimates, *t*-statistics computed using finite sample corrected standard errors (Windmeijer, 2005) in parentheses. The Hansen J-test row displays the p-values for the null hypothesis of instrument validity. The Diff-in-Hansen row displays the p-values for the validity of the additional moment restrictions imposed for the system GMM estimations. The AR(1) and AR(2) rows display the p-values for first and second order autocorrelation test in the error term of the first differences equations.

Table 7: Marginal productivities and Hicks elasticities of complementarity.

Panel A: Based on OLS estimates[a]		Midwives	Support workers	Consultants	Doctors
Marginal productivity		9.49	1.40	12.19	17.06
		(1.97)	(3.21)	(9.69)	(4.84)
Hicks elasticities	Support workers	-68.59	-	-	-
		(159.30)			
	Consultants	2.11	305.80	-	-
		(10.55)	(710.25)		
	Doctors	-2.02	120.49	-6.29	-
		(2.49)	(251.63)	(19.12)	
Panel B: Based on FE estimates[b]		Midwives	Support workers	Consultants	Doctors
Marginal productivity		4.32	-1.46	52.94	14.82
		(2.34)	(4.73)	(19.85)	(5.47)
Hicks elasticities	Support workers	56.62	-	-	-
		(224.32)			
	Consultants	-4.58	-91.38	-	-
		(8.92)	(315.06)		
	Doctors	-7.67	7.32	4.94	-
		(9.49)	(116.68)	(9.47)	
Panel C: Based on GMM estimates[c]		Midwives	Support workers	Consultants	Doctors
Marginal productivity		6.54	-3.99	66.14	48.48
		(3.36)	(8.52)	(28.16)	(12.31)
Hicks elasticities	Support workers	10.51	-	-	-
		(61.86)			
	Consultants	2.18	-17.93	-	-
		(7.48)	(71.71)		
	Doctors	-2.30	-4.38	-7.01	-
		(2.54)	(30.84)	(4.13)	
Panel D: Based on GMM estimates[d]		Midwives	Support workers	Consultants	Doctors
Marginal productivity		7.88	-1.69	73.47	39.66
		(3.04)	(9.62)	(28.78)	(11.55)
Hicks elasticities	Support workers	79.25	-	-	-
		(437.88)			
	Consultants	1.67	-135.91	-	-
		(5.03)	(814.257)		
	Doctors	-1.39	12.76	-5.46	-
		(2.47)	(128.75)	(3.48)	

Notes: [a] Calculated using the parameter estimates presented in column [4] of Table 3.
[b] Calculated using the parameter estimates presented in column [4] of Table 4.
[c] Calculated using the parameter estimates presented in column [4] of Table 5.
[d] Calculated using the parameter estimates presented in column [2] of Table 7.
Estimated standard errors (in parentheses) are obtained using the delta method.