HC
7/28/06
10:05 AM

Can pay regulation kill? Panel data evidence on the effect of labor markets and skills on hospital quality and productivity

Emma Hall

UK Department of Health

Carol Propper

University of Bristol, CMPO, and CEPR

John Van Reenen

London School of Economics, Centre for Economic Performance, and CEPR

This draft: July 5th 2006

Preliminary

Abstract

In this paper, we examine whether labor markets for health workers affect health outcomes in hospitals. We exploit the fact that wages for clinical staff in the British National Health Service are centrally regulated with little variation to reflect the substantial differences in local labour markets. Consequently, we predict that areas with higher outside wages should suffer from problems of recruiting, retaining and motivating workers and this should harm hospital performance. We construct hospital-level panel data on both quality as measured by AMI death rates (within hospital deaths within 30 days of emergency admission for acute myocardial infarction) and productivity. We present evidence that stronger local labor markets significantly *worsen* hospital outcomes in terms of quality and productivity. A 10% increase in the outside wage is associated with a 3% to 8% increase in AMI death rates. We find that an important part of this effect is operate through hospitals in high outside wage areas having to rely more on temporary "agency staff" as they are unable to increase (regulated) wages in order to attract permanent employees. We also find that hospital performance is improved by a richer skill mix, especially of physicians and to a lesser extent, nurses (as compared to less skilled staff). We quantify the magnitudes of these "hidden costs" of labour market regulation, which appear to be substantial.

JEL classification: J45, I11

Keywords: hospital quality, hospital productivity, skills, labor markets, wage regulations

Acknowledgements: We would like to thank Tim Besley, Amanda Gosling, Andrew Oswald, Paul Grout and participants in seminars at Warwick, Melbourne and the Department of Health for helpful comments. We thank South West Observatory for access to the HES data and Davidson Ho for his help with these data and Andrew Jackson for help in accessing much of the administrative data. Financial support is from the ESRC through the CEP and CMPO Research Centers.

Corresponding author: j.vanreenen@lse.ac.uk; Centre for Economic Performance, LSE, Houghton Street, London, WC1E 2AE, UK.

I. Introduction

The public provision of services is frequently accompanied by centralization of wage setting for public sector workers, even though the majority of these services are provided at local level. In many European countries health care workers, high school teachers, and the police do not negotiate directly with their local employer for pay. Instead, negotiations are carried out centrally, between the state or national level government and the employees' union.

In the UK, regional pay differences are considerable (Bulman, 2002). For example, in 1999 the average non-manual wage in Inner London (the highest wage region) was 65% higher than the average wage in the North East (the lowest wage region). These differences are mitigated after controlling for human capital and other factors, but significant regional differences persist. As in the United States (e.g. Borjas, 2004) the cross sectional dispersion of UK public sector pay is much lower than in the private sector (Disney and Gosling, 1998). The public sector has not experienced the increase in wage inequality to the same degree as the private sector over the last three decades. One reason for the greater equality of public sector pay across areas and individuals is that pay rates are usually set nationally. The combination of centrally set wages with local labour markets in which there are varying wages might be expected to affect outcomes in the public sector. A large pay gap may result in shortages of staff and difficulties in recruitment and retention particularly of high quality staff. There is evidence that for men that falling public sector relative wages have led to a decline in the quality of the public sector workforce (Nickell and Quintini, 2002). More generally, large pay gaps may lead to a reduction in motivation of staff or to the employment of lower quality of staff employed in the public sector.

The UK National Health Service (NHS) is a classic example of public sector wage setting. Pay for staff in NHS hospitals – which account for the vast majority of the hospital care provided in the UK - is set by a central review body that sets pay scales in which there is limited regional variation, and the variation that exists is unlikely to fully reflect the wages differentials in the external labour markets in which the staff are employed. We would expect to see these pay differences reflected in staffing difficulties that manifest themselves in the lower performance of hospitals operating in high outside wage labour markets. This has never been rigorously examined. While there is some evidence that nursing labour supply is responsive to the wages of nursing staff (Holmas, 2002) and responds to the difference between the wages paid at hospital level and the outside labour market wage (e.g. Elliott et al, 2005) there has been no examination of whether the quantity of medical staff (nurses and physicians) or the outside wages faced by different clinical staff affects the quality and productivity of hospital care.

In this paper we test the impact of staff inputs and the central wage setting process on the performance of NHS hospitals in England. We examine whether the quantity of various types of medical staff and the difference between hospital pay and the outside wage (in the local labour market) has an impact on the quality and quantity of care produced in NHS hospitals in England. Our measure of hospital quality is within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI)[1]. The productivity measure is the volume of clinical activity undertaken by the hospital per medical employee. Our data are from a panel of almost all acute hospitals in England and cover the period 1995 to 2001. This allows us to control for regional and time effects, to allow for lags in the impact of wages on output, and to control for unobserved heterogeneity of hospitals through fixed effects.

We find two key results. First, consistent with basic production theory, hospitals with a richer mix of skills have better quality health outcomes and higher productivity. Our second key result is that hospitals which operate in areas with high outside wages (relative to inside wages) are of lower quality (as indicated by AMI death rates) and lower productivity. This is not simply because they have trouble maintaining high enough staffing levels as we condition on labour inputs. This is more likely to be due to the difficulty of retaining high quality staff and/or the lower levels of effort that are supplied when outside wages are high. In particular, we show evidence that such hospitals have to rely disproportionately on temporary agency staff and the intensive use of agency staff is associated with worse health outcomes. This relates to an

[1] For examples of the use of AMIs to proxy hospital quality see Kessler and McClellan (2000) for the US, the review of Gaynor (2004). For the UK see Burgess et al (2004)

emerging line of literature on the causes and consequences of the use of temporary help in modern economies (e.g. Autor and Houseman, 2005; Erickek et al, 2003; Houseman et al, 2003).

The structure of the paper is as follows. In Section II we discuss the institutional background to our study, the nature of the research design and related literature. In section III we sketch the econometric model that we are estimating and issues surrounding the approach. In Section IV we discuss the data and in Section V offer a preliminary data description. Section VI presents the econometric results and section VII concluding comments.

II Institutional background and related literature

II.A. The wage setting process for UK health workers

In the UK health care is free at the point of use for all citizens and is provided through the National Health Service (NHS), a state monopoly provider[2]. Just under a million workers are employed in the NHS and the wages and conditions of clinical staff are highly regulated. Our study focuses on clinical care in acute hospitals so we focus on three main groups of staff: physicians, nurses and support to clinical staff (known as health care assistants). Exact definitions are given in the data section, but broadly speaking these cover over three quarters of all hospital staff with the residual group made up of managers, clerical assistants, nurse trainees, and other support staff. In our sample 13% of clinical staff are physicians, 65% are nurses and 22% are health care assistants.

Physicians and nurses' pay is essentially regulated to a precise national scale - a maximum and minimum scale that has little differentiation over the country, despite a wide variation in regional labor markets. Since 1984 these pay scales have been set by two "National Pay Review Bodies" (NPRBs) known as the Review Body for Nursing Staff, Midwives and Professions Allied to Medicine and the Review Body for Doctors

[2] There is a small privately funded sector, which specializes in the provision of elective services for which there are long NHS waiting lists.

and Dentists. Each year, the Review Bodies take evidence from the Department of Health, the main labor unions and other interested parties before making a recommendation on changes to the level and structure of pay. The government makes the final decision about whether or not to implement their recommendations (it generally does this in full).

Under these national scales the same terms and conditions apply across the UK and they allow only minor differences in pay between different areas. Additional allowances are paid to those who work in London and contiguous areas, but these are small relative to the differences in the external labor market (these allowances are up to about 11% higher in the highest cost area of Inner London compared to the low cost areas. The outside wage differential is closer to 60% - see Appendix C). Beyond these regional allowance hospitals have little scope for aligning the pay of qualified nurses to conditions in local labour market conditions. Pay scales are short and offer very little scope for either appointing new hires at different points on the scale, or accelerating workers up to higher grades. The centralized pay setting arrangements for do not allow pay to be easily adjusted to address staff shortages in local markets.

NHS hospital employers have some scope to adjust non-pecuniary benefits, such as the quality of the working environment, training and relocation expenses. To meet their recruitment needs employers have participated in national initiatives, such as “return to practice” schemes, which offer additional funding for re-training and recruitment of staff who have been out of the workforce for some time.

The health care assistants group is more heterogeneous. There are no Pay Review Bodies and employers have some more discretion over setting pay in response to local conditions. This flexibility should not be exaggerated, however, as pay is generally determined by bargaining with another set of unions who also have preferences for national pay setting.

From an econometric perspective, this institutional setting is attractive because it enables an examination of the impact of different local wages on health outcomes. In most labor markets equilibrium wages will be the outcome of demand and supply shocks so identifying their impact on hospital outcomes is difficult as the price is endogenous to unobserved shocks. In the UK case the inside wage is held broadly fixed as outside shocks change skill prices in the local labor market. There is a wedge between the worker's offer wage and the outside wage. Consequently, variation in the outside wage can in principle be used to analyze the effects of labor markets on health outcomes.

Figure 1 illustrates the case for two local labor markets, "North" and "South" where outside wages are much higher in the South than in the North (generating a shift to the right in the labor supply curve). Given an equal pay rate across geographic areas this will mean a lower level of employment in the South compared to the North.

In principle, the regulated wage could be set above the competitive wage so it acts as a minimum wage and thus employers shed staff. However, there are chronic shortages of clinical staff in all parts of the NHS and clinical unemployment is practically zero (e.g. Finlayson 2002). Therefore, it is more likely that the wage is being set below the competitive wage generating excess demand[3]. It is possible, however, that wages are set above the competitive wage in some low wage areas in the North and NHS staffs in these areas are enjoying rents.

We consider in more detail the consequences of regulating wages for health workers in this way rather, first analyzing employer responses then worker responses. Turning first to the firm side, employers have incentives to overcome the regulatory constraint. First, they could use "grade drift" by over-promoting identical workers to higher grades even if they do not have the requisite skills. This will help them achieve the desired quantity, but at the cost of the lower quality of the over-promoted staff. Second, they could offer various non-pecuniary benefits such as better working conditions in the high wage areas. These strategies are limited by clinical unions' power in pushing for homogeneous national conditions and governments have been reluctant to challenge this[4]. We do investigate these mechanisms in the empirical

---

[3] In the absence of pay regulation, large local hospitals may have monopsony power so the equilibrium wage will not be at the intersection of the labor demand and supply curve. But so long as the regulated wage lies below the monopsony wage employers the constraint will still be binding.

[4] The desire for nominal equality across workers in geographical areas and other dimensions has long been a mainstay of union activity. It is not obvious why this should be the case, as real wages within the NHS are made more unequal since the cost of living varies by area. If unions represent the view of the median worker, however, this worker may be better off with a more compressed wage policy.

work, using the fact that there were reforms during the period of our data (the NHS "Internal Market" period) which made local pay flexibility easier.

[Figure 1 about here]

Third, assuming the regulated wage is binding, cost-minimizing employers will try to adjust by substituting towards other factors of production. Consequently, other types of staff whose pay is less constrained by regulated pay will increase and non-labor factors will be in higher demand. Because of the specific skills required for different medical interventions, however, substitution to much less skilled workers (health care assistants and non-clinical NHS staff) will be limited. One key group of workers where substitution is easier is "agency" staff. There are a large number of nurses employed in hospitals on temporary contracts. Private sector firms "rent out" agency nurses and other staff to the NHS and such staff do not have their pay regulated by the state. Consequently, the availability of agency staff will enable NHS employers to bring their employment closer to their desired levels.

The reliance on temporary agency workers in the high wage areas may have costs. Agency staff may be less committed (for example, less likely to share the public service 'mission' of the NHS), less flexible and less well trained than permanent staff and they will have built up less job specific human capital[5]. Their presence can often cause resentment among the permanent staff (especially if they are paid more). Therefore, the presence of agency staff may be a mechanism through which stronger external labor markets can lead to worse health outcomes in hospitals and we investigate this directly in the empirical work.

Turning to the employee side, in local areas where outside opportunities are better clinical staff will supply less labor as shown in Figure 1. In a static sense, this may lead to lower participation rates, as fewer qualified staff will offer themselves for

[5] Many reports (e.g. Audit Commission 2001) find that agency nurses have little notice before working their shifts, and that they are often employed to provide cover at weekends and at night when direct supervision is less likely to be available. As a result, induction and handover may be non-existent, or inadequate, and nurses may have little time to get accustomed to the workings of the hospital. They may be unfamiliar with the patients under their care, with local procedures, practices and equipment, with their surroundings and their colleagues. The Audit Commission argued that all these factors, combined with generally poorer attendance at training sessions increase the chances of patients receiving care of a poorer quality than they would otherwise get.

work[6]. In a dynamic setting this will lead to higher rates of vacancies and turnover in the high wage areas. We discuss existing estimates of the elasticity of labor supply of medical staff in Appendix A (see also the next section)[7].

In this paper we are interested in whether there are further effects of wage regulation on hospital outcomes over and above the pure effects of reduced labor supply. A stronger local labor market may reduce the quality of staff that is prepared to work in the hospital sector as permanent staff or from the substitution into temporary agency nurses as discussed above[8]. Secondly, there may be effects on the effort of workers through various "efficiency wage" channels such as lower motivation and greater shirking. We model this formally below.

If mobility were costless and there exists a single national labor market then using local wages as a signal of labor market pressure makes little sense. Mobility in Britain is far less than in the US (for example, less than one per cent of adults move between the nine English regions in any given year). Nurses and health care assistants are predominantly female, often with childcare responsibilities and need to be close to hospitals, as they are required to work shifts. It is likely, therefore, that mobility is not perfectly responsive to wage differentials. We show some evidence for this below (if mobility was costless it is unlikely that geographic differentials of the magnitude we observe could be sustained).

II.B Related Literature

Our study relates to several different literatures. First, we show evidence for the importance of physicians and nurses on hospital quality and productivity. Although

[6] NHS employees face low switching costs into agency work, or employment in the limited private sector.

[7] Gosling and Van Reenen (2006) show, for example, that nurses respond to higher outside wages relative to inside wages by switching away from nursing to other occupations (and to a lesser extent, to non-participation). Using a long panel of regions between 1984 and 2001 when there were some significant changes in mandated regional differences they show that a 10% fall in nurse relative wages reduces nurse employment by up to 15%.

[8] This raises the question of why any nurses want to work in the public sector instead of simply becoming agency nurses and earning more. There is some stigma attached to being an agency nurse rather than a full-time employee and because nurses are "motivated agents" (see Besley and Ghatak, 2005) this may be a factor. In addition, permanent staff have other non-pecuniary benefits such as greater job security, better promotion prospects and enhanced pensions. Nurses could also work in the private sector. However, the demand for nurses in the private sector is limited by the small size of this sector (Laing and Buisson 2005) and nurses in this sector have less opportunities for promotion, training and contribution to 'public service mission'.

there is a growing literature on the importance of nurses, physicians have not been a major focus. Second, our study relates to the labor supply literature looking at the effect of wages on employment. We survey these first two literatures in Appendix A. Thirdly; our study is part of the emerging personnel economics literature on payments systems and firm performance. We find that one element of pay systems can have unexpected outcomes in terms of organizational performance. Finally, our study relates to the impact of local economic conditions on health (e.g. Ruhm, 2006). These studies focus on how economic conditions effect the demand for health – e.g. by changing people's wealth or stress levels. We suggest an alternative mechanism operating through the supply side. In our model, labor market conditions affect the supply of a key clinical input, health workers, especially when combined with rigid national pay setting.

III Empirical strategy

III.A. Modeling Approach

To motivate our empirical work, assume that we can characterize the output of a representative hospital by a Cobb–Douglas production function[9]

$$Y = AL^{\alpha} K^{\beta} \tag{1}$$

where Y is quality constant output, L is effective labour input allowing for quality and quantity dimensions, K is a vector of non-labour inputs (which for expositional simplicity we will treat as scalar) and A is a Hicks neutral efficiency parameter. We write L as the product of "effort" (E) and labour quantity ($\tilde{L}$). Effort is a catch all term for the other factors that transforms labour into efficiency units along dimensions we cannot directly observe.

$$L = \tilde{L} E \tag{2}$$

We consider disaggregating the labor quantity into different types of heterogeneous workers so that the labour quantity index can be written as

$$\tilde{L} = \sum_k \gamma_k N_k \tag{3}$$

[9] This should be viewed as a first-order approximation to a more complicated functional form. It is straightforward to generalise this to more complex functional forms such as translog and some experiments are included in the empirical results.

where N_k is labour of skill type k with relative marginal productivity γ_k (>1) and we normalize $\gamma = 1$ for $k = 0$, the lowest skill type[10]. Taking logs of equation (1) and substituting in equations (2) and (3) gives

$$\ln Y = \ln E + \ln A + \alpha \ln N + \beta \ln K + \alpha \ln\{1 + \sum_{k,k\neq 0}(\gamma_k - 1)S_k\} \qquad (4)$$

where N is the sum of employees[11] and the share of workers of skill class k is $S_k = \frac{N_k}{N}$. We model the effort function as:

$$E = e(W,W^O,Z) \qquad (5)$$

Where W is the "inside" wage (i.e. wage paid to group k in the hospital), W^O is the "outside" wage and Z are other factors affecting effort/quality of workers. We expect effort to be rising in the inside wage and falling in the outside wage other things equal. Using a first order log linear approximation for the effort function in equation (5) and substituting this into equation (4) gives:

$$\ln Y = \ln A + \alpha \ln N + \beta \ln K + \alpha \ln\{1 + \sum_{k,k\neq 0}(\gamma_k - 1)S_k\} + \delta \ln W + \varphi \ln W^O + \ln Z'\theta \qquad (6)$$

An alternative to estimating (6) directly by nonlinear least squares is by using the approximation $ln(1+x) \approx x$ which gives us:

$$\ln Y = \ln A + \alpha \ln N + \beta \ln K + \alpha \sum_{k,k\neq 0}(\gamma_k - 1)S_k + \delta \ln W + \varphi \ln W^O + \ln Z'\theta \quad (7)$$

Theoretically, the object of the left hand side of equation (7), Y, is quality-adjusted output. However, we do not observe this directly. Instead, we have various proxies for this measure of performance. We utilize a key measure of quality (D, death rates following admission for emergency AMI) and a simple measure of hospital activity

[10] See inter alia Hellerstein *et al*. (1999) or Dearden, Reed and Van Reenen (2005)
[11] See Levine (1988), Wadhwani and Wall (1990) or Machin and Manning (1995) for examples of this approach in the efficiency wage literature following this approach):

the number of "Finished Consultant Episodes" (FCEs), which are essentially (non case-mix adjusted) admissions.

III.B. interpreting the wage effects

The wage effects in equation (7) reflect any impact wages may have on (average) worker effort or worker quality. We expect φ<0 because conditional on a given "inside" wage in a hospital an increase in the outside wage should reduce E. Similarly an increase in the inside wage should increase E (δ>0). The various mechanisms have been extensively discussed in the efficiency wage literature. For example when outside wages rise relative to inside wages this may (1) induce lower effort because the effective cost of shirking has fallen as losing one's job is less important if the outside labor market is strong (Shapiro-Stiglitz, 1984); (2) demotivate staff for socio-psychological reasons (Akerlof, 1982); (3) increase turnover rates under models of search; (4) make it harder to attract higher quality workers. In the context of the publicly run UK health system this may be reflected in higher vacancies and the greater reliance on agency nurses (who are contracted on a temporary basis to NHS hospitals). We examine some of these possible mechanisms in the results sections, although it is obviously difficult to observe many of these mechanisms directly.

As discussed above wages can also have an effect on hospital performance through the *quantity* of employees of different skill types. Conditioning the production function on labour inputs in equation (7) abstracts away from these effects so we can focus on whether there is an impact of the labour markets through the *E(.)* function[12]. We will therefore be underestimating the importance of wages on hospital production and show specifications where we relax this.

III.C. Econometric models

Using lower case letters to denote natural logarithms we will estimate production functions for hospital *i* at time *t* as:

[12] Because we observe a discrete number of skill groups one concern is that we are grouping over heterogeneous skill groups within these categories. The interpretation of the inside wage becomes ambiguous because it may simply reflect unobserved labor quality. The coefficient on the outside wage, however, should be robust to this problem of interpretation as the theory implies that it should take a *negative* value.

$$(y-n)_{it} = a_{it} + \mu n_{it} + \beta_1 S_{it}^{PHYS} + \beta_2 S_{it}^{NURSES} + \delta w_{it} + \varphi w_{it}^{O} + z_{it}^{'}\theta \tag{8}$$

Compared to equation (7) we have used three main skill groups – physicians, nurses and health care assistants (the base category). S_{it}^{PHYS} is the share of physicians in total medical staff and S_{it}^{NURSES} is the share of nurses in total medical staff. Physicians receive the largest amount of training[13], nurses[14] the second highest amount and health care assistants the least. So in terms of the model we expect $\beta_1 > \beta_2 > 0$. Note that we have transformed the dependent variable from output into "productivity" (Finished Consultant Episodes per worker) so $\mu = \alpha - 1$ is a scale parameter that will be equal to zero under "constant returns". Our baseline regressions include employment as a size control but we also show the robustness to imposing constant returns to avoid an obvious division bias (employment being on the left hand side and right hand side of the regression). Note that we have absorbed the non-labor inputs into the z-vector.

Since we have panel data we decompose the unobserved total factor productivity term into its variance components: $a_{it} = \eta_i + \tau_t + v_{it}$ where η_i is a hospital effect, τ_t are a set of time dummies and v_{it} is a stochastic error term whose properties we discuss below. We present results treating η_i as a fixed effect (e.g. long-differenced results or system GMM) and treating it as uncorrelated with the right hand side variables (i.e. standard OLS).

$$(y-n)_{it} = \beta_1 S_{it}^{PHYS} + \beta_2 S_{it}^{NURSES} + \delta w_{it} + \varphi w_{it}^{O} + \theta z_{it} + \eta_i + \tau_t + v_{it} \tag{9}$$

We will use various proxies for the outside wage (w_{it}^{O}) based on average wages in the local labor market around the hospital. We experiment with measures based at a

[13] Physicians in the UK follow a five year undergraduate BA program and then spend a further 4-11 years in training, depending on the specialty. Nurses follow a three year undergraduate degree to become a registered nurse: specialist nurses then train for a further year (or more).
[14] We group Allied Health Professionals with nurses. This group includes radiologists, physiotherapists, etc. We could not reject the equality of coefficients of this group with nurses in the results.

disaggregated level (we have over 100 distinct travel to work areas in our data) and a relatively aggregate level (the nine regions of England). We focus on female non-manual wages as this is the most likely comparator group for nurses but we also consider other outside wage comparators. We view physicians as operating essentially in a national labor market so the time dummies will capture their outside wages, but we also examine this using alternative wage measures discussed in the next section. Since hospitals are a small part of the local labour market we treat the outside wage as exogenous, although we will always lag the variable to avoid any immediate feedback effects from transient area level shocks (permanent shocks are picked up by hospital fixed effects).

Identification of the inside wage coefficient is more challenging. We observe the hospital inside wage but it is likely to be problematic to interpret. First, to the extent that the hospital can partially manipulate the inside wage will respond to shocks that affect hospital performance. Secondly, higher wages may reflect a better skill mix such as a superior grading structure. Thus finding a positive coefficient would not reflect effort but simply better human capital. Consequently, we present the first set of results that do not condition on inside wage information under the assumption that the national wage is truly national. We then consider alternative methods of including the inside wage. Firstly, we use the suggestion of Gosling and Van Reenen (2005) to use the predicted nurse wage from the decisions of the National Pay Review Body for Nurses. These Pay Bodies regulate the level of nurses pay by grade and area (e.g. there is a wage supplement tied to the cost of living in the local area). We use the grade structure in the previous year to predict how wages will increase in the following period based on the regulated pay uprating. Secondly, we include observed inside wages and use the system GMM approach discussed below to allow for endogeneity.

Identifying the coefficients on the factor inputs in production functions is an old problem in econometrics (see Ackerberg et al, 2005, for a recent survey). In equation (7) the endogenous factor inputs are the numbers of employees of different skill types. Our preferred method draws on a recent contribution by Bond and Söderbom (2005) which examines the estimation of a model of a Cobb Douglas production function when inputs with differential adjustment costs are optimally chosen. In our context we make the plausible assumption that the hospital faces larger adjustment costs from changing the number of physicians relative to adjusting the number of nurses. Under reasonable parameterizations of the adjustment cost process lags of the endogenous variables will be correlated with current values and this can be used to justify the moment conditions underlying the Blundell and Bond (1998) estimator as applied in the production function context (e.g. Blundell and Bond, 2000). Essentially this estimator builds on the traditional moment conditions that lagged levels of the endogenous variables can be use to instrument the first differenced endogenous variables (Arellano and Bond, 1991). By (testable) assumptions on the initial conditions the "System GMM" approach also allows lagged differences to be used as instruments for the equation in levels. We detail the System GMM approach in Appendix B. The approach allows the current employment of all skill groups to be affected by shocks to productivity (i.e. endogenous in the production function). We compare the GMM approach to alternative methods of estimating equation (9) by long differences and by OLS in levels.

We also include controls for case mix (discussed below) and focus on acute hospitals (non-acute hospitals are a more heterogeneous set and include mental health and community hospitals so we drop them from the sample).

The hospital quality equation is estimated in a symmetric way to the production function (all the coefficients are allowed to differ, of course, as indicated by the "d" superscript):

$$d_{it} = \beta_1^d S_{it}^{PHYS} + \beta_2^d S_{it}^{NURSES} + \delta^d w_{it} + \varphi^d w_{it}^O + z_{it}^d{}'\theta^d + \eta_i^d + \tau_i^d + \nu_{it}^d \tag{10}$$

We again include controls for casemix (and we have some AMI specific information here). A concern with interpreting the outside wage term in equation (10) is that higher economic activity can increase the likelihood of death risks (as argued by Ruhm, 2006). This might be due to greater air pollution, traffic congestion or stress at work. To guard against this we condition on mortality rates in the catchment area of the hospital. If general mortality rates in the area increase because of higher economic activity or for any other reason, we should be controlling for this through mortality rates. Additionally, the use of productivity alongside AMI is useful in this respect. If the effect of improved outside labor markets was solely through decreasing health it is unclear why this would tend to depress hospital productivity.

IV Data

IV.A. Basic Information

The unit of observation in this study is the hospital, so all measures are at hospital level. We construct a panel data set of NHS hospitals (called "trusts" in the UK[15]) covering the financial years 1995/6-2002/3. The panel is unbalanced as the number of hospitals changes over the period. The number of hospitals falls over the period due to mergers and acquisitions, which increased significantly since 1991 (the start of the "Internal Market" period). The net effect of these changes is that the number of hospitals in our data is largest in 1995 and falls thereafter. In 1995 the number of acute hospitals in our data was 234. This fell to 227 by 1997 and fell further to become 175 in 2002.

We match data from a variety of sources to these hospitals for the analyses undertaken here. To ensure that the results are not driven by a handful of outliers we undertake a range of robustness checks that utilize different samples of the full data set: we outline these in the results section below. Details of the data and sources are given in Appendix C.

IV.B. Measures of quality, productivity and casemix

These are derived from hospital episode statistics (HES) data for the financial years 1991/2 through 2002/03 (we use 1995/6 in our analyses). We measure quality of output by within hospital deaths within 30 days of admission for emergency acute myocardial infarction (AMI) for patients aged 55 or over[16]. AMI was chosen because

[15] An NHS "trust" is a financial, managerial and administrative unit and may cover more than one physical hospital. It is appropriate to think of a hospital as a firm that may be single plant or multiplant. We use the term "hospital" rather than "hospital trust" for expositional convenience.

[16] Use of emergency admissions reduces the problem of patient selection.

it is a common condition; the treatment objective is generally to ensure survival; the quality of care affects survival; the infrastructure used to treat AMI is common to other hospital services[17]; and all patients with a recognized AMI are admitted, so there is little scope for selection bias to affect the decision of who gets admitted (Volpp et al, 2003). Deaths following emergency admission for AMI have been published by both US and UK governments as indicators of hospital quality. McClellan and Staiger (2000) argued that suitably adjusted measures of death rate correlate well with other measures of quality. Variants of this measure have been used widely in studies of hospital quality in the US (starting with Kessler and McClellan, 2000: see Burgess et al (2004) for the UK).

Our measure is the annual hospital-level average, derived from the hospital episode statistics data (HES). To avoid the problem of variability of rates from small denominators we only undertake analyses using hospitals with at least 150 emergency AMI admissions per annum. Without patient level data we cannot control for individual patient co-morbidity, but to allow for differences in case-mix we include three sets of controls. First, we control for unobserved hospital fixed effects, which will control for differences in case-mix that are fixed over time[18]. Second, we control for the mortality of the catchment area of the hospital (which is time varying) and will pick up the degree of ill health of the population that the hospital draws its cases from. Third, we control for the age-gender distribution of admissions for emergency AMI[19]. In data appendix C we discuss further issues in the use of our AMI measure as a measure of hospital quality. There may be some time varying, within area AMI death rates related to unobservable casemix but not captured by area mortality rates. However, this error would have to be correlated systematically with outside wages or the skill variables in order to undermine our inferences. This is possible, but in our view, is unlikely.

[17] Many of the actions to reduce deaths from emergency admissions for AMI need to be taken soon after an attack, and so the performance of a hospital in terms of AMI reflects the performance of its accident and emergency unit. Around half the patients admitted to an acute English hospital are admitted through the accident and emergency department.

[18] Hospitals in England do not compete with each other for emergency AMI patients.

[19] The proportion of AMI admissions of patients aged between 55 and 60, 60 and 65, 65 and 70, 75 and 75, 75 and 80, 80 and 85 and over 85, separately for men and women

Productivity is measured in a way similar to labour productivity in studies of other sectors – as total output per head. Our measure of total output is finished consultant episodes (FCEs). This is a standard output measure used in the NHS and indicates the total volume of medical activity. While this measure has been used in the literature (e.g. Vita 1990), it does not take into account the severity of patient inputs. No published indicators of patient severity within age or gender group are available for the whole period covered by our data (see data Appendix). To allow for variation in case mix, we again control for hospital fixed effects and for the age-gender profile of total admissions at hospital level in our main analyses[20].

IV.C. Wages

We use several measures of outside wages. Our main measure is derived from the New Earnings Survey (NES) that is a 1% sample of all employees in Great Britain covering about 300,000 individuals a year. The NES is mandatory administrative panel data provided by firms to the Department of Work and Pensions and contains information on earnings and hours. Our main measure is average annual earnings, but we also consider hourly wages. We use the area code in the NES to construct one hundred county-based travel to work areas (or boroughs in London). Using the postcodes of the headquarters of county (and borough) councils, we matched each NHS hospital to all county councils that fell within a twenty-kilometer radius from the hospital headquarters. The local area wage is constructed as the average of the county wages of all the councils that fell into this radius. Where no councils fell within the twenty kilometer radius (this occurred for roughly a quarter of the sample), the wage applicable to the nearest council was used[21].

---

[20] We used 18 five year age bands for males and females separately giving a total of 36 casemix controls.As another control we used the proportion of admissions admitted in three categories: emergencies, electives or transfers, as there will be variation in severity across these different types of admission. This gave very similar results to those reported below.

[21] Almost half of the trusts had only one council within a twenty kilometre radius from the trust, with the remaining quarter having two or more councils, from which an average was calculated. All of the trusts with more than two councils within a twenty kilometre radius are either in London or on the edge of the surrounding counties (as clearly boroughs are far smaller than counties and so more will be within the 20k limit). More details are available from the authors.

Our main measure is the average wage of non-manual females since the overwhelming bulk of nurses and health care assistants are female[22]. We also considered male average non-manual wages as there are more male physicians, but we think that physicians' labor markets are more likely to be national in scope and so local outside wages may be less important. In general, mobility across regions is much lower in the UK than in the US, only about 1% of adults move across the 12 regions of the UK in a given year (compared to 7% across US states). As an alternative to the NES we also considered the Labour Force Survey (LFS). The LFS is a self-reported household survey containing about 320, 000 individuals per year (with 80,000 observations on wages). From the LFS we can extract spatial wage differentials conditioning on more characteristics to build up the outside wage offered to a "typical" nurse. We experimented with such measures that successfully predict labor supply problems in the cross section (e.g. Elliot et al, 2005). The smaller sample size and sampling variation, however, means that such constructed variables are less useful in a panel data analysis. So for the most part, we rely on the larger sample sizes of the NES that has less measurement error as it is taken directly from employer records. We also experiment with using measures of unemployment rates and employment rates as alternative indicators of labor market "tightness". We did not find that these added explanatory power over and above the information in the wage, which in principle should fully reflect labour market conditions.

We also constructed measures of the hospital-specific "inside" wage facing health workers. One possible measure of the "inside" wage is simply the average wage paid to nurses or other health workers in the hospital. We present evidence using this measure, but it is likely to be problematic as it may simply reflect the grading structure (as discussed above in the econometric section). We therefore used a more exogenous measure of the price of labor based on the predicted regional wage for a nurse following the method of Gosling and Van Reenen (2006). We use mandated wage uprating by the National Pay Review Body (which has an area and grade specific component) to calculate the predicted wage increase for an average nurse in

[22] We also calculated the average wage for women in secretarial occupations and for all females: results using these measures are very similar to those presented here. The correlations between the wages for women in secretarial occupations, all females and female non manual workers were all above 0.95.

the trust using the regional characteristics based on the NES. This can be used as an instrumental variable for the observed inside wage.

IV.D. Skill Groups

We use data on four different main groups of staff: all physicians, nurses, qualified allied health professionals and health care assistants. In the analyses presented here, we allocate qualified allied health professionals to the nurse group and unqualified allied health professionals (AHPs) to the health care assistants group. Extensive checks show that our results are not sensitive to this summation. Our staff measure is annual whole time equivalents. We define total clinical staffing as the sum of staffing across these groups. Shares for each group of staff are defined relative to this measure of total staff[23].

Table 1 presents the data used in the analysis.

V. Preliminary Data Description

We begin with a description of the trends in AMI death rates, productivity, staffing and wage rates across time.

V.A. Variation of AMI rate and productivity over time

Figure 2 presents the distribution of AMI deaths 1995-2002 for acute hospitals. This (and subsequent figures) shows the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} quantiles of the distribution. The most striking feature of Figure 1 is the remarkable variation of death rates at any point in time between different hospitals. There are twice as many deaths in for the bottom decile as there are in the top decile. Some of this variation can be

[23] We also have a total employment measure that includes the non-clinical staff, but unfortunately does not disaggregate between highly skilled groups such as senior managers and less skilled groups such as orderlies. Consequently, our main results use total clinical staff as the main employment measure and we check the robustness of the results to conditioning on the total employment measure.

accounted for by case mix but, as we shall see, there remains much residual variation that is potentially related to the quantity and quality of labour inputs.

[Figures 2 and 3 about here]

Looking at the evolution of the distribution Figure 2 shows a gradual decrease in death rate over time suggesting improvements in survival rates. There has been a long run trend of a fall in the death rate from AMI shown in our data. Interestingly there is some convergence in death rates between hospitals at the top and bottom of the distribution towards the end of the period. This variation over time is useful to us in identifying changes in the panel.

Figure 3 shows the distribution of log productivity. As with AMI death rates there is a large dispersion in productivity: some 60 log points in 1995 between the top and bottom deciles. This huge heterogeneity mirrors the well-known findings in the productivity literature that has looked at differences between private sector firms: there is significant and persistent productivity heterogeneity even within very disaggregate sectors of the economy that cannot be explained by observable factor inputs (e.g. Foster et al, 2005)

The figure shows clearly the lack of trend in our measure of productivity. This mirrors the national pattern in total FCEs and in total admissions, which have shown little growth since 1997. Total FCEs in England were 11.983m in 1998/9 and 12.757m in 2002/3[24]. Similarly, to the AMI death rates there is evidence of a reduction in the variance across hospitals at the end of the period that could be related to the Labor government's attempts to regulate minimum standards more intensively since 1997.

V.B. Outside wages and outcomes: vacancies, labor supply and AMI death rates

We present some simple correlations across regions to see if there is anything in the raw data suggesting the relationships we have been discussing. First we plot the mean outside wage against the nurse vacancy rate in Figure 4 across the ten English regions (we use more disaggregated data in the econometric work). There is considerable

[24] http://www.hesonline.nhs.uk

variation in the outside wage across regions. A clear upward sloping pattern emerges with the highest outside wage areas London having a vacancy rate that is fourfold higher than the vacancy rate in the lowest outside wage area (like the North East).

Figure 5 plots out the correlation between nurse labor supply as measured as the proportion of women with a nursing qualification who are employed as nurses (rather than being employed in another occupation or non-participants). Unlike vacancies we were able to construct this for Wales and Scotland as well using Labor Force Survey data (similar in structure to the CPS). As we expect labor supply is lower in the regions where outside wages are higher. The relative wage measured as the log of the ratio of the nurses pay to the average non-manual female wage in the same region. In Inner London nurse wages are about 5% lower than the regional average whereas in the North of England wages are almost 30% higher[25]. A surprisingly large number of women qualified as nurses who actually work as nurses. The proportion of women with nursing with qualifications who actually work as nurses ranges from about 50% in the Southern regions with tight labor markets to almost 80% in Wales and North England that have weaker labor markets. Qualified nurses are much more likely to leave the profession in regions with higher outside wages.

[Figures 4 and 5 about here]

Figure 6 examines the intensity of using agency nurses and outside wage. Again, we find that the regions with high outside wages rely a lot more on agency nurses than the regions with low outside wages.

Finally, Figure 7 plots out the AMI death rate as a function of the outside wage for the ten regions of England. There appears to be a positive relation with London having the highest AMI death rates and the low-wage regions of the North having lower AMI death rates.

Overall then, regions with high outside wages are characterized by higher vacancy rates, lower nurse participation rates and a greater use of temporary agency staff. They

[25] It cannot be concluded from this simple comparison that nurses outside London are earning quasi-rents from the public sector as the comparison does not take into account non-pecuniary aspects of being a nurse, which may it an unattractive occupation for many people. These non-pecuniary aspects are likely to be relatively stable over time.

also tend to have higher death rates from AMI. There are of course many reasons why these figures may be misleading and there is no causal connection between high wages and poor performance. For example, there may be many other factors positively influencing the outside wage and the AMI death that we have not controlled for. To tackle this we turn to the econometric results where we look at within region variation both in the cross section (by using area and hospital trust level data) and in the time series (by controlling for hospital fixed effects) as well as conditioning on confounding variables such as casemix and local mortality rates.

[Figures 6 and 7 about here]

VI. Results

VI.A. Hospital Quality as measured by Death rates from AMI

Table 2 presents the estimates for hospital quality as measured by ln(AMI death rates). The outside wage is measured by the ln(average wage) of non-manual women in over one hundred local "travel to work" areas. Column (1) presents the pooled OLS estimates of the association of AMI death rates with staff shares and outside wages. We control for AMI specific casemix (admissions in fourteen age-gender bands), hospital type (i.e. whether the hospital was a specialist hospital, or a teaching hospital), the area mortality rates, size (as measured by the log of total employees), year dummies and ten regional dummies[26].

According to column (1) of Table 2 hospitals with better qualified employees (i.e. a higher proportion of physicians and/or nurses relative to health care assistants) have significantly lower AMI death rates (i.e. higher hospital quality of care). The coefficients are sensible being larger for physician share (the highest human capital group) compared to nurse share. The omitted base group of Health Care Assistants has the shortest training period of the three groups. Nevertheless, even after controlling for skill mix, the outside wage enters the regression with a significantly negative sign, suggesting that hospitals located in areas with a stronger labour market

[26] The results were robust to including different measures of hospital size as extra controls (e.g. total staffing, number of finished consultant episodes, total number of beds or the total admissions). These terms were never significantly different from zero.

tend to have significantly higher death rates. A 10% increase in outside pay is associated with a 3% increase in AMI death rates.

Column (2) of Table 2 considers long-differenced specifications (annualized three year differences)[27]. The patterns of signs and significance on the key variables are the same as OLS in levels, but the marginal effects are larger in magnitude. Finally, column (3) contains our preferred GMM specification that treats skill shares as endogenous. This estimator exploits the "within" information used in columns (2) and the levels information used in column (1). The key coefficients are statistically significant and the skill shares larger in absolute magnitude compared to the previous column.

The diagnostics are given at the base of Table 2. For the instrumental variables to be valid for GMM-SYS there should be no second order serial correlation in the differenced residuals and no correlation of the error term with the instruments. The high p-values on the LM(2) and Sargan test are consistent with the validity of the instrument set.

VI.B Hospital Productivity

Table 3 repeats the analysis on the same sample as Table 2 but uses productivity as the dependent variable as measured by the log of the number of finished consultant episodes in the year per whole time equivalent clinical worker. The order of the specifications is identical to Table 2 and the control variables are the same except we use a longer vector of case mix controls (admission rates to the hospital across 36 age-gender cells). Column (1) shows that the share of physicians is significantly and positively associated with higher productivity, while the share of nurses is also positive, but statistically insignificant. Importantly, outside area pay is associated with significantly *lower* productivity. A 10% increase in outside pay is associated with a 4.6% decrease in productivity.

[27] We focus on long-differences to reduce the attenuation bias associated with transitory measurement error. Including a full set of hospital dummies (within groups) leads to similar marginal effects with larger standard errors. For example the coefficient on the outside wage in an identical specification to column (1) estimated by within groups is 0.433 with a standard error of 0.324.

The long-differenced results increase the coefficient on physician share and nurses' share (which is now significant at the 5% level). The outside pay variable becomes insignificant however. The final column of Table 3 has the preferred GMM results. Here, all three key variables are correctly signed and significant. The magnitudes of outside pay and physician share are similar to OLS, but nurse share is larger. Again, the diagnostics at the base of the column fail to reject instrument validity.

In summary, and taking both tables together, we have found two key results. First, a richer skill mix, in particular using more physicians, appears to have a positive effect on raising the quality and quantity of hospital output. This is what basic human capital theory would predict although, to our knowledge, this has not been widely demonstrated before in the health sector. Secondly, higher outside wages tend to depress the quality and productivity of hospitals. This is a more controversial finding. In particular, we find that these outside wage effects exist even after conditioning on skill inputs so the outside wage coefficient is not simply reflecting a lower *quantity* of key staff. We consider the other mechanisms through which external labour markets may be having an effect on hospital quality and productivity below.

VI.C Inside Wages

According to our model effort, quality and labor supply are determined by the comparison between the inside wage in the hospital and the outside wage in the labor market. Our empirical work up until now has focused on the outside wage, however, because of the absence of exogenous variation in the inside wage. In Table 4 we look at this in more detail. In fact there is a regional and grade-specific component of the regulated wage that does vary over time and across regions, so we can use this to construct a National Pay Review Body inside wage (we call this the "NPRB predicted inside wage"). Unfortunately, a lot of the time variation across regions took place prior to our sample period in the late 1980s and early 1990s when extra increases were given to nurse grades in different parts of the high wage South-East (see Gosling and Van Reenen, 2006). Consequently, the instrument may have insufficient variation to identify performance effects.

Column (1) of Table 4 includes the (instrumented) inside wage in the AMI death rate regressions. The marginal effect is negative as expected (higher inside wages are associated with a reduction in AMI death rates) but is insignificant. A similar result is obtained in column (2) where we use the NPRB inside wage directly in the regression. The coefficient is large and negative, but insignificant at conventional levels. The final two columns use productivity as the dependent variable. The strongest evidence is in column (3) where we do find a positive and significant coefficient on the inside wage – a 10% increase in hospital pay is associated with a 2.5% increase in productivity. As expected the marginal effect of the outside wage increases in absolute magnitude (as inside and outside wages as positively correlated). The final column includes the NPRB predicted inside wage and although correctly signed the coefficient is not significant.

Overall, we do find weak evidence that inside wages matter for hospital performance, but identifying their effects is much harder than the effect of the outside wage.

VI.D Magnitudes

These results appear to be statistically significant, but are they economically significant? Using the estimates in Tables 2 and 3, we can compare the effect of a change in skill mix and a change in the outside wages on the quality and quantity of hospital output. Recall that the staffing variables are expressed as shares (the variable ranges between zero and unity) and the AMI variables as expressed in natural logarithms of rates. From Table 3, column (3), a one-percentage point increase the share of physicians is associated with a reduction in AMI deaths rates of 5.3%. A one-percentage point increase in nurse share of total staffing is predicted to decrease the AMI death rate by 2.2%. Cross-sectionally, we can compare the effect of moving from the 10^{th} to the 90^{th} percentile of the sample distribution of each of the staff share variables on AMI deaths and output. There is about a seven percentage point difference in physician share between the top and bottom decile of hospitals (using 1996 values) so a change of this size would be predicted to lead to a fall in AMI death rates of 37%. Since the worst 10% of hospitals had an AMI death rate 62% (10/16) above the best hospitals, we predict that increasing the physician share by seven percentage points could account for 60% of this difference in the "quality spread"[28].

[28] In a time series context, in the median hospital in our sample increased the share of physicians by about 2 percentage points and the death rate from AMI fell by about 32% (a 7 percentage point fall on a 22% base) between 1995 and 2002. Taken literally, this implies that the increase in the physician share

Turning to the estimated effect of labour market tightness, a 10% increase in the outside wages (holding the inside wage and labour inputs fixed) is associated with a 4.3% increase in death rates (Table 2 column (3)) and a 5% fall in productivity (Table 3 column (3)). The decile ratio of outside wages between areas is about 33% (in 1996) so a move from the worst to best decile of labour markets is associated with a 14.2% increase in death rates (or just under one quarter of the 62% quality spread quality spread).

The difference in productivity in 1996 between the bottom and top decile was 62 log points in our sample. A seven-percentage point increase in physician share (i.e. moving from the bottom to the top decile across hospitals) is associated with a 35% increase in productivity, implying that physician share could account for almost 58% of the productivity dispersion. Moving from the best to worst decile of labour markets is associated with a 16.3% increase in productivity (so over a quarter of the between hospital productivity distribution).

Although all of these calculations are very crude, they suggest that labour markets and skills are potentially could be very important in accounting for the cross sectional health inequalities that have been a focus of the existing literature.

VI.E. What is the mechanism through which higher outside wages affect hospital outcomes? The role of temporary agency staff

The estimates above show that quality and quantity of output are positively associated with the quantity of the most qualified staff and the outside labour market wage. We have made the assumption that outside wage is associated with the quality of staffing: we examine here whether these alternative measures of staff quality over the period are directly associated with outside wages. One channel that has been suggested (e.g. National Audit Office, 2001) is through the greater reliance of temporary, agency staff in the high wage areas. If these agency staff have lower general or hospital specific human capital they may depress hospital outcomes. In a related literature several

---

could account for almost a third of the fall in death rates in hospitals over this period ((2*5.3)/32=33%).

recent papers (e.g. Autor and Houseman, 2005) have suggested that temporary jobs are not "stepping stones" to better careers which is consistent with the notion that workers build up little human capital in these positions.

We have information available to construct a measure of the intensity to which hospitals rely on nursing agency staff for a sub-sample of the data. We use the proportion of total staff costs accounted for by agency nurses as our key indicator in Table 5 ("agency"). The first column simply regresses the intensity of use of agency nurses on the outside pay rate (and other controls). There is a highly significant correlation, suggesting that agency nurses are used much more intensively in the high outside wage areas, other things equal. Column (2) then includes the agency term directly in a hospital AMI regression identical to our preferred model in column (3) of Table 2[29]. A greater use of agency staff is associated with significantly higher death rates in hospitals: a doubling of the intensity of use of agency staff is associated with a 7.6% increase in the death rate. Column (3) simply repeats the preferred specification on the sub-sample with non-missing agency nurse information with only outside wages for comparison purposes. The results are very similar to the larger sample. Then column (4) has a "horse race" with both agency and outside wage measures entered simultaneously. The outside wage coefficient falls to under a third of its value in the previous column and is no longer significant at conventional levels. The coefficient on the agency variable also falls, but it remains significant at the 5% level. This suggests that a significant part (if not all) of the way that the outside labor market is affecting productivity is through greater agency staffing.

The next three columns of Table 5 repeat the experiment but use productivity instead of AMI death rates as an outcome measure. In column (5) we show that a greater use of agency staff is associated with significantly lower productivity. Column (6) shows that the marginal effect of the outside wage on productivity is higher in the sub-sample where we have agency staff information than the overall sample. In column (7) the magnitude of the outside wage coefficient has fallen by about 0.08 (or 10% of its value in the previous column), but is remains significant at the 5% level. The same is also true for agency staffing.

[29] We treat the agency variable the same as the other hospital-level variables like physician share and instrument it with past values in the GMM approach.

The results are suggestive in Table 5 that agency staffing may be part of the mechanism through which higher outside wages negatively affect hospital outcomes, although it may not be the only mechanism (at least for productivity).

VI.F. Robustness Checks

We describe here a sample of the large number of robustness checks we performed on the main results. These are summarized in Table 6. All cells report the coefficient and standard error on the outside wage from separate regressions. The first column has AMI death rates and the second column has productivity. We begin in row (1) with a baseline regression taken from the final columns of Table 2 and Table 3. The other regressions use this as the baseline in the rest of the table.

There were some changes in the policy environment facing the NHS over our sample period. In particular, the Conservative administration experimented with the "Internal Market" reforms beginning in the early 1990s. This allowed for greater local competition between hospitals encouraging hospitals to compete for patients and staff. There was some greater degree of local flexibility allowed for in pay setting. The internal market reforms were largely abolished when the Conservatives lost the election to Tony Blair in 1997. If the outside wage really reflects labor market issues, we would expect this to have a stronger effect in 1997 and beyond when hospitals had less pay flexibility relative to 1996 and before ("the Internal Market period"). Row 2 tests this idea by allowing the outside pay coefficient to be different in the Internal Market period. Consistently with our model, AMI death rates where significantly less sensitive to outside pay in the Internal market period than they were subsequently (the interaction of the outside wage with the Internal Market period dummy is negative and significant at the 10% level). The interaction is not significant in the productivity, equation, however.

An alternative explanation for the importance of the outside wage is that hospitals in high outside wage areas face sharper budgetary constraints. The funding formula for the NHS contains a "market forces factor" that should reflect the higher costs in more expensive areas, but it may not fully compensate. Consequently, hospitals in high wage areas may be chronically under-funded and this could cause worse quality and productivity. To test for this idea we included measures of the hospital's financial surplus (or deficit) as an additional control. In row 3 we show that the coefficient on the outside wage is very similar in the AMI equation and slightly lower (but still significant) in the productivity regression.

We were concerned that we may have misspecified the econometric model and not allowed for sufficient dynamics. The specification in row 4 of Table 6 includes a lagged dependent variable (treated as endogenous) and presents the long-run effects of outside wages. Although the lagged dependent variable was significant, the long-run effects of the outside wage remain significant being larger in absolute magnitude for the AMI equation and a bit smaller for the productivity equation.

According to Figure 1, the high outside wage areas may be affected more by the regulated wage than the low cost areas (for example, the regulated wage may be close to the "free market" equilibrium wage for the North). Consequently, we would expect a larger effect in London than in the rest of the country. We experimented with dropping London (Inner and Outer) from the sample in row 5. As we expect, the marginal effects are somewhat smaller in this reduced sample for both AMI and productivity, but they remain significant at the 10% level.

We also tried dropping some outliers in the change in the outside wage in row 6 and dropping the regional dummies in row 7. This generally strengthens the results. In row 8, we use an alternative measure of the outside wage – the regional wage in each of the nine regions (i.e. much more aggregated than the one hundred plus area outside wages used in our main analysis). The marginal effects are larger in magnitude but statistically insignificant unless we drop the regional dummies (row 9). Finally, we include a variable with total employees (including non-clinical workers). The variable is insignificant and the coefficient on outside wages falls, but is still significant at the 5% level for the AMI equation and the 10% level for the productivity equation.

Overall then, our results appear robust to a wide variety of experiments.

VII Conclusions

This paper has examined the impact of skills and wages on the performance of hospitals. There is a small but growing literature in labor economics econometrically estimating the impact of human capital and labour markets on firm productivity in the private sector. However, there are hardly any studies in the health sector, an important and rapidly growing part of the economy. Those studies that do exist have difficulty in identifying impacts because there is substantial unobservable heterogeneity between hospitals. To tackle this problem we have assembled a new longitudinal dataset on acute hospitals in England. We examine the impact of skills and the external labor market on hospital productivity and hospital quality (as measured by death rates from AMI).

We find two key results. Consistent with basic production theory, hospitals with a richer mix of skills have better quality health outcomes and higher productivity. Our second key result is that hospitals who operate in areas with high outside wages (relative to inside wages) suffer from lower quality and lower productivity. This is not simply because they have trouble maintaining high enough staffing levels as we condition on labour inputs. This is more likely to be due to the difficulty of retaining high quality staff and/or the lower levels of effort that are supplied when outside wages are high. The reliance on temporary agency staff instead of permanent staff is an important mechanism in generating these worse outcomes. This finding is related to the emerging literature on the quality of temporary jobs for workers (e.g. Autor and Houseman, 2005).

From a policy perspective, our study has important implications for regulated labor markets. The National Health Service, a quasi-monopoly provider, dominates the UK health system and wages for physicians and nurses are determined centrally. The local variation of wages does not fully reflect the higher outside wage in areas where the labour market is tight (such as London and the South East). The low relative wages in these high outside wage areas appear to have a direct impact on the death rates in hospitals and the level of their productivity. Changing the system of wage setting to allow wages to reflect market realities would be predicted to both improve productivity and save lives in the higher wage areas.

References

Ackerberg, Daniel, Lanier Benkard, Steven Berry and Ariel Pakes (2005) "Econometric Tools for Analyzing Market Outcomes", Harvard University mimeo prepared for *Handbook of Econometrics*, James Heckman and Ed Leamer (eds)

Akerlof, George (1982) “Labor Contracts as partial gift-exchange” *Quarterly Journal of Economics*, November

Aiken, Clarke, Sloane, Sochalski and Silber (2002) “Hospital Nurse Staffing and Patient Mortality, Nurse Burnout and Job Dissatisfaction” *Journal of the American Medical Association* 288, 19 October 23/30

Aiken, Clarke, Cheung, Sloane and Silber (2003) Education Levels of Hospital Nurses and Surgical Patient Mortality *Journal of the American Medical Association* Vol 290 No 12 September 24.

Askildsen, Jan Erik, Baltagi, Badi H and Holmas, Tor Helge (2002) “Will increased wages reduce shortages of nurses? A panel data study of nurse labour supply” Working paper 2102, Department of Economics Bergen.

Antonazzo, Emmanuel,Scott, Anthony, Skatun, Diane and Elliott, Bob (2003) "The labour market for nursing: a review of the labour supply literature" *Health Economics*, 12: 467-478

Arellano, Manuel and Meghir, Costas (1992) "Female labour supply and on the job search: an empirical model estimated using complementary datasets" *Review of Economic Studies*, 59: 537-559

Audit Commission (2001) *Brief Encounters: getting the best form Temporary Nursing Staff,* London: Audit Commission

Autor, David and Susan Houseman (2005) “Do temporary help jobs improve job prospects for low skilled workers: evidence from randomized assignment” MIT mimeo

Baltagi, Badi H, Bratberg, Espen and Holmas, Tor Helge (2003) “A panel data study of physicians’ labour supply: the case of Norway” CESIfo Working Paper 895.

Besley, Timothy and Maitreesh Ghatak (2005) “Motivated agents” *American Economic Review*

Bilodeau, Daniel, Cremieux, Pierre-Yves, Ouellette, Pierre (2000) “Hospital Cost Functions in a non-Market Health Care system” *The Review of Economics and Statistics* 82*,* 489-498.

Bilodeau, Daniel, Cremieux, Pierre-Yves, Jaumard, Brigitte, Ouellette, Pierre, Vovor Tsevi (2004) “Measuring Hospital Performance in the Presence of Quasi-Fixed Inputs: An Analysis of Quebec Hospitals” *Journal of Productivity Analysis 21,* 183-199.

Blundell, Richard, Duncan, Alan and Meghir ,Costas (1998) "Estimating labour supply responses using tax policy reforms" *Econometrica*, 66: 827-861

Blundell, Richard and Stephen Bond (2000), "GMM estimation with persistent panel data: An application to production functions", *Econometric Reviews* 19(3), 321-340

Blundell, Richard and Bond, Stephen (1998) “Initial Conditions and Moment Restrictions in dynamic panel data models” *Journal of Econometrics*, 88, 115-143.

Bond, Steven and Måns Söderbom (2005) "Adjustment costs and the identification of Cobb Douglas production functions", Institute for Fiscal Studies Working Paper W05/04.

Borjas (2004) “The Wage structure and the sorting of workers into the public sector” NBER Working paper No. 9313

Bulman J. (2002) “Patterns of Pay: results of the 2002 New Earnings Survey” *Labour Market Trends* 110;12: 651-653.

Chevalier, Arnaud and Dolton, Peter (2003) "The pay, working hours and job satisfaction of health care professionals" Centre for Economic Performance mimeo

Cowing Thomas and Alphonse Holtmann (1983) “The multi-product short-run hospital cost function: Empirical Evidence and Policy Implications from cross sectional data” *Southern Economic Journal* 49, 637-653.

Curtin (2003) “An Integrated Analysis of Nurse Staffing and Related Variables: Effects on Patient Outcomes” *Journal of Issues in Nursing* 2003

Dearden, Lorraine, Howard Reed and John Van Reenen (2005) "Training and Corporate Productivity: Evidence from a panel of UK industries" Forthcoming, *Oxford Bulletin of Economic and Social Research*.

Disney Richard and Amanda Gosling (1998) Does it pay to work in the public sector? *Fiscal Studies* 19: 347-374.

Elliott, B, Ma, A, Scott, A, Bell, D, and Roberts, E (2005) “Geographically differentiated pay in the labour market for nurses” Mimeo, University of Aberdeen.

Elliott, Bob (2003) “Labour Markets in the NHS: An Agenda for Research" *Health Economics*, 12: 797-801

Erickek, George, Susan Houseman and Arne Kalleberg (2003) “The Effects of Temporary Services and Contracting Out on Low Skilled Workers: Evidence from Auto Suppliers, Hospitals and Public Schools” in Applebaum, Eileen, Annetter Bernherdt and richard Murnane (eds) *Low wage America: How Employers aee reshaping Opportunity in the Workplace,* NY: Russel Sage

Finlayson, B, Dixon, *J*, Meadows, S Blair, G,*(* 2002) **“**Mind the gap: the extent of the NHS nursing shortage” *British Medical Journal* 325(7363): 538–541.

Frijters, Paul, Shields, Michael and Price, Stephen (2003) “Investigating the quitting decisions of nurses: Panel Data Evidence from the British National Health Service” IZA Discussion Paper No. 794

Foster, L., Haltiwanger, J. and Syverson, C. (2005) “Reallocation, Firm turnover and efficiency: Selection on productivity or profitability?” NBER Working Paper 11555

Gaynor, M (2004) “Competition and Quality in Hospital Markets. What do we know? What don’t we know?” *Economie Publique* 15,2, 3-40.

Gaynor, M, Rebitzer, M and Taylor, L (2002) “Incentives in HMOs” CMPO, University of Bristol DP 03/089.

Gaynor, Martin and Pauly, Mark V (1990) “Compensation and productive efficiency in partnerships: evidence from medical group practice,” *Journal of Political Economy* 98, 544-573.

Gosling, Amanda and John Van Reenen (2005) “Using Institutionalised pay setting to identify the effect of wages on employment: The Case of Nurses”, mimeo Centre for Economic Performance

Hellerstein, Judith, Neumark, David and Troske, Ken (1999) “Wages, productivity and worker characteristics: evidence from plant level production functions and wage equations”, *Journal of Labor Economics*, 17, 409–446.

Holmas, T (2002) "Keeping Nurses at Work: a duration analysis" *Health Economics*, 11, 493 - 503

Houseman, Susan, Arne Kallenerg and George Erickek (2003) “The Role of Temporary Help Employment in Tight Labor Markets” *Industrial and Labor Relations Review*, 571, 1, 105-127

Jones, A (2000) “Health Econometrics” in A J Culyer and J P Newhouse (eds) *Handbook of Health Economics*, Vol 1A, Chapter 6. Amsterdam: North Holland

Kessler, D P and McClellan (2002) “Is hospital competition socially wasteful?” *Quarterly Journal of Economics* 115: 577-615.

Kooreman, Peter (1994) “Data Envelope Analysis and Parametric Frontier Estimation: Complementary Tools” *Journal of Health Economics 13,* 345-346.

Laing and Buisson (2005) *Laing's Healthcare Market Review (18th edition).* London: Laing and Buisson

Levine, David (1988) "Can Wage increases pay for themselves: Tests with a production function" *Economic Journal* 102(414) 1102-1115

Machin, Stephen and Manning, Alan (1992) "Testing Dynamic Models of Worker Effort" *Journal of Labor Economics,* 10 (3) 288-305

Mark, Harless, McCue and Xu (2004) "A Longitudinal Examination of Hospital Registered Nurse Staffing and Quality of Care" *Health Services Research* 39:2 April.

McClellan, M and Staiger, D (2000) "Comparing the Quality of Health Care Providers" in A Garber (ed) *Frontiers in Health Policy Research* MIT Press, 3, 113-136

McGuire, Thomas G and Pauly, Mark V 1991, "Physician response to fee changes with multiple payers," *Journal of Health Economics* 10, 385-410.

Morris, Steven and McGuire, Alistair (2002) "The private net present value and private internal return to becoming a nurse in Great Britain" *Applied Economics*, 34, 2189-2200

Needleman, Buerhaus, Mattke, Stewart and Zelevinsky (2002) "Nurse-Staffing Levels and the Quality of Care in Hospitals" *New England Journal of Medicine* 346, 22,

Newhouse, Joseph P (1994) Frontier Estimation: How Useful a Tool for Health Economics? *Journal of Health Economics 13, 317-322.*

Nickell, Stephen and Quintini, Glenda (2002) "The consequences of the decline in public sector pay in Britain: A Little bit of evidence" *Economic Journal* 112, F107-F118

Noether, Monica (1986) "The growing supply of physicians: has the market become more competitive?" *Journal of Labor Economics*, 4, 503-537.

Office of Manpower Economics (2003) "Workforce Survey Results for nursing staff, midwives and health visitors" www.ome.gov.uk

Philips, VL. (1995) "Nurses Labour Supply: Participation, hours of work and discontinuities in the supply function" *Journal of Health Economics*, 14: 567-582

Review Body for Nursing Staff, Midwives and Professions Allied to Medicine (2004), *Annual Report* HMSO

Rice, Nigel (2003) "The labour supply of nurses in the UK: Evidence from the British Household Panel Survey", mimeo Centre for Health Economics, University of York

Rizzo, John A and Blumenthal, David 1994, "Physician labor supply: Do income effects matter?" *Journal of Health Economics* 13, 433-453.

Ruhm, Christopher (2006) "A healthy economy can break your heart" NBER Working Paper No. 12102

Shapiro, Carl and Joseph Stiglitz (1984) "Equilibrium Unemployment as a worker discipline device" *American Economic Review*, 74: 433-44

Scott, A (2000) "The Economics of General Practice" in A J Culyer and J Newhouse (eds) *The Handbook of Health Economics*. Amsterdam: North Holland

Shields, Michael. and Ward, Melanie (2001) "Improving nurse retention in the National Health Service in England: the impact of job satisfaction on intentions to quit" *Journal of Health Economics*, 20, 677-701

Showalter, Mark H and Thurston, Norman K (1997) "Taxes and labor supply of high-income physicians," *Journal of Public Economics* 66, 73-97

Sloan, Frank A (1975) "Physician supply behavior in the short run," *Industrial and Labor Relations Review* 28, 549-569.

Staiger, Douglas , Spetz, Joanne and Phibbs, Ciaran (1999) "Is there monopsony in the labour market? Evidence from a natural experiment" NBER Working Papers 7258

Vita, Michael G (1990) "Exploring Hospital Production Relationships with Flexible Functional Forms" *Journal of Health Economics 9,* 1-21.

Wadhwani, Sushil and Martin Wall (1991) "A Direct Test of the Efficiency Wage Hypothesis using UK micro data" *Oxford Economic Papers* 43(4) 529-548

Table 1: Summary of main variables

	Mean	Standard deviation	Min	Max
AMI Variables				
AMI death rate (55 plus)	21.125	4.520	2.964	36.941
Total AMI deaths (55 plus)	7993.624	3382.425	1100	29400
Total AMI admissions (55 plus)	384.958	160.261	151	1348
Productivity and FCE (finished Consultant Episodes)				
Productivity (total FCEs/ total staffing)	30.981	7.718	5.094	65.121
Total FCEs	58,620.82	2,441.15	13,490	138,984
Staffing Variables				
Total staffing (physicians+nurses+AHP+Health Care Assistants)	1909.447	774.049	432.9	4269.83
Physicians share of staffing	0.130	0.030	0.047	0.249
Nurses (plus qualified Allied Health Professionals) share of staffing	0.646	0.034	0.493	0.765
Hospital Expenditure Variables				
Share of expenditure on Agency staff as a proportion of total expenditure	0.035	0.028	0.001	0.163
Wage Variables				
Ln(Area outside wage)	9.602	0.141	9.272	9.987
Ln(Predicted NPRB wage)	9.711	0.088	9.558	9.991
Other variables				
Standardized Mortality rate in the local area (per 100,000)	723.034	77.232	518.73	944.21
Proportion of elective admissions (to total admissions)	0.416	0.187	0	1
Proportion of emergency admissions (to total admissions)	0.415	0.170	0	1
Proportion of transfer admissions (to total admissions)	0.169	0.124	0	0.919

Notes

Data for 1996-2001, all acute trusts, 907 observations over 211 hospital trusts. Other case mix variables are admissions within 5 year age-gender bands for AMI (55+) and total admissions (all ages). Staffing refers to whole time equivalent clinical staffing.

Table 2: The quality of hospitals – Skills and wage effects

Dependent variable	Ln(AMI Rate)	Ln(AMI Rate)	Ln(AMI Rate)
Estimation technique	OLS	Long Differences	GMM-SYS
	(1)	(2)	(3)
Ln (Area outside pay)	0.303**	0.823**	0.431**
	(0.150)	(0.381)	(0.188)
Physicians share	-1.107***	-2.198**	-5.267**
	(0.359)	(0.883)	(2.753)
Nurses share	-0.524*	-1.435**	-2.194*
	(0.276)	(0.638)	(1.262)
Hospital fixed effects	No	No	Yes
Casemix controls (14)	Yes	Yes	Yes
Year dummies (7)	Yes	Yes	Yes
Region dummies (10)	Yes	No	Yes
SC(1) p-value			0.002
SC(2) p-value			0.171
Hansen-Sargan p-value			0.763
No of Hospitals	211	211	211
Observations	907	348	907

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%
Quality is measured by the within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI) for over 55 year olds admitted with AMI to the hospital. Casemix controls are the proportion of total emergency admissions for AMI made up by each 5-year age-gender band from age 55 upwards. All regressions control for area mortality rates, (lagged) employment size and hospital type (i.e. whether the acute hospital was a specialist hospital, teaching hospital or "normal" hospital). Long-differences are three- year annual average growth rates. In the System-GMM estimates, one-step robust estimates are presented. Physician share, nurse share and total employment are treated as endogenous (the outside wage is lagged and treated as exogenous). In the GMM specification instruments in the differenced equations are levels of own values t-2 through t-5; instruments in the levels equations are once lagged differences. SC(k) is Arellano-Bond (1991) test of serial correlation of order k of the first differenced residuals. Sargan-Hansen is a test of the over-identifying restrictions. Time period is 1996-2001. Standard errors in parentheses under coefficients are robust to arbitrary heteroskedacity and autocorrelation.

Table 3: Productivity in hospitals – Skills and wage effects

Dependent variable	Ln(Productivity)	Ln(Productivity)	Ln(Productivity)
Estimation technique	OLS	Long Differences	GMM-SYS
	(1)	(2)	(3)
Ln (Area outside pay)	-0.454***	0.241	-0.495**
	(0.159)	(0.275)	(0.230)
Physicians share	5.552***	2.869***	4.654***
	(0.434)	(0.507)	(0.905)
Nurses share	0.149	1.071***	1.523**
	(0.225)	(0.369)	(0.701)
Hospital fixed effects	No	No	Yes
Casemix controls (36)	Yes	Yes	Yes
Year dummies (7)	Yes	Yes	Yes
Region dummies (10)	Yes	No	Yes
SC(1) p-value			0.010
SC(2) p-value			0.218
Hansen-Sargan p-value			0.107
No of Hospitals	211	211	211
Observations	907	348	907

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%. Productivity measured as the number of Finished Consultant Episodes (FCEs) per whole-time equivalent (WTE) employee; casemix controls are proportion of admissions in five year age-gender cells. Quality is measured by the within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI) for over 55 year olds admitted with AMI to the hospital. Casemix controls are the proportion of total emergency admissions for AMI made up by each 5-year age-gender band from age 55 upwards. All regressions control for area mortality rates, (lagged) employment size and hospital type (i.e. whether the acute hospital was a specialist hospital, teaching hospital or "normal" hospital). Long-differences are three- year annual average growth rates. In the System-GMM estimates, one-step robust estimates are presented. Physician share, nurse share and total employment are treated as endogenous (the outside wage is lagged and treated as exogenous). In the GMM specification instruments in the differenced equations are levels of own values t-2 through t-5; instruments in the levels equations are once lagged differences. SC(k) is Arellano-Bond (1991) test of serial correlation of order k of the first differenced residuals. Sargan-Hansen is a test of the over-identifying restrictions. Time period is 1996-2001. Standard errors in parentheses under coefficients are robust to arbitrary heteroskedacity and autocorrelation.

Table 4: Inside Wage experiments

Dependent variable	Ln(AMI)	Ln(AMI)	Ln(Productivity)	Ln(Productivity)
Estimation technique	GMM-SYS	GMM-SYS	GMM-SYS	GMM-SYS
	(1)	(2)	(3)	(4)
Average inside wage	-0.240		0.249**	
	(0.188)		(0.096)	
Predicted ln(inside wage using NPRB IV)		-0.693		0.237
		(1.113)		(0.698)
Ln (Area outside pay)	0.427**	0.449**	-0.688***	-0.476**
	(0.210)	(0.190)	(0.216)	(0.215)
Physicians share	-4.096**	-6.046**	2.911***	5.798**
	(2.228)	(2.336)	(1.008)	(1.002)
Nurses share	-1.945*	-2.375**	-0.113	1.579*
	(1.153)	(1.134)	(0.601)	(0.610)
Hospital fixed effects	Yes	Yes	Yes	Yes
Casemix controls (36)	Yes	Yes	Yes	Yes
Year dummies (7)	Yes	Yes	Yes	Yes
Region dummies (10)	Yes	Yes	Yes	Yes
SC(1) p-value				
SC(2) p-value				
Hansen-Sargan p-value				
No of Hospitals	211	211	211	211
Observations	706	706	706	706

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%. Quality is measured by the within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI) for over 55 year olds admitted with AMI to the hospital. Productivity measured as the number of Finished Consultant Episodes (FCEs) per whole-time equivalent (WTE) employee; casemix controls are proportion of admissions in five year age-gender cells. All regressions control for area mortality rates, (lagged) employment size and hospital type (i.e. whether the acute hospital was a specialist hospital, teaching hospital or "normal" hospital). System-GMM estimates with one-step robust estimates are presented. Physician share, nurse share, employment size and lagged inside wage are treated as endogenous. Outside wage is lagged and treated as exogenous. Instruments in the differenced equations are levels of own values t-2 through t-5; instruments in the levels equations are once lagged differences. We also use the lagged predicted inside wage as an instrument. SC(k) is Arellano-Bond (1991) test of serial correlation of order k of the first differenced residuals. Sargan-Hansen is a test of the over-identifying restrictions. Time period is 1996-2001. Predicted inside wage using NPRB IV uses the predicted regulated wage as described in the text (coefficients and standard errors divided by 10).

Table 5: The role of agency staff expenditure in accounting for the impact of the labour market on hospital quality and productivity

Dependent variable	Ln(Agency)	Ln(AMI)	Ln(AMI)	Ln(AMI)	Ln (productivity)	Ln (productivity)	Ln (productivity)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln (Area outside pay)	2.557*** (1.131)		-0.423** (0.189)	-0.131 (0.254)		-0.703** (0.232)	-0.622* (0.235)
Ln(Agency)		-0.091*** (0.028)		-0.076** (0.029)	-0.100*** (0.031)		-0.046** (0.021)
Hospital fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (7)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SC(1) p-value	.886	.007	0.010	.008	.007	0.010	.008
SC(2) p-value	.132	.126	0.239	.180	.126	0.239	.180
Sargan p-value	.390	.128	0.124	.161	.128	0.124	.161
No. of hospitals	177	177	177	177	177	177	177
Observations	523	523	523	523	523	523	523

Notes

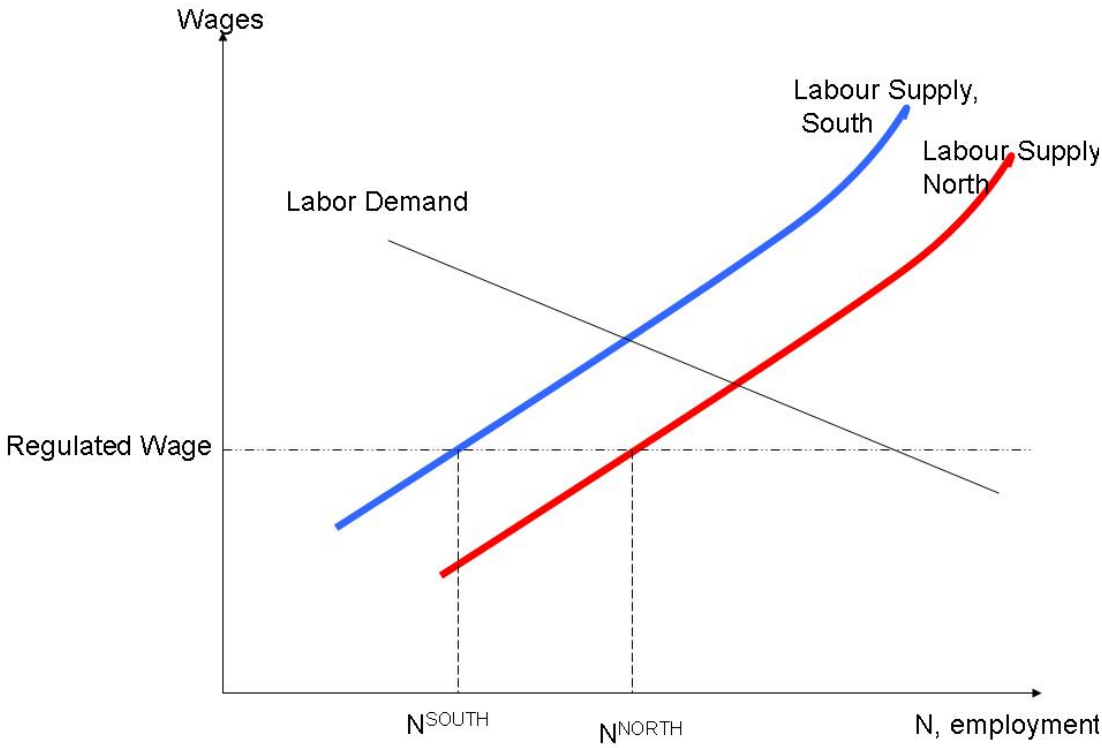
*significant at 10%; ** significant at 5%; *** significant at 1%. “Agency” Expenditure is the share of total staff expenditure that is accounted for by expenditure on non-NHS nursing staff. Productivity measured as the number of Finished Consultant Episodes (FCEs) per whole-time equivalent (WTE) employee. Quality is measured by the within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI) for over 55 year olds admitted with AMI to the hospital. All columns are estimated by System GMM (instruments in the differenced equations are levels of own values t-2 through t-5; instruments in the levels equations are once lagged differences). The specifications in columns (2), (4), (5) and (7) are identical to specifications in Tables 3 and 4 except we also include agency as an additional instrument. All standard errors are robust to arbitrary heteroskedacity and autocorrelation; in the System-GMM estimates one step robust estimates are presented and all staff variables are treated as endogenous (outside wage is lagged and treated as exogenous)

Table 6: Robustness Tests - coefficient(s.e.) on outside wage

	Dependent variable	Ln(AMI)	Ln(Productivity)	Observations
		(1)	(2)	
1	Baseline	0.432** (0.187)	-0.495** (0.230)	907
2	"Internal market" period interaction with outside wage	-0.222* (0.119)	-0.133 (0.092)	907
	Linear outside wage	0.608*** (0.204)	-0.394* (0.229)	
3	Include hospital financial surplus	0.397* (0.199)	-0.441** (0.250)	750
4	Include a lagged dependent variable: long-run effect [p-value]	0.506** [0.02]	-0.487** [0.045]	903
5	Drop Inner and Outer London	0.314* (0.174)	-0.375* (0.224)	782
6	Drop big jumps in outside wage	0.488** (0.227)	-0.549** (0.201)	891
7	Drop Regional Dummies	0.674*** (0.207)	-0.428** (0.170)	907
8	Regional outside wage	1.352 (1.104)	-0.429 (0.598)	907
9	Regional outside wage (drop regional dummies)	0.735*** (0.224)	-0.390** (0.167)	907
10	Include alternative total hospital employment measure	0.364** (0.178)	-0.383** (0.184)	907

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%. Each cell reports the coefficient and robust standard error from a separate GMM regression. Productivity measured as the number of Finished Consultant Episodes (FCEs) per whole-time equivalent (WTE) employee; Quality is measured by the within hospital deaths within 30 days of emergency admission for acute myocardial infarction (AMI) for over 55 year olds admitted with AMI to the hospital. All regressions have the same System GMM estimation and specifications as in final columns of Tables 2 (for the AMI regressions) and Table 3 (for the productivity regressions). Outside wage is (lagged) area wage unless elsewhere specified.

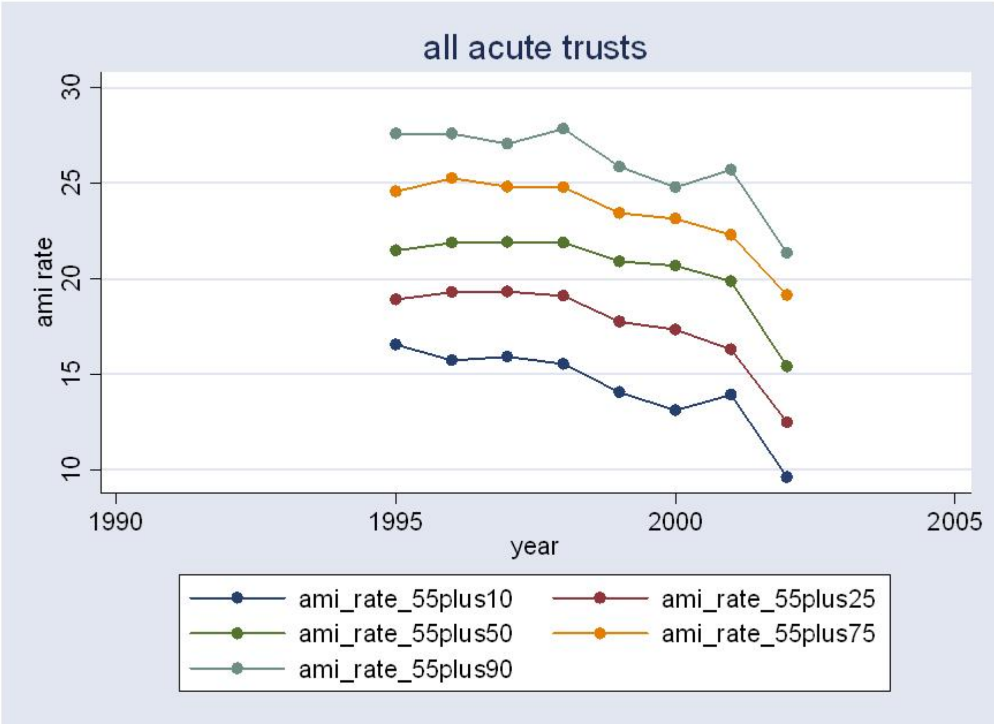
Figure 1: Hypothetical Impact of regulated wage on the labor market



Notes

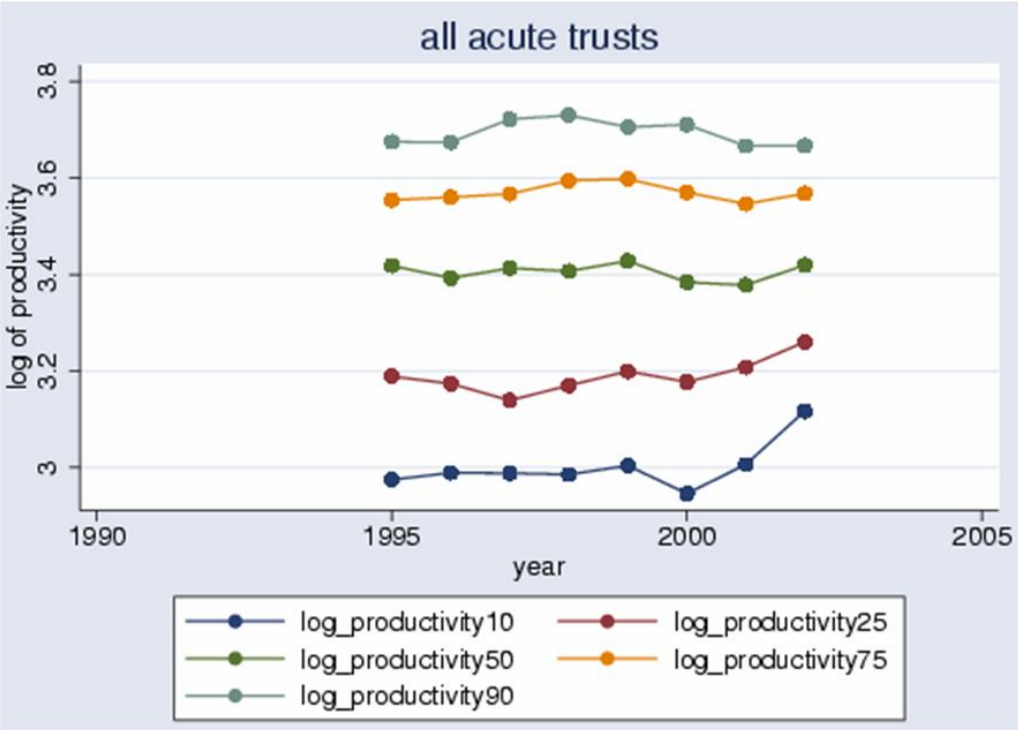
The theoretical impact of a nationally regulated wage on NHS labor supply in two areas with different outside labor markets. South (e.g. London area) has a stronger outside labor market with higher alternative wages than North, so the supply curve lies to the left. A single nationally fixed wage (so long as it is below the competitive level) will result in a lower number of employees in the South than in the North.

Figure 2: The distribution of Death rates from AMI across hospitals, 1995-2002



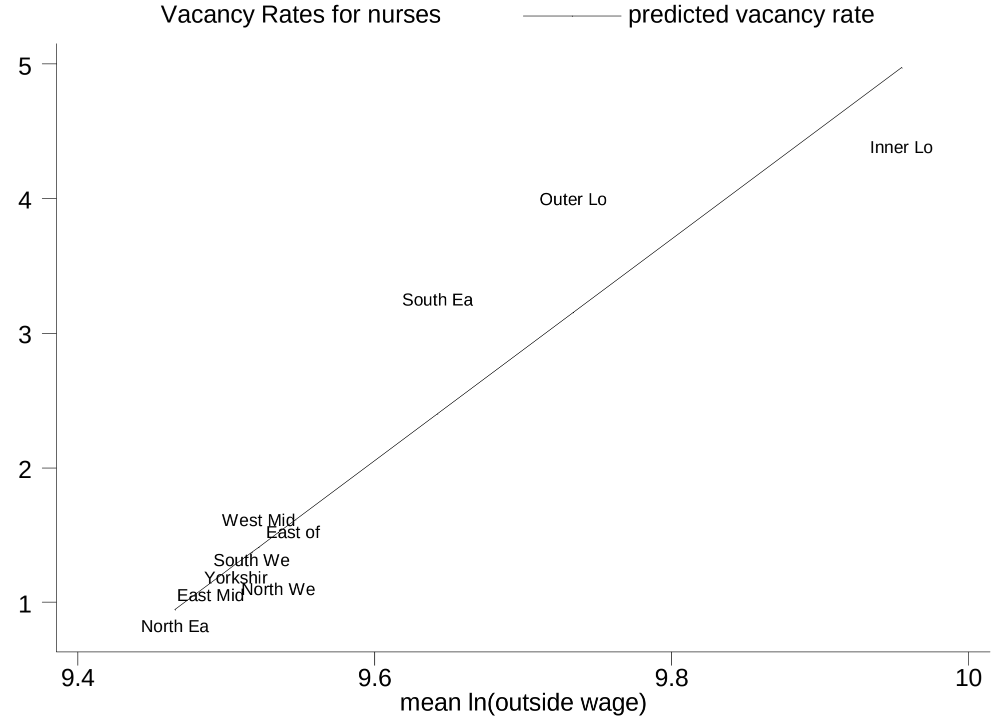
Notes: This is taken from our full panel of acute hospital trusts. The lines plot out the death rates at different percentiles of the distribution. The top line shows the evolution of AMI death rates at the 10th percentile (highest death rates) and the bottom line the evolution of death rates at the 90th percentile (lowest death rates).

Figure 3: The distribution of Productivity (Finished Consultant Episodes per clinical staff member) across hospitals, 1995-2002



Notes: This is taken from our full sample of acute hospital trusts. The lines plot out the level of ln(productivity) at different percentiles of the distribution. The top line shows the evolution of ln(productivity) at the 10th percentile (highest death rates) and the bottom line the evolution of ln(productivity) at the 90th percentile (lowest death rates).

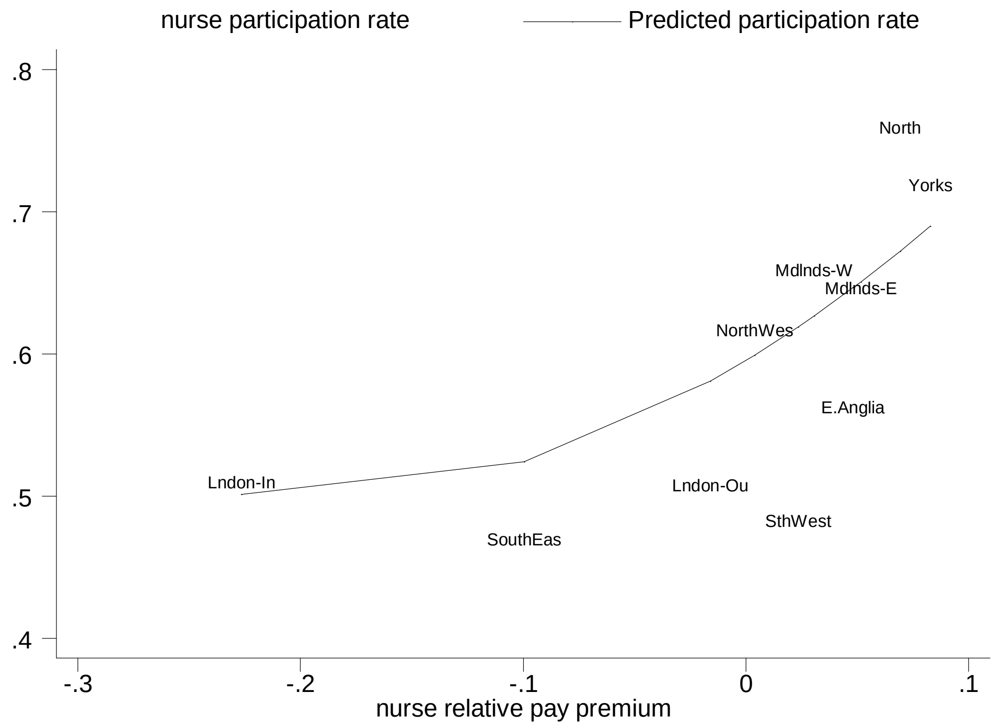
Figure 4: Nurse Vacancy Rates and outside Wages



Notes

Each observation is one of the ten regions in England (the average region has 4.9m people and we have a 1% example of workers from the New Earnings Survey). Outside pay is the average log wages of all female non-manual workers (from New Earnings survey). Vacancy rates are the proportion of nurse posts that have been vacant for three months or more (Office of Manpower Economics).

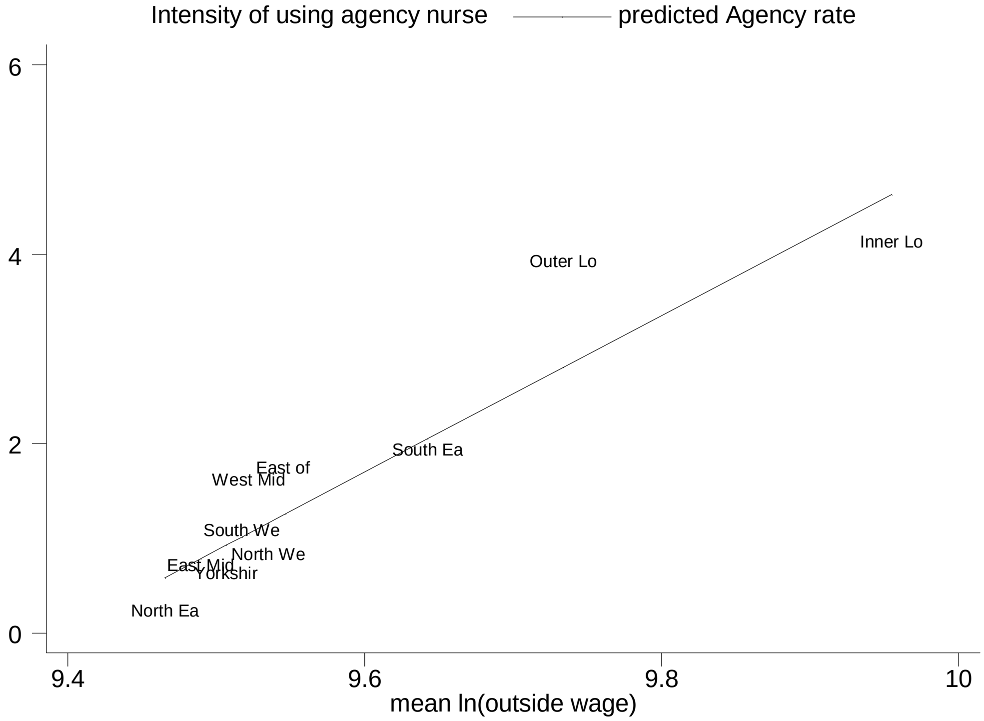
Figure 5: Nurse labor Supply and relative pay – lower relative pay is associate d with a lower nurse participation rate (whole UK)



Notes

Each observation is one of the 12 regions in Great Britain. Nurses' relative pay is the percentage difference in the average hourly pay of a female nurse relative to a non-manual female in the same region (the average region has 4.9m people and we have a 1% example of workers from the New Earnings Survey). Nurse participation rate is the average proportion of females with a nursing qualification who are working as a nurse (alternatives are not working or working in a different occupation).

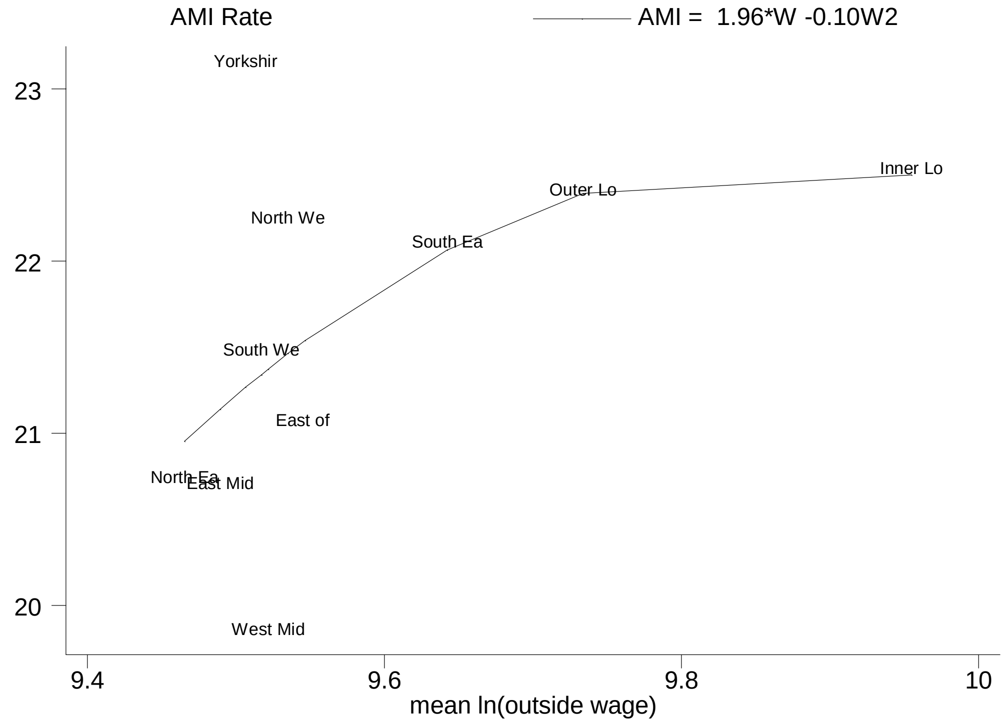
Figure 6: Intensity of use of agency nurses and outside wages



Notes

Each observation is one of the 10 regions in England. Outside pay is the average log wage of all female non-manual workers. Intensity of use of agency nurses is the expenditure on agency staff as a proportion of hospital staff expenditure.

Figure 7: AMI Death Rates and outside wages



Notes

Each observation is one of the 10 regions in England. Outside pay is the average log wage of all female non-manual workers. A.M.I. rates are within hospital deaths within 30 days of emergency admission for acute myocardial infarction for over 55 year olds admitted with AMI to the hospital.

Appendix A: Short literature view of the role of staff in hospital production and the responsiveness of supply to wages

The importance of medical staff in hospital production

There is a large and growing literature on the importance of nurses in hospital production, but somewhat oddly, there has been less focus on the impact of physicians. Most studies of hospital production estimate costs rather than production and many studies of hospital cost functions have ignored physician labour, partly on the grounds that in the US and other hospital systems physicians are not directly employed by hospitals. Bilodeau et al (2000) is a recent exception, finding that hospitals do in the short-run cost minimize with respect to quasi-fixed inputs, including physicians[30]. Jensen and Morrissey (1986) is one of the few papers that focus directly on the impact of physicians on hospital production. They find that increases in physicians increased hospital output but by less than increases in nurses (although by more than medical residents). This is surprising as physicians have longer in training and therefore are likely to have higher human capital. Since the period covered by the data in this study, the amount of training received by nurses has risen and there has been considerable technological change in hospital production.

Most of the literature on the impact of nurses on patient outcomes is from the USA (see Curtin, 2003, for a recent review). Most studies focus on the effect of staffing levels; a few examine the effect of nurse experience. Aiken et al. (2002) use a large cross section of acute care Pennsylvanian hospitals and examine the relationship between two measures of patient care (deaths within 30 days of admission and 'failure to rescue') and nurse staffing ratios (the average number of patients in a nurse's workload). They find an effect of nurse staffing on both outcomes: the odds of patient mortality increased by 7% for every additional patient in the average nurse's workload in the hospital. The same data source is used in Aiken et al. (2003), which investigate the impact the education level of nurses has on patient mortality and finds

[30] The estimation of hospital cost functions has been approached using both parametric methods (e.g. Vita 1990, Cowing and Holtman 1983 and see Jones 2000) and non-parametric methods (e.g. Kooreman 1994; Newhouse 1994; Bilodeau et al 2004).

a positive relationship between education level and the risk of poor outcomes[31]. Needleman et al. (2004) conducted a study based on over 5 million patient discharges from 799 hospitals in 11 states. They used several measure of patient outcomes and hours of nursing care per day (adjusted for the severity of the casemix) to analyze the effect of registered nursing hours on patient outcomes. They found a strong and consistent relationship between nurse staffing and five outcomes for medical patients. A higher number of Registered Nurses also were associated with a 3 to 12% reduction in the rate of adverse outcomes, while a higher staffing level for all types of nurses was associated with a decrease in adverse outcomes, ranging from two to 25%.

The studies cited above and indeed the vast majority of studies on the impact of nurses on quality of care use cross sectional data, so are unable to control for unobserved heterogeneity between hospitals. One recent study uses panel data, allowing control for heterogeneity. Mark et al (2004) examine the impact of nursing staffing levels on the mortality ratio of a hospital (number of observed deaths divided by the number of expected deaths given case mix). Using OLS and within group (WG) regressions they found few statistically significant coefficients on nurse staffing. This is in contrast to the cross sectional studies. Estimating a dynamic panel data model with a lagged dependent variable they found evidence of a positive relationship between Registered Nurses staffing levels and lower mortality ratios, but the lack of support for this in the WG estimates suggests that there may be misspecification in the model.

Is the labour supply of health workers responsive to labour market conditions?

There are relatively few studies of the impact of the labour market conditions on physicians. Most of the literature has focused on responses to different contracts[32]. The relatively few papers on supply have tended to focus on physicians in private practice. Older examples include Sloan (1975) and Noether (1986). More recently, Rizzo and Blumenthal (1994) examine the impact of wage and non-wage income for a sample of self employed US physicians. They find uncompensated wage elasticity for male physicians of 0.23. Showalter and Thurston (1997) also examine US physicians,

[31] Each 10% increase in the proportion of nurses with higher degrees decreased the risk of mortality and failure to rescue by 5%, after controlling for hospital and patient characteristics.
[32] Examples include Gaynor and Pauly (1990), Gaynor et al. (2002), Scott (2000).

distinguishing between those in self-employment and those who are employees. They find a wage elasticity of around 0.33 for all groups, but one that is essentially zero for the employed and is around 0.6 for self-employed solo practitioners. One recent paper has studied physicians employed in the public sector, in an institutional setting not dissimilar to the UK. Baltagi et al (2003) use panel data on 1303 Norwegian male hospital physicians, covering the period 1993-1997. Using GMM techniques, they find long run wage coefficients of around 0.55. These are rather higher than the US estimates: the differences may be due both to institutional settings and to the fact the Baltagi et al (2003) study is one of the few to use panel data.

Annazzo et al (2003) provide a recent update on the empirical evidence of nursing labour supply. The estimates (mainly from North American studies) display a large degree of variation. Most of the studies are of hours, rather than participation per se. The hours margin is relatively inflexible in the UK as there is little choice in the number of hours worked per week in the NHS. The three studies that did look at participation could find no significant effect of wages. Furthermore, the focus is generally on working as a nurse or not working at all. In reality, there is the margin of working as a nurse or in another occupation that is not considered in these papers.

For the UK, Skatun et al (2002) and Frijters et al (2003) both use longitudinal data from the Quarterly Labour Force Survey (QLFS). Frijters et al (2003) look at quitting decision and conclude that wages have a small effect relative to non-wage factors[33],[34]. Skatun et al (2002) look at labor market participation (but not at occupational choice between nursing and other in work options). Both these papers find that the wage elasticity of participation is below unity; the estimates in Frijters et al are less than under 0.1. Phillips (1995) uses a different data set[35] and estimates a response of participation to wages of 1.4. Rice (2003) uses the British Household Panel Survey

[33] Shields and Ward (2001) come to a similar conclusion by comparing pay with non-pay factors, though some of their controls (grade and region) are heavily collinear with pay in the cross section.
[34] Frijters et al use the equation of nurses who leave to construct predicted "outside wages" for the nurses who stay, but these nurses are unlikely to be a random sub-sample of all nurses, so the imputation is dubious. The excluded variables in the wage equation are arbitrary and could easily be included in the quitting equation For example, "unsatisfied with pay in nursing" is in the wage equation but not in the quitting equation.
[35] The Women and Employment survey 1980 is used: the final sample size is 312 observations.

(BHPS), controlling for unobserved heterogeneity by including fixed effects[36]. All these papers find the response of hours to wages to be inelastic. A recent study by Elliot et al (2005) uses a sample of English hospitals for the years 1999-2002 and examines the association between the average (over the four years) wage gap between nurses pay and that of comparable women at regional level and the nursing vacancy rate at hospital level. They find that higher levels of the wage gap are positively associated with higher vacancy rates, but do not exploit the panel nature of their data, so do not to control for heterogeneity across hospitals.

Finally, Gosling and Van Reenen (2005) use the structure of pay determination for public sector nurses to identify the impact of wages on participation in nursing. This approach deals with the endogeneity of wages by using regulatory decisions as an instrumental variable for the observed wage in a (selection adjusted) participation equation. They also allow for the participation choice to include working in another profession, rather than simply to be not-working. They find estimates of the elasticity of wage on employment elasticities are biased downwards in OLS and reasonably large in their IV results, typically around unity.

Studies in another setting in which wages for nurses are set centrally (Norway) have also shown nurses labour supply responds to wages (and other factors). Askildsen et al (2002), using a large panel data set, showed that nurses' labour supply responded to wages. Holmas (2002) using the same data found that both wages and working conditions had an impact on nurses decisions to quit the public health care system.

Appendix B: System GMM Estimation

The basic equation we wish to estimate can be written in simplified form as

$$y_{it} = \pi \; x_{it} + u_{it} \tag{A2}$$

[36] His attempts to deal with transitory shocks are also subject to the problem that wage are instrumented with variables that are arbitrarily excluded from the participation equation (such as degree qualification, experience and the regional participation rate). Exclusion of personal characteristics such as education that shift the entire wage profile can be justified under some restrictive structural assumptions, but when tested these assumptions are usually violated (e.g. Blundell et al, 1998).

Subscript *i* indicates the hospital *t* is time and θ is the parameter of interest. Assume that the stochastic error term, u_{it}, takes the form

$$u_{it} = \eta_i + \tau_t + \upsilon_{it} \tag{A3}$$

The τ_t represent macro-economic shocks captured by a series of time dummies, η_i is an individual effect, and υ_{it} is a serially uncorrelated mean zero error term.

If we allow inputs to be endogenous, we will require instrumental variables. In the absence of any obvious natural experiments, we consider moment conditions that will enable us to construct a GMM estimator for equation (A4). A common method would be to take first differences of (A4) to sweep out the fixed effects:

$$\Delta y_{it} = \pi \ \Delta x_{it} + \Delta\tau_t + \Delta\upsilon_{it} \tag{A5}$$

Since υ_{it} is serially uncorrelated the moment condition

$$E(x_{it-2}\Delta\upsilon_{it}) = 0 \tag{A6}$$

ensures that instruments dated t-2 and earlier[37] are valid and can be used to construct a GMM estimator for equation (4) in first differences (Arellano and Bond, 1991). A problem with this estimator is that variables with a high degree of persistence over time (such as capital) will have very low correlation between their first difference (Δx_{it}) and the lagged levels being used an instrument (e.g. x_{it-2}). This problem of weak instruments can lead to substantial bias in finite samples.

Blundell and Bond (1998) point out that under a restriction on the initial conditions another set of moment conditions are available[38]:

$$E(\Delta x_{it-1}(\eta_i + \upsilon_{it})) = 0 \tag{A7}$$

[37] Additional instruments dated t-3, t-4, etc. become available as the panel progresses through time.

[38] The conditions are that the initial change in productivity is uncorrelated with the fixed effect $E(\Delta y_{i2}\eta_i) = 0$ and that initial changes in the endogenous variables are also uncorrelated with the fixed effect $E(\Delta x_{i2}\eta_i) = 0$

This implies that lags of the first differences of the endogenous variables can be used to instrument the levels equation (A4) directly. The econometric strategy is then to combine the instruments implied by the moment conditions (A6) and (A7). We stack the equations in differences and levels (i.e. (A4) and (A5)). We can obtain consistent estimates of the coefficients and use these to recover the underlying structural parameters in (A2).

The estimation strategy assumes the absence of serial correlation in the levels error terms (υ_{it})[39]. We report serial correlation tests in addition to the Sargan-Hansen test of the over-identifying restrictions in all the GMM results[40].

This GMM "system" estimator has been found to perform well in Monte Carlo simulations (Blundell and Bond, 1998) and in the context of the estimation of production functions (Blundell and Bond, 2000). The procedure should also be a way of controlling for transitory measurement error (the fixed effects control for permanent measurement error).

Appendix C: Data Description

Sample of hospitals

Our panel covers the period 1991/2-2002/3. Sources of data are given in Table A1.

We do not have data on all hospitals for the whole period 91/2-2002/3.

i) Data pre-1995 is only for those hospitals that had been given freestanding financial status within the NHS prior to that date. By 1995 almost all hospitals had free standing status; before that date the finances of some were still recorded at District Health Authority level.
ii) AMI deaths were subject to recoding (a change in ICD codes) in 1994/5.

---

[39] If the process is MA(1) instead of MA(0) then the moment conditions in (6) and (7) no longer hold. Nevertheless $E(x_{it-3}\Delta\upsilon_{it}) = 0$ and $E(\Delta x_{it-2}(\eta_i + \upsilon_{it})) = 0$ remain valid so earlier dated lags could still be used as instruments. This is the situation empirically with the wage equations.

[40] These are based on the first differenced residuals so we expect significant first order serial correlation but require zero second order serial correlation for the instruments to be valid. If there is significant second order correlation we need to drop the instruments back a further time period (this happens to be the case for the wage equation in the results below).

For these reasons, we undertake analyses from 1995/6. Wherever possible we repeated our analyses using the longer time series. The results are robust to this.

In estimation, we trim the top and bottom centiles of the following variables: total staffing, nurses pay. To be used in the analysis the observations must have data on all the key variables including agency nurse and vacancy data. In the AMI regressions they must be acute hospitals. This same sample is used to estimate productivity, but we also undertake productivity analyses including non-acute hospitals.

Use of AMI as a measure of quality

We use here the '30-day' death rate for acute myocardial infarction (AMI). This measures in-hospital deaths within 30 days of emergency admission with a myocardial infarction for patients aged 55 and over. We use 55 year olds and over because of the very small numbers under 55 years old. There are several issues in using this 30-day within-hospital death rate measure. The first is the variability in rates: death rates may be quite variable over time hospital-by-hospital, reflecting, in part, small denominators (hospitals may treat relatively few patients in any one year). This noise in the measures of death rates can lead to misclassification of the quality of hospitals (McClellan and Staiger, 1999). A recent paper concludes that raw UK hospital level rates exhibit considerably less variability than the raw US data, but not than US rates which have been 'filtered' to reduce noise (Burgess et al, 2004). To reduce misclassification based on small sample sizes, we omit all hospitals with less than 150 emergency AMI admissions in any one year

The second issue is that we use the 30-day rate. McClellan and Staiger (1999) show that the 30-day rate has more estimation error than the 7-day mortality rate (largely because hospital quality appears to be more important in affecting 7-day mortality). Despite this, the English government chose the 30-day rate rather than the 7-day rate published for the first time in 1999, for English hospitals[41] and for that reason we use this here.

[41] In 2006 the OECD recommended use of the 30-day AMI mortality rate as one of its 17 key indicators of health care quality (Mattke et al 2006).

The third issue arises because our measure is the in-hospital death rate. Deaths occurring after transfer to another provider are credited to the provider where the patient was first admitted, whilst deaths following discharge are omitted. This may bias the results if hospitals have a motive to discharge early. Such incentives would have been small as these death rates were not published until 1999 and hospitals not ranked by the Department of Health in terms of outcomes until 2001, when they were ranked on a composite bundle of over twenty indicators.

Case-mix adjustment of FCEs

The Department of Health for England produces a measure of cost-adjusted output: the Reference Cost Index, but this is not available in consistent form for most of the period our data covers. This reference cost index not adjusted for the Market Forces Factors - the relevant index which is pre-adjustment for wage differences across regions - is available at the earliest date in 1999 and then in 2001 and 2002.

Wages

To get a better idea of the regulated pay structure consider nurse pay scales at 1st April 1999. Clinical grades range from A to I and correspond to spinal points 3 to 37. For example, Clinical Grade G, a "ward sister" corresponds to a grade between spinal points 12 (£20,145 per annum) to spinal point 9 (£23,300 per annum). There are allowances (or "weightings") for being in high cost areas. For Inner London this was £2205 plus 5% of salary up to a maximum of £750, for Outer London this was 1570 plus 5% of salary up to a maximum of £750 and for the "fringe" (various areas in the South East) this was £285 plus 2.5% of salary up to a maximum of £375. For a ward sister on 23,300 a year working in the most expensive area of the UK, Inner London, the extra regional allowance would be worth only 11% more salary (2205+750/ (2205+750+23,300)). Since this is capped, for a more senior nurse on a higher salary the proportional value is lower. By contrast, in 1999 in the NES the annual non-manual wage in Inner London is about 65% higher than that of the Northeast, the lowest wage region.

In calculating the instrumental variable for inside pay measure we take into account the NHS grade structure in a region in a year (using wage data from the NES). We then use the decisions of the NPRB over the changes in the wage structure taking into account all the London weightings, etc (which may differ by grade) to form the predicted wage in the next period (specific to each region).

Other variables

The mortality by area is the directly standardised (by age and gender) mortality rate from all cause mortality at county/borough level. To calculate a number for a given trust we use the (unweighted by distance) average of all counties/boroughs within 10 kilometres.

HES Data

HES data record inpatient activity in the NHS delivered by NHS hospitals. The main unit of recording is the Finished Consultant Episode (a period of admitted patient care under a consultant or allied healthcare professional within an NHS trust). This is not always the same as a single stay (spell) in hospital, because a patient may be transferred from one consultant to another during their stay. In these cases, there will be two or more episode records for the spell of treatment.

Diagnoses are currently coded according to the International Classification of Diseases, 10th Revision (ICD-10) and surgical procedures (operations) according to the Office of Population, Censuses and Surveys: Classification of Surgical Operations and Procedures, fourth Revision (OPCS-4).

HES records also contain the age of the patient and where they lived. There are further codes to identify the hospital, the length of time the patient stayed in hospital, and the specialty of the consultant who treated them. If the patient was on a waiting list (as opposed to being admitted in an emergency), the time they actually waited is also recorded.

http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=537 (accessed 4 July 2006)

Table A1 Data sources

Data type	Source of Data	Years of Data
AMI death rates and admissions rates, by age and gender	Hospital Episode Statistics	1991-2002
Number of admissions, by age, gender and method of admission	Hospital Episode Statistics	1991-2000
Total FCE's	Hospital Episode Statistics	1991-2002
Physicians staffing levels	Department of Health Medical Workforce Census	1991-2002
Qualified Nurse staffing levels	Department of Health Medical Workforce Census	1991-2002
Unqualified Nurse staffing levels	Department of Health Medical Workforce Census	1991-2002
Qualified AHPs staffing levels	Department of Health Medical Workforce Census	1991-2002
Unqualified AHPs staffing levels	Department of Health Medical Workforce Census	1991-2002
HCA staffing levels	Department of Health Medical Workforce Census	1995-2002
Expenditure on physicians (including locums)	Hospital financial returns (from Dept Health)	1991-2002
Expenditure on qualified nurses	Hospital financial returns (from Dept Health)	1991-2002
Expenditure on unqualified nurses	Hospital financial returns (from Dept Health)	1991-2002
Expenditure on AHPs	Hospital financial returns (from Dept Health)	1991-2002
Expenditure on HCAs	Hospital financial returns (from Dept Health)	1995-2002
Expenditure on agency nurses	Hospital financial returns (from Dept Health)	1991-2002
Nurses vacancy rates	Office of Manpower Economics	1996-2001
Mortality rates	Department of Health	1991-2002
Wage data – regional and area wages	New Earnings Survey	1991-2001