

## LABOUR MARKETS IN THE NHS: INSTITUTIONAL RIGIDITIES AND RENTS<sup>1</sup>

*Bob Elliott, Health Economics Research Unit, University of Aberdeen*

1. The NHS is the largest single employer in Great Britain. In 2000 the NHS in England employed 991,000 people, NHS Scotland employed 136,000, and the NHS in Wales 75,000. NHS employees accounted for 6.7% of all employees in Scotland and nearly 5% of the total for Great Britain as a whole. The largest single group are nurses, midwives and health visiting staff who, in 2000, totalled 538,000, while in that same year there were 92,000 medical and dental staff. The proposed massive injection of funds into the NHS consequent upon the Wanless report and the associated desire to increase NHS capacity means that there has probably never been a time when it has been more important to understand what attracts, retains and motivates NHS staff.

2. The NHS workforce has three significant distinguishing characteristics: size, as evidenced above; heterogeneity, as evidenced by the range of occupations employed and hence the many different occupational labour markets in which it operates; and geographical coverage, it has a substantial presence in every labour market in Great Britain. The combination of very substantial spatial and occupational heterogeneity sets the NHS apart from other employers. Devising pay and reward structures which efficiently recruit, retain and motivate labour in so many different labour markets presents a substantial challenge to the NHS and the evidence seems to be that it has not done this very well. We might expect to find this spatial and occupational diversity mirrored in considerable diversity of pay structures. But the common institutional structures of the NHS, both its organisational structure and the institutions that represent employees, the trade unions, militate against this.

3. These institutional structures explain the disequilibrium that is a striking feature of many, perhaps most, NHS labour markets. This disequilibrium manifests itself as acute and persistent regional and speciality shortages of key medical and nursing staff, and in pay structures that are perceived as failing to motivate staff to deliver the quality of service required. Such issues have traditionally been studied by labour economists, and are related below to what are understood to be the critical current concerns of the NHS.

### ***1. Labour Supply***

#### **(i) Doctors**

4. In most labour markets the supply of training places responds, with a lag due to frictions, to the demand for training places, however this does not appear to be the case in the market for doctors. One prominent feature of this market is long run disequilibria; demand persistently exceeds supply and accordingly medical professionals persistently receive rents. For many years now the market for doctors in the UK has been characterised by excess vacancies; vacancies in excess of those generated by normal turnover. These vacancies have been filled by recruiting doctors

---

<sup>1</sup> A preliminary version of this paper benefited greatly from the comments of Martin Chalkley, Alasdair Munro and Tony Scott. They are not responsible for the errors that still remain.

from overseas with the result that doctors trained overseas now comprise well over 15 per cent of the total. Yet at the same time there is a persistent oversupply of apparently suitably qualified applicants to medical school – large numbers have the required Higher and A level grades. This situation persists even though the entry requirements are virtually the highest required by any profession. So why does the supply of places not expand to satisfy demand?

5. The labour economist would respond that such persistent disequilibria can only result from the distortionary effects of institutions and, on further inquiry, would be likely to identify two institutional arrangements that might account for this. The first would be the way in which medical training is financed in the UK. In the UK the government provides the funds which finance training places in medical schools, payments by the individuals undertaking training represent only a trivially small share of the total. It might be argued that public funding has never been sufficient to provide an adequate supply of the required places, although clearly this is not a necessary feature of a publicly funded system. Other publicly funded systems, several in continental Europe for example, have on occasions produced a surplus of doctors. So why is the UK different in this respect? Perhaps the answer lies in the second institution.

6. The second institution is the trade union that represents the medical profession. The labour economist would point out that the number of medical places is determined through negotiation between government and this trade union. This negotiation produces the expected result that the number of training places is always smaller than is required to fill outstanding vacancies and this, of course, is an effective device for ensuring that the pay of medical staff is maintained above the market clearing rate. This second explanation argues that medical workforce planning should be taken out of the hands of the medical professionals.

7. Of course it is possible that neither of these two explanations is correct and that it is inappropriate to characterise the situation as one of disequilibrium. If the abilities required to perform effectively as a doctor are correlated with school performance but cannot be separately distinguished in measured performance tests it makes sense to set pay above the market clearing rate in order to attract a large applicant pool with high qualifications and to proceed to then pick the most highly qualified among these. This explanation is a variant of efficiency wage theory.

8. Alternatively Chalkley has suggested that the overpayment may result from information asymmetries where medical professionals enjoy an information advantage over both NHS managers and patients. Medical professionals may learn things during training and afterwards that enable them to extract rent. However we would not want only the best rent extractors to be attracted to a medical career for we would end up paying them much more. We need a system that encourages altruistic individuals into the medical profession and the existing system does this by offering pay below the levels that the brightest people could earn elsewhere.

9. There is also an important short run dimension to the labour supply of medical staff that needs to be addressed. I know of no studies that have sought to distinguish the own wage elasticity of the supply of hours of work of doctors. This may be because until recently it was assumed that doctors had little choice over the hours they worked and few alternative earning opportunities. However with the move to greater part-time working, most pronounced among GP's where the majority of new entrants are women, it is essential that the determinants of the supply of hours are better understood. Again the increased move to earlier retirement means that we need to understand the relative attractiveness of different aspects of doctors work and remuneration and contrast these to other labour market and non market opportunities.

*10. The task of health economists is to distinguish between the above explanations of medical pay. Is pay above the market clearing rate an efficient device for attracting the right sort of applicant or is it a distortion? At present we simply do not know. A further task for the health economist is to estimate the elasticities of hours of work for medical professionals, but, is the data up to the task?*

## **(ii) Nurses**

11. In the long run the supply of nurses is determined by the number of people choosing to acquire a nursing qualification and once acquired, work as a nurse. The factors influencing this choice appear to have changed substantially in recent years. This is because the range of jobs open to women, who still constitute the vast majority of new entrants to nursing, has expanded substantially in recent years. This in turn reflects both the wider choice of training opportunities open to women and the fact that the range of occupations open to, or perceived to be attractive to, women has grown. Further while the NHS remains the dominant buyer of the labour supplied by qualified nurses it is far from the only buyer, as the growth of the private health care sector evidences. Thus the NHS now finds itself operating in a much more competitive labour market than in the past. Understanding what is happening in this market is clearly crucial.

12. This is also true of the short run as more employers offer work to those who have qualified as nurses<sup>2</sup>. Understanding the short run elasticities of the supply of hours and participation by those qualified as nurses is also crucial and being actively researched (Morris, 2002; Skatun et al 2002). The study by Skatun et al (2002), estimated the elasticity of supply of hours of work, conditional on current employment as a nurse, at + 0.34 in 1999/2000 for the UK. This same study estimated an elasticity of participation into nursing from out of the labour market at +0.55 and further suggested that recent changes in the public funding of nursery provision may have had a positive effect on nursing labour supply. There are however still issues to be addressed by researchers working in this area. We need accurate measures of the actual work experience of nurses, and we then need to estimate models explaining nurses' decisions about non labour market activities.

---

<sup>2</sup> A recent study by Elliott et al. revealed that in 2000 almost one-third of all those people of working age holding a nursing qualification, living in Scotland and still working were not working as a nurses in either the private sector or the NHS.

13. Research has revealed that aspects of jobs other than pay are also important determinants of the attractiveness of nursing jobs. Thus a study of job satisfaction for nurses (Shields and Ward, 2001) revealed that working conditions have a significant impact on job satisfaction. More generally the hours of work, working conditions, and the time at which hours have to be supplied are all thought to be critical determinants of the short-run supply of labour to nursing, these dimensions require more systematic investigation.

14. Labour market theory, the theory of compensating differentials (Rosen, 1986) informs us that in the labour market we ‘pay’ for the attractive attributes of jobs through lower wages than otherwise and we are compensated for unattractive attributes through higher wages. It also tells us that in equilibrium all jobs should produce the same levels of satisfaction. Thus those studies that reveal that levels of satisfaction differ between jobs are informing us that there are attributes of some jobs that are not correctly ‘priced’. Research to discover just what the most important attributes are and the prices that attach to them is urgently required for nurses. This could be done by identifying the total rewards and bundles of attributes that attach to a sufficiently large number of different nursing jobs and then identifying the key distinguishing attributes and rewards of each job. Such an exercise has yet to be undertaken but is critical for the design of efficient reward packages for nurses. One reason why such an exercise can be undertaken for nurses is that there are a sufficient number of alternative, non NHS job opportunities for nurses and Trusts and Health Boards have discretion in designing the package of working conditions they offer for a fixed wage. An alternative method would be to use discrete choice experiments, stated preference methods, to capture the trade-offs that individuals would be willing to make.

*15. The task for health economists is to generate more robust estimates of the determinants of labour supply using the existing secondary datasets and to collect the primary data necessary to distinguish the prices that nurses are willing to pay for different job attributes.*

## **2. Incentives for Performance<sup>3</sup>**

16 “A persons pay should reflect their outputs, results and performance. This means the best performers and those who contribute most should be best rewarded. We should challenge systems which give automatic pay increases to poor and inefficient performers” Modernising Government, (1999). Thus incentive pay is a central feature of the present government’s modernisation agenda.

17. The appropriateness of incentive payment schemes, which tie pay to measured dimensions of performance in an attempt to induce improved performance, by health service professionals, is a continuous focus of debate. Opponents contend that health service professionals have a strong ethical interest in doing good and that there is therefore less need for strong and explicit pecuniary incentives (Mooney and Ryan,

---

<sup>3</sup> These theoretical underpinning has been very well reviewed in the chapter ‘Incentives in Health Care’ by Scott A, and Farrar S in *Advances in Health Economics*. Scott A, Maynard A and Elliott B (eds). Wiley (2003) forthcoming.

1993). They contend that tasks are complex, involve multi-tasking (Holmstrom and Milgrom, 1991) and that tying reward to the successful performance of only some of these tasks will distort effort as agents pursue more vigorously those tasks to which specific reward is attached. There is for example evidence that where fee-for-service payments have been introduced these can result in increased activity and perhaps over treatment (McGuire, 2000). Perhaps for this reason direct measures of performance have never really played a prominent role in motivating the performance of either the majority of hospital doctors or nurses.

18. Only hospital consultants could be said to be paid for performance, some of them receive distinction awards which can be thought of as rewards for lifetime performance, and the highest of these payments can almost double a consultant's salary. However the real significance of these payments is the role that they play in motivating the performance of junior hospital doctors. Until recently the effect of these payments was to produce an occupational wage structure for hospital doctors which is more unequal than that of any other public sector occupation in the UK. Until recently the highest paid consultant employed in an NHS hospital, those receiving merit awards, earned an annual salary that was almost eight times that of the lowest salary; junior hospital doctors. When expressed in terms of pay per hour the differential was even greater, as a consequence of the relatively longer hours of work by junior hospital doctors, the differential was perhaps of the order of 10 to 15 times greater. What was the rationale for this marked inequality and how have recent changes to the contracts of junior hospital doctors affected this?

19. A rationale for this substantial salary differential is offered by tournament theory (Lazear and Rosen, 1981). Not all junior hospital doctors will become consultants and earn a consultants salary and the prospect of additional merit awards this brings. The consultants salary can therefore be viewed as the prize awarded to those who win the tournament, those who gain promotion. Competition for promotion then motivates the performance of junior hospital doctors. Moreover their performance will be more highly motivated the greater the prize they stand to gain and the smaller their chance of winning. Thus the larger the salary gap between the junior hospital doctors salaries and consultants salaries and the smaller the number of consultants posts relative to junior hospital doctors posts, the greater the incentives to strive for promotion.

20. Recently however two critical features of the tournament have changed. First the size of the prize has been substantially reduced and second the number of prizes has been increased. The first of these is a consequence of the new junior hospital doctors contract and this will be further exacerbated in the future by the implementation of the Working-Time Directive. The effect of both of these has been to substantially increase the hourly salary of junior hospital doctors and thus to narrow the pay gap. Recently there has also been an increase in the number of consultants posts.

21. What are the predicted consequences of this change? First as a result of the reduced incentives to performance we should expect junior hospital doctors to strive less hard. Second it is clear that these changes have resulted in a substantial increase in the expected lifetime rewards from working as a hospital doctor. Both the expected earnings while working as a junior hospital doctor and the chances of promotion to

consultant have increased. All other things equal this would be expected to increase the number of applicants to train as a doctor.

*22. The task for health economists is both to model and then using these models predict changes in the labour market for and the performance of hospital doctors. An additional task is to research new methods for motivating the performance of NHS professionals.*

### **3. Wage Setting**

23. In the NHS medical and nursing staff are still paid under highly centralised national pay arrangements. How appropriate are the existing systems of wage setting? The persistent shortages suggest that the wage structure is not performing one of its central functions, that of allocating labour efficiently. The reform of public sector pay structures has been a prominent feature of the modernisation agenda of recent governments, as they search for greater labour market flexibility, but in the NHS reform has been slow to arrive. One approach to reform might be to introduce decentralised pay and grading arrangements, such as were introduced for the civil service over eight years ago, with NHS Trusts responsible for these decisions. Detailed analysis of the results of pay reform, and in particular decentralisation, in other parts of the public sector is needed to inform decisions in the NHS.

24. The current centralised arrangements for medical and nursing staff have produced higher pay increases than those awarded to any of the other major groups of public sector workers, save Scottish teachers. Below the increases achieved by medical and nursing staff are compared to the annual pay awards to police officers, a group often perceived to have done well in recent years. The increases received by police officers are moreover indexed to the median of the distribution of pay settlements awarded to non-manual workers in the private sector of the UK economy in the twelve months to June each year. The comparison with police pay awards therefore also informs us how doctors and nurses pay has grown relative to pay in the private sector. Despite these higher pay settlements shortages still remain. How much resource has been devoted to increasing rates of pay in areas of excess supply?

**The Size of Annual Pay Settlements Awarded Nurses, Hospital Doctors and Private Sector Non- Manuals**

	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>
Nurses	3.8	4.7	3.4	3.7	3.6
Hospital Doctors	4.2	3.5	3.3	3.9	3.6
Police (Private sector non-manuals)	4.0	3.6	3.0	3.5	3.0

25. An explanation for these generous awards can be found in the institutions of wage setting. Doctors and nurses are represented by powerful trade unions which act as monopoly suppliers of their labour. The employers, also combine to present a unified voice through the NHS Confederation and Department of Health and thus effectively act as a monopsonist in each of these markets. Under these conditions pay will be determined through a process of bargaining, and the resulting level of pay will exceed

the supply price of labour. In an attempt to mitigate the uncertainty and reduce the costs associated with pay bargaining under these conditions Review Bodies were established to set the pay of both these groups.

26. The Review Bodies that set the pay of doctors and nurses weigh the arguments of both sides, consider the claims and counterclaims in the light of the available empirical evidence and recommend an annual pay award. The higher pay settlements that have been awarded to doctors and nurses may again, in part, reflect the market power exercised by the trade unions representing these groups. However they are also likely to reflect the favourable market conditions in which doctors and nurses are perceived by the Review Body to sell their labour. Good health is a normal, perhaps even a superior good, and as a result the demand for the services provided by doctors and nurses has grown and is expected to grow strongly in the next few years. The demand for the services provided by the medical and nursing workforce will also be further ratcheted upwards, as the extra resources resulting from Wanless flow into the NHS. However it does not follow that the appropriate response to these developments is to continue to award the uniform, across-the-board, increases in pay to all doctors and nurses.

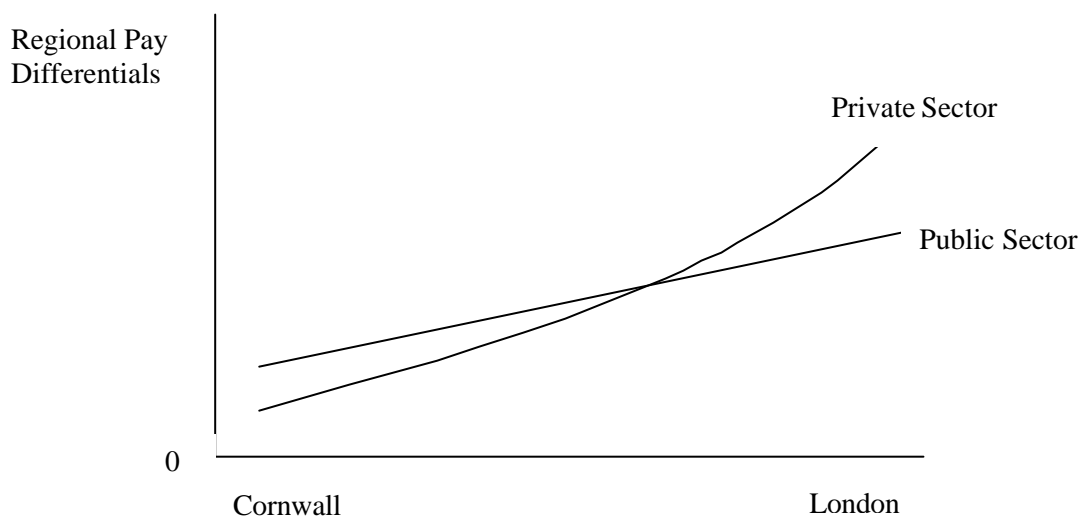
27. In the next few years this growth in demand will be concentrated in certain geographical areas and specialties. Centralised wage fixing arrangements cannot be expected to respond flexibly to these changing patterns of demand, a more decentralised system for setting pay would appear to be required. If the existing centralised arrangements for setting pay are retained the conservative assumptions about pay that underpin the Wanless report seem likely to be exposed. Pay inflation will be greater than forecast and the resulting increase in activity far less. How much greater than forecast by Wanless will depend on the degree to which pay setting in the NHS is decentralised.

*27. The task for health economists is first to seek to explain the existing system of wage setting and explore more rigorously the consequences of this system. Further to provide an analysis of the flexibilities within the present system and an evaluation of alternative decentralised mechanisms of wage setting.*

#### **4. Spatial Wage Differentials**

28. One consequence of the system of wage setting is that the labour market for both doctors and nurses is characterised by severe geographical imbalance. Vacancy rates for nurses are very high in London and some other metropolitan areas of England and very low in Scotland and most rural counties of England. There is also substantial spatial variation in vacancy rates for doctors. This geographical imbalance signals that the wage structure for these groups is not sufficiently attuned to local market conditions. This can happen where trade unions exert a powerful influence over pay setting because their concerns about equity and fairness will lead them to advocate uniform, national, rates of pay. Moreover an institutional mechanism for setting pay such as a national Review Body is likely to reinforce this tendency. National rates take no account of the particular balance of supply and demand in local labour markets. The effect is to produce a much flatter pay structure, across the relevant domain, than would occur in the absence of trade unions.

29. Trade Unions and national pay structures play a less prominent role in pay setting in the private sector where employers are motivated to offer rates of pay sufficient to attract, retain and motivate the required labour. The result is greater responsiveness of pay rates to local market conditions and greater geographical pay variation. The spectrum of regional pay differentials in the public and private sectors can be represented as in Diagram 1 below. In the diagram the lines represent the locus of a set of coefficients on dummy variables indicating the area in which the individual works which are included in a human capital earnings equation estimated for each sector. The human capital equations seek to explain differences in the pay of individuals by reference to all the measurable productivity enhancing attributes of these individuals and the characteristics of the jobs they do, which for example will include measures of the individuals hours of work, education, age, tenure in the job, occupation, and industry.



30. It can be seen that the private sector spatial wage structure exhibits much greater geographical variation than the public sector pay structure. Thus even where the average rate of pay in the two sectors is equal the consequences of this pay structure are easy to see. In London and the South East, the extreme right of the horizontal axis, the public sector pays less than required to attract and retain labour and the public sector labour market is thus characterised by shortage. In the more rural areas of England and in Scotland and Wales, as depicted towards the origin of the horizontal axis, public sector pay is higher than is necessary and there is excess demand for public sector jobs. A carefully chosen private sector comparator occupation can serve as a benchmark by which to judge the efficiency of the pay rates for any group of public sector workers. Where the two pay structures appear as depicted above the public sector pay structure is revealed to be inefficient for it overpays in some areas and underpays in others.

31. *The task for health economists is to seek to quantify the magnitude of over (under) payment for each region and each of the major professional groups employed by the NHS and then to calculate the savings (costs) that result from moving pay rates to the market clearing rate.*



## 5. *Conclusions*

32. The research agenda which emerges within each of the above four sections has been detailed. The challenge now is to blend the best of labour economics, formal modelling of hypotheses and robust estimation resulting from the application of advanced econometric techniques to large data sets, with the best of health economics, a concern with the rigorous quantification of the costs and benefits of alternative activities.

33. One general lesson that emerges from empirical work in the field of labour economics is that the data is rarely adequate to fully test our theories. In labour economics the focus is upon matching employer data sets, which accurately record pay, occupation and type of industry, to employee reported data, which more accurately records hours of work, family characteristics and other sources of income. Data sets, which enable the researcher to compile accurate measures of individuals working lives, are also critical. Public access data sets such as the New Earnings Panel Data set, the Labour Force Survey and the British Household Panel Survey, which researchers use, each have some but none has all of these desirable characteristics. Moreover the sample size is such that when we use these data we can only ever study the largest occupational groups in the NHS, nurses and sometime hospital doctors. The NHS, will be the largest beneficiary of the above research agenda, and it is uniquely placed to generate the information necessary to undertake this research. The conclusion should be obvious!

## 6. *References*

Elliott R.F, Mavromaras K, Scott A, Bell DNF, Antonazzo E, Gerova V, Van der Pol M *Labour Markets and NHS Scotland*. Report to the Scottish Executive Central Research Unit. University of Aberdeen: Aberdeen, 2002.

Holmstrom B, Milgrom P (1991) Multitask principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organisation*, **7**, 24-52.

Lazear E, Rosen S (1981) Rank order tournaments as optimal labour contracts. *Journal of Political Economy*, **89**, 841-864.

McGuire T (2000) *Physician agency*. In: Culyer AJ, Newhouse JP. *Handbook of Health Economics*, Volume 1A. Elsevier.

Mooney G, Ryan M (1993) Agency in health care: getting beyond first principles. *Journal of Health Economics*, **12**, 125-238.

Modernising Government, (1999) London, HMSO.

Morris, S (2002) Unpublished PhD, University of London

Rosen S (1986) The theory of equalising differences. In Ashenfelter O and Layard PRG (eds) *Handbook of Labor Economics*, **11**, 641 – 92, North Holland.

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

Skatun D, Antonazzo E, Scott A and Elliott RF (2002) *Attracting qualified nurses back into nursing: an econometric analysis of nurses labour supply*. Paper presented at Scottish Health Economics Network, December.