Paper presented at the Health Economists' Study Group Meeting

University of Kent, Canterbury, 9-11 July

THE EFFECTS OF RURALITY AND REMOTNESS ON HOSPITAL COSTS IN SCOTLAND

By

Patricia Fernandes

HERU, University of Aberdeen and EEG, Universidade do Minho

Anthony Scott

HERU, University of Aberdeen

Abstract

Resource allocation formulae in Scotland include an adjustment for remoteness and rurality. However, there is little empirical evidence about the precise effects of remoteness and rurality on hospital costs. The main objective of this paper is to identify and examine the theoretical reasons as to why hospitals in remote and rural areas incur higher costs, than comparable hospitals in urban areas. This will be accomplished by examining the relative impact on costs of different determinants associated with remoteness and rurality. This paper focuses on the hypothesis to be tested namely the influence on hospital costs of factors such as:

- The existence of diseconomies of scale in smaller size hospitals (such as one would expect to find in rural and remote areas of Scotland), which has been repeatedly pointed out as the major reason for those hospitals to experience higher costs.
- High unit costs of non-labour inputs. The price of materials and supplies has been considered as a relevant factor of hospital costs and it might be possible that hospitals located in rural and remote areas experience higher costs due to higher transportation costs of supplies (especially relevant in the case of hospitals located in Islands).
- The possible inexistence of a net of social support, such as the ones existing in urban hospitals, might be one of the sources of rural/remote hospital higher costs because of higher lengths of stay per patient.
- Labour market issues. Factors such as wages, turnover costs, high unit labour costs and hospital staff composition might play an important part on the determination of hospital costs.

Acknowledgements: Patricia Fernandes is funded by the Fundação para a Ciência e a Tecnologia-Ministério para a Ciência e Ensino Superior, Portugal and by The Institute of Applied Health Sciences, University of Aberdeen. HERU is funded by the Chief Scientist Office of the Scottish Executive Health Department. The views are those of the authors.

1. Introduction

The problem of scarce resources and increasing and unlimited demand is at the core of economic theory and is particularly poignant when one tries to address the question of the financing and delivery of health care services. Choices regarding the methods of sharing health resources have to be made. The purpose is choosing the most efficient method, but, simultaneously, factors such as equity and quality need being taken into consideration. Equity is a common objective of modern systems of provision of health care and is not, necessarily, coincident with efficiency objectives.

The possible trade-off between efficiency and equity is easily exemplified if one thinks, for instance, of a situation where, for equity reasons, there is need to maintain a certain level of provision of health services (in order to guarantee access to what is considered to be a minimum level of care) which are considered to be beyond what it would be regarded as efficient. This provision of "ready access" to populations not only raises efficiency questions but also questions such as the relation between quantity (volume of clinical activity) and quality (outcomes of care).

For instance, if we consider the case of acute services, such as Accident and Emergency Services, due to the difficulties created by geography and isolation, delays in starting treatment for some seriously injured people may result in a significant increase in mortality and therefore these type of services should exist in less accessible areas. But, compared to urban A&E services, the services provided in more isolated areas usually see fewer patients and are generally within smaller hospitals with a more modest scale of activities, not able to support a full range of specialist staff (Scottish Health Service Advisory Council, 1995), therefore possibly compromising the quality of care provided[1]. This is a very important problem because according to the Department of

[1] The relationship between volume of clinical activity and clinical quality is of dubious nature. It is usually assumed to be a positive one, but a review carried out by the York NHS Centre of Reviews and Dissemination (1997) casted some doubts about that assumption by questioning the quality, and therefore the validity, of many of the studies reviewed. In a study dealing specifically with small rural hospitals (Schlenker *et al*., 1996 in Rickets, 2000) it was found that outcomes typically made in such hospitals had little variation associated with volume.

Health (1998), *"local access to services must never be provided at the expense of quality"*[2].

These questions of equal accessibility to health services, at a uniformly high quality, are a crucial issue, especially when regarding remote or/and rural areas, as well as islands, areas where equilibrium between accessibility, cost and quality of care is somewhat difficult to achieve.

Therefore, the problems associated with the provision of health care in rural and remotes areas have become, in the last few years, an increasingly important area of policy. In Scotland, the Acute Services Review done by the Department of Health (1998) emphasised the importance of supporting and developing services in remote and rural areas, specifically stating the concern *"to ensure ready access through local portals to the entire network of acute service provision*".

To deal with this problem, several formulae of health care resource allocation included adjustments in order to take into account the higher unit costs of providing health care in rural and remote areas. Since the 90s both Wales and Scotland included sparsity weighting in their allocation formulas regarding the provision of community health care. In both cases the influence of the sparsity weighting had very small impact in the community health service budget and no adjustment was made for hospital services which made the process somewhat arbitrary (Watt and Sheldon, 1993).

More recently, in 2000, the National Review of Resource Allocation in Scotland recommended the introduction of an adjustment to reflect the additional costs faced by some Health Boards in delivering hospital services to populations living in remote and rural areas. That adjustment was based on the relationship between an indicator of remoteness[3] and the costs faced by the Health Boards in delivering hospital services in those areas. The actual expenditure on services in each Board was compared to the level of expenditure that it would have incurred if these services were purchased at Scottish

[2] It may be, of course, that some would be willing to accept the trade-off in order to have easier access to a minimum level of care. Like in the efficiency/equity trade-off, a judgement has to be made.
[3] Number of road kilometres per 1,000 population over the three year period 1996-98.

average costs, thus calculating the "relative" level of expenses in each Health Board and making adjustments according to that relative level.

It was recognised that, besides taking in account factors such as the size and the profile (age, sex, morbidity and life circumstances) of the population, an additional adjustment had to be made regarding remoteness and rurality.

It should be pointed out that what is being assessed in these adjustments is not how health "needs" vary between rural and urban areas but instead how the unit costs incurred by providing health care services in both areas differ. Also, these costs should regard only "unavoidable" costs, in the sense that they are caused by external circumstances and thus should not reflect any type of inefficiency in the provision of those services. A very important distinction has to be made between the "*impact of market forces from other factors which might affect provider costs, such as technology, economies of scale, quality variations and efficiency*[4]".

The efficiency question, that is, that hospitals produce the maximum level of output with the minimum of inputs consumed, is a particularly important one. It is often argued that hospitals are not efficient since the neo-classical firm optimisation behaviour of cost minimisation does not apply to hospitals, due to the special features of this industry (Jacobs, 2000). The provision of health care services can be characterised by market failure and non-profit maximising firm behaviour both of which are sources of disruption from optimal (efficient) performance (Maniadakis and Hollingsworth, 1998).

It is thus fundamental to only take into account the inefficiency due to factors outside of the hospitals' control as opposite to technical inefficiency (too many inputs to produce a certain level of output), so as not to distort incentives and compensate financially hospitals for inefficient practice.

In terms of academic research, the rurality and remoteness dimension of costs has been absent in most of the hospital cost literature. Some of the studies in the hospital cost

[4] Department of Health, 1998

function literature used some type of location dummies[5] or even an Index of urban size area[6] but the main purpose of the inclusion of these variables was to determine the impact of the factor price variation in hospital costs and most of them concluded that hospitals located in urban or metropolitan areas have higher costs.

Only a very small number of studies tried to specifically analyse the reasons, other than factor prices, why hospital costs in rural/urban areas differ. The main aim of this thesis is therefore to analyse the influence of remoteness and rurality on Scottish hospital costs.

1.1 Literature Review

Thorpe (1988) tried to answer the question of why urban hospital costs are so high. He concluded that the excess costs arise from six factors: major teaching programs[7], a more severely ill case-mix[8], cost-increasing competition[9], staffed excess capacity[10], higher input prices and costs resulting from patients admitted through emergency room[11], with teaching status being the most important determinant of high costs.

Finch and Christianson (1981) examined the reduction of rural hospital capacity through closures in order to assess if the magnitude of the potential cost savings from the more intensive use of hospital capacity, and from exploration of economies of size, would justify the closure and then merger of rural hospitals. Quadratic and logarithmic cost functions for rural hospitals were estimated and it was concluded that economies of scale occurred over the range of hospital sizes observable in rural areas (for both cost function specifications) and that hospital size which minimised average cost per patient

[5] Lave and Lave (1970), Lave et al. (1972), Culyer *et al.* (1978), Pauly (1978), Friedman and Pauly (1981), Grannemann *et al.* (1986), Koop and Carey (1991), Fournier and Mitchell (1992), Carey (1997), Sinay (1998), Carey and Burgness JR (1999), Folland and Hofler (2001) and Li and Rosenman (2001)

[6] Vitaliano (1987).

[7] Excess costs derived from University teaching status are probably related with variations in physician and resident practice patterns, including more intensive use of ancillary services.

[8] Case-mix indices were not considered to justly reflect the excess costs of treating a disproportionate number of low-income patients.

[9] American hospitals have traditionally competed for physicians and their patients on the basis of specialised clinical services and amenities rather than price and as a result, this competition has lead to increased costs for hospitals.

[10] Hospitals that deal with a greater number of emergency cases will have tendency to staff more conservatively, maintaining an excess of staff capacity to deal with possible peak periods.

day was around 113 beds. It was concluded that, "*providing hospital care in rural areas with fewer beds concentrated in larger, better utilised facilities would lower average costs per patient day*".

In fact, many of these studies focused mainly on the analysis of possible closures or mergers between small hospitals in rural areas in order to take advantage of economies of scale and therefore decrease unit costs.

Mayer *et al.* (1987) conducted a survival study that analysed the patterns of rural hospital closures and concluded that closure was explained by a combination of external and internal variables, with the ownership of the hospital being consistently the most significant variable. It was also concluded that the rural hospitals most likely to close were the ones located in areas adjacent to metropolitan areas and with fewer admission and lower marginal occupancy rates.

Sinay (1998) estimated a multi-product hybrid translog cost function to investigate the operating efficiencies of merging and closing hospitals and concluded that rural hospitals were operating at a size level that had great potential for achieving scope and scale efficiencies through mergers.

Simpson (1995) on the other hand, focused his study in the analysis of hospital minimum efficient size and found that of the twenty hospitals that had opened in California in the period in analysis, fourteen had under 100 licensed beds and eighteen had fewer that 200. The idea that sub-100 bed hospitals are inefficient and cannot efficiently provide a wide range of hospital services is therefore contradicted. Still, although focusing their analysis on small hospitals efficiency, it should be noted that these hospitals were not necessarily located in rural or remote areas[12].

A few studies incorporating the rural/urban feature studied the cost of a specific diagnosis and not the hospital as a whole.

[11] It was considered that higher inpatient costs resulting from admissions through the emergency room resulted mainly from diseconomies of scope and/or unmeasured differences in patient ill severity.
[12] Some of them were located in Los Angeles and San Diego for instance.

Culler *et al.* (1995) analysed the relative cost of providing a specific procedure, knee replacement surgery, to rural residents in rural community-based hospitals rather than in urban hospitals. A total costs function, having the patient as the unit of analysis, was estimated, taking into account variables such as individual demographic and health status variables, hospital characteristics and environmental variables. The results indicated that the predicted cost of this procedure for any given type of patient was less when delivered in a "typical" rural than when delivered in a "typical" urban hospital.

Lave *et al.* (1996) undertook a retrospective database study examining patients hospitalised with community-acquired pneumonia and concluded that the treatment of residents from rural areas differed from that of residents in more urban areas. Rural residents were more likely to be treated by family doctors while the urban residents were mainly treated by internal medicine specialists and also, in rural areas admission to intensive care units and the use of mechanical ventilation was less likely to occur than in urban areas. Therefore, this use of fewer resources and shorter lengths of stay resulted in the costs of treating urban residents being higher than that of treating rural residents. Still, it was stated that one of the limitations of the study was that the data on rural patients was incomplete and therefore could had led to an underestimation of hospitalisation costs for rural patients.

Bronstein *et al.* (1997) compared the cost of care for two common childhood diseases across providers practising in rural, small town and urban counties in Alabama and, after controlling for factors such as physician characteristics and patient demographic and utilisation factors, the costs of treating urban residents were found to be higher than that of rural ones. Possible explanations for this could be the lesser number of rural episodes that included outpatient facilities charges and the fewer ancillary services that were provided in rural settings. Still, this implied that rural physicians had a potentially unfavourable care content profile, which raises questions of care quality.

Heller *et al.* (2000) examined the difference in use of cardiac procedures during and after hospital admission for acute chest pain in 47 metropolitan or non-metropolitan hospitals across New South Wales. It was observed that patients admitted to metropolitan hospitals had higher rates of most cardiac procedures than did patients in non-metropolitan hospitals. Those differences disappeared during the 22 months after

discharge, when the rates were similar between hospitals. It should be noted that the influence of other factors such as geographical, racial or gender differences in cardiac procedures rates were taken into account in this study, although a systematic influence of age on all procedure groups examined was reported.

Finally, Gould and Moon (2000), in one of the very few non-American studies undertaken in this specific area, analysed the problems of providing health care in British Island communities through an observational study (telephone survey). A case study regarding the Isle of Wight offered no conclusive proof that the cost per capita of providing care was significantly higher than in other areas of the U.K. and therefore the authors concluded that the precise magnitude of the diseconomies of scale stemming from island status was an issue that waited detailed empirical study.

This study therefore, although not making any kind of analysis of the level of costs in rural hospitals relative to urban ones, did partly to corroborate the idea of the existence of economies of scale at a relatively low level.

2. Hypothesis Formulation

Although the study of hospital's production possibilities has been essentially constituted of empirical estimations of costs functions, the estimation of these cost functions still brings about a range of problems, from the measurement of output to the actual specification of an appropriate functional form[13].

Nevertheless, even if we are not able to specify the "true" output measure or the "true" functional form, it is all the same quite unambiguous to state that the hospital's output can be described as the technological relationship that allows a certain level of output (Q) to be obtained, given a set of inputs such as labour (L) and capital (K):

$$Q = f (L, K)$$

[13] For a review on these questions see Mann and Yett (1968), Lave and Lave (1970), Cowing *et al.* (1983), Tatchell (1983), Breyer (1987) and Ellis (1991)

And thus, we can think of hospital costs, in a very simple and straightforward way, as the cost of producing that level of output Q, given a certain level of the inputs K and L and their prices, represented by a vector *p*:

$$\text{Cost} = C (Q, p)$$

We can then start thinking how these factors that shape hospital costs are potentially important in the differentiation between hospitals located in rural and remote areas. When thinking about the prices of inputs, one has to consider on one hand the price of the input labour, given the peculiar features of the labour market associated with hospitals, but also the price of other inputs, e.g. capital.

In addition, one should take into account factors related with the level and nature of output, which will too have an impact on hospital costs. One factor is the possibility that the size of the hospital (one will expect to find smaller size hospitals in rural and remote areas), and thus the level of output, will lead to the existence of diseconomies of scale and therefore to higher level of costs.

There are also other dimensions, not directly associated with the hospital in itself but with the external factors, such as the possibility that the lack of a network of social support in rural and remote areas might lead to delayed discharges and thus affect the level of output and costs.

2.1 Prices of Inputs

The same amount of health expenditure will buy different amounts of care in different locations and such differentials in provider prices might be caused by local variations in market forces. We can think of Market Forces Factor (MFF) as the extra cost in the provision of health care derived from the higher costs, in certain locations, of resources such as labour and capital (made of land, buildings and equipment), caused by local variations in market forces.

2.1.1 *Labour*

Labour costs account for significant proportion of the total hospital running expenses, with pays accounting for around 70% of acute hospital expenditure (Scott *et al*., 2003).

Therefore, since they represent such an important part of labour and thus hospital costs, wage rates were considered an essential determinant of hospital costs from the early hospital cost studies.

In one of the first studies of hospital costs, Cohen (1967) stated that wage cost differentials inflated costs of urban hospitals relative to the costs of hospitals in nonurbanised areas.

In 1983[14], when ad-hoc specifications of cost functions started to be replaced by production-theoretic specifications, the importance of wages was reinstated trough their specific inclusion in the input prices vector.

One can describe the labour market as the context in which the employees sell the services of their labour and buy the attributes of their jobs, whilst, simultaneously, the employers buy the services and characteristics of workers and sell the attributes of the jobs offered. When the choices of both the employer and employee match, regarding the worker characteristics and the job attributes, then market equilibrium is achieved, with a corresponding distribution of wages. It would then be expected that, *"If the market works without hindrance, wages of all employers would equalise*[15]*"*.

But, not all jobs or employees are the same. There is much more to a job than the wage associated with it. The non-pecuniary aspects of employment also influence the valuation of that same job. *Ceteris paribus*, the wage will vary according to the job attributes. This means that, everything else equal[16], different jobs, with different characteristics, will be rewarded with different wages rates.

[14] Cowing *et al*., (1983) and Conrad and Strauss (1983).
[15] Ehrenberg and Smith, 1988, page 251.
[16] We are assuming that workers are comparable, in the sense that all other conditions of work that might influence wages, such as worker characteristics (skill level, age race, gender etc), are kept constant.

When examining this issue, the theoretical framework of compensating wage differentials[17] is usually employed.

The reference work on this theory is Adam Smith's "Wealth of Nations", where he describes his theory of net advantages of different jobs, resulting in that "*the whole of the advantages and disadvantages of different employments of labour and stock must in the same neighbourhood be either perfectly equal or continually tending to equality. If in the same neighbourhood, there was any employment evidently either more or less advantageous than the rest, so many people would crowd into it in the one case, and so many would desert it in the other, that its advantages would soon return to the level of other employments*".

According to the theory of compensating wage differentials, individuals can then be persuaded to accept less attractive jobs by compensating differences in their wage rates, i.e., any non-monetary disadvantages in a given job can be compensated through higher wages. On the contrary, favourable working conditions will be associated with the payment of lower than average wages. Basically, this means that positive wage differentials, that is higher wages, will arise with jobs associated with "unpleasant characteristics", whilst negative wage differentials (lower wages) will be associated with "pleasant characteristics".

There are several "pleasant/unpleasant" characteristics, or job attributes, which might influence wage differentials (Rosen, 1986)[18]. The first, and probably most obvious one, is the working environment, or what Smith (1776) called the "*agreeableness of the job*". If a worker has to work in displeasing working conditions, where, for instance, risks to life and death exist, one would expect that a premium would have to be paid to compensate for those harsh conditions.

Another important attribute is the existence of local amenities, such as intercity and interregional differences in climate, crime, pollution and crowding. This attribute is

[17] See Rosen (1986).
[18] See also the survey made by Rosen (1986) on the attributes that have been empirically proved to determine wage differentials.

highly subjective in the sense that some of these factors are highly dependent on personal tastes and vary from person to person[19].

The potential for training and learning, as well as uncertainty associated with some jobs[20] are also crucial. One will expect to get compensated for jobs where there is few possibilities for training and learning and where the probabilities of success are more insecure.

Differences in working schedules, such as the possibility of shift work and flexible work schedules will be expected to influence wages differentials as well, depending, once again, on personal preferences.

Finally, the "*constancy of employment*"[21] or unemployment and income risk also has to be taken into account since workers will need to be compensated for jobs where unemployment risk is higher.

Which of these job attributes will then determine the existence of compensating wage differentials and therefore affect the wages of hospitals located in remote and rural areas?

The decision to work in a rural/remote hospital might, potentially, be influenced by nonpecuniary aspects of jobs, namely work-time characteristics such as a higher proportion of "unsocial" hours of work (night shifts) and inflexible work schedules. In theory, the workers will, therefore, be in a position where they are offered a fixed "package" of work hours and have to decide accordingly. One would, thus, expect that workers in rural and remote hospitals would have to be compensated for that inflexibility and therefore be paid a higher wage[22].

Furthermore, when working in a smaller rural hospital, the opportunity for training and specialisation would be considerably lower and therefore it might be difficult to attract

[19] One may prefer live in a busy city while others prefer the peaceful countryside.
[20] The so-called "*probability of success*", described by Smith (1776).
[21] Smith (1776).
[22] Certain non-monetary characteristics, such as out of hours work, was found to be more important than income in GPs' job choices (Scott, 2001; Sibbald *et al*. 2003).

qualified staff, without some kind of monetary compensation[23]. All of this would lead us to think that wages should be higher in remote and rural hospitals.

On the other hand, it can also prove difficult to attract workers in big city hospitals, because of higher housing costs, high standard of living and higher crime rates and thus we could expect higher wages in such hospitals. In fact, although it depends on personal preferences, it is usually considered more pleasant to live in smaller cities[24] and it might be harder to retain staff in a local labour market that offers many other work opportunities, at similar wage levels[25].

It is therefore difficult to predict, theoretically, where will wages rates be higher, as this depends on the combination of relative advantages/disadvantages each worker values.

If hospitals in rural and remote areas do face recruitment and retention difficulties, one would expect that a positive wage differential would be needed to attract workers. But do hospitals have the necessary degree of flexibility in order to set wage rates as a response to the characteristics of the local labour market[26]?

In reality pay is set at a national level in the NHS and although an internal market was introduced, in 1991, where trusts were given powers to adjust pay and conditions of

[23] In 2002, the Advisory Group Commissioned by the Scottish Executive to Review the Scottish Medical Workforce emphasised the need "(...) *to ensure that flexible opportunities for working and training are relevant and effective for rural and remote environment. Remote and rural versions of training, continuing professional development, career planing and imaginative uses of IT are needed to handle the difficulties created by travel, distance and isolation from colleagues*".

[24] Smaller crime rates, crowding and pollution levels, lower costs of living.

[25] Advisory Group Commissioned by the Scottish Executive to Review the Scottish Medical Workforce suggested the need for the development and promotion of a "rural" package for doctors and their families in order to "*improve awareness of the distinctive, and in many ways positive; aspects of practice in such an environment*" (2002).

[26] It should be pointed out that, when talking about local labour markets for the NHS, one should distinguish between the major professionals groups (doctors, dentists, nurses) and other non-professional staff (administrative and clerical, ancillary and unqualified nurses). When dealing with non-professional staff, the NHS has to compete with other local employers since that staff group can easily find "proxy" jobs in the local labour market. On the other hand, in the case of the professional groups, it can be considered that an internal labour market (defined by Filer *et al*, 1996, as "*all the ways in which employment relationships are structured to create ands maintain incentives for workers to produce when there are investments in specific human capital that would be lost if the employment relationships were severed*") exists, with the NHS being basically the only employer, since the employees have sector specific-skills and cannot move easily to another local sector, unless other employers exist. In this case the NHS would not have to compete directly with the local labour market and thus it can be argued that factor price variation does not apply to doctors, dentist and nurses since they have extreme low mobility, at local level.

employment, by the mid-90s local pay agreements were still rare, with about two thirds of the trusts having less than 20% of staff on local contracts (Gray and Buchan, 1998).

But, even when health authorities, in a framework of national pay-scales, are not able to respond to labour market conditions by adjustments to pay and conditions, as the theory of compensating differentials would suggest, this pressure might lead to an impact on other aspects of labour costs.

The theory of wage differentials still applies even in conditions where hospitals do not have the flexibility to set pay conditions, that is, when there is a deviation of the actual wage offered to the one that would result from the market equilibrium. For instance, local managers might try to compensate for any job disadvantages by promoting grade or wage drift, i.e., recruiting staff to higher grades than it would be usual[27].

Another non-monetary device used to compensate employees for working in less pleasurable jobs is what is called "fringe benefits", which may include vacations and pensions, that are used as a substitute for direct wage compensations.

If these solutions prove to be insufficient to compensate the employee, then the employer will end up being faced with consequences such as recruitment and retention problems[28] and also greater staff wastage and high turnover rates, which may lead to increases in costs.

"The probability of worker turnover in a job is (...) inversely related with the deviation between the wage paid and the wage predicted from the market wage-attributes regression line[29]*"*

[27] Allowing this way employees to be monetarily compensated without an actual increase in wage rates. For example, hospital consultants given distinction awards see their salary almost doubled.

[28] "(...) *we found that the status and location of a hospital may affect its ability to attract recruits. On the whole, the main teaching centres tend to fare better than smaller community or district hospitals*", Accounts Commission for Scotland, 1998; *"Recruitment and retention difficulties can be most intensive in remote and rural areas*", Advisory Group Commissioned by the Scottish Executive to Review the Scottish Medical Workforce, 2002.

[29] Rosen (1986).

Labour turnover in the NHS, especially that connected with nursing, have for long been pointed out as one of NHS biggest problems since high rates of turnover can have a significant impact on hospital costs.

Turnover costs derive from three broad categories (Gray *et al*., 1996): the costs of recruiting new staff (advertisement, selection, interviewing); the costs of filling the vacancy while a new staff member is not hired[30] (overtime payments, agency fees) and productivity losses incurred since new staff take some time to achieve the average level of productivity as the previous workers.

It is not clear if turnover rates are higher in hospitals in remote and rural areas.

Gray and Phillips (1996) found significant association between turnover rates, across a range of staff groups, and the size of the private health care sector (number of beds in private hospitals and nursing homes) and the pay of the staff group relative to the local average for equivalent workers. This could lead us to think that rural and remote hospitals should have lower turnover rates as they face less competition from the local private sector.

However, a review paper (Tai *et al*., 1998), dealing with nursing turnover, found that "most researchers did not find a significant relationship between staff turnover rates and facility location".

The effect of turnover rates on labour costs in rural and remote areas is therefore not straightforward.

2.1.2 *Capital*

Finally, the other input one should consider is related with the price of land and of physical estate (rent and rates), the purchase of equipment and of goods and services.

[30] "*Overall, expenditure on locums by Scottish trusts has almost doubled over the last three years (...) for most long-term locum appointments, for example, there will be compensating savings in the salaries budget from unfilled vacancies (...) by contrast, trusts have to meet the full cost of providing short-term locum cover*", Account Commission for Scotland, 1998.

Although land costs are likely to be higher in urban areas, the effect on hospital costs depends on whether individual hospitals face the true price.

Especially relevant in the case of hospitals located in Islands, it might be possible that hospitals located in rural and remote areas experience higher unit costs due to higher transportation costs of supplies.

The Review of the Area Cost Adjustment (Elliot *et al*., 1996), which developed an Area Cost Adjustment (ACA) in order to *"compensate local authorities in parts of England for the higher costs of inputs they encounter"*. The ACA was designed so as to reflect differences in labour costs but also in what it called the Rates Cost Adjustment, although recognising that it represented a relatively small proportion of local authority expenditure. Possible non-labour expenditure included the cost of occupying land and buildings (rent and rates); the purchase of equipment and vehicles; cost of insurance and the supply of good and services such as water, gas and electricity[31].

The English allocation formula to primary care trusts includes a land index[32] and a building cost location factor[33], admitting the fact that land/building costs might vary between inner cities and rural sites (Department of Health, 2003).

In the Scottish case, Fair Shares for All considered the possibility that factors such as the price of land and the value of physical estate, might, for example, very between inner city areas and rural sites. It eventually decided that since 61% of total hospital running expenses comprised staff salaries and capital charges accounted for less than 11% of the running costs of hospitals in 1997/98, a land/buildings factor could not make a sufficient material difference to justify the work needed to develop it as an element in the proposed formula. Still, it emphasised that more research should be made in order to consider its inclusion in future refinements of the formula (The National Review of Resource Allocation for the NHS in Scotland, 2000).

[31] The specific case of the Islands of Wight and Scilly was also analysed, since they might have been faced with extra costs due to additional transportation costs due to their island and remote location.
[32] Using data from the Valuation Office's valuation of the NHS estate.
[33] Based on a rolling average of tender prices for all public and private contracts, provided by the Building Cost Information Service to the Valuation Office.

In empirical terms, particularly in the last wave of hospital cost studies, that make use of production-theoretic specifications of cost functions, the price of materials and supplies has been considered as a relevant factor of hospital costs, and have often been included in the input prices vector[34].

2.2 Output related features

2.2.1 *Diseconomies of scale and scope*

The existence of diseconomies of scale in smaller size hospitals (such as one would expect to find in rural and remote areas of Scotland) has been repeatedly pointed out as the major reason for those hospitals experiencing higher costs. The Scottish formulae of hospital resource allocation specifically admitts the existence of diseconomies of scale and scope in hospitals located in rural and remote areas and thus adjusts the formulae accordingly.

Despite that, strong and irrefutable evidence that economies of scale occur from a certain threshold and that rural and remote hospitals incur in diseconomies due to their smaller size has yet to be found.

Economic theory suggests several reasons for the existence of economies of scale[35] that can be applied to the specific case of hospitals (Aletras *et al.*, 1997 and Harrison and Prentice, 1996):

---

[34] Still, in most of these studies (Cowing and Holtman, 1983; Eakin and Kniesner, 1988; Vita, 1990; Fournier and Mitchell, 1992) mostly because of unavailability of data on these prices, it was assumed that the price of non-labour inputs was constant, since it was bought at uniform prices across hospitals, or that they were highly correlated with local wages (Carey, 1997). In the specific case of long-run hospital cost functions (Grannemann et al., 1986) capital prices were assumed either to be constant or that they were adequately controlled for by profit/non for-profit and system/non-system dummies.

[35] In the estimation of economies of scale one should make a distinction between short-run and long-run economies. *Short-run* economies of scale can be estimated by relating cost effects to output changes, keeping the capacity constant, that is to say, not allowing the fixed factor (e.g. beds) to change (i.e. we will be measuring what effect will output increases have on hospital variable costs, for a given hospital size). However, this does not truly satisfy the economic definition of economies of scale, which is by nature a long run phenomenon. The ideal approach would be to derive *long-run* scale economies from the true long-run cost function, which can be obtained by estimating short-run cost functions as a base and then derive the long-run cost-function by using the envelope theorem. See Aletras (1999) for a more detailed discussion about the problem of long-run/short-run scale estimation.

- a large scale of production allows for specialisation to occur (larger hospitals provide more opportunities for staff to specialise in specific functions) which may lead to greater efficiency and reduced unit costs
- technical factors such as the existence of significant fixed costs that can be spread as production increases (expensive fixed assets such as operating theatres or diagnostic equipment may be used more intensively in large hospitals)
- economies on the reserves of labour or materials existing in bigger hospitals (need for a proportionally smaller number of beds or staffing to respond to unexpected demand for instance)
- pecuniary economies such as quantity discounts or lower interest rates on borrowed capital

Regarding the possible economies on the reserves of labour in bigger hospitals, it should be noted that the influence of a bigger number of staff on hospital unit costs is debatable. One could assume that the scope for the existent staff to cover for absentees in smaller size hospitals is less and thus will lead to higher unit labour costs due to out of hour's work or the need to resort to agency work. Alternatively, it might be that in rural and remote hospitals staff are hired to perform tasks that in other hospitals would be performed by more than one person and would thus lead to costs saving in the sense that only one wage will be paid for the performance of a wider range of tasks than usual.

A systematic review of literature was carried out by the NHS Centre for Reviews and Dissemination (1996) and after analysing more than 100 studies about the existence of economies of scale or scope, they concluded that, *"the more reliable studies find constant returns or even diseconomies for the average hospital: one with roughly 200-300 beds*" and "*if any economies exist they appear to be quickly exhausted or outweighed by diseconomies*". The existence of economies of scope, "*was not in general empirically validated for the average hospital, but this does not necessarily mean they are absent*".

This level of doubt regarding the existence of economies of scale and/or scope seems to suggest the need for further investigation in order to access if it determines in any way the level of costs. However, the estimation of economies of scale and scope poses several problems, which have been the main source of the divergent results in literature.

Among these problems one can enumerate questions associated with the difficulty of measuring hospital output, namely its definition and multiple nature[36]; the multiplicity of factors that influence hospital costs and the difficulty to account for the contribution of each of them, and also the non existence of a unarguable methodology in the estimation of cost functions[37] (Oliveira e Bevan, 2003).

2.2.2 *Lack of social support in rural and remote areas*

The inexistence of a network of social support, such as those existing in urban hospitals, might be one of the sources of rural/remote hospital higher costs because of higher lengths of stay per patient.

A report by the National Audit Office (2003) found that a significant number of older patients occupying adult and general NHS acute beds[38], and already declared fit and safe, had their discharged delayed, and in a third of cases, discharge was delayed for more than a month. In fact, in September 2002 around 8.9 per cent of older patients occupying NHS acute beds had their discharge from hospital delayed. The rates of delay in Scotland were very similar. These delays are not only prejudicial for patients but also have a negative impact on costs (the NHS Confederation calculated that delayed discharges account for around 2.2 million lost bed-days for all type of NHS bed each year[39]).

The reasons for delayed discharges are related not only with internal hospital processes (co-ordination of discharge planning and availability to transport the patients back

[36] Like other multi-product firms, which are characterised by settings with hundreds of outputs, hospitals also produce various types of output (inpatient and outpatient care as well as education and research services), with the added complexity associated with the fact that none of those outputs is homogeneous across hospitals, raising problems of case-mix control.

[37] Vita (1990) considered that none of the possible functional forms was without problems and that the choice of the functional form depended mainly on the research objective.

[38] People aged 65 years and over are the main users of the NHS, especially acute care in hospitals, occupying almost two-thirds of general and acute beds (Department of Health, 2000). Findings of the National Bed Inquiry.

[39] House of Commons Health Committee (2002). Delayed Discharges. Third Report of Session 2001-02, HC 617 – paragraph 31.

home) but, predominantly, with external factors such as the provision of care in the community.

It was found that the major cause of delay was the existence of resource constraint in the provision of nursing and residential places. One could then hypothesize that in rural and remote areas, the constraint on the provision of care in the community would be greater, thus increasing the delay in discharges and consequently hospital costs. One example would be the possibility that hospitals in rural and remote areas might treat as inpatients potential outpatients, due to less scope to perform day surgery (due to bigger travelling distances).

2.5 Conclusion

The analysis of existing literature on hospital costs and the way it deals with rurality and remoteness makes one conclude that such dimension of costs has not yet been dealt with in a systematic way. Some studies have tried to address this question indirectly, through the analysis of factor price variation on hospital costs, while others merely studied the pattern of closures/mergers between hospitals in rural areas.

It was thus considered that one should go back to theory of production and cost functions and try to access how the basic components of a cost function, inputs and output, would be affected by the remoteness and rurality dimension. Input prices and output related features were therefore examined.

The main component of input prices for hospitals is labour and special attention was hence paid to labour market issues, specifically trying to ascertain in which way the theory of compensating wage differentials applied, when dealing with the specific question of rurality and remoteness. It was found that the impact of this location dimension on wage rates was uncertain as it depended on the combination of relative advantages and disadvantages each worker valued.

The framework of national pay scales that characterizes the NHS made it necessary to also take into account labour market features, other than wage rates, such as turnover

rates. However, it was considered that their effect on labour costs was also not straightforward.

Non-labour input prices associated with capital costs were considered, as well, and hypothesised to be relevant, although have been difficult to include in empirical work due to data unavailability.

Regarding output features, the question of economies of scale and scope was addressed, as well as the problems associated with their estimation and the consequent inconsistence of the results obtained in the existing studies. Finally, it was hypothesised that hospitals might suffer from a lack of social support in rural and remote areas, which would lead to higher lengths of stay and thus higher costs.

In summary, although reasons were pointed for the existence of higher costs in rural and remote hospitals, other influencing factors have dubious effect on the level of costs and some might even contribute to lower costs.

The purpose of the following analysis will therefore be the assessment of the net effect of all the hospital cost determinants listed above.

References

Accounts Comission for Scotland 1998, *Doing the Rounds: the use of locum doctors in Scotland's hospitals*.

Advisory Group Comissioned by the Scottish Executive to Review the Scottish Medical Workforce 2002, *Future Practice: A review of the Scottish Medical Workforce*, Scottish Executive.

Aletras, V. H. 1999, "A comparison of hospital scale effects in short-run and long-run cost functions", *Health Economics*, vol. 8, no. 6, pp. 521-530.

Aletras, V., Jones, A., & Sheldon, T. A. 1997, "Economies of scale and scope," in *Concentration and choice in health care*, T. A. Sheldon, B. Ferguson, & J. Posnett, eds., The Royal Society of Medicine Press Limited.

Breyer, F. 1987, "The specification of a hospital cost function: a comment on the recent literature", *Journal of Health Economics*, vol. 6, pp. 147-157.

Bronstein, J. M., Johnson, V. A., & Fargason Jr, C. A. 1997, "How rural physicians compare on cost and quality measures for Medicaid ambulatory care episodes", *The Journal of Rural Health*, vol. 13, no. 2, pp. 126-135.

Carey, K. 1997, "A panel data design for estimation of hospital cost functions", *The Review of Economics and Statistics*, vol. 79, no. 3, pp. 443-453.

Carey, K. & Burgess JR., J. F. 1999, "On measuring the hospital cost/quality trade-off", *Health Economics*, vol. 8, pp. 509-520.

Cohen, H. A. 1967, "Variations in cost among hospitals of different sizes", *Southern Economic Journal*, vol. 33, pp. 355-366.

Conrad, R. F. & Strauss Robert P. 1983, "A multiple-output multiple-input model of the hospital industry in North Carolina", *Applied Economics*, vol. 15, pp. 341-352.

Cowing, T. G., Holtmann, A. G., & Anderson, G. F. 1983, "Multiproduct short-run hospital cost functions: empirical evidence and policy implications from cross-section data", *Southern Economic Journal*, vol. 49, pp. 637-653.

Cowing, T. G., Holtmann, A. G., & Powers, S. 1983, "Hospital Cost Analysis: a survey and evaluation of recent studies", *Advances in Health Economics and Health Services Research*, vol. 4, pp. 257-303.

Culler, S. D., Holmes, A. M., & Gutierrez, B. 1995, "Expected Hospital costs of knee replacement for rural residents by location of service", *Medical Care*, vol. 33, no. 12, pp. 1188-1209.

Culyer, A. J., Wiseman, J., Drummond, M. F., & West, P. A. 1978, "What accounts for the higher costs of teaching hospitals?", *Social and Economic Administration*, vol. 12, no. 1, pp. 20-30.

Department of Health. History of the Staff MFF. www.doh.gov.uk/pub/docs/doh/rawp1.pdf . 1998, access 5-5-2003.

Department of Health 1998, *Acute Services*, The Scottish Office.

Department of Health. Resource Allocation: Weighted Capitation Formula. www.doh.gov.uk/pub/docs/doh/capitation-formulas.pdf . 2003 access 10-5-2003.

Eakin, B. K. & Kniesner, T. J. 1988, "Estimating a non-minimum cost function for hospitals", *Southern Economic Journal*, vol. 54, pp. 583-597.

Ehrenberg, R. G. & Smith, R. S. 1988, *Modern Labor Economics: theory and public policy*, 3 edn.

Elliot, R. F., McDonald, D., & MacIver, R. 2003, *Local Government Finance: Review of the Area Cost Adjustment*, University of Aberdeen.

Ellis, R. P. Hospital cost function estimation when firms may not try to minimize total costs. Discussion Paper # 29, Department of Economics, Boston University . 1993.

Filer, R. K., Hamermesh, D. S., & Rees, A. E. 1996, *The economics of work and pay*, 6 edn, Harper Collins College Publishers.

Finch, L. E. & Christianson, J. B. 1981, "Rural hospitals costs: an analysis with policy implications", *Public Health Reports*, vol. 96, no. 5, pp. 423-433.

Folland, S. T. & Hofler, R. A. 2002, "How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production", *Health Economics*, vol. 10, no. 8, pp. 683-698.

Fournier, G. M. & Mitchell, J. M. 1992, "Hospital costs and competition for services: a multiproduct analysis", *The Review of Economics and Statistics*, vol. 74, no. 4, pp. 627-634.

Friedman, B. & Pauly, M. V. 1981, "Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals", *The Review of Economics and Statistics*, vol. 63, no. 3, pp. 620-624.

Gould, M. I. & Moon, G. 2000, "Problems of providing health care in British islands communities", *Social Science and Medicine*, vol. 50, pp. 1081-1090.

Grannemann, T. W., Brown, R. S., & Pauly, M. V. 1986, "Estimating hospital costs: a multiple-output analysis", *Journal of Health Economics*, vol. 5, pp. 107-127.

Gray, A. M., Phillips, V. L., & Normand, C. 1996, "The costs of nursing turnover: evidence from the British National Health Service", *Health Policy*, vol. 38, pp. 117-128.

Gray, A. M. & Phillips, V. L. 1996, "Labour turnover in the British National Health Service", *Health Policy*, vol. 36, pp. 273-289.

Gray, A. M. & Buchan, J. 1998, "Pay in the British NHS: a local solution for a national service?", *Journal of Health Services Research Policy*, vol. 3, no. 2, pp. 113-120.

Harrison, A. & Prentice, S. 1996, *Acute Futures* King's Fund.

Heller, R. F., O'Connell, R. L., D'Este, C., Lim, L. L., & Fletcher, P. J. 2000, "Differences in cardiac procedures among patients in metropolitan and non-metropolitan hospitals in New South Wales after acute myocardial infarction and angina", *Australian Journal of Rural Health*, vol. 8, no. 6, pp. 310-317.

Jacobs, R. Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. [177]. 2000. The University of York. Discussion Paper. Centre for Health Economics.

Koop, G. & Carey Kathleen. Uncertain demand and hospital cost function estimation. Management Sciences group, Department of Veteran Affairs. 1991.

Lave, J. R. & Lave, L. B. 1970, "Hospital cost functions", *American Economic Review*, vol. 60, pp. 379-395.

Lave, J. R. & Lave, L. B. 1970, "Economic analysis for health service efficiency: a review article", *Applied Economics*, vol. 1, pp. 293-305.

Lave, J. R., Lave, L. B., & Silverman Lester P. 1972, "Hospital cost estimation controlling for case-mix", *Applied Economics* , vol. 4, pp. 165-180.

Lave, J. R., Fine, M. J., Sankey, S. S., Hanusa, B. H., Weissfeld, L. A., & Kapoor, W. N. 1996, "Hospitalized Pneumonia: Outcomes, Treatment Patters, and Costs in Urban and Rural Areas", *Journal of General Internal Medicine*, vol. 11, pp. 415-421.

Li, T. & Rosenman, R. 2001, "Estimating hospital costs with a generalized Leontief function", *Health Economics*, vol. 10, no. 6, pp. 523-538.

Maniadakis, N. & Hollingsworth, B. The impact of the internal market on hospital efficiency productivity and service quality. Paper presented at HESG. 1998.

Mann, J. K. & Yett, D. E. 1968, "The analysis of hospital costs: a review article", *Journal of Business*, vol. 41, pp. 191-202.

Mayer, J. D., Kohlenberg, E. R., Sieferman, G. E., & Rosenblatt, R. A. 1987, "Patterns of rural hospital closure in the United States", *Social Science & Medicine*, vol. 24, no. 4, pp. 327-334.

NHS Centre for Reviews and Dissemination 1997, *The Relationship between hospital volume and quality of health outcomes*, The University of York, CRD Report 8 Part I.

NHS Centre for Reviews and Dissemination 1997, *The Relationship between volume and the scope of activity and hospital costs* CRD Report 8 Part II.

Oliveira, M., Bevan, G., & . A multilevel model to estimate unavoidable costs and to disentangle allocative inefficiences of hospital care. Paper presented at HESG. 2003.

Pauly, M. V. 1978, "Medical Staff characteristics and hospital costs", *The Journal of Human Resources*, vol. 13, no. supplement, pp. 77-111.

Ricketts, T. C. 2000, "The changing nature of rural health care", *Annual Review Public Health*, vol. 21, pp. 639-657.

Scott, A. 2001, "Eliciting GPs preferences for pecuniary and non-pecuniary job characteristics", *Journal of Health Economics* , vol. 20, pp. 329-347.

Scott, A., Gillespie, G., Major, K., & Morris, S. Pay in National Health Service: an evidence-based policy? Report for Scotecon. 2003.

Scottish Health Service Advisory Group 1995, *Health Services in Remote and Island Areas in Scotland* The Scottish Office Home and Health Department.

Sibbald, B., Bojke, C., & Gravelle, H. 2003, "National Survey of job satisfaction and retirement intentions among general practitioners in England", *British Medical Journal*, vol. 326.

Simpson, J. 1995, "A note on entry by small hospitals", *Journal of Health Economics*, vol. 14, pp. 107-113.

Sinay, U. T. 1998, "Hospital mergers and closures : survival of rural hospitals", *The Journal of Rural Health*, vol. 14, no. 4, pp. 357-365.

Smith, A. 1970, "Of Wages and Proffit in the different Employments of Labour and Stock," in *The Wealth of Nations*, Pelikan Books, pp. 201-247.

Tai, T. W. C., Bame, S. I., & Robinson, C. D. 1998, "Review of Nursing Turnover Research, 1977-1996", *Social Science and Medicine*, vol. 47, no. 12, pp. 1905-1924.

Tatchell, M. 1983, "Measuring hospital output: a review of the service mix and case mix approaches", *Social Science & Medicine*, vol. 17, no. 13, pp. 871-883.

The Comptroller and Auditor General 2003, *Ensuring the effective discharge of older patients from NHS acute hospitals*, National Audit Office.

The National Review of Resource Allocation for the NHS in Scotland 2000, *Fair Shares for All: Final Report*, Scottish Executive.

Thorpe, K. E. 1988, "Why are urban hospital costs so high? The relative importance of patient source of admission, teaching, competition, and case-mix", *Health Services Research*, vol. 22, no. 6, pp. 821-834.

Vita, M. G. 1990, "Exploring Hospital Production Relationships with Flexible Functional Forms", *Journal of Health Economics*, vol. 9, pp. 1-21.

Vitaliano, D. F. 1987, "On the estimation of hospital costs functions", *Journal of Health Economics*, vol. 6, pp. 305-318.

Watt, I. S. & Sheldon, T. A. 1993, "Rurality and resource allocation in the Uk", *Health Policy*, vol. 26, pp. 19-27.