



# Competition and contestability between acute hospitals

Jon Sussex  
*Office of Health Economics*



Office of Health Economics  
12 Whitehall London SW1A 2DY

© March 1998, Office of Health Economics

Price £10.00

ISBN 1899040 35 8

Printed by BSC Print Ltd., London

## About the author

Jon Sussex is Associate Director of the Office of Health Economics.

## Office of Health Economics

The Office of Health Economics was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The Office of Health Economics is supported by an annual grant from the Association of the British Pharmaceutical Industry and by sales of publications, and welcomes financial support from other bodies interested in its work.

For recent OHE publications see inside back cover.

# Contents

Executive summary	4
1. Introduction	6
2. Why competition?	8
3. Does competition between hospitals exist in the NHS?	11
4. Economies of scale	14
5. Economies of scope	18
6. Medical training and practice – requirements and guidelines	20
7. Other barriers to competition	22
8. Contestability	27
9. How contestable are acute hospital services?	29
10. Conclusions	35
References	37
Appendix 1 – The Hirschman-Herfindahl Index (HHI)	40

# Executive summary

The UK government has said that it wants to replace competition between health care providers in the National Health Service (NHS) with co-operation, backed up in extreme cases by ultimate recourse to contestability. Contestability relies, not on the existence of a rival service provider, but on the credibility of the threat that sub-standard performance will ultimately be punished by a new entrant appearing to take over from the under-achiever. Before such an extreme is reached, however, contestability implies that, at the very least, the efficiency and effectiveness of health care providers is compared with that of their peers.

This monograph assesses the expected and revealed scope for competition in the provision of acute hospital services and the extent to which contestability is likely to offer anything more. The main points that arise from this evaluation are:

- the use of inter-hospital competition as a stimulus to improved efficiency and quality of care, and hence as a tool for advancing patients' interests, was presented by the then Conservative government as an important constituent of its 1989-1991 NHS reforms package;
- empirical evidence from the USA during the 1980s suggested that aspects of service quality had been the dominant field of competition between health care providers, but that the growth of managed care at the end of this period was leading to some price competition too. Even after the 1989-1991 reforms however, the structure of the NHS differs significantly from the US health care system. Consequently, it is uncertain how far the US experience would be relevant in the UK context;
- when viewed overall, studies attempting to determine the existence and impact of competition between acute hospitals in the post-1991 NHS have proven indeterminate. The existence of some localised monopolies in the provision of acute hospital services is apparent. Indeed this is partly a result of deliberate planning of the NHS in the past, aimed at achieving a network of 'district general hospitals'. However, there have also been some manifestations of competition for individual

specialties and procedures in the field of non-emergency care. The proportion of acute hospital services for which rival providers are actively in competition is nonetheless probably small;

- the obstacles to more widespread and effective inter-hospital competition in the NHS are many:
  - some limited economies of scale, reinforced by the requirements and guidelines of the medical professions, which mean that the minimum efficient scale of an acute hospital with an Accident and Emergency (A&E) department requires it to serve a population of hundreds, rather than tens, of thousands. Acute hospital care remains a labour-intensive process, however, so that economies of scale are unlikely to be large;
  - probably more important than economies of scale are the economies of scope that exist in an A&E-based acute hospital. Once the A&E services have been set up in accordance with medical professional guidelines, other elective or more specialised services may be added at the expense of relatively low marginal costs;
  - limited willingness by patients to travel beyond the nearest acute hospital; at least where the treatment or diagnosis involved is for something non-life-threatening and not particularly specialised;
  - the numerous ways in which the Department of Health in England and its counterparts in Northern Ireland, Scotland and Wales, regulate and restrict the NHS internal market. These include: controlling the supply of capital available to hospitals, limiting the numbers of doctors trained, achieving near-uniformity in the pay of hospital professionals nationally, requiring all providers to price at average cost including a six per cent real return on capital, and preventing providers from retaining any surpluses earned;
  - lack of good, timely, reliable information on quality and cost upon which to base contracts for health care services;

- the absence of direct rewards for the managers and providers of hospital services for competing with others. There is a similar lack of direct rewards for their counterpart managers in the health authorities that commission most of their services, for promoting that competition;
- the extreme public, and hence political, unpopularity of changing the provider of acute hospital services if this will involve the closure of an existing hospital facility;
- as these constraints to competition have become increasingly apparent, so the concept of contestability has been given greater attention. Competition requires the existence of more than one hospital serving a given population, which may be seen as wasteful over-provision outside the most densely-populated urban areas. Contestability does not require such over-provision, merely the possibility of a new entrant being ready to take over if the incumbent's performance slips to unacceptably poor levels;
- contestability is subject to some, but not all, of the constraints that limit competition between hospitals. Contestability can occur even in the presence of economies of scale and scope, and despite the medical training and practice requirements of the professions. Patients' willingness to travel for hospital care, and popular opposition to hospital closures, constrain contestability of the kind which would require the location of services to move physically from one hospital site to another but would not prevent contests to provide better and/or cheaper services from the same site. However, contestability is weakened, just as competition is, by Department of Health regulation of hospitals' output prices, return on capital, input supply and input prices; by the primitiveness of price and quality information; and by the extent to which providers and purchasers of care adopt satisficing, rather than maximising, behaviour;
- contestability is not practical where it would require a new entrant either to replicate an entire acute hospital, or to replace its entire staff. Individual specialties or procedures may more plausibly be subject to the threat of a new entrant, although almost certainly only by a hospital which is already providing such services to a different population. But there is little evidence of this happening in the NHS to date;
- the most plausible form of contestability appears to be that which would subject a hospital's management and administration, but not its front line care staff or facilities, to the threat of being replaced by new entrants. Something like this has happened with a small number of non-acute NHS Trusts, but only as a last resort following a protracted period of (ultimately failed) negotiation between the management of the Trust and the main commissioners of care services from it. So far, this has not happened to an acute hospital;
- reliance on contestability alone as a stimulus to greater efficiency and higher quality in the provision of acute hospital care, therefore appears inadvisable. Greater success towards achieving the same ends sought from contestability might be realised by concentrating on the creation and dissemination of appropriate benchmark information. An appeal to professional pride may, in a profession-driven public service like the UK NHS, ultimately achieve more than an appeal to competitive instincts. Clinicians and managers will be keen to establish and maintain a reputation for their hospital as high quality and not wasteful. Damage to such a reputation, once earned, could be fatal to the careers of all concerned;
- the threat of lost business or replacement of the management team could remain as the last resort for use only in the most extreme or intractable circumstances;
- the reforms of the NHS proposed by the current government in its December 1997 White Papers explicitly recognise the need for benchmarking and if carried through successfully should provide the foundations for greater use of that tool for improved performance.

# Introduction

The 1989-1991 Conservative reforms of the National Health Service (NHS) introduced fundamental changes in the organisation of health care in the UK. At the same time they emphasised for the first time in the context of the NHS the potential benefits to patients and taxpayers of competition between care providers. In December 1997, the publication of the new Labour government's own NHS reform White Papers for England and Scotland<sup>1</sup> marked the end of this promotion of competition and its replacement instead by exhortation to all parts of the NHS to cooperate with one another to provide 'integrated care'.

Lying behind the replacement of competition by co-operation, however, remains the explicit threat that health care commissioners 'as a last resort will be able to change provider if, over time, performance does not meet the required standard' (Cm 3807, para. 3.13). That is, provision of health care services is to be contestable, a point which was made even more explicitly by the then Labour spokesperson on health shortly before the 1997 general election: 'we envisage commissioning groups and health authorities using a system of

contestability to force improvements in standards' (Smith, December 1996). The division between providers of health care on the one hand and commissioners (formerly 'purchasers') of it on the other, which was introduced by the Conservatives as the core of their 1989-1991 NHS reforms, is to be retained.

The purpose of this monograph is to assess the extent to which competition between acute hospitals was ever possible in the UK and to what extent contestability has any greater applicability.

Hospital services in general account for around 54 per cent of total NHS expenditure, compared with around 33 per cent for all community and primary health services taken together (including the cost of dispensed medicines) and 13 per cent for central administration and other centrally run services (source: OHE, 1997, Table 2.16). Acute hospitals account in turn for the large majority of hospital expenditure. Of the £22 billion spent on English NHS Trusts in total in 1995/96, nearly two-thirds, or £14 billion, was for acute hospital services.

The structure of the NHS in 1997/98 and of the acute hospital services within it is the direct result of the 1989-1991 Conservative reforms. The January 1989 White Paper Cm 555 'Working for Patients' laid out the principles of the then proposed reform of the NHS. These principles were adopted essentially unchanged and given form by the NHS and Community Care Act of June 1990. This Act established, with effect from 1 April 1991, an organisational structure in which health authorities (HAs)<sup>2</sup> and fundholding general medical practitioners (GPs)<sup>3</sup> were responsible for determining and paying for the health care services to be provided to the populations they serve. NHS provision of health care was to be by separately-managed entities: NHS Trusts,<sup>4</sup> in addition to the primary care services which had since the creation of the NHS been provided by independent contractors – GPs, pharmacists, dentists and opticians – working in the community. The relationship between purchasers and providers of health care services is expressed in the form of one-year 'contracts'.<sup>5</sup> The private

1 For England Cm 3807, *The new NHS: modern, dependable*; and for Scotland Cm 3811, *Designed to care: renewing the National Health Service in Scotland*.

2 District Health Authorities in England and Wales, Health Boards in Scotland, Health and Social Services Boards in Northern Ireland. For brevity, the term 'health authorities' will be used throughout this paper to apply to all of these bodies.

3 The proportion of GPs volunteering to hold practice budgets for purchasing non-emergency health care services for their patients has increased steadily since 1991. By April 1997, approximately 60 per cent of the UK population was registered with a fundholding practice.

4 All NHS provision in the hospital and community health services (i.e. excluding the 'family health services': GPs and community pharmacists, dentists and opticians) is now by NHS Trusts, of which there are over 500 in the UK in total. Trusts are not just acute hospitals. They may provide community and/or mental health services or ambulance services.

5 Although NHS 'contracts' are not legally enforceable in the sense that one party may not sue another in the courts for breach of such a contract, they do specify the health care services to be provided and the revenue that these will attract to the provider and are the yardstick against which purchasers start to assess the performance of their providers. Failure by a provider to meet the terms of an NHS contract can lead to a loss of business in just the same way as would breach of a legally enforceable contract.

sector could also compete with the Trusts for contracts to provide health care to NHS patients.

The proposals set out in the English NHS White Paper of December 1997 will preserve the purchaser/provider split established in 1991. GP fundholding will be abolished from April 1999 but all GPs, whether currently fundholding or not, are to be involved in commissioning health care (commissioning being the present government's preferred euphemism for purchasing in the NHS). The GPs, organised into geographical groups (average size 50 GPs), will either advise their local HA who will then do the purchasing, or they will hold and operate their own group purchasing budgets. Annual contracts with NHS Trusts are to be replaced by three-year, or longer, 'service agreements'.<sup>6</sup>

The essentials of the NHS internal market structure will therefore remain, at least in England: health care purchasers/commissioners holding budgets on behalf of large populations and negotiating with separate providers (Trusts) on service volume, quality and price. Thus it remains, in addition to the patients themselves, the HAs, GPs, NHS Trusts and private hospitals who are the relevant parties in the discussion of competition and contestability in the provision of acute hospital services.

Against this background, Section 2 of this paper describes the rationale for the attempt to stimulate competition between acute hospitals that was a major constituent of the NHS reforms implemented in 1991. Section 3 discusses how far the intended objectives of competition have proved achievable. Sections 4 to 7 assess the numerous barriers to inter-hospital competition which exist in the NHS. The concept of contestability, as distinct from competition, is defined in Section 8 and its ability to overcome the problems faced by competition to achieve efficiency and quality benefits in the NHS acute hospital sector is assessed in Section 9. The final section summarises the conclusions that may be drawn from this, suggests some policy implications that result and compares these with the proposals in the December 1997 White Papers.

<sup>6</sup> The proposals in the Scottish White Paper imply more of a return to the pre-1991 position. For although Trusts will remain separate from Health Boards, acute Trusts are to be merged so that there exists only one per Board area, with the exception of Glasgow and Lothian (Edinburgh and surrounding area), and the Trusts and Boards are to draw up a service strategy together. All GPs in Scotland are to be formed into geographically-determined 'local health care cooperatives', which will in turn will join with the local community and mental health services to form 'Primary Care Trusts'. The Scottish White Paper makes no mention of service agreements (contracts) nor of Boards being able to switch to a different provider if the existing Trust proves unsatisfactory.

# Why competition?

The promotion of competition in all areas of economic activity was an overriding economic policy aim of successive Conservative governments between 1979 and 1997. In the public sector this policy was furthered by the privatisation of the nationalised industries and concurrent restructuring of the markets in which they operated; and by requirements in the remaining areas of public sector activity for provision of many services as possible to be market tested. Compulsory competitive tendering of catering, cleaning, laundry and other non-clinical support services had already been introduced into the NHS some years before the 1991 creation of the internal market. The aim of creating an NHS internal market could therefore be seen as trying to bring to clinical services an element of the competition that already existed in the provision of non-clinical support services within hospitals.

In health care, as in other sectors, competition was expected to stimulate:

- *allocative efficiency.* That is, the mix of health care services actually provided would come to reflect the mix demanded by purchasers on behalf of the populations they served, taking into account the costs of providing those services. The previously non-competitive hospital sector was characterised as producing a balance of services determined principally by what it was convenient for providers to supply (which might in turn be characterised as the same as they had supplied last year, plus or minus some minor variations stimulated by the wishes of their doctors, or other senior health care staff). This rebalancing of services would be expected to happen as a result of contract revenues being specifically linked by purchasers to individual services and individual modes of service delivery (e.g. day case as opposed to inpatient treatment) so that the needs of financial viability would dictate abandonment of unwanted services/modes of delivery and investment in those that were wanted but as yet were under-supplied. A purchaser faced by a monopolist hospital provider would be faced with an 'all or nothing' type of decision and so might

have to accept within the overall service package bought, some services it would rather not purchase if it had the choice. The existing of a competing hospital or hospitals would provide that element of choice;

- *technical efficiency.* The avoidance of unnecessary expenditure, i.e. of wasteful inefficiency, would be made even more imperative than hitherto. It has become a truism when comparing the NHS with other health services in the developed world, to observe how relatively inexpensive it is. In the UK, universal coverage of the population is achieved at the cost of just 6.9 per cent of GDP being spent on health care (of which the large majority is on the NHS) compared with 9.6 per cent in Germany and 14.5 per cent in the USA, for example (1995 data, taken from OHE, 1997, Table 2.3). Cost control has been a major preoccupation in the NHS for most of its existence. The traditional ways of achieving this have been fixed budgets for hospital and community health services, and 'gatekeepers' (GPs and the medical and nursing staffs of hospital A&E departments) who strictly control patient access to specialist (i.e. expensive) services. The NHS is undoubtedly low cost relative to other countries' health care systems, but not necessarily more efficient. The introduction of competition, if effective, provides the incentive to be more efficient than your rival. For in that way a hospital might not only defend its current activities but also expand them at the expense of a less efficient provider, if it could demonstrate this value for money advantage to the purchasers of health care;

- *higher quality.* As described above, competition between health care providers in the USA has manifested itself as much through attempts to demonstrate higher quality services as through lower prices for a given service package. A similar emphasis on service quality improvement was expected to be encouraged by inter-provider competition in the new NHS internal market created by the 1989-1991 reforms. Better quality providers would win increased revenues at the expense of their lower quality competitors. The



relative stress to be placed on higher quality services or lower priced services would be a matter for the purchasers to decide and to communicate through their contract negotiations with providers.

The desirability of applying competitive forces to the provision of health care as one means of pursuing these objectives attracted wide, if qualified, support among economists. For example, Culyer and Posnett (1990) stated 'our general conclusion is that it is indeed possible to harness competitive forces in order to promote greater hospital efficiency' but made this 'conditional on the government creating an environment that exploits competitive processes wherever possible' (page 12). Whether such an environment has been, or could be, created in the UK is discussed in detail in Sections 3 to 7 below.

Much of the empirical literature assessing the nature, extent and impact of competition comes from the USA. It has produced a wide range of findings. The existence of competition between providers in the USA has not, in general, been questioned. Instead the debate there has been about the principal ways in which competition has manifested itself. Where hospitals are reimbursed on a fee-for-service basis, inter-hospital competition has frequently been characterised by escalating attempts to demonstrate higher quality of patient care rather than lower prices. Some authors have gone so far as to suggest that in the US hospital market product quality had replaced price as the main area of competition (see for example Brooks et al, 1997). This picture has changed somewhat in recent years however, as managed care organisations purchasing health care have exerted pressure for price reductions as well as for quality improvements. In the USA in the 1980s and 1990s, increasing evidence has been found of active, price-based competition (e.g. Brooks et al, 1997).

The UK NHS has always had, and continued to have even after the 1989-1991 reforms, a very different structure from health care in the USA. Hospital services in the UK have long been planned and organised on a geographical basis with the explicit government intention of avoiding multiple providers of any given service in any one area. Multiple provision has been equated with wasteful duplication and unnecessary cost, rather than with engendering an environment of competition between providers of health care. The 'district general hospital' as the single provider of acute hospital services in any area was the

concept pursued by successive British governments since the 1962 Ministry of Health 'Hospital Plan for England and Wales', (Harrison and Prentice, 1996, provide a clear description of this policy and its history). In 1996, the Audit Commission identified 226 hospitals in England and Wales as having a full, 24-hour Accident and Emergency (A&E) service, the core of a district general hospital. The populations covered by these hospitals range widely in size from around 100,000 to 500,000 with an average of 230,000.

On the purchasing side of the health care system, the NHS may be likened to a single large managed care organisation, albeit one with local geographical offices (the HAs and, until April 1999, the GP fundholders) who are able to exert some autonomy at the margins in determining the services they buy on behalf of the populations they serve. Providers are remunerated predominantly on what is effectively a capitation basis (so called 'block contracts') rather than fee-for-service (or 'cost per case'). Other than for emergencies, access to hospital and other specialist services can only be via a gatekeeping GP. Cost control is pervasive. Quality control has been less well developed but is evolving, including the use of audit mechanisms and by publication of performance measures. However, unlike the US health care market, there is no significant competition for 'customers' between purchasers in the NHS. HAs and commissioning groups of GPs are and will be explicitly limited to serving only the population within their defined geographical boundaries. Although since 1991 people have in principle had the opportunity to 'shop-around' between GP fundholders if they wished, there is no evidence of this happening to any significant degree. Thomas et al (1995) found from a survey of five HA areas that the average proportion of patients who changed their GP in a year for reasons other than having moved house was just 1.6 per cent. The existing workloads of most GPs, with list sizes close to or above the average UK figure of 1,900 patients, may also deter many GPs from actively seeking additional 'customers'. The average time per patient/GP consultation is already down to around eight minutes (Wilson 1991) in the face of the current demand pressures.

The relevance of US data on the scope for, and impact of, competition between health care providers to an assessment of the UK situation is therefore limited. Nevertheless, the 1991 NHS reforms were intended by the government to produce the favourable outcomes of competition

discernible in parts of the US health care system (increased responsiveness to consumers' wishes and improved quality of service) while, it was hoped, avoiding the unfavourable (much higher costs). As the government put it in 'Working for Patients' (1989), hospitals were to be incentivised 'to attract patients, so they will make sure that the service they offer is what their patients want. And in turn they will stimulate other NHS hospitals to respond to what people want locally.'

Both price and quality competition were expected by the then UK government to result from the creation of separate providers and purchasers of health care in April 1991. Indeed, one of the explicit objectives of creating that purchaser/provider split was to stimulate competition between health care providers. Competition was not the only objective, however. The overall aim of the split was to ensure the accountability of providers to purchasers and to clarify the assessment of the population's health care needs by trying to divorce it from the narrower interests of hospitals' clinicians and managers. This reassertion of the interests of the population in general, and of patients in particular, over those of health care providers was to be given force by the hoped-for competition between providers seeking to respond ever more effectively and efficiently to the demands made by purchasers, in the manner of competitive providers of any other good or service.

Evaluation of the existence and impact, beneficial or otherwise, of competition in acute hospital service provision is still going on. However, by late 1994, the NHS Executive was already asserting, in 'The Operation of the NHS Internal Market: Local Freedoms, National Responsibilities', that:

'By encouraging efficiency and by giving health professionals and managers the freedom and the incentives to respond to patients' needs, the NHS internal market has shown itself to be a powerful tool for putting patients' interests first.'

Notwithstanding that assertion, evidence of whether the hoped-for competition and its associated expected benefits have been achieved remains scarce. Furthermore, there can be no presumption of success because there exist numerous constraints which severely limit the scope for competition between acute hospitals. These constraints derive both from the general nature of hospital services and from the specific structure and rules of operation of the NHS.

# Does competition between hospitals exist in the NHS?

Determining the existence and impact of inter-hospital competition is problematic. Some studies have estimated the potential for competition in acute hospital services to exist by measuring the concentration of hospitals in any area. Specifically they have used the Hirschman-Herfindahl Index of market concentration (HHI) in a district. The HHI measure reflects both the number of providers in a market area and the evenness of distribution of market share across those providers. The smaller the number of providers in the market and the more uneven their market share, the higher is the HHI value on a scale from zero to 10,000. The higher the HHI value the greater is the assumed degree of monopoly power vested in the providers. In a purely monopolistic market the HHI would have the value 10,000. As the market becomes less concentrated, i.e. the more providers there are and the more equal their market shares are, so the HHI approaches zero. US investigations of monopoly power commonly assume that an HHI value of 1,800 or more indicates some degree of monopoly power, although no single threshold value has been consistently applied in the case of hospitals. Appendix 1 sets out a full definition of the HHI.

The test for the existence of competition has traditionally been whether areas with greater numbers of nearby alternative providers (lower HHI values) have lower prices for health care. In the USA, Melnick et al (1992, quoted in Brooks et al, 1997) found that hospitals in less concentrated, more competitive, markets appeared to receive lower prices. Brooks et al (1997) reached a similar conclusion when analysing specifically the bargaining behaviour between hospitals and insurers over the prices to be paid for appendectomies: lower prices were paid in areas where the HHI was lower, i.e. where competition was assumed to be greater.

In the UK, Appleby et al (1994) and Propper (1996) have pursued different variants of this same line of enquiry. Appleby and colleagues investigated the scope for competition, via a combination of a

national survey of health authorities and more detailed analysis of the contracting behaviour of 22 health authorities and 33 major acute units in the West Midlands. They defined a 'market' as equating to the area covered by a single health authority. The size of population covered by a health authority varied but at that time (1991/92) averaged around a quarter of a million. Appleby et al then calculated HHIs for each of the 36 general surgery services (some units have more than one hospital site from which they provide general surgery) and concluded that two-thirds of them (treating 62 per cent of all NHS general surgery patients in the West Midlands) were in potentially competitive markets because their HHIs were below 1,800 – the threshold above which the US Department of Justice would take there to be a potential monopoly in existence. Appleby et al then qualify their conclusion, however, by suggesting that they might be overestimating the amount of potential competition because they have chosen an area as large as an HA district as their definition of a market and because they did not differentiate between different surgical treatments but just looked at all general surgery as a homogeneous service. If effective markets cover smaller geographical areas, because of patients' unwillingness to travel or residents' unwillingness on whatever grounds to see the closure of a local hospital service, then more hospitals would appear monopolistic. Also, for those 'general surgery' treatments which are not provided by all hospitals' general surgery services, the degree of potential competition will be reduced. Furthermore, the reason for studying general surgery services in particular was that the authors considered them more likely to be capable of competition than most other acute hospital patient services. Hence the potential for competition would probably be lower for most other hospital-based medical and surgical services.

Propper (1996) measured the degree of potential competition in general surgery, orthopaedics, ENT (ear, nose and throat surgery) and gynaecology, by the percentage each hospital service had of the total activity in that specialty for the population

lying within an estimated 30-minute travel time of the hospital. She found that in the eight (out of 14) English regions for which complete data were available, covering 118 non-specialist acute hospitals, relatively few providers did not have another provider located within 30 minutes' travel. Only 8.5 per cent of the hospitals were free of competition within 30 minutes' journey in all four surgical specialties. On average, a hospital would provide between 30 per cent (orthopaedics) and 37 per cent (gynaecology) of the total hospital activity in the specialty within its 30-minute travel zone; although the range was from a low of just three per cent, up to 100 per cent. Thus the potential for competition, at least for these surgical specialties, appears great. This is borne out by the Audit Commission (1996a) which found that 83 per cent of English and Welsh hospitals with full A&E departments (and hence also with a range of associated acute hospital services) were less than 15 miles from the next nearest A&E hospital.

The demonstration of the potential for competition naturally leads to the question whether that potential has been realised? Appleby et al (1994) looked at whether contracts had changed at all since the introduction of the NHS internal market in April 1991 and found that: 'evidence from our case study districts revealed that by and large contractual changes in the first two years of the reforms were small even though a high proportion of contracts were affected.' They did not report whether such changes as there had been had involved any providers in losing contracts but the implication is that this was a rare event if it happened at all. The absence of contract shifting could, of course, indicate a high level of purchaser satisfaction rather than a lack of effective competition.

More recent analyses of purchasing have found that acute hospitals typically receive the large majority of their income from a single source: their local HA. Paton et al (1997) found from a survey of English and Welsh Trusts (non-acute as well as acute) and HAs conducted in 1995/96 that most Trusts have at least 70-80 per cent of their business funded locally, i.e. by the local HA and, to a much smaller extent, by local fundholding GPs. They concluded that: 'competitive markets involving a range of providers, both local and non-local, are largely absent'. A very similar picture is evident from the work of the Accounts Commission for Scotland (1997), which reviewed contracting in each year from 1994/95 to 1996/97 in the 15 Scottish health boards. The Accounts Commission concluded that 'planning and

delivering health services within the internal market remains predominantly a local affair'. Ninety-two per cent of expenditure on secondary care contracts (which includes acute hospital services) in Scotland in 1996/97 was on NHS providers within the local health board area, a proportion hardly changed since 1994/95 when it was 93 per cent. Furthermore, contracts with non-NHS providers, who might have been expected to be competing keenly for NHS purchasers' money, remained on a very small scale at less than 1.5 per cent of total contract expenditure.

The studies referred to in the previous paragraph do not distinguish the behaviour of acute hospitals from that of other types of Trust. The NHS Executive's Third National Review of Contracting 1994/95, looked at acute hospital contracts separately and found that the 107 HAs in England in the autumn of 1994 held a total of 392 main contracts with 251 acute providers. Thus each acute provider had an average of only one and a half main contracts, which implies that 50 per cent or more of acute providers can only have had one such contract. ('Main' contracts were defined as those which, starting with the largest value contract, cumulatively accounted for 80 per cent of an HA's total spend on acute hospital services).

There appears to be a general lack of evidence about the extent of switching of contracts by purchasers as a 'market' response to dissatisfaction with the mix, qualities or prices of services obtained from existing providers. As will be discussed later, there are strong reasons why HAs find it difficult to switch any but the smallest contracts between alternative providers without threatening the financial viability of all or a large part of the losing hospital's services. There is ad hoc and anecdotal evidence about fundholding GPs' readiness to switch contracts between providers (e.g. Glennerster et al 1994 and Propper 1996) but no authoritative data appear to have been collected on the extent of this with clinical services, as opposed to diagnostic pathology tests or imaging. Indeed, the Audit Commission (1996) in its, admittedly possibly unrepresentative, sample of fundholding practices found that almost no practices in the part of the country it studied had moved their elective surgical contracts from one hospital to another, despite the presence of several alternative hospitals nearby and despite treatment modes and prices differing widely between those hospitals.

However, even if there were more information about the frequency with which contracts have

been moved, this would not necessarily be conclusive. Lack of contract switching could indicate either that purchasers were failing to take advantage of the potential for competition between their providers, or that, on the contrary, the threat of losing contracts was being effective in stimulating incumbent providers to improve their performance so as not to lose their existing customers. In the latter case, competition would be not only active but also effective. A number of studies of fundholding GPs have indeed found that they have been able to achieve improvements in service quality (such as more rapid return of diagnostic test results, reduced waiting times, provision of new services such as physiotherapy and counselling) from their incumbent health care providers and hence without actually having to switch to new suppliers (see for example the chapter on GP fundholding in West (1997) for a good summary of this).

Having established the potential for competition for some surgical specialties, Propper (1996) tested for its existence and possible impact on hospital efficiency by examining the prices posted by NHS providers for extra-contractual referrals (ECRs) in those specialties – i.e. the prices charged for unplanned referrals made by fundholding GPs and (non-local) HAs who did not have prior agreed contracts with the provider for the service concerned. ECRs generally account for a very small percentage of a Trust's total income, around one or two per cent, with the vast majority of revenue being earned under the terms of contracts. However, the prices paid under contracts are not generally publicly available. Propper found that ECR prices posted for ENT and gynaecology were significantly lower, the greater was the number of competitors within 30 minutes travel time of a hospital. However this relationship was not found for general surgery or orthopaedics. Furthermore, no conclusions can be drawn from a study of ECR prices about the relationship, if any, between contract prices and the number of competitors, as there is no necessary relationship between contract prices and ECR prices. (See the discussion below of Department of Health regulation of hospitals' output prices).

A similar analysis by Propper, Wilson and Söderlund (1997) of prices charged to GP fundholders for eight elective surgical treatments found that although the Department of Health's regulatory rules appeared not to be observed, prices did not appear to be reflecting the extent of competition locally. Prices were not clearly lower

where the number of local competitors for a hospital service was higher.

Other writers have expressed doubt as to whether competition in the provision of acute hospital services is effective even if it is possible (see for example: Ferguson and Palmer, 1994, Maynard, 1994, Spurgeon et al, 1997). The reasons for this scepticism vary but are drawn from among the numerous and powerful factors which appear to place strict, and in many cases probably suffocating, limits on effective competition. These obstacles to competition providing effective stimuli for efficient and high quality hospital performance responsive to the requirements of purchasers, are listed in Box 1. They are described more fully in the following pages.

#### **BOX 1 Barriers to competition between NHS Acute Hospitals**

- Economies of scale
- Economies of scope
- Medical training requirements: Calman, Royal Colleges' accreditation requirements
- Medical practice requirements: the New Deal for junior doctors, Royal College guidelines
- Patients' willingness to travel
- Department of Health regulation of hospitals' output prices and return on capital
- Department of Health control of hospital inputs' – labour and capital – supply and prices
- Primitiveness of contracting and of service quality and price information
- Satisficing by providers and/or purchasers
- Public and political opposition to hospital service relocation and hospital closure

# Economies of scale

The policy conclusion that competition is desirable because it will lead to the greatest allocative and technical efficiency and to a level of service quality which matches the wishes of the service users, depends on numerous restrictive assumptions. Among these is the assumption that the scale of activity at which the average cost per unit of (service) output is minimised will be small relative to the total volume of demand for the service in the relevant market if the service were to be sold at a price equal to that minimum average cost. If it is met, this condition means that the total demand for the service may be satisfied by a number of competing providers, each supplying only a part of the total market, without this resulting in higher costs than if a single provider supplied the whole market.

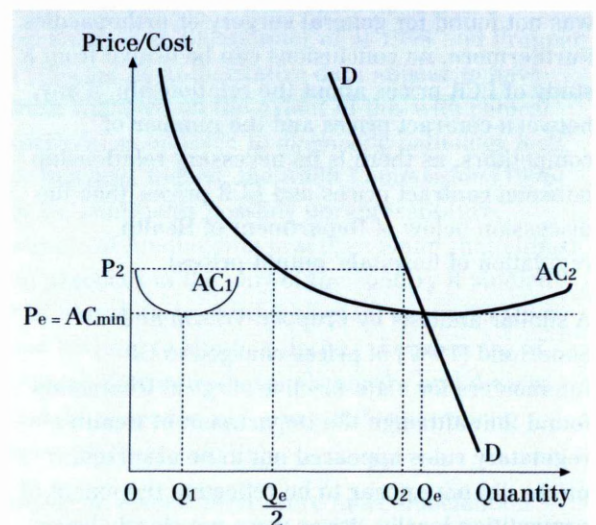
The presence of economies of scale to the extent that the total demand for the service could be met at lowest cost by a single provider (pricing at average cost) would however imply that competition would be wasteful. Figure 1 illustrates this point. The market demand for the service in question is represented by the curve DD. The minimum average cost of producing the service is  $AC_{min}$ . If the service were to be sold at market price  $P_e$ , then a total volume of service equal to  $Q_e$  would be demanded. The average cost curve of a single producer shown as  $AC_1$  illustrates the position where economies of scale are not significant because minimum average cost is achieved at a scale of output,  $Q_1$ , which is small relative to total market demand  $Q_e$ . In this case several producers each supplying  $Q_1$  could meet total demand while minimising the total cost of doing so. However, were the technology of producing the service such that the average cost curve of a single producer is  $AC_2$ , then costs could only be minimised by there being just the one provider supplying the entire market. For example: if two providers were each to try and supply  $Q_e/2$  of the service to the market, then this would be at a higher average cost, higher by the amount  $P_2 - P_e$ , than a single supplier could achieve. Furthermore, because demand for the service is assumed to decline as its price increases,

the higher price of  $P_2$  which would have to be charged to cover average costs would result in less of the service being consumed: namely  $Q_2$  rather than  $Q_e$ . Thus, where there are economies of scale which are large relative to the size of the market, competition is inefficient, both technically (production costs are no longer at the lowest possible level) and allocatively (less of the service is consumed than if it had been available at a price equalling the lowest possible average cost).

The question is, therefore, whether acute hospital services are subject to significant economies of scale at levels of activity corresponding to the size of market typically faced by UK acute hospitals. The evidence available is, unfortunately, contradictory.

The size of market which an acute hospital can serve is limited by the willingness and ability of patients to travel to it. Both willingness to travel, for non-emergency ('elective') care, and ability to travel without detriment to health to receive emergency care, decline as distance from patient to hospital increases. The rapidity and impact of this decline are discussed later on in this paper. It

Figure 1 Economies of scale



is, however, the case that in some parts of the country there exist single acute hospitals effectively serving a particular population as a local monopoly with no effective competitor for that local market. Indeed, in the 1960s and 1970s this was an explicit goal of NHS planning. That this trend has continued over the last 18 years, despite the declining role of central 'planning' of provision in the NHS, suggests that there is a minimum efficient scale for an acute hospital. In the extreme, it would clearly be unacceptably costly to have a hospital on every corner or in every neighbourhood. This is clearly the case in sparsely-populated, predominantly-rural areas. However, in large, densely-populated conurbations, competing acute hospitals may be observed to exist only a couple of miles apart. The unanswered question is where the cross-over takes place; that is, what proportion of the UK's population live in areas where competition between hospitals would not obviously be inefficient.

It may be observed that for many years the trend in the NHS, even in urban and suburban areas, has been for acute hospital services to be concentrated onto ever fewer sites. This has been ostensibly in the pursuit of both higher quality care and more care delivered per taxpayer's pound spent (i.e. lower average costs). Furthermore, managers responsible for NHS hospitals which operate across two or more sites, may commonly be observed to have organised services across those sites in such a way as to remove duplication and minimise costs spent on staff, materials and even patients being transferred between sites; i.e. any one service will normally now be provided on only one of the two sites even if in the past it had been available at both. This behaviour is consistent with the view that concentration of acute hospital services will reduce costs for a given volume of output, i.e. economies of scale will be achieved. The emphasis of the government on merging NHS Trusts to realise savings in 'bureaucracy' implies a similar faith in the existence of, managerial at least, economies of scale in acute hospitals.

Seemingly in contradiction of this 'received wisdom' of the benefits of concentration, however, the authoritative literature review undertaken for the NHS Executive by the NHS Centre for Reviews and Dissemination (Ferguson et al, 1997) found that: 'The majority of studies report constant or decreasing returns to scale for acute hospital services... If economies of scale are evident these

economies appear to be fully exploited at a relatively low level (in the range 100-200 beds).' However, the natures of the studies found are such that Ferguson et al were forced to conclude that: 'The extent or size of any economies or diseconomies of scale cannot be reliably estimated from the literature.' The same literature review found that while increased scale sometimes appeared to be associated with better patient outcomes, it was unclear whether this was because large hospitals had different case mixes from small ones rather than being inherently capable of better quality care. In addition to this failing, the hospital economies of scale literature suffers from numerous other weaknesses when it is attempted to apply it to the current position in the UK, including that:

- it is largely based on US studies, where the structure of hospital services is quite different from that in the UK and so too is the mix of inputs used;
- the indicator of scale almost always used has been the number of beds. Thus 'scale' has been equated with the quantity of one particular input, and it is perhaps not surprising to find that costs tend to increase roughly in proportion with the number of staffed beds made available. Part of the efficiency of a larger hospital is its ability to make better use of its bed stock in the face of a variable flow of demand over time. Measuring scale by the number of patients treated, or treatments delivered, would be a more appropriate approach;
- furthermore, beds have been treated as homogeneous, thereby abstracting from the fact that the type of patient being treated and the treatment being offered differ between beds and are major determinants of cost. If small hospitals predominantly provide basic treatments to uncomplicated cases, while the more difficult cases and more complex treatments are only provided in the larger hospitals, this would mask any underlying economies of scale that might otherwise be found to exist;
- the technology for providing acute hospital treatment is changing rapidly at present in the face of new requirements and guidance from the government and the medical professional bodies. The specific consequences of these are discussed separately below, but in essence they imply that the minimum efficient scale of an acute hospital is now somewhat above the scale of some existing NHS hospitals, i.e. significant economies of scale now appear to exist in their markets. However, the

point to be made here is that the studies reviewed by Ferguson et al (1997) all pre-date these changes and so their findings may not hold for the future, even if they are a correct representation of the past.

A further complication is that while the existence of substantial economies of scale in the long-term may be doubtful – and this was the issue reviewed by Ferguson et al (1997) – they may appear to exist in the short-term as a result of inflexibilities in employing or releasing resource inputs. Within a time horizon of a year or so, it may be rather difficult for a hospital to substantially alter either up or down its capacity. This short run inflexibility applies both to physical capacity (buildings and equipment) and to the numbers of key professional staff employed, particularly given the inflexibility of the contracts under which consultants (fully trained medical specialists) are employed. Hence, hospital managers may act as if there are economies of scale because of these inflexibilities. Spare existing capacity can be put to use at little marginal cost and, conversely, loss of marginal activity permits few costs to be avoided. Over the longer term, however, when capacity can be permanently adjusted to the level of demand, substantial economies of scale may no longer be available.

Acute hospitals' cost functions are labour intensive. A typical cost breakdown for a major acute hospital is shown in Table 1. The numbers in the first column of this table are rounded and based on the 1994/95 costs of four different acute Trusts (two large teaching Trusts, a medium-sized general acute hospital and a small acute hospital, all three including 24-hour A&E departments). The final column provides a verification for these acute

hospital numbers by showing the aggregate cost breakdown for all (acute and non-acute) English NHS Trusts in 1995/96. These combined acute plus non-acute figures reflect the fact that the costs of community and mental health services are even more dominated by staff costs and less dependent on capital than are acute hospitals'.

Thus the costs of capital (taking a six per cent rate of return on current cost assets to represent normal profits in this sector) are only around one tenth of an acute hospital's total annualised costs. It is these costs of capital which are most likely to be fixed and/or to display significant indivisibilities and so lead to economies of scale. However, even the amount of capital required will vary, in the long run, positively with the amount of activity which a hospital undertakes. More patients means not only more care staff and more medicines but also more wards, operating theatres, outpatient treatment/examination rooms and support areas, with all of their equipment. Thus by no means all of the costs of capital are fixed.

Other non-pay costs include medical supplies and other materials and bought-in services, and are essentially variable. So too are most pay costs. The only indivisible elements of staff costs, i.e. where economies of scale are to be expected a priori, are probably the costs of the most senior managers. Roughly the same amount of top management will be required whether a Trust has an annual income of £50 million or £100 million. In an average-sized acute Trust these senior management pay costs would form less than one per cent of total Trust costs. The employment costs associated with senior medical staff (i.e. consultants) are not fixed – doubling the number of patients treated will broadly require doubling the number of consultants – but they can be difficult to reduce, even over a period of several years, owing to the permanent contracts that consultants hold and the expense of buying these out. Thus it may, for this institutional reason, be difficult to reduce senior medical staff costs at an existing hospital other than by natural staff wastage or the transfer of the consultants to another hospital which is expanding its workload or has experienced unwanted loss of consultant staff. The existence of this cost inflexibility does not impart economies of scale as such, but does make it difficult to adjust scale to changed patterns of demand.

Overall, fixed costs beyond a one-year horizon are likely to be limited to some elements of capital plus the salary costs of the most senior management layer. Fixed costs should therefore be expected to

**Table 1 Typical acute hospital cost structure: percentages of total costs**

<i>Cost category</i>	<i>Acute hospitals<sup>1</sup></i>	<i>All trusts<sup>2</sup></i>
Pay	55-60%	64%
<i>of which: consultants</i>	8%	
<i>junior doctors</i>	6%	
Non-pay	30-35%	29%
Depreciation and 6% return on capital	10%	7%
<b>TOTAL</b>	<b>100%</b>	<b>100%</b>

*Sources:*

1. Calculated from the accounts of four acute units and rounded.
2. HC 127. NHS (England) summarised accounts 1995/96. The Stationery Office.



be less than 10 per cent of total acute hospital costs, and certainly no more than 15 per cent even if all consultant costs are considered fixed. Thus, economies of scale, other than in the short-run before (non-consultant) staff numbers can be adjusted, should be expected to exist but to be relatively small percentages of total costs. For example, a 10 per cent increase in a hospital's activity might be expected (leaving aside any technological change) to provoke a nine per cent or greater increase in costs.

To summarise the economies of scale issue: cost economies of scale may well exist – although the evidence is not conclusive either way – but will probably be small owing to the labour-intensive nature of health care services. If so, to have competing acute hospitals in many parts of the country could imply some waste (unnecessary costs). Although the scale of that waste would be only a small fraction of total hospital costs (of the order of a few percentage points) it could amount to a large absolute sum when aggregated across the UK.

The question of whether larger hospitals may be able to deliver higher quality care than smaller hospitals is addressed below in the discussion of medical practice. Although the available evidence on such benefits from scale is equivocal, the medical professional bodies that issue guidance on appropriate modes of delivering acute hospital care and which accredit hospitals to undertake the training of junior doctors, evidently believe that there are such quality economies of scale.

# Economies of scope

**E**conomies of scope exist where an organisation which supplies two services simultaneously is able to do so at a cost less than the combined expenditure of two separate organisations each producing just one of the services. Where such economies are significant, competition will be limited to those organisations capable of providing the whole range of jointly produced services. For acute hospital services there are many examples of where such economies of scope should be expected to exist, for instance:

- a hospital which already provides general surgery would probably be able additionally to provide a specified volume and quality of ENT surgery at lower cost than a hospital which only provided ENT surgery. This is because the same operating theatres, diagnostic equipment, anaesthetists, theatre staff (apart from the surgeons themselves), diagnostic staff (pathology, radiology), beds, nurses, support staff and management who are required if general surgery is provided would be equally capable of providing ENT surgery. Unless all of these facilities and staff were being used at 100 per cent of their capacity at all times, some of them could support ENT too. A hospital supplying ENT alone would have to fund these fully, rather than sharing the costs with general surgery;
- a hospital providing emergency surgical services to meet the variable and unpredictable needs of patients, has the capacity in periods where emergency demand is not at its peak level also to provide elective surgery. A consultant surgeon seeing planned outpatient attendees can be called away in an emergency to treat, or advise on the treatment of, a newly arrived surgical emergency patient.

The *a priori* likelihood of economies of scope existing in the provision of acute hospital services stems in large part from the fact that many of the resources provided for acute care are, in effect, on permanent stand-by. They are there to be called upon to diagnose and/or treat individual patients whose precise requirements and the timing of those requirements cannot be predicted exactly. A

consultant surgeon conducting an outpatient clinic can be called to theatre if required in an emergency. A general physician looking after existing inpatients in a hospital can be called upon to see particular emergencies as they arrive in the A&E department. A pathologist conducting routine testing for much of the day can divert at short notice to provide an urgent diagnostic response if a particular emergency or unusual case arrives; and so on. An acute hospital represents a collection of an extremely large variety of specialised resources, each of which is capable of being used in combination with other inputs to help any from a wide range of patients and illnesses. The precise use to which the resources are put at any moment depends on which patients with which illnesses have arrived at the hospital.

Economies of scope may be manifested not only by lower costs but also by higher quality in a hospital providing multiple services than in separate hospitals only providing individual services. Medical professional guidance makes frequent reference to the need for, or desirability of, close links between various hospital services, i.e. co-location of them on the same hospital site and this is discussed below in the section on medical practice and training.

However, similarly to the question of economies of scale, the empirical evidence on the existence of economies of scope, whether cost or quality economies, is weak. The same NHS Centre for Reviews and Dissemination study which investigated economies of scale also reviewed the literature for empirical evidence of economies of scope and a similar lack of clear evidence was found (Ferguson et al, 1997). This lack of evidence is perhaps unsurprising for two reasons. First, very few single specialty hospitals exist for comparison with general hospitals, and those that do focus on highly specialised rather than more common services (e.g. providing only cardiac care, or neurology and neurosurgery, or specialist orthopaedic surgery, or ophthalmology). Nearly all acute hospitals deliver a range of services based around emergency general medicine and surgery,

which in itself is strong circumstantial evidence for the existence of significant economies of scope, with respect either to service quality or cost or both.

The second possible explanation for the lack of empirical evidence of economies of scope is that, although real, their magnitudes may not be large. For example, an acute hospital which only provided non-emergency, planned, care might, if the demand were there, achieve 90-100 per cent bed occupancy through the year. (Private hospitals in the UK, although providing only elective care, have typically achieved much lower occupancies than this, due to insufficient demand for them in the face of competition from the NHS). In contrast, even an efficiently run emergency facility might achieve only 70 per cent occupancy on average, owing to fluctuations in demand, day by day. If no extra capacity purely for elective admissions were to be provided, then this would imply that the amount of low cost elective work the emergency hospital could carry out would be at best around 25 per cent of its total bed days.

Overall, therefore, acute hospital services appear to offer wide-ranging economies of scope, although the magnitude of those economies is once again uncertain and may not be large once a core basket of services has been established.

# Medical training and practice: requirements and guidelines

Whether or not economies of scale existed in the past in the provision of acute hospital care, recent changes to the requirements for training hospital doctors in the UK have created some. These changes have been prompted by European Union laws requiring mutual recognition of different member states' specialist medical training, and are set out in the April 1993 report of the Department of Health's Working Group on Specialist Medical Training (Calman et al, 1993). As a result of this report, not only was the length of postgraduate training required by hospital doctors to achieve the status of certified specialist (a prerequisite for becoming a hospital consultant) reduced to around seven or eight years (from what was often ten or more), but also the proportion of time spent in unsupervised practice by doctors in training was to be severely curtailed. Both of these changes imply that a greater proportion of the medical input in hospital care is now having to be provided by consultants and less by junior and middle grade doctors. This has increased the cost of care in smaller hospitals and has increased the minimum size of the consultant team in any specialty which is sustainable (i.e. to which it will be possible to recruit consultants) by requiring a greater contribution by consultants than previously to providing night time and weekend cover for emergencies.

The medical Royal Colleges are responsible for accrediting training posts for junior doctors. If a post is not accredited for training, it will be practically impossible for a hospital to find a competent junior doctor willing to fill it. If there are no junior doctors in a specialty, consultants will not be willing to work in it. Consequently, removal of training accreditation can make it impossible for a hospital to continue to provide the affected specialty. Important among the Royal Colleges' accreditation criteria are that doctors in training should have exposure both to a large volume and a wide variety of cases, and also that they should train at hospitals which provide a wide range of linked specialty and clinical support services. These requirements imply that small hospitals seeing relatively few patients or with a

limited range of specialties and support services will find it difficult to survive and will have to form partnerships with other hospitals and share key clinical staff with them. Thus, economies of scale and scope have been created/increased, so that competition is limited to markets which can potentially support larger hospitals.

In response to growing criticisms of the excessive hours worked by some junior hospital medical staff, the government agreed with the Joint Negotiating Committee for Hospital Medical and Dental Staff in 1991 a so-called 'New Deal' on junior doctors' hours and conditions (NHS Management Executive, 1991). The New Deal should have been fully implemented by the end of 1995. It has imposed maxima on the weekly hours that junior doctors may be required to work or be on-call for, both in total and in any single stretch. This has put an upward cost pressure on acute hospitals and reinforced the necessity to avoid small units where it would be impossible to rota staff so as to meet the New Deal requirements.

The medical Royal Colleges and other medical professional associations have produced a range of guidance on what they deem to be the appropriate minimum staffing level, caseload, casemix and range of on-site linked services required of an acute hospital which wishes to provide any particular specialty. Box 2 shows, for example, that professional guidance indicates a requirement for a wide range of services to be present on-site, if a hospital is to run an A&E department. The ostensible grounds for recommendations such as these are generally ones of patient safety.

The NHS Centre for Reviews and Dissemination has undertaken a major systematic survey of the recent literature on the relationship between the volume of activity of a hospital and patient outcomes (CRD Report 2, 1995). They found that many studies report a correlation between increased hospital volume of activity and better patient outcomes (lower mortality, principally) for a range of, mainly surgical, procedures. However, the CRD also found that these studies generally had not adjusted adequately for differences in case

**BOX 2 Recommended co-location of specialties and supporting services with A&E departments**

The Audit Commission in its 1996 report on A&E services in England and Wales collected the following examples of guidance on essential and desirable services to have on-site at a district general hospital with an A&E department – E = essential, D = desirable:

	<i>RCS</i> <sup>1</sup>	<i>BOA</i> <sup>2</sup>	<i>CSAG</i> <sup>3</sup>
Anaesthetics	E	E	E
General medicine	E	E	E
General surgery	E	E	E
Orthopaedic surgery	E	E	E
Cardiology	E	–	D
Geriatric medicine	–	–	E
Gynaecology	–	–	E
Psychiatry	–	–	E
Intensive therapy unit	D	E	E
Paediatric medicine	D	E	E
Maxillo-facial surgery	D	D	–
Nephrology	D	–	–
Ophthalmology	D	–	–
ENT	D	–	–
Paediatric surgery	D	–	–
Plastic surgery	D	–	–
Urology	D	–	D
24-hour staffed operating theatre	D	E	E
24-hour diagnostic imaging	E	E	E
24-hour diagnostic pathology	E	–	E

*Sources:*

1. Royal College of Surgeons of England (November 1988). Report of the Working Party on the Management of Patients with Major Injuries.
2. British Orthopaedic Association (November 1992). The management of skeletal trauma in the UK.
3. Clinical Standards Advisory Group (January 1995). Urgent and emergency admissions to hospital.

by volume as a key argument for centralisation of services’.

Despite this inconclusive result, the trend of professional pressure for fewer, larger hospitals in the UK is continuing. The Senate of Surgery and the Royal College of Surgeons have each produced reports in July 1997 arguing for acute hospitals’ surgical services to be based on resident populations of 450,000-500,000 (about twice the current average size of acute hospitals in the NHS) on grounds of providing emergency services of adequate quality, increasing specialisation, and meeting the new working and training arrangements described above (Senate of Surgery of Great Britain and Ireland, 1997 and Royal College of Surgeons of England, 1997).

Although this guidance cannot strictly be enforced, the professional pressure is great enough to make it happen. If so, it means that economies of scope and scale have been increased (or created if they did not exist before) in the last few years.

mix seen at different hospitals (high volume hospitals may be treating a higher proportion of simpler cases). Because they also found ‘a paucity of information as to why high volume might be associated with better outcomes’, the CRD concluded that ‘the main recommendation is that policy-makers should be cautious when invoking the assumed improvements in outcome achieved

# Other barriers to competition

## 7.1 Patients' willingness to travel

The size of market available to an acute hospital is obviously affected by the size of catchment area from which it is able to draw in patients. If patients are unwilling to travel larger than strictly necessary distances, that is they express strong preferences for being treated at the nearest hospital where the service they require is provided, then competition between hospitals becomes less effective. Patients are unlikely to be willing to travel for cost-saving reasons as they have no incentive to seek out a lower cost but more distant provider of services because they do not have to pay any additional costs of the more convenient hospital. Against that, attending a more distant hospital is likely to involve higher travel costs and greater travel times both for the patient themselves and for their family and friends who accompany or visit them. This effect is so strong that even where the choice is between treatment at a particular hospital or no treatment, the difficulty of accessing a service will affect the rate at which a population makes use of it. The greater the cost or inconvenience involved in getting to where a service is provided, the greater must the expected benefit of that service be before a patient will seek referral to it. Thus willingness to travel is related to the magnitude of the expected health benefit from the intended treatment. Consequently, willingness to travel may be expected not to be great for non-urgent treatments of non-life-threatening conditions, i.e. for most elective surgery.

HAs are constrained in their ability to enforce a lower cost choice of provider as they cannot prevent GPs referring patients to the hospital of their choice. Fundholding GPs are also constrained, by their patients' wishes, but as it is they who make the referral to hospital they have the opportunity of persuading the patient face to face to accept (on whatever grounds) referral to the lower cost but more distant provider. Surveys of GPs' referral criteria imply however that quality of care and waiting times are the dominant factors and that price barely figures (see the discussion in

Section 7.5 below). Although it is to be expected that GPs would say that so as not to appear to be allowing financial considerations to enter their relationship with their patients, there is no evidence of significant shopping around between hospitals by fundholders, let alone non-fundholders, on cost grounds.

Where competition might be less constrained by patients being unwilling to travel, and hence where competition may be expected to be more prevalent, is over quality of service rather than price. Patients, and hence their GPs, face a strong incentive, which is directly correlated with the expected benefit of the treatment required, to seek referral to the highest quality provider of care. Shorter waiting times for first outpatient appointment and subsequent admission or day case treatment may also encourage patients to travel further than the nearest provider. Furthermore, HAs, however financially constrained, are unlikely to try and force patients to go to a lower cost provider if to do so would be to accept lower standards of care or longer waits than were on offer at another feasible but more costly provider.

The evidence on patients' willingness to travel is, perhaps inevitably, mixed. The Centre for Reviews and Dissemination (Ferguson et al, 1997) found that there was empirical evidence of what they term a distance-decay relationship. In both urban and rural populations, rates of usage of hospital services are lower the greater is the distance from hospital, although this effect is less evident the more important a service is perceived to be. Thus, for treatment of a minor, non-life-threatening illness patients are less willing to travel than for treatment of major, life-threatening conditions.

An opinion survey by Stewart and Donaldson (1991, quoted in Harrison and Prentice, 1996) found that willingness to travel was clearly and positively correlated with the expected wait for treatment at the nearest potential provider. Around half of patients would rather travel than wait for two months, but nearly all (96 per cent) would travel if the alternative were a one-year

wait. Howell et al (1990, and also quoted in Harrison and Prentice, 1996) found that distance was not seen as a major obstacle by most patients facing routine elective surgery; 90 per cent were willing to travel up to 50 miles, and two-thirds of the patients interviewed were willing to be referred 300 miles away. However, Harrison and Prentice also note that other studies have apparently found a rather greater reluctance to travel. For example, Mahon et al (1994), based on two surveys of patients in north west England, reported that a substantial minority (35-40 per cent) would not be prepared to travel further to be seen more quickly and that fewer than a quarter would be willing to travel more than 30 miles.

Thus, while the magnitude of the relationship between distance and willingness to be referred remains unclear, there evidently is such a relationship and its existence in combination with the other obstacles discussed in this paper, implies a diminution of the effectiveness of inter-hospital competition. Some patients, even if only a minority of the total, can be expected to resist strongly any attempt to move the point of acute hospital service provision further away, especially if the justification for the move is to save the NHS (rather than the patient) costs. This issue is returned to below in the discussion of public and political opposition to hospital closure.

## 7.2 Department of Health regulation of hospitals' output prices and return on capital

Even where the incentive for hospitals to compete exists, some of the means of competition common in other markets may be hindered by government regulation of the NHS market for hospital services. Regulations govern both the basis upon which Trusts may set prices and on the rate of return that may be earned, not only in aggregate, but on each service individually. Thus flexible pricing, e.g. differential pricing for different purchasers, is in principle excluded. An existing acute hospital which decided to add, say, urology to the surgical services it already offered, would not be permitted to price those services at the level implied by the incremental costs of adding them to the existing infrastructure of the hospital and its support services. Instead, a proportion of those already-existing infrastructure costs would have also to be included in the urology prices.

According to Department of Health regulations, the price of any service offered by a Trust should

be set equal to its average cost including a six per cent rate of return on (the current cost value of) the assets employed. The only officially-sanctioned exception to this rule is where a Trust finds itself with short-run, unplanned spare capacity which it may offer at a, low, price equal to short-run marginal cost. In this way, additional amounts of activity negotiated within a financial year may be priced cheaply, but such amounts can, by definition, only be a small proportion of a hospital's total activity.

The restraining effect on competition of the government's price regulation may not be great in practice, however. Hitherto, prices for the same services have varied widely between providers, probably because of poor cost information and varying allocation rules as much as because of differing efficiency levels; and also because Department of Health regulation of prices has been loose and, other than at the whole hospital level, not intrusive. In a review of NHS pricing, Ellwood (1996) concluded that 'prices were not providing appropriate signals to facilitate the market's efficiency aims. Prices were neither reasonable measures of resource consumption nor capable of being compared meaningfully between alternative providers'. Propper and Bartlett (1997) have suggested that Trusts are taking advantage of this and evading the average cost pricing rule when it suits them. Thus, price regulation as practised hitherto has probably not done much to restrict competition, despite its implicit capacity to do so.

## 7.3 Department of Health control of hospital inputs' supply and prices

A possibly more restrictive type of market regulation is, however, applied by the government through the central control of hospital inputs. The government places limits on: the numbers of trained medical staff available, the pay levels of all NHS professionals and the quantity of capital investment that Trusts are permitted to undertake.

Despite minor variations due to the introduction of local pay bargaining in recent years (i.e. differences of the order of one or two percentage points) the pay rates of NHS doctors, nurses and other professions allied to medicine are still largely the same everywhere in the country (with the exception of London weighting). They follow closely the findings of national pay review bodies and national pay settlement norms. Furthermore,

the current government is abandoning the local pay bargaining policy. Thus hospital staff costs, if the same mix and volume of services are to be delivered, will only vary to the extent that different hospitals use different mixes of staff or have more or less efficient workforces.

Perhaps connected with this but also, in the case of doctors, because of restrictions on the numbers of medical school and training places, some key hospital staff are in nationally short supply. Yet other groups although not in short supply nationally are very difficult to recruit and retain in certain local areas, such as inner cities, which are relatively unattractive or expensive to live and work in. Where these shortages are binding constraints, effective competition in the services that depend on those key staff will be prevented.

Perhaps the greatest single restriction, however, is on the availability of capital for investment. Trusts require regional or central approval from the NHS Executive (and for the largest building schemes from the Treasury as well) for any significant capital investment such as would probably be required to expand capacity to serve a new market. This restriction has remained even though the majority of NHS capital expenditure in future is to be privately financed. The queue of Trusts proposing capital investment schemes is long but the number of schemes being cleared to proceed is small. It appears most unlikely that any NHS Trust would ever be permitted to borrow to finance investment in increased capacity to be used to compete with an existing provider already providing the services in question.

## 7.4 Primitiveness of contracting and of service quality and price information

For a competitive market to work effectively, purchasers have to be able to know the price and quality of what they are buying and the price/quality combinations available from alternative providers. Without this information they can neither judge the appropriateness of their purchasing decisions between more of one service and less of another, nor determine the most cost-effective sources of supply between alternative provider hospitals. The lack of good information about service quality and price in the NHS 'internal market' has been widely and consistently reported. Furthermore, what information there is, is controlled by the providers. Purchasers have few independent sources of information and

benchmarks with which to assess the reasonableness of the quality and cost of the services they receive. Propper (1993a) observed that, even in the US health care market, competitive bidding between providers for the right to supply purchasers has been limited. She suggests that the high costs of obtaining information about the quality of care being provided have led purchasers to seek long-term relationships with a small number of providers rather than to solicit widespread competition. The current government is reinforcing this by now requiring purchasers and Trusts to sign 'service agreements' to last at least three years, instead of annual contracts (Cm 3807, para. 2.21).

In the UK NHS a consequence of the lack of market information has been continuing heavy reliance by purchasers on large block contracts with small numbers of providers. In such an environment of ignorance about precisely what will be delivered under a contract and about what the price consequences would be of marginal changes to service volumes, mix or qualities, competition is likely to be stifled. The Accounts Commission for Scotland (1997) found that 'the contracting process does not provide robust information on the quantity, cost and quality of services delivered and it is not possible to assess value for money'. The Commission's finding that even after five years of the NHS 'internal market' simple and 'sophisticated' block contracts (see Box 3) still accounted for 61 per cent of expenditure in Scotland in 1996/97, and that this had actually increased from 57 per cent in 1994/95. The lack of specificity in contracts appears to be even greater in other parts of the UK than in Scotland, although the reasons for these differences are obscure. Paton et al (1997) found in England and Wales in 1995/96 that 84 per cent of income in the Trusts surveyed came via block contracts. This proportion represents comparatively little change from the position found by Appleby et al (1994) in the West Midlands region in 1992/93, where 89 per cent of acute hospital contracts by value were simple or sophisticated block; or from the position reported by the NHS Executive's Third National Review of Contracting in the whole of England in 1994 when 74 per cent of HAs' main contracts with acute providers were sophisticated or simple block. In the interests of reduced bureaucracy, the proposals in the 1997 White Papers explicitly reject the use of cost per case contracts, and so are likely to reinforce the prevalence of block contracts.



### BOX 3 Types of Contracts in the NHS 1991-1998

**Simple block** – Purchaser pays provider an annual fee (usually in monthly instalments) for access for the population for which the purchaser is responsible to a defined range of services. No activity levels are specified. Simple block contracts are effectively a continuation of pre-1991 funding arrangements, when health authorities combined health care provision with purchasing and received an annual sum to provide all of the hospital and other secondary health care – volume unspecified – required by their population.

**Sophisticated block** – Simple block contracts modified by the addition of activity ceilings and floors (e.g. a projected level of activity plus or minus five per cent). If actual activity exceeds the ceiling, the provider is explicitly entitled to renegotiate with the purchaser for more resources. If outturn activity falls short of the floor, the purchaser is entitled to negotiate to claw back some of its expenditure from the provider.

**Cost and volume** – Provider receives a fixed sum for treating a specified number of cases and a pre-determined fee per case treated beyond that number. This is, therefore, a two-part tariff, such as is common in utilities' pricing (i.e. for telecommunications, electricity, gas, etc.).

**Cost per case** – Provider receives a fee for each patient treated or item of service (e.g. diagnostic test, surgical procedure, outpatient consultation) provided.

While services continue to be bought as a single bundle because of the difficulty and cost involved in identifying separate cost and quality data for individual services, it is not obvious how competition for provision of those individual services can be effective.

## 7.5 'Satisficing' by providers and/or purchasers

Even where competition is possible, the agents responsible for enacting competitive behaviour may prefer instead to 'satisfice' and so leave the hoped-for benefits of competitive incentives unrealised. Satisficing is behaviour which does not attempt to achieve the best but is content merely to achieve acceptable levels of performance. That is, the managers or clinicians of a hospital may prefer a 'quiet life', operating at an acceptable level of efficiency and quality without constantly striving to win additional work or plaudits by being that bit more competitive. Alternatively they may prefer to pursue goals other than service

efficiency and/or activity maximisation, perhaps devoting energy and resources at the margin to research or education activities.

The same kind of approach might also be taken by purchasers. The managers of HAs stand to gain little personally (in a narrow, material sense) even if they are successful in extracting small amounts of additional activity for the same outlay of authority funds or marginal improvements in service mix or service quality by playing providers off against one another. On the contrary, making a reality of competition may well bring them a more stressful, confrontational working life as a result. Doing enough to avoid evident waste or poor quality services, without pursuing maximum quality and/or minimum cost, may be an attractive and viable option for purchasers. This is especially likely if, as in the case of HAs, purchasers are not themselves exposed to competition.

Satisficing behaviour may exist to some extent, but there are nevertheless indirect pressures which may push Trust and HA managers towards encouraging some competitive behaviour. Part of managers' success as perceived by their superiors and hence perhaps also their performance-related-pay (where effective) and promotion prospects, may be seen to depend on winning more work for a Trust or on squeezing more activity or higher quality services out of an HA's budget.

Fundholding GPs may have a more direct material incentive to pursue cost-effective service provision by hospitals, because any financial surpluses earned by the practice may be ploughed back, in part, into raising the value of the practice's premises and equipment. Notwithstanding that incentive, Miller (1997) has shown that fundholders in Yorkshire are far from taking full advantage of large price differences between alternative nearby (i.e. lying within the same health authority district) provider hospitals. Miller's survey of fundholders found them citing the (largely subjective) reputation of the hospital consultant as the single most important factor determining referral patterns. Yet many GPs were sending different patients to several different consultants in different hospitals for the same treatment, which suggests either that quality of care was the same everywhere, or that other factors (such as proximity to the patient's home, or shortness of wait for diagnosis or treatment) were decisive in the referral decision, rather than the maximisation of care quality received by patients or the minimisation of expense to the fundholder's budget.

## 7.6 Public and political opposition to hospital closure and service relocation

Spurgeon et al (1997) suggest that 'perhaps the most powerful countervailing influence to market principles is the requirement, reinforced by local views, to maintain locally accessible services. This is in part a political issue with few purchasers willing to risk the political and public fall-out of threatening local providers'. Whatever the strength of objective argument that might be marshalled and presented to the public in order to explain the net benefits of a proposed relocation of a hospital service or, in the extreme, the closure of a hospital, public opposition to such change can be guaranteed. Change is commonly perceived as a loss of service not an improvement. It is a familiar public choice problem that the losers from any proposed change (e.g. those living near to an existing service which is to be relocated) are vocal but the gainers (the patients who will receive higher quality care from the relocated service) tend not to be. Furthermore, the 'losing' Trust may well actively work up local popular support in order to defend its services. 'Local freedoms, national responsibilities' (NHS Executive, 1994a) gave explicit recognition to making 'retaining units [i.e. hospitals] which are popular with the public' a criterion by which major service changes such as mergers between NHS Trusts should be judged by health authorities and NHS Executive regional offices.

Hospital closures require formal public consultation to be undertaken and may subsequently be subject to judicial review. These are costly processes and also impose several months unavoidable delay before a purchaser may implement the strategy it has decided on. Given the cost (mainly in terms of managerial and administrative time) involved in preparing and conducting a public consultation exercise, not to mention the emotional strain on health authority managers of acting the role of local hate figure for several months, the threat of closure of a hospital will never be made lightly and is most unlikely to be pursued in the interests of winning only marginal quality improvements or cost reductions for the services purchased. The potential prize has to be large to warrant such action.

For all of the reasons set out in Sections 4 to 7, the scope for effective competition in the provision of acute hospital services is likely to be limited to:

- opportunistic, short-term provision of small, marginal quantities of non-emergency treatment

and diagnostic services from temporary slack within hospitals' existing physical and staff capacities; and

- where major hospital rationalisation is being forced by medical training and doctors' hours requirements, reinforced by guidelines which link improved service quality and safety to higher patient volumes per clinician and a wider range of on-site back-up services, 'knock-out' competition to determine which hospital will survive and which will close or shrink to a community hospital role. Once the winner has been chosen and the loser closed/reduced the scope for competition once more reverts only to the short-term, marginal variety for the foreseeable future.

It is not surprising, therefore, that the initial reforming zeal of the early 1990s has in the last couple of years given way increasingly to a downplaying of the need for competition between hospitals, even by the last government. Ham (1996) captured the essence of the new policy mood: 'quietly in the night, competition in British health care has slipped away, its passing unremarked and little noticed by those who brought it into this world'.

# Contestability

As the limitations of competition have become increasingly apparent, so the concept of contestability has been accorded greater attention. The meaning of 'contestability' is discussed further below but the essence of the concept is that the incentives to efficient and high quality performance which the existence of actively competing alternative providers is supposed to deliver can, under certain circumstances, be equally well achieved even where there is only a single provider with no existing rivals. For contestability to be effective the existing provider must fear that a new entrant would come into the market and take away their business if they were to underperform in any way. Thus, if hospital services are contestable, the currently observed existence of local monopoly providers (i.e. high degrees of provider concentration, indicated by high Hirschman-Herfindahl indices) would not be a bar to market forces stimulating hospitals to provide high quality services efficiently.

The first explicit official reference to the possibility of contestability being more relevant for some NHS hospital services than competition, appears to have been in the NHS Executive's 1994 guidance paper on the operation of the NHS internal market 'Local Freedoms, National Responsibilities'. In this, the NHS Executive recognised the view that while competition might be possible for many services, nevertheless:

'For other services (for example where there are economies of scale) it may be more efficient to have just one provider, whose behaviour is stimulated by the knowledge that another provider could replace it. In this case the system is contestable if not directly competitive.'

In the UK since then, contestability as an alternative to competition, has increasingly been associated with health care provision and with hospital services in particular. The term is now in common usage in policy statements about improving service provision in the NHS although as an almost inevitable result of this currency, the meaning attributed to 'contestability' has become less precise.

The attraction of contestability has both populist and technical roots. Competition requires the existence of more than one hospital provider of a service to an area's population, each with the capacity to increase its supply of services at the expense of the other, which implies the existence of 'wasteful' overcapacity. Contestability has the readily graspable advantage that the existence of a second provider is not required, because merely the credible potential for one to enter the market will suffice, and hence the wasteful overcapacity is avoided.

At a more technical level, the advocates of competition in provision of hospital care are faced with the uncomfortable possibility that the existence of economies of scope and scale could (as described earlier in this paper) mean that cost minimisation would only be achieved by there being a single acute hospital in any locality. That is, there would, at a local level, be a natural monopoly. 'Local' here means the area covered by a population that a fully-employed A&E hospital would be able to serve. This is of the order of anything from 200,000 (just below the current average size of A&E hospitals in England and Wales) to 500,000 (based on current Royal College of Surgeons guidance – see Royal College of Surgeons, 1997). Contestability has more general applicability than traditional notions of competition because it can in principle prevail even where there is a natural monopoly due to economies of scale and/or scope, so long as it is possible for a new entrant to become that monopolist by displacing the incumbent.

The theory of contestability was first set out in 1982 by Baumol, Panzar and Willig in their seminal work: 'Contestable markets and the theory of industry structure'. While the theory has been the subject of some debate among economists since, and has been refined slightly in the process, its basic tenets remain unchanged. Baumol defined a contestable market as 'one into which entry is absolutely free, and exit is absolutely costless' (Baumol, 1982). Free entry here means not that no costs have to be incurred, but rather that the entrant suffers no disadvantage in terms of

production technique or perceived product quality relative to the incumbent. Free exit means that any organisation could leave the market without impediment and in the process of departure can recoup any costs incurred in the entry process. Thus 'the crucial feature of a contestable market is its vulnerability to hit-and-run entry' (Baumol, 1982).

In Figure 1, with market demand DD and a technology of production with average cost curve AC2, in a contestable market an attempt by the incumbent producer to price its output above  $P_e$  and so earn supernormal profits (equal to  $[P_2 - P_e] \cdot Q_2$  if the incumbent imagines pricing at  $P_2$ , say) would immediately attract a new entrant who could undercut the incumbent by charging a price lower than  $P_2$  but greater than or equal to  $P_e$  and so take the entire market away from the incumbent. As long as the price remained above  $P_e$  there would be an incentive for a new entrant to take the market away from the incumbent. Knowing this, the incumbent will be forced by the threat of new entry to price at  $P_e$ , that is at the level of minimum average cost. The actualisation of this story depends on it being possible for a new entrant to do so costlessly, i.e. hit and run entry must be possible. That is, if the incumbent were to react to new entry by a rival by reducing its own price to  $P_e$  (or below, temporarily) the new entrant must be able to exit the market costlessly. If that is not possible, new entry will be deterred by the fear of incurring unrecoverable costs should, as is indeed likely, the incumbent respond to the advent of the new entrant. Such unrecoverable costs are referred to as sunk costs and are, in effect, the costs of investments in specialised physical or human capital for which the second-hand market will yield only a poor price. If the assets, physical or human, which a new entrant must invest in, may easily be re-sold without loss if the incumbent's reaction makes the entrant wish to withdraw from the market, then sunk costs are zero. This condition rarely if ever holds in reality, any more than the conditions required for perfect competition. The empirical and policy question, therefore, is whether sunk costs are small enough in the market of interest for contestability to be practicable.

To put the same arguments a different way: contestability exists where an incumbent faces a credible threat of losing its business to a new entrant if it acts inefficiently or with disregard to its customers' wishes. With contestability it is the threat of competition which incentivises providers

rather than the fact of competition. The degree of contestability, and hence the strength of the incentive for good performance, is a function of the credibility of the threat of entry. That threat is more credible the lower are the sunk costs faced by a potential new entrant.

The government's usage of the term 'contestability' is vaguer: it is a process 'in which the performance [of a provider] could be examined, discussed and proposals for improvement agreed' (Smith, 1997). This kind of 'contestability' is less explicit about how the threat of potential entry is made credible and emphasises instead contestability as the exercise of peer-group pressure in a kind of benchmarking exercise by commissioners of health services (groups of GPs working together with HAs), with loss of business to a better-performance-promising potential new entrant as a last resort only. However, even this notion of contestability must rest on the credibility of potential new entry as a threat to incentivise an inefficient or poor quality incumbent provider. If the existing acute hospital service provider perceives no such threat, the stimulus to it to improve its performance is reduced, so that more ingenuity may be expended by its managers or its senior clinicians on explaining why its apparently poor performance is an illusion brought about by its atypical circumstances, than on actually improving its efficiency and quality of service. Hence, the more rigorous, economist's definition of 'contestability' is that which is applied in the next section of this paper.

The following section considers to what extent markets for acute hospital services may be contestable.

# How contestable are acute hospital services?

Given the new emphasis being given both by politicians and some economists (e.g. Smith, 1996 and 1997, and Ham, 1996a and 1996b) to contestability as an alternative pressure for efficiency and quality where competition is impractical, the potential extent of contestability for acute hospital services needs to be established. Contestability can overcome some of the barriers to competition discussed in Sections 4 to 7 but not all. This section suggests to what extent and where, based on currently available evidence, contestability is likely to be of help.

The degree of contestability possible is determined by the extent of sunk costs that have to be incurred to enter the market in question. In the absence of sunk costs, acute hospital markets could be contestable even if the existence of economies of scale and scope require local monopolies. As was described earlier, current medical professional guidance requires an acute general hospital with a full A&E service to be serving a population of 450,000-500,000 if expensive specialist resources are not to be left under-utilised. But even if a number around 200,000, being a little below the current average 230,000 population served by an A&E hospital in England and Wales, is taken as a more conservative estimate of the minimum efficient scale of an acute hospital, this requires the existence of an accessible population of (2x200,000 =) 400,000 or more for competition to be possible in the sense of their being two or more providers. The question is, therefore, if not competitive in the multiple-provider sense unless there is a concentrated population of around 400,000 or more in an area and possibly not even then (because of the unavailability of spare capacity or the unwillingness of purchasers to contemplate the close of part or all of a hospital), are acute hospital services nevertheless potentially contestable? In other words, could the incentives of a credible competitive threat be obtained even though there is currently a local monopoly provider?

In answering this question, it is necessary to consider the scope for a credible threat of new entry at any of several different levels, namely:

1. **whole hospital** – the new entrant builds, equips and staffs a complete new acute hospital with a view to replacing the incumbent hospital and its staff;
2. **all hospital staff** – take-over of the physical infrastructure of the hospital and buildings by a new entrant who replaces the existing hospital staff (medical and nursing professionals, managers and support staff) with its own people;
3. **individual specialties** – e.g. provision of just cardiac surgery or neurology or ENT, or dermatology, or whatever; covering both elective and emergency work; inpatient, day case and outpatient. The new entrant would provide the specialist staff and facilities needed to deliver just the specialty in question, either by adding to such staff/facilities currently employed by it but serving another market, or by establishing a new specialty service;
4. **individual procedures** – providing only specified elective operations, diagnostic or therapy services, or groups of these, e.g. terminations of pregnancy, cataract surgery, MRI scanning, pathology tests, physiotherapy, speech therapy;
5. **management and administration only** – i.e. the hospital buildings and equipment remain the same; the medical, nursing and other professional care staff are the same; only the management and administration of the service is contested.

The following paragraphs assess the extent to which the existence of sunk costs or other difficulties might weaken the potential for contestability at each of the levels listed above. Contestability is not in general prevented by the presence of economies of scale or scope, but some of the other barriers to competition which were described earlier in this paper are also obstacles to contestability. The potential barriers to contestability are:

- existence of significant sunk costs;
- Department of Health regulation of hospitals' output prices and return on capital;

- Department of Health control of hospital inputs' – labour and capital – supply and prices;
- primitiveness of contracting and of service quality and price information;
- satisficing behaviour by providers and/or purchasers.

Contestability at the level of building and equipping from scratch a whole new acute hospital to take over the activity currently undertaken by the incumbent, is impractical. The sunk costs are much too large to permit hit and run entry. Also, the time lag between deciding to enter and having a hospital available to compete would be several years, giving the incumbent plenty of time to react to the emerging threat by reducing prices and/or improving quality and so making it unprofitable for the new entrant to supply that market. A new acute hospital with an A&E department and all associated specialist medical, surgical, anaesthetic and diagnostic services will have a total capital cost of over £100 million and will take three to five years to complete. For example, the new Dartford and Gravesham 400-bed A&E hospital, which is to be built in Kent and is the first new hospital construction deal to have been signed in the NHS for several years, is expected to have a total capital cost of £143 million, excluding land costs (Dartford and Gravesham NHS Trust, 1997). Construction time is planned to be three years, but this is in addition to the lengthy negotiation and design process (of at least one year) which preceded the signing of any contracts to build. The new 800-bed Norfolk and Norwich Hospital is to cost £214 million (Health Service Journal, 15 January 1998). Even major refurbishments of existing hospitals (should a partly redundant hospital be available to be taken in hand by a new entrant) cost anything from £40 million to £120 million in capital terms. (See for example Healthcare Market News of July 1997 for a summary of 14 major acute hospital capital schemes cleared to proceed by the new Labour government).

Furthermore, once built, an acute hospital represents a highly specialised collection of physical assets which on their own, separate from contracts and patients, will typically have an open market value, if it subsequently has to be sold off, which is a fraction of the sum originally invested. An infamous example of this in Britain was the Health Care International private hospital in Clydebank, Glasgow. Completed in 1994 it has since failed to find a market for its services to the

extent that in 1996 it was still losing a reported £20 million a year on a turnover of just over £5 million (Health Service Journal, 21 August 1997) and was operating at just 10 per cent occupancy. Far from being saleable as a going concern, the hospital had to be rescued by new owners in 1995 and has still to make anything other than large losses.

For many medical and surgical specialties and some types of specialist nursing (e.g. intensive therapy nursing), there are persistent shortages of staff in the UK labour market, even with the European Union-driven freeing up of the European health care labour market. While this is in principle remediable, provided that government limits on the funding available for medical and nurse training and government limits on the pay levels of these groups were not binding constraints, it would take years to train such specialists. The main issue here is the time that would elapse before a fully trained team could be assembled, during which time the incumbent hospital could take steps to improve its performance and so remove the opportunity for a new provider to enter the market, rather than the magnitude of the training costs themselves. This investment in human capital will not represent a sunk cost as long as the putative new entrant is able to use these resources in another market apart from the one it was initially targeting. If, however, there is no such alternative market, then the training costs would be non-recoverable. This would then add another deterrent to new entry. All in all, these specialist labour supply issues reinforce the impracticality of whole hospital contestability. They make contestability at the second level suggested above, all hospital staff, equally unlikely.

In the context of the NHS, an existing hospital's managers and staff might therefore rest safe in the knowledge that these kinds of entry by a new provider with a new hospital or new staff will not happen, however inefficient or poor quality the existing service. Contestability may, however, be more practicable if considered at a level below that of an entire acute hospital: i.e. individual specialties or procedures, or just contesting the management of services rather than their direct provision.

The question of which of these sub-hospital levels of contestability is more relevant depends on the source of the dissatisfaction with incumbent providers which contestability is supposed to overcome. Where the source of the poor

performance is perceived by commissioners of services, and by potential new entrant providers, as the clinicians or other health care professional staff, then contestability at the specialty or procedure level is required. If, however, it is the incumbent provider's management which is perceived to be the source of its underperformance, then contestability just of that management would be all that was needed.

Considering first the possibility of a new entrant providing a single hospital specialty or an individual procedure, this could be realistic only if the specialty concerned is discretionary for an A&E hospital and relatively discrete (e.g. like the examples quoted in the list above). Where the specialty is a necessary element of an A&E hospital (see Box 2) or shares many common costs with A&E hospital services (e.g. elective activity in the fields of general and orthopaedic surgery, pathology, imaging), an incumbent is unlikely to consider a threat of new entrant credible and so will not be motivated to perform better by such a threat. Services within these categories could be provided by an A&E hospital at a marginal cost well below the cost achievable by a provider of the individual service on its own. Thus, threatened new entry would have to be across the range of core A&E-related services, rather than for individual elements of such services, to be credible. But no new entrant would be likely to want to do that, for the sunk cost and time lapse reasons described above in the discussion of whole hospital or all hospital staff new entry.

Contested provision of non-A&E related acute hospital specialties or procedures may be possible. Although the on-site presence of such services may not be essential to the running of an A&E-based hospital, some of them they may nevertheless be more cost-effectively provided by such a hospital, for example because of elements of shared costs or advantages in attracting the best staff. New entrant provision of individual elective specialties or procedures is therefore most likely, if at all, to be by a nearby hospital which either already provides such services or could readily add them to its core of A&E related activities. Private acute hospitals, which in the UK deal only in elective surgery, could in principle also provide a credible threat to an incumbent A&E hospital for elective services which share few common costs with A&E-related services, e.g. day surgery.

Providers of similar services in other locations are generally recognised as the most credible potential sources of competition in most or all economic

fields (Cairns and Mahabir, 1988). There are, however, a number of reasons to doubt whether many incumbent providers of acute hospital care in the UK, or the organisations which commission services from them, perceive more than a small fraction of their total acute care budgets as actually contestable in this way. (It is interesting to note here that in its 1994 guidance on the operation of the NHS internal market, the NHS Executive defined a Trust as being in a degree of financial difficulty severe enough to warrant central intervention if it suffered either a two per cent unplanned fall in revenue or a 10 per cent planned fall).

One major source of this doubt is that NHS providers have little incentive to become new entrants in contest with an incumbent that wishes to remain. The lack of incentives for other NHS providers to contest a market results from the government's regulation of hospitals' output prices and return on capital (no profits may be earned in excess of the required six per cent real return on capital); and the attraction to many NHS managers in consequence of satisficing behaviour. Furthermore, the government's control over labour input prices, limits the scope for achieving lower costs than an incumbent other than by employing fewer, more junior or less highly trained staff. Even if feasible, any of these avenues is going to be hard to sell to a commissioner who judges service quality largely on the basis of the perceived quality of the inputs used to produce it, as is still largely the case in the NHS.

Private hospitals, while they have the incentive to contest, are handicapped by having higher marginal costs of capital than the taxpayer-funded NHS and by lacking the economies of scope which may be derived from providing emergency care services (which are the sole preserve of the NHS in the UK) alongside elective services. Furthermore, HAs may be reluctant to increase their purchases from private hospitals if they fear this will result in NHS hospital consultants who have private practices spending a reduced amount of their time in NHS hospitals in order to cope with the increased workload in the private hospitals where they work part-time.

Lack of good cost, price and service quality information also makes contestability difficult to achieve. A new entrant will not know the incumbent's true, case-mix-adjusted costs for the services of interest, nor their quality. The purchaser who is dissatisfied with an incumbent has the incentive to make the information about

the cause of the dissatisfaction available to potential new providers. But, even so, the information available (i.e. even to the purchasers) may be too poor to be of much help to the potential entrant (see the discussion in Section 2 above). Contestability can in principle be made more achievable with longer-term contracts, as these help to reduce the extent of sunk costs (an alternative use for fixed capital or specialist skills may be easier to find if there are years available in which to look). The government is indeed now promoting three to five year, rather than annual, agreements between commissioners and providers (Cm 3807). However, longer-term contracts limit the contest to once every three or five years (in this case) and so dilute the incentive for the contract winner to maintain efficiency and effectiveness between contests. Where the physical assets are unchanging and it is only the people using them who potentially change (in effect, franchising), there is the additional problem that an incumbent who wishes to withdraw at the next contest, or who expects to lose it, may inadequately maintain the assets under their stewardship in order to reduce costs until their contract expires.

In the face of these disincentives, it becomes an empirical question whether contestability exists over a significant range of acute hospital services and markets. The existence of contestability in practice is hard to prove and evidence is limited. The NHS Executive's Trust Finance Unit, in its 1996 Review of the Trust Financial Regime (Tinston, 1996) failed to find significant traces:

'few Trusts have sought to gain market share from rivals by innovating new services or by radical restructuring to achieve a cost advantage. Although historic relations between purchasers and providers have been maintained with rare exceptions, and genuinely open tendering has been rare, few excluded Trusts have challenged health authorities to open up their tendering procedures.'

One of the few examples of an HA conducting an explicit contest to try and improve the cost-effectiveness of the acute hospital services it purchased is reported by Florida-James (1997). He describes Walsall Health Authority's market testing of urology services for its 260,000 resident population in 1993. The focus of the market test was on improving the quality of the service: reducing waiting times for first outpatient appointment, having some specialist clinics set up which the incumbent was unwilling or unable to

provide, and having the service provided by a two-consultant unit rather than a single-handed urologist, which the health authority considered to be undesirable practice. Interestingly the tenderers were restricted by the health authority to the three nearest existing hospital urology services and in the end the contract was awarded to the incumbent who was pressured by the process into appointing a second consultant and thereby able to deliver shorter waits and the desired special clinics. This was at the expense of an increased cost, however. This can be seen as successful contestability in action. However, it was just for one service (costing £1.15 million, or roughly one per cent of the health authority's total budget) and despite this small scope was 'difficult and time consuming' for senior management and was feared to have damaged relations between health authority and its main A&E hospital provider (Florida-James, 1997). It is perhaps unsurprising, therefore, that there is little evidence of contestability being used actively by HAs and no evidence that Trusts are actively seeking to enter new markets.

As was noted in Section 3 above, there is greater anecdotal evidence of GP fundholders moving contracts between providers, particularly for diagnostic services, than of health authorities doing so. But again it is the general stability of service provision that is remarkable rather than the degree of change, given that fundholders cover 60 per cent of the UK population. This does not necessarily mean that contest is not happening. The mere threat of recourse to market testing may, behind closed doors, be producing the efficiency and quality improvements which health care commissioners desire. Unsurprisingly, there is no evidence about this.

Over time the credibility of a threat that is never carried out inevitably diminishes. The NHS internal market has been in operation for several years but there is still little evidence of bluffs being called. Where contracts have been moved it has generally been by HAs 'repatriating' patient flows to their local general hospital which in the past had gone to out-of-district hospitals for historical reasons (such as personal links between some GPs and individual consultants at the more distant hospital). The motive for this may have as much to do with ensuring the financial viability of the local provider as with improving the patients' experience or reducing the cost of care.

Non-hospital services, such as community nursing and primary care, may be more contestable than



acute hospital care. For example, in 1996/97 a Total Purchasing Pilot group of eight GP practices in Merton, Sutton & Wandsworth held a winner-takes-all contest between its two existing community Trust providers. As a result, the total purchasing practices transferred all of their community nursing and related business to the Trust which had hitherto supplied just 20 per cent of their combined needs. Premier Health NHS Trust, a community Trust in the English midlands has been willing to hold the contracts of community nurses working in other parts of the country. Some contestability in these non-hospital services is, therefore, apparent. It is interesting to note that the motive for this contestability has not been to change the community nurses actually providing the care, but rather to replace their managers.

In the field of mental health services, the Scarborough and North East Yorkshire Health Care NHS Trust has been reported in the press as being about to lose to a neighbouring community and mental health Trust its £6.8 million p.a. contract to provide mental health care for the residents of North Yorkshire Health Authority (Health Service Journal, 18 September 1997). The HA is to contract in the new financial year with a different Trust which already provides comparable services in a neighbouring area and is intending to re-employ the same nurses and other care staff who currently deliver services in North Yorkshire. Only the management of the services is therefore to change. The stimulus for the contract switch was apparently dissatisfaction with quality, not with costs, which have in fact risen as a result.

The examples of Merton, Sutton and Wandsworth, Premier Health and North Yorkshire suggest that while contestability appears improbable at other levels, it may be more achievable at the level of a hospital's management/administration team. It could be argued that de facto 'contestability' of senior management is already generally attained in the sense that individual senior managers do lose their jobs if the performance of their hospitals is perceived to have been poor in some noticeable regard. For example, the chief executive of an acute Trust in southern England resigned in 1997 following his hospital's successful prosecution for failure to meet nationally required health and safety standards. It would be stretching the point, however, to interpret replacement of individual managers as evidence of effective contestability of acute hospital services in action, rather than as the normal dynamics of the labour market. Newly

arriving NHS Trust chief executives have sometimes brought with them their own 'team' of one or two other senior managers, but wholesale replacement of a management team at the behest of purchasers has hitherto only happened once, at Anglian Harbours NHS Trust. Although several Trusts have been merged or reconfigured (e.g. in London: Guy's and St Thomas' Hospitals have merged while Lewisham Hospital has been de-merged from Guy's; or the merger of St Bartholomew's and the Royal London Hospitals while Homerton Hospital was de-merged from St Bartholomew's), these changes have been in response to centrally planned reconfigurations of hospital services rather than to failures of performance by any of the hospitals concerned. These changes were initiated by the Department of Health, not the local purchasers.

The demise of Anglian Harbours Trust, in contrast, was the direct result of action by its two main purchasers: East Norfolk Health Authority and Suffolk Health Authority. These purchasers announced in September 1996 that as from August 1997 they would be transferring their business, which represented 90 per cent of the Trust's total income, to other providers (Health Service Journal, 26 September 1996). The grounds for this move were, ostensibly at least, a history of resistance to the purchasers' requirements, including: lack of co-operation by Anglian Harbours with a neighbouring acute Trust leading to unnecessarily blocked beds at that Trust; failure to recruit key medical staff; and a perception by the purchasers that the Trust was, to quote the Health Service Journal, 'not effective or efficient either financially or clinically. Basically it is too small'. In effect, a hundred or more managers of Anglian Harbours Trust, including the most senior, have lost their jobs to managers in other Trusts while the front-line care staff have been left alone (Health Service Journal, 4 September 1997).

The case of Anglian Harbours is interesting for the examination of contestability, for the following reasons:

- so far it is unique as an example of a contest for anything more than a couple of million pounds worth of annual business, after six years of the NHS internal market;
- the Trust does not run an acute hospital but delivers community and cottage hospital services;
- the replacement providers have been sought and recruited by the purchasers and are existing NHS Trusts (four of them) operating overlapping

services in neighbouring areas, rather than being aggressive new entrant providers intent on forcing an incumbent out of business;

- the transfer to new providers was not a rapid, let alone a 'hit and run', event. Rather it was the end result of a cumulative process of growing dissatisfaction, following repeated and prolonged negotiation between the purchasers, Anglian Harbours Trust and the neighbouring Trusts;

- cost savings are expected to be relatively small (about £1 million per annum out of total annual costs of approximately £30 million) and to be achieved by cutting management and administrative staff. No specific failure of service quality has been cited by the purchasers, who indeed have stressed that 'patients would not notice any difference in services. They would continue to be seen by the same staff' (Health Service Journal, 4 September 1997).

The experience of the Suffolk and East Norfolk Health Authorities with Anglian Harbours NHS Trust, although an isolated example, seems to show that the ultimate sanction for a purchaser to demonstrate the contestability of a local health care market is indeed real but only as a last resort. The process of extracting improved, or at least changed, performance from the local provider Trust will be likely to involve literally years of negotiation and pressure, such as diversion of incremental funding, including commissioning of new services, to other Trusts as a sign of displeasure. Only when these channels have been exhausted will the 'nuclear deterrent' of seeking a more responsive management team (or, in the Anglian Harbours case, teams – four other Trusts each taking on the management of a different part of the incumbent's services) to replace the incumbent be launched. All of this diminishes both the short and medium term practicability of contestability. In the long term, after a few years, the threat of replacement by a new management team may be more credible. So far there have been no public examples of contestability being used overtly to stimulate the improved performance of an acute Trust's management team, but it cannot be ruled out.

# Conclusions

**F**rustration with the increasing evidence of barriers to competition between acute hospitals, combined with a change in the political climate so that emphasis is now placed on co-operative rather than competitive behaviour, have brought the concept of contestability to the fore in the NHS policy debate. The attraction of contestability, that is potential rather than actual competition, is its apparent ability to yield the efficiency and quality incentives traditionally associated with competition while avoiding either wasteful duplication of acute hospital capacity or, at least in some forms such as contestability of service management rather than of direct provision, unpopular hospital closures.

However, the existence of sunk costs, the nature of government regulation of NHS providers, the lack of information on service costs and qualities, and the attractiveness to NHS managers of satisficing behaviour, together imply that contestability is likely to be of restricted relevance as a policy tool for achieving more efficient and effective acute hospital care. Isolated examples of small scale contests for an individual elective specialty or for the management of a service exist, but these are far from widespread. The difficulty of achieving a meaningful contest for the management of an entire acute Trust's services make it a last resort, to be called upon only after all negotiating stages have been exhausted.

The December 1997 White Paper 'The new NHS' perhaps recognises this:

'If there are problems with performance, the first step will be for the Primary Care Group [i.e. the commissioner of health care services] and the NHS Trust to explore the difficulty and plan to put it right.' (Cm 3807, para. 9.13).

But nevertheless, ultimate reliance is still placed on commissioners being able to seek health care services from a new source:

'[Primary Care Groups] as a last resort will be able to change provider if, over time, performance does not meet the required standard.' (Cm 3807, para. 3.13).

The principle of attempting to bring all acute hospital services up to the efficiency and quality levels of the best is obviously a sound one. But contestability relies on the credibility of the ultimate threat of redundancy for the managers and (perhaps) staff of an underperforming acute service or Trust, to stimulate improved performance. This monograph has indicated that in many circumstances such a threat may not be credible.

In that case, an appeal to professional standards and pride through peer group benchmarking might, in a profession-driven public service such as the NHS, go a long way towards achieving the same ends as are expected from contestability but are unlikely to be achieved by it. For benchmarking information to be effective as a tool for highlighting underperformance and stimulating improvements, by the threat of public shaming of inefficient or ineffective performers in front of their professional peers (medical, nursing, health service managerial, whatever), it must be:

- recognised as valid by those peer groups;
- specifically related to the activities of the management and staff of the service providers;
- specific to case mix and to the social circumstances in which the care services are provided;
- up-to-date; and
- disseminated, i.e. known to commissioners of health care services and potentially available to the wider public.

These are clearly difficult requirements to fulfil. Comparative service quality and cost information, which takes full account of differences in the mix of patients served by different providers, is the goal and much work has already been undertaken on clinical and financial audit and casemix-related costing, and on the creation of associated databases. However, more work on improving the benchmarking tools available is undoubtedly required, and the value of using benchmarking will need to be demonstrated in practice.

The December 1997 White Paper 'The new NHS' makes a number of important proposals which, if carried through successfully over the coming years, will provide the foundations for achieving that. In particular the White Paper proposes to establish:

- 'clinical governance' requirements on all NHS Trusts. That is, Trusts will be given a new statutory duty for the quality of the care they provide and will have to demonstrate that they have arrangements in place to audit, assure and improve the quality of their services;
- a new National Institute for Clinical Excellence to draw up guidelines on clinical and cost-effectiveness and disseminate them to all parts of the NHS;
- a requirement for NHS Trusts to publish and benchmark their costs. Using this information, a national schedule of 'reference costs' for individual treatments is to be established and published, against which to compare the costs of different providers.

Contestability may be of limited help for some services in some areas, but the challenge is to incorporate the quality and cost benchmarking of existing acute hospital services into the clinical and managerial culture of the NHS. Only in that way may performance be improved without recourse, in most cases, to the upheaval of avoidable and publicly unpopular service relocations.

# References

Accounts Commission for Scotland (1997). *Expanding on contracting*. Health board commissioning: contracts and contracting. Bulletin, August 1997.

Appleby J, Smith P, Ranade W, Little V and Robinson R (1994). Monitoring managed competition. In: Robinson R and Le Grand J, eds. *Evaluating the NHS reforms*. King's Fund.

Audit Commission (1996a). *By accident or design: improving A&E services in England and Wales*. London, HMSO.

Audit Commission (1996b). *What the doctor ordered?* London, HMSO.

Audretsch DB and Acs ZJ (1991). Innovation as a means of entry: an overview. In: Geroski PA and Schwalbach J, eds. *Entry and market contestability: an international comparison*. Blackwell. pp. 222-243.

Bartlett W and Le Grand J (1993). The theory of quasi-markets. In: Le Grand J and Bartlett W, eds. *Quasi-markets and social policy*. Macmillan.

Baumol WJ (1982). *Contestable markets: an uprising in the theory of industry structure*. American Economic Review. Volume 72, Number 1, pp. 1-15.

Baumol WJ, Panzar JC and Willig RD (1988). *Contestable markets and the theory of industry structure*. Revised edition. Harcourt Brace Jovanovich.

Baumol WJ and Willig RD (1986). Contestability: developments since the book. *Oxford Economic Papers*. New Series Volume 38, November 1986 Supplement, pp. 9-36.

Brooks JM, Dor A and Wong HS (1997). Hospital-insurer bargaining: an empirical investigation of appendectomy pricing. *Journal of Health Economics*. 16, pp. 417-434.

Cairns RD (1996). Uncertain contestability. *Journal of Economic Behavior and Organization*. Volume 30, Number 1, pp. 125-131.

Cairns RD and Mahabir D (1988). Contestability: a revisionist view. *Economica*. Volume 55(218), pp. 269-276.

Calman K et al (1993). *Hospital doctors: training for the future. The report of the Working Group on Specialist Medical Training*. Department of Health, April 1993.

Chalkley M and Malcomson JM (1996). Contracts for the National Health Service. *Economic Journal*. Volume 106 (439), pp. 1691-1701.

CIPFA/HFM (1995). *Health database 1995: health services financial database*. CIPFA.

Cm 555 (1989). *Working for patients*. HMSO.

Cm 3807 (1997). *The new NHS: modern, dependable*. Stationery Office.

Cm 3811 (1997). *Designed to care: renewing the National Health Service in Scotland*. Stationery Office.

Croxson B (1997). Co-operation, competition and the NHS internal market. *HERU Discussion Paper 02/97*. Health Economics Research Unit, University of Aberdeen.

Culyer AJ and Posnett J (1990). Hospital behaviour and competition. In: Culyer AJ, Maynard AK and Posnett J, eds. *Competition in health care: reforming the NHS*. Macmillan.

Dartford and Gravesham NHS Trust (1997). First PFI hospital deal is completed. *Dartford and Gravesham NHS Trust Press Release*, 29 July 1997.

Dawson D (1995). Regulating competition in the NHS: the Department of Health guide on mergers and anti-competitive behaviour. *Discussion Paper 131*. Centre for Health Economics, University of York.

Ellwood S (1996). Pricing services in the UK National Health Service. *Financial Accountability and Management*. Volume 12, Number 4, pp. 281-301.

Ferguson B and Palmer S (1994). Markets and the NHSME guidelines: costs and prices in the NHS internal market. *Discussion Paper 120*. Centre for Health Economics, University of York.

Ferguson B, Posnett J and Sheldon T (1997). *Concentration and choice in the provision of hospital services*. CRD Report 8. University of York, NHS Centre for Reviews and Dissemination.

Florida-James PJ (1997). The internal market exposing clinical services to competitive tendering. *Journal of Public Health Medicine*. Volume 19, Number 1, March 1997, pp. 23-28.

Glennerster H, Matsaganis M, Owens P and Hancock S (1994). GP fundholding: wild card or winning hand? In: Robinson R and Le Grand J, eds. *Evaluating the NHS Reforms*. King's Fund.

- Goddard M and Smith P (1996). *Developing competition between providers: the UK experience*. Paper presented to Symposium on Competition and Solidarity in Health Care, Association International de la Mutualité, Brussels, June 13-14 1996.
- Ham C (1996a). Contestability: a middle path for health care. *British Medical Journal*. 312, pp. 70-71.
- Ham C (1996b). Public, private or community: what next for the NHS?. *Demos Paper Number 24*. London.
- Harrison A and Prentice S (1996). *Acute futures*. King's Fund.
- HC 127 (1997). *NHS (England) summarised accounts 1995/96*. The Stationery Office. London. 18 July 1997.
- Health Service Journal*. 26 September 1996. That sinking feeling. p. 11.
- Health Service Journal*. 21 August 1997. Private hospital loses £20m despite £37m regional aid rescue package. p. 7.
- Health Service Journal*. 4 September 1997. Top-down job losses as trust goes to the wall. p. 9.
- Health Service Journal*. 18 September 1997. HA shifts 'deficient' mental health services to rival trust. p. 8.
- Health Service Journal*. 15 January 1998. Work set to begin on biggest PFI hospital. p. 4.
- Healthcare Market News*. July 1997, p. 96.
- Heywood JS and Pal D (1993). Contestability and two-part pricing. *Review of Industrial Organization*. Volume 8, Number 5, pp. 557-565.
- Hornblow A (1997). New Zealand's health reforms: a clash of cultures. *British Medical Journal*. Volume 314, 28 June 1997, pp. 1892-1894.
- Kessides IN (1990). Market concentration, contestability and sunk costs. *Review of Economics and Statistics*. Volume 72, Number 4, pp. 614-622.
- Labour Party (1997). *New Labour – because Britain deserves better*. Labour Party General Election Manifesto.
- Leyland AH and Boddy FA (1997). Measuring performance in hospital care: length of stay in gynaecology. *European Journal of Public Health*. Volume 7, pp. 136-143.
- Mahon A, Wilkin D and Whitehouse C (1994). Choice of elective surgery referrals. In: Robinson R and Le Grand J, eds. *Evaluating the NHS reforms*. King's Fund.
- Mantzavinos C (1992). Contestable Markets, das neoklassische Marktmodell und die Wettbewerbstheorie. *Jahrbücher für Nationalökonomie und Statistik*. Volume 209(1-2), pp. 60-66.
- Maynard A (1994). Can competition enhance efficiency in healthcare? lessons from the reform of the UK National Health Service. *Social Science and Medicine*. Volume 39, Number 10, pp 1433-1455.
- Miller P (1997). Are GP fundholders wasting money? *Health Service Journal*. 6 March 1997.
- Morris DJ, Sinclair PJN, Slater MDE and Vickers JS (1986). Strategic behaviour and industrial competition: an introduction. *Oxford Economic Papers*. New Series, Volume 38, November 1986 Supplement, pp. 1-8.
- NHS Centre for Reviews and Dissemination (1995). *Relationship between volume and quality of health care: a review of the literature*. CRD Report 2. University of York.
- NHS Executive (1994a). *The operation of the NHS internal market: local freedoms, national responsibilities*. NHS Executive.
- NHS Executive (1994b). *The third national review of contracting 1994*. NHS Purchasing Unit. NHS Executive.
- NHS Management Executive (1991). *Junior doctors: the new deal. Hours of work of doctors in training. Working arrangements of doctors and dentists in training*.
- Office of Health Economics (1997). *Compendium of health statistics 1997 (10th edition)*. Office of Health Economics. London.
- Paton C, Birch K, Hunt K and Jordan K (1997). Counting the costs. *Health Service Journal*. 21 August 1997.
- Propper C (1993a). Quasi-markets, contracts and quality in health and social care: the US experience. In: Le Grand J and Bartlett W, eds. *Quasi-markets and Social Policy*. Macmillan.
- Propper C (1993b). Quasi-markets and regulation. In: Le Grand J and Bartlett W, eds. *Quasi-markets and Social Policy*. Macmillan.
- Propper C (1995). Regulatory reform of the NHS internal market. *Health Economics*. Volume 4, pp. 77-83.
- Propper C (1996). Market structure and prices: the responses of hospitals in the UK National Health Service to competition. *Journal of Public Economics*. Volume 61, Number 3, pp. 307-335.
- Propper C and Bartlett W (1997). The impact of competition on the behaviour of National Health Service Trusts. In: Flynn R and Williams G, eds. *Contracting for health: quasi-markets and the National Health Service*.
- Propper C, Wilson D and Söderlund N (1997). *The effects of regulation and competition in the NHS internal market: the case of GP fundholder prices*. University of Bristol, Department of Economics Discussion Paper, February 1997.
- Raftery J, Robinson R, Mulligan J and Forrest S (1996). Contracting in the NHS quasi-market. *Health Economics*. Volume 5, Number 4, pp. 353-362.

- Robinson R (1996). The impact of the NHS reforms 1991-1995: a review of research evidence. *Journal of Public Health Medicine*. Volume 18, Number 3, pp. 337-342.
- Robinson R and Le Grand J (1995). Contracting and the purchaser provider split. In: Saltman R and Von Otter C, eds. *Implementing planned markets in health care*. Oxford University Press.
- Royal College of Surgeons of England (1997). *The provision of emergency surgical services: an organisational framework*. June 1997. London.
- Schwartz M (1986). The nature and scope of contestability theory. *Oxford Economic Papers*. New Series, Volume 38, November 1986 Supplement, pp. 37-57.
- Senate of Surgery of Great Britain and Ireland (1997). *Consultant surgical practice and training in the UK*. Senate Paper 1, June 1997. London.
- Shepherd WG (1984). Contestability versus competition. *American Economic Review*. Volume 74, Number 4, pp. 572-587.
- Smith C (1996). *A health service for a new century*. Speech by Chris Smith MP, Shadow Secretary of State for Health, 3 December 1996.
- Smith C (1997). *Putting quality at the heart of health care*. Speech by Chris Smith MP, Shadow Secretary of State for Health, 6 February 1997.
- Spurgeon P, Smith P, Straker M, Deakin N, Thomas N and Walsh K (1997). The experience of contracting in health care. In: Flynn R and Williams G, eds. *Contracting for health: quasi-markets and the National Health Service*.
- Thomas K, Nicholl J and Coleman P (1995). Assessing the outcome of making it easier for patients to change general practitioner: practice characteristics associated with patient movements. *British Journal of General Practice*. November 1995. Volume 45, pp. 581-586.
- Tinston R (1996). *Review of the Trust financial regime*. NHS Executive Trust Finance Unit, Leeds.
- Van Wegberg M and Van Witteloostuijn A (1992). Credible entry threats into contestable markets: a symmetric multi-market model of contestability. *Economica*. Volume 59(236), pp. 437-452.
- West P (1997). *Understanding the National Health Service reforms: the creation of incentives?* Open University Press.
- Wilson A (1991). Consultation length in General Practice: a review. *British Journal of General Practice*. 41(344), pp. 119-122.

# Appendix 1

## The Hirschman-Herfindahl Index (HHI)

The HHI of provider concentration in a particular market is calculated by summing the squares of the market shares of all providers within the specified market area. For clarity it is conventional to multiply this sum by 10,000 and so to show the index lying in a range from near zero (very low concentration) to 10,000 (a single provider, i.e. total monopoly).

The mathematical definition of the HHI of market concentration in a district 'j' is as follows:

$$HHId_j = [n \sum_{i=1}^n (P_{ij} / \sum_{i=1}^n P_{ij})^2] \cdot 10,000$$

where:

HHId<sub>j</sub> = Hirschman-Herfindahl index of market concentration for district j

P<sub>ij</sub> = number of patients from district j being served by provider i

n = number of providers serving patients from district j

### Example

A single acute hospital 'i' in district 'j' serves the entire population of 500,000. That hospital also serves 50,000 of the 450,000 people in the neighbouring district 'k'. 200,000 of the people in district 'k' are served by hospital 'h' and 200,000 by hospital 'g'.

$$\begin{aligned} \text{Therefore: } HHId_j &= (500,000/500,000)^2 \cdot 10,000 \\ &= 10,000 \quad \text{i.e. pure local monopoly;} \end{aligned}$$

$$\begin{aligned} HHId_k &= [(50,000/450,000)^2 + \\ &\quad (200,000/450,000)^2 + \\ &\quad (200,000/450,000)^2] \cdot 10,000 \\ &= [0.4074] \cdot 10,000 \\ &= 4,074 \end{aligned}$$

The fewer the providers and the more equal their shares, the higher is the HHId.

An HHI may also be calculated from the perspective of an individual provider rather than of a geographical district. The Hirschman-Herfindahl index of market concentration for provider 'i' takes account of the market shares it enjoys in all of the markets it supplies and is defined as follows:

$$HHIc_i = (1/M_i) \cdot \{M_i \sum_{j=1}^{M_i} [(P_{ij} / \sum_{i=1}^n P_{ij}) \cdot HHId_j]\}$$

where:

HHIc<sub>i</sub> = Hirschman-Herfindahl index of market concentration for provider i

M<sub>i</sub> = number of districts in provider i's market area

### Example

The example given above implies the following HHI of market concentration for hospital i, assuming that it only serves districts j and k and no others:

$$\begin{aligned} HHIc_i &= (1/2) \cdot \{[(500,000/500,000) \cdot 10,000] + \\ &\quad [(50,000/450,000) \cdot 4,074]\} \\ &= (1/2) \cdot \{10,000 + 453\} \\ &= 5,226 \end{aligned}$$

Thus, because of its dominant position in district j, hospital i enjoys considerable, but not complete, market power overall.