‘EXPLICIT RATIONING WITHIN THE NHS QUASI-MARKET: THE EXPERIENCE OF HEALTH AUTHORITY PURCHASERS 1996-97’

PhD Thesis

London School of Economics

by

Louise Locock

March 1998

Abstract

This thesis analyses the findings of empirical research carried out in three case study UK health authorities in 1996-97, using repeat interviewing of senior managers. It aimed to test three competing hypotheses:

i. Markets are one possible system for allocating scarce resources. The process of contract specification in a complex quasi-market is likely to make rationing more explicit than it would be in a hierarchical system
ii. In the complex context of the NHS the quasi-market may fail to produce clear contracts and unambiguous allocations, because of prohibitive transaction costs, political costs and ethical costs of greater explicitness
iii. Other pressures in favour of explicitness (e.g. rising expenditure, effectiveness evidence and the Patient's Charter) may be irresistible, whatever structural form the NHS takes.

The complex relationship between explicit rationing, the internal market and other factors is discussed. Results suggest the quasi-market has contributed to the growth in explicit rationing, notably by decoupling purchasers and providers from their previously shared responsibility to manage resources. In other respects the market has speeded up or magnified the effect of other factors which would or could have happened anyway. Concern to control rising expenditure has led to more explicit decisions but is now rekindling interest in the value of fixed budgets for providers and implicit clinical decision-making. Factors such as the Patient's Charter have also had an independent effect on greater explicitness. Implicit rationing remains significant.

The implications for health care rationing of government proposals to abolish the internal market are examined. The results suggest that explicit rationing will probably continue to grow, but with a greater emphasis on explicit criteria to guide clinicians in determining who gets treatment, rather than the exclusion of whole services. The retention of some form of commissioner provider split may also exercise continuing pressure towards explicitness.

'EXPLICIT RATIONING WITHIN THE NHS QUASI-MARKET - THE EXPERIENCE OF HEALTH AUTHORITY PURCHASERS 1996-97'

'As we become aware of what we, as a society, are doing, we bear responsibility for those allocations that will be made as well as for what has been done in our names. If one understands more than before for having read this essay, one can still appreciate that tragic decisions need be made and are not the easier for the understanding.'

'Tragic Choices', Guido Calabresi and Philip Bobbitt, 1978 (p.199)

Contents

Thanks:

to the ESRC for their funding, and for their flexibility in accommodating maternity leave

to my supervisors, Professors Howard Glennerster and Julian Le Grand at LSE, and Julia Parker at Oxford, for their support, advice and insight

to all the interviewees who have taken part, for their generosity in giving me so much time and thoughtful discussion

to Sue Dopson and Rosemary Stewart at Templeton College, Oxford, for giving me something else to do to take my mind off the thesis, and for their tolerance while I finished the details

to Steve, Alexander and Freddie for love, distraction and a proper sense of priorities

to Syan and the staff of Balliol Day Nursery for making it all possible

Chapter One

Introduction

Health care rationing is not a new subject; there is a broad consensus that it has always existed both within the NHS and before its inception. However, much of this rationing has been achieved through mechanisms such as waiting lists, global budget setting, unequal distribution of resources between specialties and geographical areas, and individual clinical decisions about whether or not to treat a particular patient. These mechanisms mean that the nature of the choices being made between individuals and groups of patients is kept largely concealed or implicit, even if it is evident that some rationing is taking place.

More recently, there has been pressure from a number of quarters to bring health care rationing out into the open and make it more explicit and systematic. This pressure is being experienced internationally; countries such as New Zealand, the Netherlands, Sweden and the State of Oregon in the US have all been involved in attempts to clarify the priority-setting process by which limited funds are allocated. The question of explicit rationing has generated a considerable amount of literature; it has also resulted in heated debate in academic, managerial, medical and political circles. Health authorities which have made explicit decisions often attract enormous media attention, as in the case of Cambridge and Huntingdon which refused to pay for further leukaemia treatment for Child B. There is a fundamental public concern that such decisions threaten the comprehensive nature of the NHS, as well as a distaste for the ethical consequences in individual cases.

In these circumstances it comes as no surprise that even the meaning of explicit rationing is disputed, let alone whether it should be done, and if so how and by whom. These issues remain unresolved and a discussion of them will form part of this thesis. However, they have been the subject of some academic enquiry already. In contrast, an issue which has received very little attention to date is the relationship between explicit rationing and the quasi-market reforms within the NHS introduced in April 1991. It is this issue which is explored in this thesis.

The thesis is divided into three parts. Part One, 'Theory and International Experience', begins with a review of theories of resource allocation and their application to health care. This chapter takes as its starting point the argument that all resources are ultimately scarce and must be rationed in some form. The question is therefore not *whether* a particular health care system rations, but *how* it rations. It examines different methods of resource allocation, including markets, with particular reference to Williamson's work on 'organisational failures' and Bartlett's application of this to quasi-markets in health care (Williamson, 1975 and 1985; Bartlett, 1991)

From the theoretical analysis of the potential relationship between quasi-markets and explicit rationing, the following competing hypotheses emerge:

i. Markets are one possible system for allocating scarce resources. The process of contract specification inherent in a complex quasi-market system constitutes a form of rationing, which formalises resource allocation decisions and makes them more explicit than they would be in a hierarchical system. The theory that lies behind the move from hierarchy to quasi-market in health care would thus predict that rationing in the NHS will become more explicit as a result of the introduction of the internal market.

ii. An alternative view is that in the complex context of the NHS the quasi-market may fail to produce clear contracts and unambiguous allocations, for two main reasons: the contracting process is likely to be inefficient, because of high transaction costs, and the explicitness of clear contracts is likely to carry high political and ethical costs. This view predicts either that the contracting process will be conducted in form only, with a continuation of more implicit rationing mechanisms, or that it will be abandoned altogether.

iii. A third possibility could emerge; this is that other pressures in favour of greater explicitness - particularly resource pressures, outcomes research and the Patient's Charter - are so strong that explicit rationing will continue to increase, either through the costly quasi-market contracting system, or through some other mechanism if contracting fails to provide an adequate means of achieving it or is abandoned.

These hypotheses provide the framework for fieldwork carried out for this thesis.

Chapter Three returns to the history of rationing in health care and the current debate about what form it should take. Addressing he two questions 'how much' and 'for whom?' is fundamental to the management of public services, although the use of the word 'rationing' to describe this process is politically sensitive and does not go unchallenged. Examples are given from pre-NHS days, such as the Victorian principle of less eligibility and deserving/undeserving poor and the National Insurance Act of 1911. Examples from the post 1948 NHS, but before the market reforms, include charges and budget capping, waiting lists, rationing by specialty or care group, geographical rationing, the operation of clinical freedom to ration implicitly, explicit denial, age limits for screening programmes and the use of QALYs (quality adjusted life-years) to make priority resource decisions.

The analysis of the current debate examines whether different commentators believe a) rationing health care is necessary and b) even if it is, does it have to be explicit? The impact of evidence-based health care will be assessed - is it an alternative to rationing or a sensible means of achieving it? Is political reluctance to undertake or even acknowledge rationing evolving slowly towards acceptance and encouragement of purchasers?

The very fact that there is such disagreement about what rationing is and whether to be explicit could be seen as reluctance to accept the logic of contracting in some quarters, because of concern about its high political and ethical costs. On the other hand, there is a very strong movement in support of explicitness in priority setting, despite frequent reluctance to call this rationing.

Chapter Four summarises international experience of both rationing and quasi-market experimentation, and concludes that it is tax-funded systems which are especially prone to the resource pressures which give rise to both explicit rationing, as a way to handle scarcity, and quasi-markets, as a way of trying to improve efficiency and value for money. Quasi-markets also provide their own impetus towards explicitness, as governments experimenting with market forces want to protect consumers against possible inequities and therefore seek to define a basic minimum package of services. This assessment of international experience sets the context for an examination of the UK reforms.

Part Two of the thesis reports fieldwork findings. The methods used are described in Chapter Five.

Chapter Six summarises early experiences of explicit rationing and the internal market. It begins with a pilot fieldwork survey of eight health authority purchasers carried out in 1994. It also includes a summary of evidence presented to the health select committee investigation of purchasing and priority setting (House of Commons, 1994a, 1994b), and compares these findings with the analyses of purchasing plans carried out over several years by Klein's team at Bath University (Klein and Redmayne, 1992; Redmayne, Klein and Day, 1993; Redmayne, 1995; Redmayne, 1996). This early fieldwork concluded that explicit rationing was increasing, although it was still relatively marginal, and that the market was one important cause of this, but not the only one.

Following the pilot survey, three new health authority case studies were designed and carried out during 1996-97, to explore in more depth the complex relationship between the market and explicit rationing. The evidence from these case studies is presented in Chapters Seven and Eight. This examines what evidence there is of growing explicitness in rationing in the three districts concerned. It places particular emphasis on the contracting round for 1997/98 contracts, but also looks back at previous years' rounds. Interviewees' perceptions of the driving forces in favour of explicitness are presented, as well as their views of the obstacles they face in being explicit and the pressures in favour of remaining implicit. Respondents were also asked to compare how resource allocation decisions were made before the market was introduced with current decision-making.

In Part Three, Chapter Nine draws together the evidence in a discussion of the relationship between explicit rationing, the NHS quasi-market and other factors. The thesis concludes with an assessment of the implications of the White Paper (Secretary of State for Health, 1997) in Chapter Ten.

Part One

Theory and International Experience

Chapter Two

Theories of resource allocation

The line of argument to be followed in this chapter can be summarised as follows:

1. All forms of resource allocation are rationing
2. The question is therefore not *whether* a system does or does not ration, but *how* it rations, and to what extent it does so in an efficient and sustainable way
3. In orthodox economic theory, the market (which rations by price and willingness to pay) is the most efficient method
4. Even an efficient market may produce socially unsustainable results, particularly if the outcome is judged to be unacceptably inequitable
5. A pure market in health care is likely to fail on efficiency grounds, as well as producing high levels of inequity
6. A quasi-market is intended to overcome some of the problems of a pure market (particularly imperfect information and inequity), using tax-funded purchasing by an informed agency acting on behalf of consumers, whilst using supply-side competition to improve value for money and consumer responsiveness
7. In a complex environment such as health, a quasi-market can operate only through written contracts, which can be expected to formalise resource allocation decisions and make them more explicit than they would be in a hierarchical system
8. However, such contracts may be unsustainable, for two main reasons: achieving compliance may involve high transaction costs, and their explicitness may carry high political and ethical costs. This could result in the contracting process being carried out in form only, with a continuation of more implicit rationing mechanisms, or in it being abandoned altogether.
9. A third possibility is that other pressures in favour of greater explicitness (particularly rising expenditure, improved understanding of outcomes and the Patient's Charter) may be so strong that explicit rationing will continue to increase despite these costs, either through a quasi-market contracting process or through some other mechanism if the quasi-market is abandoned.

General theories of resource allocation

Economics is 'a science which studies human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins, 1932, p.15).

'Economics ... centres upon the issue of scarcity ... It starts from the fundamental proposition that resources are too few to satisfy all the wants of mankind' (Culyer, 1985, p.1).

The question of how scarce resources will be allocated is fundamental to human society. (As suggested by Culyer, the word 'scarce' here has a specific meaning, which differs from the popular meaning of 'rare'; to the economist, it means that all goods, however plentiful, are ultimately in limited supply. Therefore not all demands can be met, and to choose one option is to rule out another).

Economists have tended to use the word rationing to mean the allocation of scarce resources by means other than the market. According to this view, it is merely *one form* of resource allocation, generally associated with government intervention and clearly stated maximum consumption levels. More broadly, however, it can be argued that rationing and sharing out limited resources are the same thing, so that *all forms* of resource allocation, whether through the market or not, are a means of achieving rationing. Rationing in this sense will sometimes be explicit and sometimes implicit or indirect.

It is this latter definition of rationing which will form the basis for this thesis. In looking at different resource allocation theories, therefore, I shall not be questioning *whether* goods are rationed but rather *how* they are rationed and to what extent the method of rationing is efficient and sustainable in particular circumstances.

There is no doubt, however, that the popular image of rationing reflects the first of the two definitions, and also tends to equate scarcity with rarity. It is indelibly associated with crisis, especially war-time, when basics such as food, clothing and fuel are in such short supply that government regulation is required to ensure a fair distribution and prevent over-consumption by those with a greater ability to pay.

In his work on the structures of kinship, Lévi-Strauss demonstrates that such action is part of a much older tradition. He describes primitive societies' rules about incest and marriage as an intervention designed to cope with 'the insufficiency or the risky distribution of a valuable of fundamental importance', in this particular case the 'distribution of women', but equally applicable to valuables such as food or land. He continues:

> 'Certain forms of rationing are new to our society and arouse surprise in minds cast in the traditions of economic liberalism. Thus we are prompted to see collective intervention, when it affects commodities vital to our way of life, as a bold and somewhat scandalous innovation. Because the control of distribution and consumption affects gasoline, we readily think that its formulation was only contemporaneous with the motor-car. But nothing is less true. 'The system of the scarce product' constitutes an extremely general model. In this and many other cases these periods of crisis, to which until recently our society was so unaccustomed, merely re-establish, in a crucial form, a state of affairs regarded as virtually normal in primitive society. Thus, 'the system of the scarce product', as expressed in collective measures of control, is much less an innovation, due to modern conditions of warfare and the worldwide nature of our economy, than the resurgence of a set of procedures which are familiar to primitive societies and necessary to the group if its coherence is not to be continually compromised' (Lévi-Strauss, 1969, p.32).

There are several points to note here. Firstly, that rationing is not simply a modern phenomenon, and secondly, that collective intervention to achieve it is a broader concept than government regulation as we know it. However, it is not entirely clear whether Lévi-Strauss sees all forms of resource allocation as rationing; he seems to tend towards the definition of scarcity as rarity, in implying that until recently the western industrialised world has not really had to confront the problem of scarce resources, and suggests that the liberal market tradition does not itself constitute rationing. Rather he equates rationing with collective intervention of some sort, in the face of crisis.

It is true that when the extreme scarcity of war-time is over, and goods are more readily available, the government can relinquish control over distribution and revert to allowing the market to determine how resources will be allocated. However, this does not mean that scarcity according to the economist's definition has been eliminated. In these circumstances, the market rations access to goods by willingness to pay, which itself is made up of ability to pay and the exercise of preferences - those with more money can afford to buy more expensive types of food, buy more clothes, or use fuel more freely, if they choose to do so. The oil crisis in the mid-1970s provides a dramatic example of how restricted supply can lead to rationing by price at international level. If that leads to unsustainable general inflation, then administrative rationing may follow, as it did for short periods in many countries as a result of the oil crisis.

Calabresi and Bobbitt (1978) argue that society deceives itself (deliberately, in many cases) if it thinks that markets and all other methods of resource allocation are not in the end about rationing. They examine what they call first and second-order determinations or decisions about 'the scarce good'; first-order decisions concern how much of the scarce good will be produced, within the limits of natural scarcity, whilst second-order decisions determine who will get what is made. They emphasise therefore that scarcity is not only about the natural availability of goods, but is also frequently self-imposed by 'the decision by society that it is not prepared to forgo other goods and benefits in a number sufficient to remove the scarcity' (p.22). A second-order outcome, such as who dies in a road accident, may appear to be an unavoidable and fatal misfortune, but it can in fact conceal an original first-order determination to limit the amount of money spent on road safety, which has led to the accident. This appearance of random bad luck protects society from the implications of the tragic choice it has made.

Calabresi and Bobbitt identify four main allocation processes: markets, accountable political approaches (e.g. war-time rationing), lotteries and the customary or evolutionary approach. The latter process is in some ways a non-process, in that it allows allocations to evolve with no explicit selection - an example given is that we have intuitively decided an acceptable level of child-bearing, and that a system of incentives has evolved which reinforces this, although the example is not without complications, such as the effect of market forces on how many children individuals can afford. All of these processes can be modified and used in combination - for instance, a first-order decision may be political, whilst the second-order allocation operates through a market. In the road safety example above, the first-order decision is political, whilst the second order outcome is determined by lottery. This can enable society to hold conflicting values simultaneously at the two different levels.

None of these approaches can evade the necessity of first and second-order decisions, although some are better than others at concealing the real nature of what is going on. An accountable political process is often transparent in its rationing activities, whereas markets and lotteries, by presenting resource allocation as the result of 'thousands of independent, atomistic actions' (p.29) or of chance, seem to 'absolve societies from responsibility for outcomes' (p.31). On the other hand, the apparent randomness of market outcomes may also become unacceptable; a market which allows the rich to buy kidney dialysis and the poor to die will lay bare tensions between conflicting values which society may not be able to tolerate. Society is therefore constantly seeking to recast its allocation decisions in ways which conceal these tensions and apparently avoid tragic choices, until the conflict becomes evident again and a new shift takes place.

Calabresi and Bobbitt do not claim that any one of these allocation processes is more efficient than the others; they seek rather to demonstrate that society needs all of them. Using one approach and then rejecting it in favour of another is a rational way of preserving essential but conflicting values and keeping in check the potentially destructive impact on society of openly acknowledging tragic choices. They therefore cannot regard the market as a uniquely successful way of distributing resources - quite apart from its frequent failures ('the arthritic invisible hand' (p.83), as they describe it), even a supposedly perfect market will produce outcomes that society cannot accept.

The government's choice of a form of market system for health care in this country was at least partly founded on the more orthodox economic view that markets are indeed the most efficient way of allocating resources, provided they are working perfectly. (The government's motives in setting up the NHS market are discussed in more detail later in this chapter). It should be reiterated at this point that achieving efficiency in distribution does not mean avoiding rationing. Even if one rejects the Calabresi and Bobbitt line that the market involves more deliberate choices than we might care to believe, in favour of the Hayek view that it is an impersonal force which creates its own 'spontaneous order' (Hayek, 1976, p.109), it still results in the rationing of scarce resources, albeit in theoretically the most efficient way possible.

In practice, of course, markets do not always work perfectly. According to the 'Invisible Hand Theorem', markets will be efficient if, and only if, a number of assumptions hold. These are: perfect competition, no market failures and perfect information (Barr, 1993). The difficulties in ensuring a perfect market in health care will be dealt with later in this chapter. In general terms, when a market does not work perfectly there are a number of possible options to correct the problems which arise. These options are identified by Barr as regulation (such as compulsory insurance), financial intervention (in the form of price subsidies or taxes, for example on pollution) or public production (for instance the NHS), all of which intervene in the mechanism of the market to restore efficiency. If, however, none of the standard assumptions fail, there can be no justification on efficiency grounds for intervention.

Barr stresses that, although equal power for all participants is a pre-condition for perfect competition, this principle 'is not violated if some individuals have higher incomes than others' (p.81). The market will still achieve efficiency (in the sense of maximisation of social welfare), no matter how unequal the distribution of income. If we are concerned that some people are therefore not able to take part in the market as fully as others, this is a question of social injustice, not of inefficiency, and can be resolved by redistribution of income, leaving the market mechanism itself intact. The extent to which redistribution is supported by society depends on the prevailing view of what constitutes social justice, and on the relative weight accorded to it. (It should be noted that a degree of inequity in distribution may be tolerated by society for the sake of preserving efficiency, and the point at which intervention to redress inequity is judged necessary will differ from society to society. The United States, for example, tolerates higher levels of inequity in health care arising from a market distribution than would probably be acceptable in Britain).

The basic problem is that the *maximisation* of welfare may well conflict with what is regarded as a just *distribution* of welfare. This underlines Calabresi and Bobbitt's point that even a perfect market can produce unacceptable results. For instance, rationing access to housing via the market could be efficient in the sense of maximising the total quantity and quality of housing available, but could create socially unacceptable levels of homelessness amongst the poorest.

A utilitarian response is possible, by arguing that it may satisfy the quasi-altruistic preferences of wealthier people to sacrifice some of their income to alleviate poverty, or in this case specifically homelessness. This could make a degree of non-market redistribution consistent with efficient maximisation, although significantly the motive is not justice but an increase in utility.

An alternative approach is to argue that efficiency is not itself a primary aim, but is secondary to agreed policy aims; in other words, a particular system of allocating resources is only efficient to the extent that it moves society towards its primary social objectives (Le Grand, 1991). This means that if the outcome of the market is regarded as unacceptably unjust, it cannot by definition be efficient. In practice, Le Grand tends to agree with Barr that the best way to rectify this position is often by cash transfers to facilitate more equitable participation in the market, rather than intervening in the market itself, but this is nonetheless a response to market *failure*. Thus the housing market becomes inefficient ***because*** it is unjust, as opposed to Barr's view that it is efficient ***but*** unjust.

Hayek disagrees strongly that redistribution is desirable in the interests of social justice, and believes that it cannot take place without affecting the operation of the market. He describes social justice as a mirage, the pursuit of which endangers the overriding goal of personal liberty. He argues only subsistence levels of welfare redistribution can be justified; otherwise the market should hold sway, and will spontaneously allocate resources in the best possible way, although its very spontaneity makes it a nonsense to describe the outcome as 'just' or 'unjust'. Allowing anything more than minimal redistribution to take place outside the market will prevent it from working efficiently, notably by creating perverse incentives. For example, someone receiving benefits while not working may find it financially disadvantageous to take up work because of high marginal tax rates. Thus, he suggests, even cash transfer redistribution *does* interfere with the mechanism of the market and is a *cause* of market failure, as well as posing a threat to liberty because of the centralisation of power it requires (Hayek, 1976).

Libertarians such as Hayek and Friedman, whilst accepting a limited role for government in ensuring perfect market conditions are maintained, are also concerned about the effects of traditional responses to market failure such as regulation and public production (Hayek, 1944; Friedman, 1962). The results of government action, far from correcting failure, are often a cause of it, and are almost bound to be inefficient. (An illustration suggested by Barr (1993, p.103) is that 'if there were a competitive market for health care, people would acquire better information, in part because market institutions would arise to supply it.' Suppressing competition simply makes the problem of imperfect information a self-fulfilling prophecy).

This forms part of a wider analysis of public choice theory and government failure. At the level of electoral politics, competition for votes may affect the way a government manages the economy, including public services; seeking to maximise popularity may take precedence over pure economic efficiency, especially as an election approaches. Individual ministers also have to strike a balance between action to ensure re-election and the desire to maximise the power of their own departments (Mueller, 1989). Bureaucrats may have personal or professional interests which conflict both with ministerial ambitions and with the public interest, either in the form of maximising their departmental budgets (Niskanen, 1971) or of diffusing and distancing themselves from uncomfortably hard decisions (Dunleavy, 1991). These factors can result in inefficient and self-perpetuating expansion of government activity, and necessitate caution in assuming government action will always correct market failures. However, as Barr (1993) points out, assuming that the private sector will therefore automatically be more efficient in raising social welfare is equally dangerous.

A specific approach to market failures with particular relevance for the NHS market is found in Williamson's work on organisational hierarchies and markets as alternative allocative mechanisms. Rather than looking at government intervention as the response to market failures, Williamson seeks to explain how firms react to market failure by creating an internal hierarchical organisation instead (or alternatively how they react to internal organisational failure by shifting transactions to the market-place, although this is not the main focus of his work). He summarises his approach as follows:

> '(1) Markets and firms are alternative instruments for completing a related set of transactions; (2) whether a set of transactions ought to be executed across markets or within a firm depends on the relative efficiency of each mode; (3) the costs of writing and executing complex contracts across a market *vary with the*

> *characteristics of the human decision makers who are involved with the transaction on the one hand, and the objective properties of the market on the other*; and (4) although the human and environmental factors that impede exchanges between firms (across a market) manifest themselves somewhat differently within the firm, the same set of factors apply to both. A symmetrical analysis of trading thus requires that we acknowledge the transactional limits of internal organisation as well as the sources of market failure.' (Williamson, 1975, p.8-9)

The question of contracting and the associated transaction costs are central to Williamson's 'organisational failures framework'. Contracts are an essential component of all systems of exchange, although in many cases the nature of the transaction is straightforward enough for the contract to remain implicit or verbal only. Williamson, on the other hand, identifies circumstances in which explicit written contracts are required to underpin the exchange between market agents but in which the transaction costs are so high that the system becomes inefficient. Key factors include 'bounded rationality', uncertainty/complexity, opportunism and small numbers exchange. On their own, these factors may not be significant enough to cause a market to fail, but in combination with each other they can have a powerful effect.

Bounded rationality refers to the fact that human beings, whilst trying to behave rationally, have only limited powers of receiving, storing, retrieving and processing information and limited linguistic abilities to 'articulate their knowledge or feelings ... in ways which permit them to be understood by others' (p.22). Bounded rationality is not a problem when the environment or information to be mastered is relatively simple, but combined with conditions of complexity and uncertainty it becomes very difficult to assemble information about all possible future contingencies. In these circumstances, it is not feasible to rely on an implicit or verbal contract, as in simple market exchanges. However, specifying an indisputably clear contingent claims contract (i.e. where the outcome is contingent upon an uncertain environment) becomes virtually impossible or at least prohibitively costly. This is clearly a variant on the theme of imperfect information resulting in market failure.

Internal organisation within a firm can help overcome this difficulty because it 'permits the parties to deal with uncertainty/complexity in an adaptive, sequential fashion ... Rather than specifying the decision tree exhaustively in advance, and deriving the corresponding contingent prices, events are permitted to unfold and attention is restricted to only the actual rather than all possible outcomes' (p.25).

Opportunism and small numbers exchange also interact to create high costs. If it proves impossible to put together a detailed contingent claims contract, an alternative might be to agree an incomplete contract, which does not spell out every detail and includes a clause to the effect that both parties should co-operate 'in a joint-profit maximising way, when unforeseen contingencies develop' (p.91). However, this offers a temptation to behave opportunistically (which Williamson defines as 'self-interest seeking with guile', p.26) and can manifest itself as 'selective or distorted information disclosure or ... self-disbelieved promises regarding future conduct' (p.26) in the interests of gaining an advantage over the other party. This can lead to unanticipated price increases or inefficient performance; trying to combat opportunism by more elaborate contract monitoring is also costly. The problem is even worse when combined with uncertainty, which leads to what Williamson calls 'information impactedness' (p.31) - this is when one of the parties has much better information than the other, and the second party can neither achieve information parity without great cost nor rely on the first party to reveal the information.

Opportunism would not in itself be a very great problem in a competitive market with plenty of other competitors, as companies behaving opportunistically would soon find themselves unable to renew their contracts. In a small market, however, especially where new entrants to the market would need to invest heavily in specialised equipment and skills, there is little scope for countering opportunism. In a subsequent work, Williamson elaborates on this problem of asset specificity in combination with bounded rationality and opportunism, and argues that it is a crucial obstacle to efficient competition (Williamson, 1985). A more promising way to avoid this kind of domination of a complex market is to opt for internal hierarchical organisation instead.

Application of resource allocation theories to health care

Health care is a scarce resource which has to be allocated in the same way as any other good. Its scarcity derives from both the amount that society and individuals are willing to spend on it, and from the availability of sufficient professional skills, knowledge and equipment to deliver it, which is only partly determined by the amount of money available.

If, as argued in the previous section, rationing and sharing out resources are the same thing, all forms of resource allocation in health care constitute a means of achieving rationing, whether implicitly or explicitly. Denying or limiting access to potentially life-saving interventions is a prime example of a 'tragic choice', making the selection of rationing method in health care a more contentious matter than for most other goods.

The chief debate in health care resource allocation is whether to opt for a market system or what Calabresi and Bobbitt would call an accountable political system, although there may also be an element of lottery involved in either. The NHS reforms reflect the current government's preference for moving away from purely political systems towards a market allocation of resources, although still within a regulated political framework. (This corresponds with Calabresi and Bobbitt's view that it is possible to combine different allocation methods simultaneously at different levels).

Will a market system in health care in fact be an efficient and socially acceptable way of rationing resources. Before addressing this question, it is worth identifying what efficiency means in the health care context.

Barr (1993) identifies three main types of efficiency: efficiency in production, in product mix and in consumption. In health care, efficiency in production is the maximum amount of care that can be provided within available resources. Efficiency in product mix means offering the optimal mixture of types of treatment to meet patients' needs, given what is known about medical technology. Efficiency in consumption implies equilibrium between the resources we are prepared to spend on health and the amount of treatment we need; one of the most common objections to publicly provided health services is that they distort this relationship between willingness to pay and consumption and encourage excessive use of services. Again, it should be noted that this concept of efficiency makes no comment on the justice or otherwise of the *distribution* of services.

Imperfect markets

Barr states that 'health care conforms only minimally with the assumptions necessary for market efficiency' (p.332), that is to say perfect competition, no market failures and perfect information. Furthermore, the complexity of health care and the subjectivity of questions about the value and quality of life make defining and measuring efficiency extremely difficult.

Perhaps the greatest problem for markets in health care is imperfect information, which significantly affects efficiency in consumption. This may occur in any of the following ways:

- people may not realise they are ill and need treatment
- health technology is highly specialised; even doctors struggle to keep up with new developments, let alone individual lay people
- this makes consumers reliant on the supplier for information; they lack the knowledge to judge between advice from different professionals, and in cases of acute illness do not have time to shop around
- illness is an emotional and frightening subject, compromising the consumer's ability to make rational judgements
- patients have little knowledge of appropriate prices for each treatment
- knowledge about future health care needs is generally not available

Insurance is only a partial solution to the uncertainty of future needs, and has its own problems, particularly the inability to cover all risks and the danger of over-consumption resulting from third party payment. The US provides a salutary example of how costs can spiral under an insurance-based system, especially where insurers can pass on higher premiums to employers rather than individuals - almost a system of fourth party payment, one might say. Third party payment may also create incentives to over-produce, as illustrated by the duplication of high technology and low occupancy rates found in hospitals in the US.

There are also problems with maintaining perfect competition in health care - doctors exercise a monopoly over the practice of medicine and are more likely to seek the approval of their colleagues than their consumers, argues Barr. As well as affecting efficiency of production and regulating entry to the market by new practitioners, this is likely to result in an inefficient product mix, as doctors gravitate towards more glamorous, high technology forms of care. At a broader level, hospitals require significant advance investment in skills and equipment. This can make it difficult for new competitors to gain entry to the market and succeed in operating efficiently. Perfect competition also relies on equal power between producers and consumers; this in turn relies on consumers having access to perfect information, which has already been seen to be lacking.

Barr (1993) does not consider market failures related to externalities a significant efficiency problem in health care; although they do exist (e.g. controlling communicable diseases through vaccination), they can be solved by regulation without recourse to public production (e.g. compulsory vaccination). It could also be argued that general good health amongst the population as a whole produces external benefits, as it ensures people remain economically productive tax-payers rather than expensive social security consumers. This argument is complicated by the role of other factors such as good housing, sanitation, education or socio-economic status in producing good health, with health care itself playing a relatively small part. (See, for example, Townsend and Davidson, 1982; Office of Health Economics, 1985).

Although Barr is not convinced non-market alternatives provide a better solution to these problems, he offers the following view: 'Information failures and the lack of competition justify regulation: the externality, coupled with major insurance problems, may justify public funding; and a strong (though not overriding) argument for public production and allocation arises out of the serious problems with both consumer information and private insurance' (p.86).

McGuire *et al.* (1994) have also produced an extensive economic analysis of market failures in health care. They emphasise the problems involved in trying to measure output, define efficiency and monitor performance, and suggest these require at least the regulation of finance and provision. They are critical of the belief that institutional reform of the NHS can avoid the problem of trying to ration services to match available funding, and conclude there is no simple way of defining the correct level of health expenditure.

From the above discussion, it can be concluded that pure markets are unlikely to allocate health care efficiently. The inequity of a market distribution has already been mentioned. Some societies may be able to tolerate a situation where money determines the extent to which preferences for health care can be exercised, but the fact that health care affects such basic aspects of life as survival, pain and physical capacities makes inequity in its distribution a contentious and emotive matter. Even the US does not allow the market to take its full natural course and intervenes on behalf of the poorest in society through Medicaid, although this still leaves millions of low-paid uninsured people to struggle with market forces.

The inequity arising from choices being determined by ability to pay combines with the inefficiency arising from those choices being poorly informed and subject to monopolistic provision to make pure markets an unsuccessful means of rationing health care.

Quasi-markets

The Conservative government to some extent acknowledged the problems that exist in health care markets, both by opting to retain public funding and public production and by describing the system it established in 1991 as a 'managed market'. There has been significant intervention to regulate the NHS market, and ministers have strongly denied that NHS Trusts represented the first stage of wholesale privatisation as their critics argued.

There are two main components of quasi-markets which seek to reap the potential benefits of a market system whilst retaining the control of an accountable political process: these components are the separate identification of purchasers and providers, and the creation of competition between providers.

Glennerster and Le Grand (1994) have examined possible explanations for the government's choice of quasi-markets in several areas of social services, including education as well as health, and conclude that the main driving force was not an ideological commitment to markets, a desire to cut public expenditure and state intervention, or to privatise welfare, although all these elements were undoubtedly present. They conclude rather that the government was trying to find a way of making services more responsive to rising consumer expectations without increasing taxes or abandoning the popularly supported principle of public funding. Having exhausted the potential for greater managerial efficiency during the 1980s, the government then turned to competition as a way to squeeze out greater efficiency and more responsiveness to consumer requirements. The White Paper 'Working for Patients' (Secretaries of State etc., 1989) in which the reforms were first announced is indeed shot through with the rhetoric of improving patient choice.

For competition to work effectively there must be an identifiable purchaser who responds to the competing offers made by providers. A market-place with no customers is simply a contradiction in terms. Given the inefficiency and inequity of health care markets where individuals are the consumers, the solution of the quasi-market is to substitute purchasing by an agency, in the form of either health authorities or GP fundholders.

Quasi-markets only function efficiently if certain conditions are met, like real markets (Glennerster, 1992); quasi-markets in social services face a number of common difficulties, and the NHS market has its own specific problems, the result of which may be called 'quasi-market failure'. These problems fall into two main categories: imperfect competition and imperfect information. I shall deal with imperfect competition first, before moving on to imperfect information, which is more directly relevant to the quasi-market's impact on rationing, although issues of competition also have some relevance.

Perfect competition in the NHS is threatened by a number of monopolies and monopsonies, in addition to the monopoly of the medical profession as a whole. In many areas of the country, following years of centralised planning, there is only one possible hospital or community service to turn to. District general hospitals were developed to ensure local accessibility and a more even distribution of facilities around the country.

The evidence about whether people are prepared to travel long distances to obtain better health care is conflicting. Mahon *et al.* (1994) found that most people do not wish to be far from family and friends whilst in hospital, although responses differed by specialty. Middle class patients have been prepared to travel to famous London hospitals to obtain better quality care, but this has been a limited phenomenon both socially and geographically and is illustrates the unequal distribution of information about services within the population. Recent waiting list initiatives have been used to send people to more distant hospitals to get their operation performed more quickly, and this may reduce resistance to the idea amongst the population more widely, but it seems unlikely to become the norm. It would also not be practicable for emergency needs. To some extent, therefore, competition could actually conflict with patient preferences for local care, even more so if it resulted in the closure of some uneconomic but locally valued services.

In highly specialised services, it has been argued that local accessibility is less important than maintaining high quality services which concentrate expertise and patients in one place. This creates a different kind of geographical monopoly. (See, for example, the recent proposals for rationalising cancer services, Department of Health, 1994). However, recent work by the NHS Centre for Reviews and Dissemination (1995) has queried this supposed relationship between volume and quality.

These monopolistic obstacles to competition can be overcome in conurbations with a large number of health care institutions in one area. However, the majority of NHS consumers do not live in such areas.

In addition to the problem of provider monopolies, there is the problem of purchaser monopolies and monopsonies. Unless an individual is registered with a GP fundholding practice, all health care services are currently purchased on his or her behalf by the local health authority. Unlike an insurance-based system, there is little choice of purchasers; the only possible choices for an individual consumer would be either to move to another health authority, or to change GP to a fundholding practice.

A health authority monopsony occurs when the authority is the only major purchaser of services from providers in that area. This could mean that providers have little choice but to fall in with purchaser requirements. Even where GP fundholding has become widespread, undermining purchasing monopsonies for the elective and out-patient services which fundholders are able to buy, the monopsony may persist in other services which do not form part of fund-holding, such as emergency services or very expensive procedures, unless a total purchasing pilot has been operating.

In the above discussion of imperfect competition, the question of consumer or patient choice has already cropped up more than once. The problem in arguing that a quasi-market will improve patient choice is that in a quasi-market the consumer and the purchaser are not the same person. This in turn is linked to imperfect information as a major cause of market failure in health care. Where the inequality of market knowledge between consumer and provider is too great, quasi-markets offer an *alternative* to individual purchasing by introducing supposedly more informed purchasing agencies, namely health authorities and GP fundholders. Thus quasi-markets are not in themselves designed to enhance individual consumer choice. Given the lack of competition between purchasers, the extent to which services become more responsive to consumer preference is dependent upon how seriously committed purchasers are to acting on consumer views. Realistically, even this must be largely at the level of collective views and preferences, rather than individual choices. This is particularly true at health authority level, where the scale of purchasing means the majority of decisions are at group or specialty level; GP fundholders may reflect individual patient wishes more accurately, although even they must ultimately manage their budget for the collective benefit of all the practice's patients.

Quasi-market purchasing is similar to social insurance models, in trying to pool risk and uncertainty about future individual needs. Like any third party payment system, however, it disrupts the relationship between willingness to pay and consumption. Some might argue health authorities have even less control over utilisation than insurance companies, particularly as referrals are controlled by GPs. However, health authorities' inability to pass on spiralling costs in the form of increased insurance premiums does create some budgetary incentive to restrain over-consumption. Ate least GP fundholding offers congruence between the purchaser and the referrer, if not the patient.

Does the introduction of agency purchasing really do anything to solve the problem of imperfect information? To some extent it does. Obviously, health authorities can never replicate the understanding each individual has about their own health or illness, but they are in many ways better informed than an individual member of the public can be. They can predict levels of different types of illness in their local population, build up knowledge of the variety of conditions and potential treatments that exist, and gather information about potential providers.

However, there is still a huge gap between their level of knowledge and the goal of perfect information. There is only limited understanding of how to measure need, outcomes and the effectiveness of different treatments for the population as a whole. Health authorities do not know enough about the costs and quality of services offered by different providers - partly because of the difficulties of measurement, and partly because of the difficulties of gaining access to information held by providers. This latter point illustrates the combination of opportunism and small numbers exchange described by Williamson (1975), demonstrating that imperfect competition also has a bearing on imperfect information. The relevance of Williamson's work will be discussed in more detail shortly.

GP fundholders have some information advantages over health authorities, stemming from their closeness to their practice populations, which enables them to assess need and evaluate treatments provided to individuals more effectively, and from their experience as referrers to different hospitals. However, they still have to contend with the more general difficulties in measuring need and outcomes, and perhaps have less experience of costing hospital services than health authorities. The Audit Commission (1996) has criticised the majority of GP fundholders for making purchasing decisions which are not informed by the latest research on clinical effectiveness; health authorities may stand a better chance of keeping up with research through their public health departments than an individual GP.

Fundholders may be able to overcome provider opportunism more successfully than health authorities through their inside knowledge of hospital medicine and by their ability to be more flexible about changing contracts away from an unsatisfactory provider, especially given the elective nature of most of the services. This is supported by the finding that 40% of GP fundholders felt the reforms had increased their freedom of choice in referral, compared to only 5% of non-fundholding GPs (Mahon *et al.*, 1994). None of the fundholders felt their freedom of referral had been reduced, compared to 17% of non-fundholders.

In summary, purchasing agencies may go some way to reduce the problems of imperfect information, but the complexity and uncertainty of health care is such that only limited success is possible. The continuing impact of imperfect information leads on to the issue of contracts.

Contracts

The very complexity and uncertainty of health care means that market exchange in this field cannot easily be left to implicit or verbal contracts. The quasi-market therefore relies on written contracts.

Theoretically, the specification of contracts between purchaser and provider should formalise resource allocation decisions and make them more explicit than they would be in a hierarchical system, bringing to the surface previously implicit or covert choices. The discretion for providers to decide how to spend their budgets has been considerably circumscribed, and the market culture has encouraged providers to view their workload as fixed and to believe that they should take on more patients only if additional funding can also be agreed. Purchasers are under pressure to carry out needs assessments for their population and make clearer choices about what type and quantity of treatments they need to buy (Hunter, 1993a).

There are three levels of rationing decision which could be made more explicit by contracting: *what*, *how much* and *for whom*. The corollary of specifying more clearly what will be provided is that one must also specify what, if anything, will not be provided; examples of treatments which have been excluded from contracts altogether by some health authorities include tattoo removal, sterilisation reversal and *in vitro* fertilisation (IVF). Having once made clear what will or will not be provided, contracts should theoretically specify how much of particular services or treatments will be provided. In practice, the level of detail to which this aspect of contracting is pursued could vary from a statement of the total volume of in-patient episodes expected in one hospital to a break-down of the volume of episodes in each specialty or even the number of different operations or courses of treatment for patients with different conditions. Together, these two levels ('what' and 'how much') constitute what Calabresi and Bobbitt would describe as a first-order decision (Calabresi and Bobbitt, 1978).

The second-order decision is who will receive those treatments which are purchased. In many contracts, this may not be specified and the selection of recipients will be left to frontline clinical staff, very much as it was before the introduction of the market. However, some authorities have started using contracts to specify how limited numbers of treatments should be allocated. Most frequently this is based on effectiveness evidence (e.g. age limits for IVF) or on the grounds that not treating will result in significant psychological morbidity (e.g. exceptions made for tattoo removal). Occasionally social criteria are applied (e.g. IVF only for married couples with no children by a previous marriage). All of these examples will be discussed in greater depth later on.

When cost-per-case contracts are used, the first and second order decisions are simultaneous - this is more common in GP fundholding, although it is used by health authorities, particularly for extra contractual referrals and for some low volume specialised services.

Although the logic of quasi-market contracting should produce more explicit rationing decisions, two potential problems may prevent this logic from being fully realised. Firstly, contracts may be inefficient, in that they result in high transaction costs, or even impossible to draw up. Secondly, explicitness carries high political and ethical costs. This could result in the contracting process being carried out in form only, with a continuation of more implicit rationing mechanisms. These two problem areas will be examined in turn.

The efficiency of contracting

Bartlett (1991) has used Williamson's organisational failures framework to analyse the theory of quasi-markets and contracts in the context of the NHS. He suggests that the NHS provides an excellent illustration of Williamson's theories in practice.

Firstly, he examines the difficulties of writing, implementing and enforcing clear contingent claims contracts in circumstances where bounded rationality interacts with a complex and uncertain environment. To foresee and describe all possible eventualities in all specialties and set up an appropriate pricing structure in a complex field such as health care is a formidable task. Bartlett argues that the information-gathering and monitoring effort required to specify such contracts will lead to spiralling transaction costs; indeed, it is probable that no matter how much is spent on the contracting process it will in any case prove impossible to establish truly satisfactory contracts. In these circumstances it is likely that purchasers will revert to incomplete 'block' contracts instead, which do not make such clear allocative decisions and which show more similarities with the internal hierarchical organisation of the pre-reforms NHS than a quasi-market structure.

As already pointed out, the problem of imperfect information also stems partly from the difficulties for purchasers of obtaining accurate information from providers. This illustrates the interaction between small numbers exchange (i.e. very limited competition) and opportunism. Thus block contracts are made more likely both because of the inherent complexity of health care and because of obstacles to competition in the NHS. They may reduce the transaction costs of information gathering on the one hand, but they perpetuate the problem of opportunistic behaviour by providers. Opportunism has its own implications both for costs and for the control of allocative decisions. As Bartlett says, 'this effect on costs could arise both where providers operate to a less stringent set of working practices than would be required if effective monitoring, and penalties for poor performance, were available, and where unconstrained pursuit of professional excellence biases activity towards prestige treatments to the detriment of cheaper, more mundane but equally effective ones.' (p.58)

Lack of competition is thus an important influence on contracting, both because of its direct impact on efficiency and costs, and because of its implications for the broader problem of imperfect information. Health care is a highly asset-specific market, where the buildings, equipment and skills required are not easily transferable to alternative uses. Given the high initial outlay and subsequent risk of financial difficulties for new entrants to such a market, it is unlikely that the small numbers exchange problem will go away, so competition will remain weak.

Some commentators argue small numbers exchange does not represent a major problem for the NHS market. Ham (1996) suggests contestability is the key to ensuring that it remains efficient despite the absence of strong competition; contestability relies on *comparing* performance rather than competition and on the threat of moving contracts if providers do not respond to unfavourable comparisons by improving their own performance. He accepts that, for the threat to remain credible, contracts must actually be moved on occasion, but this may be the point at which the theory of contestability falls down, if no competitors are available within easy reach and potential new entrants to the market face insuperable financial or other barriers. Williamson (1985) is sceptical of the claims of contestability theory, as developed originally by Baumol *et al.* (1982). He argues that it 'reduces asset specificity to insignificance, so that hit-and-run entry is easy. Transaction cost economics, by contrast, magnifies the condition of asset specificity. The existence of durable, firm specific assets is held to be widespread, and accordingly hit-and-run entry is often infeasible' (p.31, footnote).

Appleby *et al.* (1994) agree that the NHS may not exhibit the characteristics of a contestable market. However, their research in West Midlands Region has suggested that the level of monopoly enjoyed by acute providers was not as great as popularly imagined. In general surgery, their sample specialty, only 38% of patients were treated in hospitals which they defined as monopolistic. They interpret this finding with caution, firstly because 38% still represents quite a sizeable element of monopoly and secondly because the degree of monopoly would almost certainly be higher at sub-specialty level, in other specialties and in a more detailed break-down of market catchment areas.

They acknowledge Bartlett's points, but say that 'the significance of this for the internal market is that it points to the desirability of long term stable contractual relationships between purchasers and providers. If long term contracts exist, competition is likely to take place *for* markets at periodic stages of contract negotiation, rather than *in* markets on a day-to-day basis.' (p.26)

Long term relationships with preferred providers have a well-established place in the commercial market-place, so their presence in the NHS need not demonstrate the failure of the quasi-market. On the other hand, it could be argued that the language of partnership and long term contracts evident in much recent discussion of the NHS, and the interest in contestability, were a way of circumventing the imperative of contingent claims contracts and were simply proving Williamson and Bartlett right. Failing an active and formal decision to return to internal vertical integration as a more efficient model, has the NHS been drifting back towards it unofficially, if indeed it ever left it? Management consultant Kingsley Manning (1996) stated in a conference speech that 'central direction is still very strong. Trusts are more like branches of a franchised operation than independent autonomous organisations. There is still a tendency for trusts to reinforce the status quo'. If this is true, the expectation that the logic of the internal market would lead to more explicit rationing decisions may have been misplaced.

The Conservative government itself acknowledged the heavy administrative burden created by the contracting process, despite the fact that most contracts remain incomplete block contracts; in the terms of reference for the NHS efficiency scrutiny, the first task for the scrutiny team was to 'consider the scope for simplifying the processes and transaction costs associated with contracting and invoicing between purchasers and providers' (Department of Health, 1995). This certainly supports Bartlett's predictions.

The political and ethical costs of contracting

Bartlett's focus is on the transaction costs of contracting in a quasi-market. In addition, the 'life and death' nature of health care and the level of popular support for the NHS may give rise to political and ethical costs which are as great or greater than the transaction costs.

Firstly, transaction costs are themselves a political issue for any government committed to containing public expenditure and keen to demonstrate that additional investment in the NHS is supporting frontline clinical staff. Media reports of large increases in finance, administration and information technology staff needed to run the internal market have provided opposition parties with plenty of embarrassing ammunition. The new Labour Government is now experiencing the same pressure to demonstrate investment in clinical care.

Furthermore, increasing explicitness in purchasing plans and contracts has created tension and ambivalence at political level. It is unclear whether the Conservative government intended its reforms to increase the explicitness of rationing or even foresaw this is as a natural consequence. Glennerster and Le Grand (1994) take the view that the government's main aim was to improve consumer responsiveness, an aim distinctly at odds with the possibility that consumers might increasingly be refused access to certain services. Ham *et al.* (1990) stress the crucial role of the 1987 NHS funding crisis in triggering the review, and concur with Glennerster and Le Grand that the government hoped a market-based system would squeeze more efficiency out of the same level of resources. This would suggest they hoped the market would avoid the need for more explicit rationing rather than bringing it up to the surface.

Alternatively, the openly stated requirement that health authorities should carry out needs assessment and priority-setting as a precursor to establishing contracts does suggest some degree of rationing was expected, albeit it by another name. Crucially, however, this was expected to be done at local level, leaving politicians clear of uncomfortable decisions and able to lay the blame on the natural operation of the market or to scapegoat health authority managers for making the 'wrong' choices. Glennerster and Le Grand (1994) suggest devolution of power and responsibility in priority setting may also be in the interests of top Department of Health civil servants, whom one might otherwise expect to be reluctant to relinquish control - as they say, 'If there is dirty and unpopular work to do, let someone else do it' (p.15).

There are strong reasons why the government should be anxious to avoid direct involvement in explicit rationing, even if in private it is acknowledged privately. The National Health Service Act 1946 was explicitly designed to put an end to the previous system of rationing by ability to pay, and established

> '...a comprehensive health service designed to secure improvement in the physical and mental health of the people of England and Wales and the prevention, diagnosis and treatment of illness, and for that purpose to provide or secure the effective provision of services in accordance with the following provisions of this Act.
>
> The services so provided shall be free of charge, except where any provision of this Act expressly provides for the making and recovery of charges'.

The aims of the Act are crucial in understanding the attitude of successive governments to rationing. Although it quickly became apparent that the NHS could not meet all the demands upon it and charges were introduced around the margins to try to control spiralling expenditure, the basic principle of a comprehensive, publicly funded service free to all citizens still has a powerful hold over the views of politicians and public alike. In opinion surveys, the NHS has consistently been found to be one of the most popular and fiercely defended elements of the British welfare state (See, for example, Taylor-Gooby, 1991). Only the Thatcher government gave any significant consideration to alternative methods of funding and in the face of enormous opposition quickly reverted to the principle of funding from general taxation, although Mrs Thatcher is believed to have insisted personally on including tax relief for health insurance for elderly people in the NHS reforms (Ham *et al.*, 1990).

Loyalty to the concept of a free and comprehensive service makes it difficult to countenance explicit rationing. Any attempt by government to specify elements of health care for which people are not eligible under the NHS is seen by the public as a contravention of their basic rights under the 1946 NHS Act. (In contrast the courts do not generally support the idea of an absolute right to be treated and have dismissed cases brought by individuals who have been refused care, notably in the recent case of the 10-year-old 'Child B' with leukaemia. In making such judgements, the courts have relied particularly on the wording of the NHS Act 1977, which says: 'The Secretary of State is under a duty to provide services to such an extent that he considers necessary to meet all reasonable requirements').

Whatever the strict legal position, the fact remains that successive governments have been reluctant to be seen to restrict access to the NHS for fear of jeopardising their electoral position, and any rationing has generally been implicit.

Unfortunately for politicians, the devolution of priority setting has not shielded them entirely. Although theoretically they could allow the market or health authorities to take the blame, in practice they have often felt unable to let this happen. There is considerable anecdotal evidence of ministerial intervention to reverse health authority decisions to refuse to pay for certain referrals.

By their very nature, these incidents are not well documented, but a verbal account of one such case was given at a lecture in 1991. A health authority decided not to allow a patient needing a joint replacement to be referred to another provider with a lower waiting time than the local hospital, but a significantly higher cost per case. The chief executive's argument was that the referral would have been both inequitable and an inefficient use of money with an opportunity cost for other patients on the waiting list. The patient contacted the local MP, setting off a chain of communication which culminated in a minister telephoning the health authority chairman and the chairman vetoing the chief executive's decision (Institute of Health Services Management, 1991).

There were signs of growing acceptance of explicit rationing by the Conservative government. The Child B case in 1995 was notable for the lack of political intervention and the government's willingness to allow it to go to court. Official documents such as 'The National Health Service: A Service With Ambitions' (Secretary of State for Health, 1996) were becoming more forthright in acknowledging the need for priority setting, although continuing to emphasise that this should be carried out at local level. The report of a working party set up jointly by the NHSE even used the word 'rationing' (Academy of Medical Royal Colleges *et al.*, 1997, p.6). However, the Labour government is evidently uneasy with this, a point discussed in greater detail in the final chapter.

Political unease is of course influenced by concern about the ethical costs of explicit rationing. The managers of Cambridge and Huntingdon Health Commission, who decided not to fund a private referral for Child B requested by her family are perhaps unusual for the degree of ethical confidence they have displayed in their decision. A significant factor in this must be the fact that the child's own doctors advised the health authority against further treatment on the grounds that the slim chance of a successful outcome did not justify the additional suffering involved in further treatment. This could be regarded as a purely clinical decision, with a sound ethical basis of pursuing the patient's best interests - this may incidentally help account for the government's lack of intervention.

However, the patient and her family wished to proceed with treatment, and found a clinician willing to help them. At this stage, in the words of the judge at the first hearing, the health authority began to 'toll the bell of limited resources', arguing that the level of suffering and the poor chances of survival did not justify the level of expenditure. This changes the ethical picture considerably. This is not to say that a decision which includes cost is necessarily unethical, but that it brings a different ethical perspective into play - the utilitarian perspective. Utilitarianism sets the individual patient's wishes and rights against the interests of other existing and potential patients who could benefit more from the resources, with the ultimate aim of maximising the total benefit. One could argue that the child was wrong; that her real interests were in not being treated and were therefore not in conflict with the interests of other people; and that the health authority purchasing on her behalf was simply correcting her own imperfect information. However, not only does this challenge her autonomous judgement of where her interests lay, but the introduction of cost into the equation does suggest that at least to some extent her interests were being set against the opportunity costs for other patients.

If the case of Child B poses major ethical dilemmas, despite agreement between local clinicians and managers, it may be imagined that ethical debate will be much stronger where no consensus exists between them, and where the evidence about ability to benefit from treatment is less clear-cut. All forms of rationing essentially trade off some people's interests against others, but this may be achieved in covert or random ways; the problem with more explicit forms of rationing is that they involve a deliberate and overt trade-off and lay bare the tragic choices being made. Collective purchasing by an agency, which aims to maximise benefit for whole groups of people, is likely to end up taking a utilitarian position. Detailed contracts would make this position abundantly clear.

Ethical dilemmas of this kind bear an emotional cost for individuals making the decisions, a cost to politicians in terms of electoral popularity, a cost to society in confronting the conflict of values and not least a cost to individuals harmed by the decisions made. These ethical costs may be high enough to threaten the development of explicit rationing through the quasi-market.

Other pressures in favour of explicit rationing

So far two possible scenarios have been considered: firstly, that the contracting processes of the quasi-market will make rationing decisions more explicit, or secondly, that contracting will fail to realise this expectation as a result of the high transactional, political and ethical costs of greater explicitness.

Is there a third possibility? The first two hypotheses assume that the successful or unsuccessful functioning of the contracting process is the dominant factor in determining whether rationing decisions will become more explicit. However, there are other factors which may influence events more strongly than contracting, especially rising expenditure, improved understanding of outcomes and the Patient's Charter.

Rising expenditure

It is often taken as indisputable fact that demand for health care is infinite, and that therefore health care expenditure has infinite potential for growth. There are those who disagree with this position, and their views will be discussed later. Whatever the truth of the matter, there is a demonstrable trend of rising health expenditure in many Western countries, accompanied by a widespread perception that the cost of meeting all demands for health care is unaffordable and that expenditure must be brought under control.

Total health care expenditure in all OECD countries more than doubled as a share of GDP over the period 1960 to 1992, from 3.9% to 8.4 %. Much of this increase took place up to 1980, when the average reached 7.2%, with a slower rate of increase between 1980 and 1990 up to 7.9%. The OECD expresses some caution about the apparent increase between 1990 and 1992, as it is affected by the recession and resulting weak position of GDP. The rate of growth varies between countries; for example, the US has continued to have a faster growth rate than others (from 9.2% in 1980, to 12.4% in 1990 and 14% in 1992), whilst the slow-down in expenditure has become ever sharper in Europe. The European average rose from 7.1% in 1980 to only 7.6% in 1990 and 8.0% in 1992 (OECD, 1995).

The interrelation between demand, cost and expenditure is not straightforward. Rising demand may lead to higher expenditure, either because more people want what has always been on offer, or because people want new, more sophisticated or better quality services. However, greater efficiencies may enable rising demand to be accommodated within existing expenditure. Rising costs may result from inflation, or from the development of more expensive techniques, which is in turn closely related to rising demand. However, higher costs may simply result in more stringent rationing rather than higher expenditure, if this is what society chooses. Rising expenditure may also have a circular effect of increasing previously suppressed demand, and perhaps encouraging providers to increase their costs.

The recent <u>deceleration</u> in the growth of health's share of GDP may be as significant for explicit rationing as the continued growth in expenditure itself. Health care providers accustomed to relatively generous increases in funding will have raised expectations of development and technological achievement. As governments seek to rein in this level of growth, providers' perception of the gap between desirable and actual levels of funding will be all the sharper, even though funding is still increasing. As a result, they may feel more inclined to make very clear statements of what can no longer be afforded.

Three factors in particular play a role in this complex equation: the development of expensive high technology, demographic change and rising consumer expectations. These pressures are found internationally, regardless of the structure of health care provision in each country.

Technological innovation has the potential to reduce costs in some areas - for example, the introduction of less invasive forms of surgery results in quicker recovery and discharge from hospital. However, it often requires high capital investment, and earlier discharge results in increased capacity, so more patients can be treated and expenditure rises even if the cost per patient may be slightly lower.

New technology often enables conditions to be treated for which no treatment was previously available, greatly extending the scope for medical intervention and thereby increasing demand. In vitro fertilisation is a good example - couples whose infertility would previously have been regarded as a sad fact of life can now do something about it. Even though the rates of success may be statistically low, at an individual level it is difficult to forego the chance. Health care professionals will also find it hard to resist applying new technology, not only on the principle that they should do the best they can for every patient but also because it is only by experiment and experience that success rates will be improved.

The impact of demographic change on health care demand has been extensively documented. The most important element is the increased number of elderly people, who tend to have more serious conditions and longer and more frequent stays in hospital than younger people, all of which increases expenditure (Ermisch, 1990; Johnson and Falkingham, 1992). Of particular concern is the proportion of very elderly people, whose health needs are even greater. The percentage of those aged 80 and over rose from 1.9% in 1961 to 3.7% in 1991, and is projected to be 5.2% by 2021 (Central Statistical Office, 1994). Increased survival partly accounts for the emergence of different types of health needs, such as the growing incidence of Alzheimer's Disease.

Recent increases in joint replacements illustrate the combined impact of technological development and rising numbers of elderly people, increasing the demand both in terms of volume and in terms of the sophistication of treatment. In this case, the growing number of elderly people suffering from serious joint problems has also acted as a stimulus to technological development. The growth in expenditure has been partly contained by lengthening waiting lists.

Care for elderly people has always been at the forefront of disputes between local authorities and the health service (see, for example, the Boucher Report, 1957); as strain on their limited budgets increases, each agency has sought to define more clearly the limitations of its own responsibilities and the extent of the other agency's duties. Increasingly the NHS has withdrawn from long term care, so that many people who would have expected to obtain free NHS care have been faced with means tested local authority care as the only alternative. Local authorities are beginning to respond that they cannot afford the costs of community care for elderly people. The national requirement to establish jointly agreed eligibility criteria for continuing care in each health and local authority has so far done little to clarify responsibilities.

However, some analysts question the assumption that elderly people will inevitably impose a burden on health services. Thane (1989) argues that much ill health and dependency is shaped by society's expectations and that there is little value in transposing the existing health of elderly people onto a very different future population. This may be one area where new technologies may reduce dependency and therefore costs; preventive health care and better social conditions mean successive generations of elderly people are staying healthier for longer. Taylor-Gooby (1991) supports this argument by pointing out that the number of elderly people has in fact been growing since the 1940s and that the 'burden' has so far successfully been absorbed. On the other hand, research in the US has found mean Medicare payments in the last year of life are seven times as high as the average yearly payment for all Medicare payments; payments in the last month of life make up 40% of the total in the last year (Lubitz and Riley, 1993). Staying healthier for longer cannot prevent final fatal illness.

Rising consumer expectations operate at a number of levels, sometimes with conflicting effects. Patients increasingly expect improved standards of accommodation and other services when in hospital, and demand shorter waiting times. They may seek more domiciliary or community care (e.g. for maternity care and terminal illness) and expect professionals to come to them; on the other hand, there is continuing public support for high technology care and expensive interventions for a few emotive cases. People expect better quality of clinical care, believing modern medicine should be able to cure everything; simultaneously, they are more sceptical about doctors' status as infallible experts, and more likely to begin costly legal proceedings against them. They also demand access to a greater range of services, as new technologies develop. People in this country born since the introduction of the NHS have come to regard free and universal access to health care as their right, unlike previous generations (Blaxter and Paterson, 1982).

Some of these changing demands may be cost neutral or even reduce costs, but evidence (where it exists) is often conflicting. Community care, for example, generates great controversy as to whether it will prove cheaper or more expensive than institutional care. Improvements in efficiency may offset some of the effect of rising demand. Overall, however, it is likely that rising demand will cause expenditure to rise unless it is countered by more stringent rationing.

In addition to the pressure of consumer expectations on expenditure, greater consumer involvement in purchasing decisions may exercise a more direct effect on the likelihood of rationing becoming more explicit. Firstly, there is some evidence that public opinion is in favour of rationing health care to some groups perceived as self-harming. A 1994 Gallup survey of 1,000 people found that 32% favoured discrimination against heavy drinkers on hospital waiting lists, 25% supported discrimination against smokers and 11% against people who were overweight (Health Service Journal, 1994a).

Whilst there may be some public pressure in favour of rationing care to certain self-harming groups, consumers within such groups have expressed concern that they are already discriminated against. Paradoxically their complaints may add to the momentum behind more explicit rationing, as hitherto covert policies are forced out into the open and debated.

Although, as noted earlier, the quasi-market was at least partly intended to improve consumer responsiveness, pressure for greater consumer involvement pre-dates the NHS reforms and has its own international momentum outside any particular structure of provision. The Patient's Charter is an illustration of this, and will be discussed shortly.

Development of outcomes research

The rapid development of outcomes research is also linked to the development of explicit rationing. This includes research into effectiveness, measures of cost-benefit in health, and consumer satisfaction measures. Outcomes research offers tools to those carrying out rationing, whether or not this is the intention of the researchers.

In some cases it clearly is the intention of researchers, an example being the economists at the University of York who advocate the use of their QALY system to develop cost-benefit league tables of different treatments to assist managers and doctors in deciding priorities (Williams, 1985; Maynard, 1994). (QALYs are Quality Adjusted Life Years, a combined measure of survival and quality of life resulting from a given treatment). At the other end of the spectrum, research identifying effective and ineffective treatments may be intended to prevent patients undergoing unnecessary trauma and finding the best way of making them better, regardless of costs and without trying to rank the benefits gained against other types of treatment for completely different conditions. Many would argue that ceasing ineffective treatments is not in fact rationing, a question which is explored in the next chapter. In practice, however, most research does not uncover absolute ineffectiveness, but rather identifies relative levels of effectiveness. It may demonstrate that treatment X is better for some people than treatment Y, but not that Y is therefore pointless.

Analysing outcomes is also an important factor in researching variation in treatment rates between different areas of one country and between different countries. If outcomes are not significantly different between groups with high or low intervention rates, this raises the question of whether higher rates are justified and challenges the assumption that more is always better.

The explosion of information about outcomes and improved understanding of effectiveness offers decision-makers a new and apparently powerful tool with which to adjudicate competing demands for limited resources. It holds out the tempting prospect of a more rational, neutral and scientifically based decision-making process. However, the complexity of comparisons between different treatments for different conditions in individual patients is immense, and there are still vast areas of health care which have not yet been properly evaluated. There is a danger that formulae such as QALYs or any other kind of ranking mechanism will be used over-simplistically by decision-makers desperate to bring some order into priority-setting.

Carr-Hill (1991) suggests QALYs are potentially dangerous because they appear to offer a straightforward technical solution to what is in fact a political problem. Hunter, whilst recognising the value of greater understanding of effectiveness, has also argued that it becomes 'corrupted or problematic...where it becomes entangled with the current preoccupation with the rationing of health services' (Hunter, 1993a, p.29). He notes that 'numbers have a curiously mesmerising effect on managers with unfounded assumptions of certainty and precision underpinning their very hardness', and suggests 'we should resist abandoning an admittedly imperfect though workable irrationality in favour of a quite spurious rationality which is probably unattainable and certainly undesirable'(p.31-2).

Despite these reservations, managers are under pressure to find ways to improve the allocation of resources, and some academics are keen to encourage them to use outcomes-based techniques. The availability of information on outcomes gives additional momentum to the development of more explicit rationing mechanisms.

Patient's Charter

In this country, the role of the Patient's Charter (Department of Health, 1991) in forcing some rationing decisions has been crucial and evidence for this will be presented in later chapters. The Patient's Charter was launched by the government independently of the NHS reforms in October 1991. Although its emphasis on empowering the consumer does give it clear links with quasi-market ideology, it tries to achieve this through the imposition of central directives to create a set of patient 'rights' (of a non-legal kind). This is distinctly at odds with market assumptions that consumers are empowered by their ability to choose or ultimately to exit from the market. The fact that the government feels additional regulation is necessary to empower consumers serves to confirm the idea that the quasi-market is not in itself a very good vehicle for consumer choice.

The Patient's Charter sets out several rights for patients, including maximum waiting times for admission to hospital. Since its introduction, the maximum waiting time target has been steadily lowered, with some regions (notably West Midlands) choosing to impose even stricter standards on health authorities.

In elective surgery long waiting times have traditionally been used as an implicit rationing mechanism based on an assessment of clinical need, for instance giving people requesting cosmetic surgery lower priority than others. The Patient's Charter had two effects. Firstly, it brought to light the availability on the NHS of operations such as tattoo removal which managers were not necessarily aware of and which they now thought should be challenged. Secondly, by changing the criterion for admission from clinical need to length of wait, it closed off the possibility of using very long waiting times as a way of rationing access to these procedures, and meant that some alternative, more explicit rationing mechanism had to be found. It is also possible that the earmarking of funding for waiting list initiatives diverted funds that could have been used for other things; this may have resulted in more rationing in other, non-elective areas, although if so this has probably been implicit.

(There have been reports of some clinicians rejecting the distortion of clinical priorities caused by the Patient's Charter (Health Service Journal, 1996). At St. George's, doctors insisted on admitting patients according to the urgency of their condition, rather than the length of time they had been waiting, and 27 clinical directors from South Thames wrote to the Secretary of State to express their concerns).

All of the above additional pressures in favour of explicit rationing - rising expenditure, outcomes research and the Patient's Charter - may be so strong that they will prove irresistible, whatever structural form the NHS takes. If this hypothesis is correct, one would predict that explicit rationing will continue to increase, either through the quasi-market contracting system, regardless of its high transactional, political and ethical costs, or through some other mechanism if contracting fails to provide an adequate means of achieving it or is abandoned altogether. This scenario has particular relevance given Labour's stated commitment to abolishing the NHS market (Secretary of State for Health, 1997).

Statement of hypotheses

To summarise, the research carried out in support of this thesis aimed to test three different hypotheses:

i. Markets are one possible system for allocating scarce resources. The process of contract specification inherent in a complex quasi-market system constitutes a form of rationing, which formalises resource allocation decisions and makes them more explicit than they would be in a hierarchical system. The theory that lies behind the move from hierarchy to quasi-market in health care would thus predict that rationing in the NHS will become more explicit as a result of the introduction of the internal market.

ii. However, in the complex context of the NHS the quasi-market may fail to produce clear contracts and unambiguous allocations. There are several reasons why this may happen:

- Williamson's 'organisational failures' thesis suggests that the transaction costs of contracting will be too high because the knowledge base is too low, and that a drift back to hierarchy may occur to enable the system to cope more efficiently;
- it is likely that politicians will be reluctant to accept the consequences of their creation (although they may perhaps be evolving towards acceptance);working without political support creates difficulty for others;
- there may be ethical unwillingness to bring tragic choices out into the open.

This view predicts either that contracting will be conducted in form only, with a continuation of more implicit rationing mechanisms, or that it will be abandoned altogether.

iii. Alternatively, other pressures in favour of explicit rationing - particularly rising expenditure, outcomes research and the Patient's Charter - may be irresistible, whatever structural form the NHS takes. This hypothesis would predict that explicit rationing will continue to increase, whether through the quasi-market contracting system, regardless of its high transactional, political and ethical costs, or through some other mechanism if contracting fails to provide an adequate means of achieving it or is abandoned altogether.

Chapter Three

The history of health care rationing and the current debate

What does rationing mean?

Whilst the previous chapter dealt with the general theory of resource allocation, this chapter looks in more detail at the history of real resource allocation decisions in health care and at the current debate about rationing.

The starting point remains the assumption that, like any resource, the amount of health care available is limited, and therefore choices must be made about its allocation, whether implicitly or explicitly. However, the use of the word 'rationing' to describe this process is politically sensitive, and does not go unchallenged; furthermore, there are those who question the need to make choices at all. These debates will be examined in depth later in this chapter.

Even the meaning of 'explicit' and 'implicit' in the context of rationing is subject to individual interpretation, and it is therefore necessary to clarify the interpretation which will be used in this chapter. Implicit rationing is sometimes taken to mean rationing by clinicians. Whilst this is one very important component of implicit rationing, it is not a comprehensive definition. Building on the 'explicit-unrecognised' and 'open-closed' continua proposed by Glennerster (1975) (see below), I suggest that explicitness consists in a decision being planned and evaluated (or at least measured), with a clear attempt to distinguish who will receive what, and in being understood and agreed by a group of people, not just the individual clinician. (This does not necessarily include the public or the patient, however). Mechanic (1995) describes it as 'trying to establish all the rules beforehand' (p.1659). Implicit rationing is a process in which the reasoning involved is not clearly stated to anyone except (or possibly including) the person making the decisions, or in which active decision-making is avoided altogether. Thus waiting lists are an implicit mechanism not because they embody decisions made by individual clinicians, but because their reasoning remains private; the distinguishing feature is not *who* makes the decision but *how* they do it. It is perfectly possible for an individual doctor to make an explicit rationing decision and share the criteria used with colleagues and the patient. Random rationing, perhaps a lottery or a waiting list run purely on a first-come, first-served basis, is also implicit because it avoids deliberate decision-making about who will receive what.

In fact much of the current disagreement about rationing focuses on *explicit* forms of rationing; there is a considerable degree of consensus that *implicit* rationing has always gone on in health care, even if this has been unintended.

The existence of various kinds of rationing has been consistently acknowledged in social policy analysis (e.g. Parker, 1967; Glennerster, 1975; Cooper, 1975), and different analytical frameworks abound.

Glennerster (1975) emphasises that the two questions 'how much and for whom?' are fundamental to all social administration and that 'all of those concerned with the delivery of social services, whether at the centre or the periphery, are caught up in the rationing process' (p.11-12)

In Chapter Two of this thesis, three questions were in fact identified, starting with *what*, as well as ***how much*** and ***for whom***. This may reflect the fact that explicit total exclusion of some services, albeit at the margins, is now a reality in the NHS, whereas previously it would not have been contemplated so starkly.

Glennerster identifies four dimensions or continua by which the 'locus and nature' of allocation decisions can be analysed (p.38):

Central - peripheral: rationing decisions can be made at any level from the Cabinet down to local service providers.
Explicit - unrecognised (implicit): an explicit decision is one which has been planned, where there is some attempt to distinguish who is receiving what and to evaluate the outcome. An unrecognised decision emerges without such awareness and deliberate action.
Open - closed: this concerns the degree to which the knowledge on which decisions are based is openly known and debated. An allocation can be made explicitly but on the basis of information to which access is restricted.

Technical - political: decisions may be made on technical or professional grounds, or they may be taken in response to political pressures. Taking political in its broadest sense, this continuum could also include bureaucratic self-interest.

The last three categories have much in common with the components of 'rational rationing' identified by New and Le Grand (1996): explicitness, democratic participation and systematic decision-making processes (p.23).

To this could be added the following continuum:

Population - individual: this to some extent parallels the central - peripheral continuum, but concerns the people affected by the rationing decisions rather than those making them. In theory, central decisions could affect quite small groups of people - if, for example, the government implemented a national ban on tattoo removal - although in practice it is more likely that central decisions will affect larger groups of the population and peripheral decisions smaller groups or individuals.

Different rationing mechanisms can take place at points along all these continua. Building on work by Parker (1975), Harrison and Hunter (1994) identify the following mechanisms: *deterrence, delay, deflection, dilution and denial* (p.25-30).

Deterrence includes introducing charges (price rationing), making access inconvenient and creating psychological or social barriers to using the service - the higher utilisation rates amongst the middle classes noted by Le Grand (1982) may for instance be partly accounted for by the fact that most health care providers are themselves middle class and may subconsciously make it more difficult for people from lower socio-economic groups to approach the service. The ability to manipulate one's GP to obtain the level of service desired is a rationing device which favours the better off and better educated.

Delay operates most obviously through waiting lists (time rationing), while ***deflection*** is illustrated by GPs choosing not to refer a patient to secondary care services or diverting them to an alternative agency such as social services. Harrison and Hunter also explore the idea of deflecting demand by giving patients more information, the assumption being that if patients knew more about the effectiveness (or otherwise) and side-effects of particular treatments, they might more often choose not to seek care.

Responses to health care demand can be ***diluted*** by a reduction in quality, which could mean giving an individual less treatment or fewer drugs, for example, or permitting a whole area such as mental health services to remain at a lower quality level than, say, acute services. A less obvious means of dilution explored by Harrison and Hunter is the use of doctors' clinical freedom to ration care at an individual level.

Denial generally takes the form of explicitly refusing care, either to certain groups in the population or to the population as a whole. It can sometimes also be implicit, for example when waiting lists are so long that the person never in fact gets treatment.

Harrison and Hunter do not discuss how **quantity rationing** - explicitly limited amounts of treatment - would fit into their framework. In a more general context this would be best exemplified by the ration book, allowing each person a fixed quantity of certain goods. Some health authorities which pay for IVF allow each woman only a limited number of attempts. Cervical and breast cancer screening are offered at fixed intervals. It is possible to view this kind of rationing as a form of denial; although this may appear paradoxical, the corollary of giving someone a fixed quantity of care is that they are explicitly denied any further amounts. Quantity rationing can also operate in combination with delay: a certain number of treatments are allowed for the financial year and allocated either on the basis of clinical priority or on a first-come first-served basis, with a waiting list for those denied care that year.

Klein *et al.* (1996) also identify a list of rationing mechanisms; this is similar to Harrison and Hunter's, but adds rationing by selection or by termination. **Rationing by selection**, they argue, 'is the converse of rationing by denial but can have the same outcome' (p.11) Providers select those clients most likely to benefit, or those who are most deserving, or those least likely to cause any problems. **Rationing by termination** is when someone is discharged from care and treatment is terminated.

The authors explore the levels of decision-making, identifying a macro, meso and micro level on Glennerster's central-peripheral continuum, and argue that only micro level decisions should be described as rationing in the strict sense. This is the service delivery level, where decisions are made about which individuals will receive treatment. At the other levels, decisions are about collective priority-setting, determining largely 'what' and 'how much' will be provided. They may begin to address the 'for whom' question, but only by setting out general entitlements or eligibility criteria; the interpretation of this in individual cases remains a micro decision.

In addition to the above dimensions and mechanisms of rationing, three further analytical categories may be useful:

Who rations? Are decisions made by doctors (consultants or GPs), managers, politicians, the courts, or the public? This links back to the macro-meso-micro distinction.

What is rationed? Is it capital (buildings and equipment), revenue, staff time or particular treatments or services? This has implications for how visible the decision is.

Which criteria are used to make rationing decisions? Harrison and Hunter identify equity, effectiveness, cost-effectiveness/cost-utility, individual rights, need and random selection (such as a lottery). New and Le Grand (1996) describe two headings of *'need-related characteristics,* such as illness or health deficit, effectiveness and cost-effectiveness, and *characteristics unrelated to need,* such as age, gender, race, socio-economic status, waiting time, number of dependants, and desert' (p.63). This category gets to the heart of the ethical debate.

Finally, one may look at rationing decisions in terms of **whose interests** they serve, although this is by no means a simple question to answer in many cases. This question is explored further in Chapter Nine.

The history of health care rationing

It is important to set out the history of rationing for two main reasons: firstly, it is relevant to understanding whether there are forces independent of contracting and the quasi-market which are making explicit rationing more likely. Secondly, it helps to establish a baseline against which new developments in rationing can be assessed.

Rationing in Britain before the introduction of the NHS

In Victorian Britain, health care was rationed both implicitly and explicitly. Firstly, access was controlled implicitly by ability to pay; those who could afford private doctors had better access to care. The poor were left to fall back on what voluntary, municipal or Poor Law services were available in their area. In the latter part of the nineteenth century, they were also subjected to a more explicit form of rationing by being categorised as 'deserving' or 'undeserving' poor. Care provided to the undeserving poor was limited both in quantity and quality; they faced additional deterrents in the form of harsh and unsympathetic treatment even if they did gain access to health care, to discourage abuse of the system. Whereas those who could afford private doctors were usually treated at home, those reliant on public or voluntary services were more likely to end up in hospital; hospital mortality was high and people feared being admitted.

Whether those involved in such allocation decisions perceived what they were doing as rationing is another matter; Glennerster (1975) suggests that the quite deliberate application of the principle of less eligibility (which ensured those receiving benefits were left worse off than those in employment) was indeed 'a classic example of a rationing device' (p.39), but was probably not perceived as such at the time. The principle could have been applied even if there had been no financial need to ration - and probably would have been - since its primary aim was to counter immorality and dependence rather than to control expenditure. On the other hand, it did reflect a sense that there was a *moral* need to ration benefits, to which end a situation of artificial scarcity was created. The prevailing views on personal responsibility and desert were not conducive to the idea of publicly funded services, beyond philanthropic giving.

The Lloyd George National Insurance Act of 1911, building on the model of contributory health insurance schemes set up by Friendly Societies in the nineteenth century, was intended amongst other things to improve access to medical care. However, even this left more than half the population uninsured, notably the dependants of contributing working men, as well as higher earners and self-employed people. The Act also did nothing to tackle the uneven distribution and quality of services in different parts of the country, as the emphasis was still firmly on personal responsibility. However, it was an important turning point in the development of health services and the increase in coverage which it brought about must have helped create stronger expectations amongst the public that they should have access to health care.

Rationing since the introduction of the NHS

Charges and budget capping

As noted in the previous chapter, the National Health Service Act of 1946 was supposed to put an end to rationing by ability to pay. Indeed, Beveridge's famous assumption that the NHS would get rid of a 'backlog of ill health' suggests he believed it would do away with the need for any kind of rationing. The dream of a free, comprehensive service was in fact under threat from the start by the inclusion in the Act of the words 'except where any provision of this Act expressly provides for the making and recovery of charges.'

Originally, charges were only to be made if the patient wished to have an appliance more expensive than the standard one, or to have the privacy of a single room in hospital, known as an 'amenity bed'. This did not involve any extra nursing or medical care. Even so, this is an example of dilution through inability to pay.

However, it very quickly became apparent that the NHS could not meet the demands made upon it and that charges would need to be introduced. In 1949 a one shilling prescription charge was established, followed in 1951 by charges for specific dental and optical services. The aim was both to raise additional revenue and to control expenditure by deterrence; expenditure control was further aided by the introduction in 1950 of a budgetary ceiling for NHS, which had previously been limited only by demand (Klein, 1995).

The introduction of charges inevitably began to threaten the principle of equity on which the NHS was meant to be founded. Although charges have remained marginal, and have never raised enough revenue to cover the whole cost of the services involved, they can exercise a real deterrent effect on individuals. Research into the effect of introducing charges for eye-testing in 1991 demonstrated a reduction of up to 19% in both referrals to ophthalmologists and identified cases of glaucoma (Laidlaw *et al.*, 1994).

Waiting lists

In parallel with charges, probably the single most important implicit rationing mechanism in the history of the NHS began to take effect - the waiting list. As Klein (1995) points out, waiting lists have been a remarkably stable feature of the NHS, and have commanded surprising levels of support or at least tolerance until recently. He suggests the reason lies in the atmosphere of the post-war socialist era in which the NHS was born, when 'rationing and queuing were symbols not of inadequacy but of fairness in the distribution of scarce resources' (p.133).

Waiting lists do not involve an outright denial of care, but rely on delay and tacit judgements by clinicians about individual patients, the criteria for which are never articulated openly. Even though someone may have to wait so long that in practice they are never treated, the system offers some hope that finally their turn will come. Because it is aimed at non-urgent conditions, it does not pose any apparent threat to access to life-saving treatment. (In reality, however, having to wait too long for what was originally a non-urgent problem may result in deterioration of the condition to the point of urgency - delays in elective heart surgery are a case in point).

Waiting lists may also encourage deflection, by making people seek the alternative of the private sector - evidence that this is in fact the case is presented by Besley *et al.* (1996) - or they may deter people from seeking care altogether. Anecdotal accounts suggest long waiting times even to be seen at out-patients for varicose veins have suppressed demand for referral from GPs and from individual patients.

Increasingly, the use of long waiting times has been closed off by the Patient's Charter, as discussed in Chapter Two. This does not mean that waiting lists have been abandoned altogether as a mechanism; indeed, the government's own figures demonstrate that the total number of people waiting for an operation rose substantially during the period of most intense activity to reduce long waiting times (Health Service Journal, 1994b). Independent research has shown that patients in several specialties who would previously have been treated in less than one month were having to wait one to three months as a result of the focus on long waiters (Harley, 1995).

Rationing by specialty/care group

The inequality of distribution of resources between specialties and care groups is given by Harrison and Hunter (1994) as an example of rationing by dilution. Despite successive attempts by central government to give priority to the so-called 'Cinderella' services such as mental health care and services for elderly people or people with learning disabilities, it has proved a hard struggle to actually implement the policies and improve the quality of care. As Ham (1992) notes, 'the claims of non-priority groups, particularly in the acute hospital sector, may be pressed strongly at the local level, and may push service development in a different direction from that desired by central government' (p.210). The fact that people with chronic illness are often more vulnerable and less able to complain about the effects of rationing by dilution must also be significant. The failure to invest adequately in alternative community services and a number of highly publicised murders involving people with a history of psychiatric problems have served only to strengthen public opinion against innovation in these areas.

Levenson (1996) notes that the lack of funding for mental health care results in only the most severely ill people receiving any care at all. She adds, 'It is a curious fact, and no doubt a reflection of the lack of public awareness about mental health, that such extreme rationing of preventive and non-emergency services can take place in mental healthcare while rationing elsewhere in the NHS generally provokes indignation' (p.29).

NHS withdrawal from long term care for elderly people has already been mentioned as an illustration of rationing arising from demographic pressures. To some extent it is in line with community care principles supported by the government that people should be enabled to lead as independent and non-institutionalised a life as possible, but it also provides evidence of deflection of responsibility to another agency, generally without an accompanying transfer of resources. The Government has tried to reduce the amount of rationing in this field by requiring health authorities to specify a set of explicit eligibility criteria for continuing care. It should be noted, however, that eligibility criteria are in themselves an explicit rationing mechanism; their intention in this case may be to make services available to more people, but one cannot logically define eligibility without at the same time creating an explicit category of ineligibility.

Geographical rationing

The creation of the NHS did not in itself do anything to redistribute resources around the country; financial stringency during the 1950s left inherited inequalities in the distribution of hospital beds virtually untouched, by default targeting rationing at poorer areas of the country. This, too, is a form of implicit dilution, in that services of inferior quality and quantity are allowed to remain through lack of positive action to alter the situation. It is also a form of deterrence; there is evidence that proximity to hospitals explains some of the variation in utilisation rates (Clarke *et al.*, 1993). Over the years, various attempts have been made to intervene more explicitly and target rationing at better off areas.

The Hospital Plan of 1962 was an early effort to inject capital into a national hospital building programme, to cope with deteriorating building stock and ensure all parts of the country had access to a standard district general hospital or DGH, although in the event implementation was slow and patchy (Klein, 1995).

Even if hospitals are spread more evenly around the country, they tend to be sited in middle class districts. This means higher travel costs for people from lower socio-economic groups, who are also more likely than middle class people to have to take unpaid leave from work to attend appointments, an additional financial deterrent (Le Grand, 1982).

From the inception of the NHS, the Medical Practices Committee had power to restrict the number of GPs setting up practice in areas already well supplied with doctors and designate special areas in need of more doctors. In 1966, the designated area allowance was introduced as a financial inducement to encourage more doctors to set up in designated areas, which appeared to reduce inequalities of distribution (Ham, 1992).

A major exercise in revenue redistribution was initiated in 1970, when a new weighted population formula was introduced for the allocation of resources to regions. This was subsequently refined following the report of the Resource Allocation Working Party (RAWP) (Department of Health and Social Security, 1976). RAWP softened its rationing effect on the better off areas of the country in two significant ways. Firstly, it was committed to achieving redistribution not by taking resources away from any regions over target, but only by ensuring growth money was aimed primarily at under target regions, a strategy which seemed promising in times of economic growth. Secondly, it gave some protection to the major teaching hospitals (the majority of which were in London and the South East) by means of SIFT, the service increment for teaching. This calculation of the surplus costs incurred by hospitals with teaching responsibilities was removed from the weighted allocation formula.

Efforts to find an improved allocation formula based on levels of need in the population have been underway for several years; as part of this process, the NHS Executive commissioned a report from the Centre for Health Economics at York University (Carr-Hill *et al.*, 1994). The report proposed two separate weightings for acute and psychiatric service needs, which would have had the effect of moving more resources to inner cities and the north of the country, where standardised mortality ratios and other health indicators from the 1991 census indicate health needs are greatest. The government decided to weaken the proposals by applying no weighting at all to community and administrative services (24% of the total hospital and community health services budget), on the basis that there was no evidence this expenditure was affected by different levels of deprivation. The York health economists responded by demonstrating that the result of this alteration of their work switched resources back to London and the South, incidentally benefiting the then Secretary of State for Health's constituency (Smith and Peacock, 1995).

The government also introduced a market forces factor which allows for higher rates of staff pay in London and the South East than are currently found expected as a result of local pay negotiation. This also resulted in financial losses to northern health authorities.

RAWP and its successors are explicit, centralised rationing mechanisms in one sense, whether they aim genuinely to improve equity of resource distribution, or to preserve existing distributions for political reasons. However, the fact that they set only global expenditure targets means that the implications of the financial losses for certain regions or districts continue to be decided at local level, often on an implicit, clinician-led basis.

Clinical freedom

Similarly, although the total limited budget for the NHS is an explicit rationing mechanism at macro level, it does not address the micro rationing decisions about which individuals will get the limited resources available. This has traditionally been done by doctors exercising their clinical freedom by making judgements about need in individual cases or in groups of patients.

The description by Harrison and Hunter (1994) of clinical freedom as a means of rationing by dilution may at first sight appear paradoxical; doctors in the UK defend clinical freedom on the grounds that it enables them to act purely in the best interests of individual patients. Patients also tend to view it as a protection - hence the public anxiety that GP fund-holding, by introducing budgetary responsibility into the traditional doctor-patient relationship, could result in treatment decisions made on the grounds of cost rather than individual benefit.

However, the intricate relationship between clinical freedom and rationing has been examined by a number of commentators. Cooper (1975) states that 'the NHS attempts to ration ... scarce resources ... not in accordance with the individual's ability and willingness to pay but in accordance with each individual's relative need' (p.50). However, he continues, 'need is not an absolute state but a matter of judgement and opinion' (p.51) and the very existence of clinical freedom means that it is doctors who have to make this judgement. 'Although the medical and allied professions have never seen their function as anything other than aiding the sick, they are nonetheless implicitly or explicitly daily involved in rationing decisions' (p.52). 'Rationing in the NHS has never been explicitly organised but has hidden behind each doctor's clinical freedom to act solely in the interests of his individual patient. Any conflict of interest between patients competing for scarce resources has been implicitly resolved by the doctor's judgements as to their relative need for care and attention' (p.59).

Aaron and Schwartz (1984) agree that doctors in Britain use their individual clinical judgement to make rationing decisions. They believe the majority justify this to themselves and their patients by recasting what are essentially questions of resource scarcity as clinical decisions about patients' ability to benefit from treatment. Aaron and Schwartz also believe, however, that there are a few doctors who recognise their role as rationing agents more explicitly. (It should also be noted that others have disagreed with their assessment; Miller and Miller (1986) argue that British reluctance to use technology so liberally may reflect genuine concerns about clinical inappropriateness, regardless of affordability).

Klein (1989) states that 'rationing by consultants had always been a fact of life in the NHS', as a way of 'disguising political decisions about resource allocation as professional decisions about clinical policy' (p.235). He argues that the medical profession entered an implicit concordat with the government and accepted the task of implicit rationing within a fixed budget in return for continued professional autonomy. The increasing reluctance of doctors to maintain this role in the late 1980s and early 1990s he ascribes at least in part to the Thatcher government's challenge to clinical freedom and demands for more managerial control, which they perceived as the Government reneging on its half of the bargain. It is arguable that, in turn, the politicians perceived doctors' persistent demands for more funding as reneging on *their* half of the bargain.

The role of clinicians in rationing remains a significant element of the current debate, and is examined later in this chapter.

Explicit denial

It is possible for denial of treatment to take place implicitly, for example when waiting lists are so long that the person never receives treatment.

In most cases, however, denial takes the form of explicitly refusing care, either to certain groups or to the population as a whole. Harrison and Hunter (1994) maintain that rationing through the exercise of clinical freedom does not constitute full denial, since some treatment is usually provided, albeit at a reduced level. They go on to say that 'the possibility of complete denial of treatment has arrived on the policy agenda with the creation of the "purchaser/provider split" in the NHS' (p.29). This conclusion can be questioned on two counts: firstly, the exercise of clinical judgement has led to denial of care and secondly, this was happening before the introduction of the NHS market. The example of renal dialysis provides an illustration of both points.

As early as the 1970s renal dialysis was being explicitly rationed in this country; some treatment centres developed strict eligibility criteria, which resulted in some groups - such as blind diabetics - being turned away altogether, even though they would have benefited from dialysis and in many other countries would have been treated without question (Wing, 1983).

Denial of access to renal dialysis for people with kidney failure has a stark result: death. Klein (1995) notes, 'The remarkable fact that the NHS can get away with this politically - that a refusal to save lives does not raise a storm of political protest - demonstrates the positive advantages that central policy-makers can derive from the doctrine of clinical autonomy. For, of course, it is not ministers or civil servants who decide who shall be treated. It is the clinicians concerned' (p.78). Crucially, however, this is not the kind of hidden, individually based kind of rationing described as dilution by Harrison and Hunter; it is rationing by an organised set of criteria, aimed at whole groups of the population, still within the confines of medical judgement but jointly agreed by groups of doctors rather than individual professionals. Although renal specialists may have hoped to operate these criteria without the public becoming too aware of them, they were certainly explicit within the NHS and became explicit outside it as knowledge of the existence of such internal policies spread and individual cases were highlighted in the media.

A further example of explicit rationing before the NHS market is the lack of availability of *in vitro* fertilisation (IVF), which also happens to be one of the most often-cited examples of rationing since the market came into effect. Many health authorities took a definite decision not to provide IVF in the hospitals which they managed, because it was felt to be an expensive and experimental new technology of unproven effectiveness and/or because they did not regard the treatment of infertility as a high enough priority in which to invest new resources.

Although IVF was thus not available in many health authorities, it was still possible for a woman living within such a health authority to be referred to an NHS facility in another district which did provide free IVF. If the hospital accepted the referral, she could thus by-pass the lack of provision in her local health authority and obtain treatment elsewhere. Once the market was introduced, she had to obtain the agreement of her health authority of residence to pay for her NHS treatment, wherever it was provided; thus if a district had an explicit policy not to fund IVF care, all residents were bound by it and there were no geographical loopholes. Nonetheless, pre-market decisions not to fund NHS IVF treatment are still a clear example of explicit rationing, even if some people managed to evade their effects.

Finally, limited list prescribing - preventing GPs from prescribing certain remedies - has been in place for several years now. This differs from the previous examples in that it defines the boundary of NHS provision on a *national* basis, although the criteria for exclusion appear somewhat confused (New and Le Grand, 1996).

Age-limited screening programmes.

The screening programmes for cervical and breast cancer are available routinely to women within specified age groups. In the case of cervical cancer, the age group is broadly defined (16-64), although there is still a clear decision not to provide screening to more elderly women. This is presented as a clinical, technical decision on the basis of effectiveness, in that 16-64 is the age group in which cervical cancer is most likely to occur. However, this is not to say that women over the age of 65 *never* get cervical cancer - it may be less common but it is certainly possible. Thus concealed beneath the explicit technical reasoning is a less explicit (at least to the public) concern with cost-effectiveness. At a population level, the cost of screening women over 65 is judged to outweigh the clinical benefits to a few individuals.

Breast screening is offered to a much more restricted age group. A comment in the Minerva column of the British Medical Journal says, 'Minerva has been following with interest the dispute about agist discrimination in the NHS. What about mammography, currently offered to women aged 50-65? Why not to the over 65s, given that all cancers become more common with age? The lack of research evidence showing a benefit is simply the result of *no research having been done*.' (British Medical Journal, 1994, p.1178) The argument usually given is that in the 50-65 age group breast cancer is more aggressive and therefore early detection is more important than in the over 65 age group. It might be argued in response, however, that early detection of slower growing cancers might achieve better survival rates than the current programme - indeed, there remains considerable technical controversy as to whether earlier detection improves survival at all for the screened age group and concern about the anxiety caused by its high level of false positive results. The political need to be seen to be doing something to counter deaths from breast cancer has an important bearing on the decision to set up a screening programme.

Women under 50 also suffer from aggressive forms of breast cancer, but less commonly than in the 50-65 age group. As with cervical cancer, a decision has been made that it is not cost-effective to screen people with a lower chance of getting the disease in question, even though the outcome for an individual is just as devastating.

Screening intervals are also an illustration of rationing on the basis of cost-effectiveness. The decision to screen for cervical cancer every three years in the former Oxford Regional Health Authority rather than at the national interval of five years reflects a judgement that the national interval rations too much. Yet even a three-yearly interval rations; it simply puts the trade-off between cost and saving lives at a different point.

QALYs

QALYs were being advocated as a systematic rationing method some years before the introduction of the market (e.g. Williams, 1985), although their practical application to real decision-making has lagged far behind the amount of theoretical discussion devoted to them. In 1992, the King's Fund Institute carried out a survey of English health authorities to assess whether the use of QALYs had increased following the reforms (Robinson and New, 1992). This found that 21% of the health authorities which responded had already used QALYs and a further 17% planned to do so. Whether there is any causal link between this finding and the reforms is open to question - it is possible that the use of QALYs reflected purchasers' concern to find tools to help them with their new role of needs assessment. Alternatively, it might have happened anyway as other pressures in favour of explicit rationing built up. New and Le Grand (1996) also note that the analysis of purchasing plans carried out for NAHAT over the same period (see Klein *et al.*, 1996) failed to find significant use of QALYs, although this might have been due to reluctance to acknowledge their use in public.

The current debate

'Talking about rationing has become fashionable again; it's the healthcare equivalent of politicians talking about the Spice Girls. It's contemporary, tough, sassy. But like the Spice Girls, talking about rationing has become the latest chic largely because everyone is talking about it. There is no longer a debate about rationing; it's inevitable - that's a fact. If enough people say it, then it must be true... We are already in danger of creating a whole industry or writers, academics and commentators involved in the science of rationing.' (Hancock, 1997, p.24)

Attempting to summarise the current debate on rationing (let alone contributing to it) is a daunting undertaking. The number of publications has increased dramatically over the past four years, and this section cannot hope to do justice to the quality and complexity of the arguments put forward. However, it is important to understand the main strands of the debate. Much of it is focused on either confirming or challenging the view that rationing is essential, and if it is essential whether it should be explicit or not. Increasingly a view is emerging that there is a role for both explicit and implicit approaches alongside each other, and there is considerable interest in the idea of procedural explicitness. A number of practical proposals for rationing mechanisms have been put forward, although the emphasis of the academic debate is still largely on establishing the theoretical position and the principles on which more rational explicit rationing could be undertaken.

Is rationing essential?

'The doomsday scenario touted by rationing fans cannot be left unchallenged. If we are not careful, we will end up believing the hype - we will be victims of a political rhetoric which says we cannot afford the NHS. We will accept healthcare rationing without any rational arguments.' (Hancock, 1997, p.24)

Although there is widespread agreement that the rationing of health care is essential and inevitable, it is possible to find a few dissenting voices. Those who believe rationing is avoidable do so for two main reasons - firstly that too much money is wasted on ineffective care and secondly that not enough money is spent on health care.

Ineffective care

The first of these arguments is that if only effective treatments were offered, there would be no need to ration at all. (Etzioni, 1991; Roberts *et al.*, 1995) It is argued, on the basis that a need for health care only exists when there is an effective treatment available, that the NHS has ample resources to meet all needs in the population; it has to ration at present only because it wastes so much money on ineffective or unproven therapies for which there is no need.

This argument relies on two fundamental assumptions: firstly, that ceasing to provide ineffective care would not constitute rationing, and secondly, that it would release substantial resources. These are points of such fundamental and entrenched disagreement that they warrant detailed consideration.

With regard to the first assumption, it has already been argued in this thesis that all methods of allocating scarce resources constitute rationing. Not providing treatments which are ineffective could be seen simply as using a sensible criterion for making necessary rationing decisions - Cooper, for example, puts forward an early case for evidence-based care as a way of doing unavoidable rationing 'more rationally, consistently and efficiently' (Cooper, 1975, p.109). The proponents of evidence-based care, however, would argue that we are not here talking about *scarce* resources at all - something which is of no benefit whatsoever cannot be needed and cannot therefore be a scarce good. As Klein and colleagues put it, 'We would not describe the NHS's refusal to allow snake oil to be prescribed as rationing' (Klein *et al*, 1996, p.75). However, the economic definition of scarcity is couched in terms of *demand*, rather than *need* - it is perfectly possible that a treatment could be pronounced ineffective and yet still be demanded or desired by consumers. Their reasons could include disbelief that it is ineffective, superstition, or a feeling that every possible avenue must be tried, however unlikely to work. The prescription of antibiotics for viral sore throats is one common example - anecdotal reports from GPs suggest many patients believe the true reason for denying them antibiotics is to save money, rather than that they are useless against a viral infection. Furthermore, as New and Le Grand (1996) point out, even need and the ability to benefit can be problematic areas: 'Undertaking a procedure which does not improve health may improve well-being in other ways....being "cared for" is valued by a patient even if the treatment is not improving health' (p.36). The placebo effect of technically ineffective treatments is a well-observed phenomenon.

These are perhaps minor objections; if there is genuine consensus that a service is neither needed nor demanded by anyone, it would be fair to argue that its exclusion does not constitute rationing. However, a much more substantial challenge can be mounted to the second assumption, namely that ineffective practice is so widespread and costly that its elimination will put an end to the scarcity of resources. This challenge involves four questions: **Do we know for certain? Who? How much? When?**

Do we know for certain? Whilst there is good evidence that some routinely performed interventions are worthless (see, for example, Enkin *et al.*, 1989), the number proved beyond doubt to have no value is fairly small. Cooper, even though a strong supporter of the principle of eradicating ineffective practice, reports that the New Zealand government has found no justification for the complete exclusion of any interventions currently funded publicly on the grounds of ineffectiveness (Cooper, 1995). The effectiveness of the great bulk of clinical care is neither proven nor disproven and relies on such factors as custom and practice, accumulated experience, and trial and error.

It is argued by the more ruthless proponents of evidence-based care (e.g. Roberts *et al.*, 1995) that only 10-20% of current practice has been proved positively to be effective, and that the rest should be regarded as ineffective until proven otherwise. In fact a greater proportion of interventions may be of proven benefit than this argument suggests. In a retrospective review of all the patients they had diagnosed and treated in one month, Ellis *et al.* (1995) found that '82% were evidence-based (i.e. there was RCT support [53%] or unanimity on the team about the existence of convincing non-experimental evidence [29%])' (p.407). Even if such evidence is lacking, ceasing to provide any kind of care other than that approved by a randomised controlled trial (RCT) would no doubt result in the loss of many effective but formally unproven treatments, in turn resulting in unmet need.

More often than not, evidence about effectiveness is conflicting, studies are poorly designed and there is professional dispute about what the results really mean. Moreover, even the scientific certainties of one generation can give way to doubt in another. Cooper (1975, p.59) cites as evidence of irrational practice the fact that 'the leech bottle managed to survive the First World War in some British hospitals', unaware of future research which would demonstrate leeches' value in reducing blood clotting and swelling, especially in skin grafting.

Who? Even when the evidence commands widespread agreement, it is usually complex and rarely concludes that a particular procedure is of no benefit to anyone. The same Department of Health Effective Health Care Bulletins have led some purchasers to exclude IVF altogether and others to introduce a new IVF service, both in the name of the effectiveness. As Dworkin (1994) points out, there is a world of difference between excluding low effectiveness or 'low yield' treatments and 'no yield' treatments; he argues that those who support the former are in fact accepting a need for rationing. Furthermore, he does not feel there are adequate methods for making such decisions in practice.

Trying to predict likely individual outcomes on the basis of statistical averages is a difficult task (McKee and Clarke, 1995; Bion, 1995; New and Le Grand, 1996). An assessment of effectiveness must also take account of the aim for each individual - Weijer and Elliott (1995) point out that the same intervention may be regarded as ineffective 'if the aim is to cure an underlying disease, but effective if the aim is to keep the patient alive' (p.684).

Interpretation of the evidence in the light of individual patient characteristics is contentious, and may include surreptitious social judgements. Smoking and obesity are both used to justify denying people certain treatments (e.g. heart surgery) on the grounds that the presence of one of these factors would make the treatment less effective than for other people. Equally, treatment for respiratory problems is not very effective if the person lives in damp housing, but denying treatment for this reason would be seen by many as unfair discrimination against people from lower socio-economic groups. There is a danger that we may find ourselves reacting differently according to the degree to which the condition is self-inflicted: the smoker may seem to be a less deserving recipient of respiratory treatment than the resident of damp housing. It may be no coincidence that two of the most common explicitly rationed procedures in this country - sterilisation reversal and tattoo removal - combine low effectiveness or low health gain with the possibility of blaming the patient for their own condition.

Calabresi and Bobbitt (1978) use dialysis as an illustration of this potential for effectiveness and social judgements to interact. 'It does not appear to be tragic in the United States to fail to provide dialysis for a person in whom such treatment is unlikely to work...[this decision determines] who is given a greater chance to live. Yet [it does not seem] to implicate a conflict in American values; it would be otherwise if the kidney were given to a wealthy recipient rather than to a poorer patient on the ground that the kidney was more likely to work in a wealthier man because he could afford to rest, have private nurses, and so forth....It seemed as if the value accorded efficiency would give us a neutral criterion whose application would not direct attention to any inadequate first-order or improper second-order determinations, but would rouse no more than pity and sorrow for those who would suffer. Yet the criterion failed and the ameliorative quality ceased with it, when efficiency-guided distributions correlated with wealth difference' (p.22-4)

How much? Given that there are very few treatments which never work, how much benefit will be gained is another difficult question. Do we say something that has only a 1% chance of success is ineffective? or 10%? or 25%? How do we trade off these chances of benefit with the unpleasantness of treatment that is often involved in such cases? The Child B case is a good illustration of these dilemmas, and of the fact that a success rate of 1% is still 100% success for one lucky individual.

When? To a lesser extent, the question of when benefit will be obtained can be significant. The insertion of grommets for glue ear is often criticised as unnecessary because glue ear will usually get better on its own after a few months. But a few months of poor hearing, falling behind at school and being teased can seem a very long time to a child and his or her parents.

The upshot of these difficulties in establishing conclusively the effectiveness of different interventions is that eliminating ineffective practices does not offer an easy route to avoiding hard allocative choices. Indeed, there is some persuasive argument that pursuing evidence-based care may result in increased expenditure rather than savings, as clinicians realise there are more people they could help (McKee and Clarke, 1995; Sackett *et al.*, 1996). Alan Maynard, himself an ardent proponent of rooting out 'useless interventions', is critical of those who attempt to assess effectiveness without consideration of *cost-*

effectiveness - for him, using effectiveness as a criterion for allocating resources is certainly a means of achieving rationing, not of avoiding it. (Maynard, 1996a, p.21)

An inadequate budget

A second argument is that rationing would be unnecessary if we spent more on health care - society has created a situation of artificial scarcity which it could eliminate. John Harris suggests: 'Any rubric for resource allocation should examine the national budget afresh to see whether there are any headings of expenditure that are more important to the community than rescuing citizens in mortal danger. For only if all other claims on funding are plausibly more important than that, is it true that resources for life-saving are limited.' (Harris, 1987, p.22) This suggests the need for survival should trump every other kind of social need, including education, poverty relief and housing. Harris is clear that his argument refers to life or death situations; when it comes to life-enhancing treatments, he does allow that nations may not be able to afford everything they would like and that some hard choices will therefore have to be made (see below for further discussion). Rawles extends the approach to all kinds of treatments, saying, 'The ethical problem of how to apportion limited resources amongst the needy has been forced on us by arbitrary limitation of health expenditure. Its solution would not be required if health expenditure were higher.' (Rawles, 1989, p.143).

Mullen (1995) argues that a more generous budget would at least 'involve fewer hard choices', and that 'the unquestioning acceptance of the necessity of rationing...is leading to a climate of defeatism' (p.26).

Although it may seem attractive to suggest that extra spending would solve all the problems, given the UK's relatively low expenditure on health, other countries with much higher expenditure are also embarking on rationing. The US is a case in point, with double the UK's share of a much higher GDP devoted to health. This suggests that more money alone is not the answer; it might defer the need to ration for a while, but the mismatch between demand and supply would probably resurface later.

There are those who believe the inevitability of demand outstripping supply is exaggerated. Williams and Frankel (1993) describe 'the myth of infinite demand' (p.13) and suggest that 'the pessimistic belief that the satisfaction of demand is in truth an unrealistic goal' should be 'questioned and abandoned in favour of a more rational determination of health care requirements with the assumption that there may be no need to ration those interventions of undoubted efficacy' (p.17).

Even if demand were finite, this would not necessarily mean all problems of a mismatch with supply could be solved. At one level, the problem is illustrated by kidney transplantation, where rationing is necessitated at least partly by the shortage of supply of donor organs. The utilitarian solution proposed by Harris (1986) that we could at random kill one healthy person in order to save two people who need a kidney transplant, is unlikely to be morally acceptable to most people, as he acknowledges. At a more general level, the necessary provision to meet even finite demands might still not be affordable by the nation. This kind of shortfall in supply would also lead to rationing.

Wordsworth *et al.* (1996) do believe rationing of scarce resources is inevitable, but they argue that the conventional belief that demographic trends and new technology are increasing the pressure on resources is not as well founded as is often assumed. Even if resource pressures are increasing, this is not necessarily an argument for spending more on the NHS. The decision to spend more on health depends both on the benefits which would be gained and the opportunity cost of not spending that money on other public services, such as housing or education. If the benefits are less than the opportunity costs, 'then resources may generate greater benefit overall if invested elsewhere' (p.32).

Rationing is essential

Finding commentators who do believe rationing is essential is much easier than finding those who do not. The majority of books published on rationing in recent years take as a given the scarcity of resources and the inevitability of rationing. Many take the standard line, criticised by Wordsworth *et al.* (1996), that demographic trends, new technology and rising consumer expectations are making the scarcity of resources worse, but the basic argument that society will never be able to afford all possible health care needs and wants is independent of this point. A selection of comments will suffice to illustrate.

'The assertion that "costs and benefits are irrelevant where human life is concerned" reflects an irresponsible approach to health policy. Such an attitude lacks merit in the real world because choices must be made .... Because resources are scarce relative to wants, we do not have the option of evaluating or not evaluating. The only option is whether to evaluate explicitly, systematically and openly, as economics forces us to do, or whether to evaluate implicitly, haphazardly and secretly, as has been done so often in the past.' (Fuchs, 1983, p.48)

'Increasing demands on health services, coupled with limited resources, have created the need to make choices on which services should be developed and which held back. Essentially, this entails the rationing of scarce resources between different geographical areas, types of service and client groups.' (Ham, 1982, p.58)

'There are two certainties in life: the scarcity of resources and death. In the health care industry, the issue is not whether to prioritise but how, i.e. what criteria should be used to decide who will be treated, who will live in pain and discomfort, and who will be left to die.' (Maynard, 1994, p.1)

'There is now a growing awareness that the rationing of health care is unavoidable. With the advance of medical technology, rising incomes and a general growth in health awareness, the demand for health care is growing faster than the resources allocated to it.' (New and Le Grand, 1996, p.1)

'Rationing is a characteristic of all those publicly funded or provided services...where constrained budgets meet unconstrained demands for services.' (Klein at el., 1996, p.9)

'Some claim that the gap between demand and supply will grow ever wider...There is insufficient evidence to support this view. But while there may not be a "crisis" looming, there will always be a need to ensure the fair distribution of finite resources.' (Lenaghan, 1996, p.i)

Is *explicit* rationing essential?

'We should resist abandoning an admittedly imperfect though workable irrationality in favour of a quite spurious rationality which is probably unattainable and certainly undesirable.' (Hunter, 1993a, p.31-2)

The case for implicit rationing

To accept that rationing is unavoidable is not in itself to argue that greater explicitness is desirable, although in practice the two positions often go together. Some arguments in favour of leaving it to more implicit means are considered below.

At a general level, the views of Calabresi and Bobbitt (1978) that society simply cannot live with explicit decision-making have already been discussed. Within a health care context, Appleyard (1992) has adopted a similar view: 'The moral choices of rationing health care are too brutal for society to contemplate' (p.14). These comments do not explore which approach to rationing might produce the best outcome in terms of justice or overall utility; they focus rather on the process itself and question society's ability to live with the ethics of explicit choices.

Coast (1997) has recently argued that society's unease with the consequences of rationing means that utility may be maximised more effectively through continued implicit decision-making. She identifies two sources of 'disutility' associated with explicit rationing: denial disutility and deprivation disutility. Denial disutility may be experienced by citizens taking part in the process of denying care to others; we feel uncomfortable with the choices we are being forced to make out in the open. Secondly, if we ourselves are in need of treatment, we may suffer more if we know a treatment is being withheld from which we could have benefited, than if we remained in ignorance of its availability. Thus we would be happier to be told by a doctor that treatment was not appropriate in our case. This deprivation disutility may extend beyond the individual patient to others who feel altruistically on our behalf, especially close family and friends. Coast argues that, particularly at the micro level, 'greater total utility may therefore result from the equivocation associated with implicit rationing than from the openness and honesty of explicitness' (p.1121)

Harris (1988) explores the outcome of different approaches with specific reference to distributive justice. Should the case arise where no more of the national budget can be allocated to health care and rationing of access to life-saving treatments therefore becomes necessary, he proposes the drawing of lots as a possible way to avoid unjust discrimination. Although this is explicit in the sense that it results in an open decision about who shall be saved, it is implicit in that it relies on chance as the deciding mechanism, rather than a written set of criteria against which a person's claim to health care is judged.

When it comes to treatments which are life-enhancing rather than life-saving, Harris accepts that prioritising on the basis of quality of life may be morally defensible. Whilst he does not commit himself to any one solution for rationing life-enhancing treatments, he continues to reject the QALY approach as inherently ageist and sexist, and therefore unjust. He argues that some people may have a strong claim on resources simply because they have suffered accumulated disadvantage and neglect by society. Justice may dictate that someone who has endured years of remediable pain and immobility should take priority, even though their life expectancy may be low. Harris remains reluctant, however, to make positive practical recommendations for how rationing should be carried out, perhaps indicating the difficulties of reconciling his emphasis on the supremacy of individual rights with acknowledging a need to rank individual claims in some sort of priority order. (It should be noted that his emphasis on the value of each individual life paradoxically leads him to endorse a utilitarian policy of maximising lives - saving two lives must always be better than saving one, even if this involves actively killing one person (Harris, 1985, 1986). His views on the value of life and its dependence on conscious awareness in the person concerned are challenged by the anti-utilitarian Ann Maclean (1993)).

Bagust (1994) defends waiting lists as an acceptable rationing mechanism; they 'could only be eliminated in a service that was heavily over-resourced and therefore inefficient', and the current political obsession with eradicating them is short-sighted and unrealistic. Purchasers and providers need to work together to 'establish realistic expectations of what can be delivered within the limited resources available. Of course, this leads to some difficult and unpleasant choices between competing needs and between groups of needy patients. Attempts to restrict the health service's obligation to treat certain conditions has proved politically indigestible, but the alternative is to continue to rely on our traditional waiting-line system of rationing and to accept extended waiting lists for some conditions as a fact of life. This is indeed a practical and proven approach but needs a determined change of political culture at both local and national level to allow the waiting list to be rehabilitated as a rationing device' (p.17).

Doyal (1993) sees waiting lists in themselves as a form of drawing lots. He says 'patients who possess morally similar prognoses and ages should be randomised. In other words, those who are deemed to be morally equal from the perspective of medical need should also have an equal chance of not receiving treatment for reasons of resource scarcity. If properly administered within a national health service like the one in the UK, waiting lists provide a rational and effective way for this to occur. Here the randomisation is created, so to speak, by the lottery of life' (p.52-3). It could be argued in response that the operation of waiting lists may not always be fair and random, and runs the same risk of systematic discrimination identified for the NHS as a whole by the Black Report (Townsend and Davidson, 1982) and Le Grand (1982).

Doyal continues, 'there will obviously come a point when prognosis and age are so different that randomisation through the administration of fair waiting lists will itself be seen by almost everyone as unfair' (p.53). In these circumstances, the 'fair innings' argument may need to be applied, whereby those who have already lived beyond average life expectancy may be given a lower priority than those who have not. Equally, those who are beyond the reaches of curative care should give way to those who could still benefit from it, although they should still be given appropriate palliative care.

The 'fair innings' approach is also considered by Harris, although he is concerned about where to draw the line - would one argue, for example, that a 30-year-old had more right to treatment than a 35-year-old? (Harris, 1988). It should be noted that in considering the fair innings approach, both Doyal and Harris are in effect contemplating explicit rationing in some circumstances. Indeed, Doyal has more recently been advocating explicit rationing at all levels of NHS decision-making (Doyal, 1997).

Both Mechanic and Hunter challenge the assumption that explicit rationing mechanisms will be more equitable than implicit mechanisms. Mechanic (1995) argues implicit decision-making at clinical level is more sensitive to the complexity of real medical decisions and the needs and preferences of individual patients. He also argues that explicit decisions may prove unsustainable in practice, because they will be resented and challenged by those affected. Hunter is well-known for defending what he describes as 'muddling through elegantly' (Hunter 1993b, p.28) - in other words allowing the current situation of implicit and incremental decision-making to continue. He suggests it would be preferable to leave things as they are until we have developed processes which can capture the complexity identified by Mechanic (Hunter, 1995), and both writers agree that explicit mechanisms are just as vulnerable to subversion by particular interest groups as current methods, if not more so. They therefore reject the claim that explicit methods will by definition be more equitable.

The case for explicit rationing

Arguments in favour of explicitness are commoner than those against. Many of the authors quoted earlier as arguing that rationing is inevitable also take the position that implicit methods are largely unjust and unacceptable; some focus more on their potential inefficiency. However, over the last five years, the debate has developed in range and complexity. As noted earlier, there is a growing body of opinion that some combination of implicit and explicit approaches may represent the best way forward, depending on the level at which rationing is being conducted. There is also increasing interest in making a distinction between explicit rationing decisions and an explicit decision-making process. The extent to which the public should be involved is another recurring issue. Some key examples of these arguments are summarised here, drawing particularly on a recent series of articles in the British Medical Journal.

The philosophical position of those who advocate explicit rationing is not always clear, a source of some confusion and failures in communication. Some generalisations can be attempted: as Harvey (1996) puts it, 'the most likely groups to promote explicit rationing are those holding consequentialist views: in practice the most active in this field have been health economists. Groups with such ethical principles are not only likely to promote explicit rationing, but also technical methodologies upon which to base this rationing. For example, health economists tend to promote technical methods based on efficiency. Those holding deontological views are the least likely to accept that choices about health care must be made' (p.106).

Proponents of QALYs are clear examples of this utilitarian advocacy of technical cost-effectiveness methodologies (e.g. Maynard 1994, 1996a, 1996b; Williams 1985, 1997) Maynard argues that criticisms levelled at QALYs, such as inherent ageism or a lack of concern for distributive equity, can be dealt with by making an equity adjustment to the formula as an explicit political decision (Maynard 1996a, 1996b). Philosophically, this attaches negative utility to the inequitable distribution of health care, and thus elaborates utilitarianism to incorporate distributive concerns. However, the presentational emphasis on efficiency can sometimes give the impression that QALYs themselves are a philosophically neutral technical device, which can be adjusted to take into account specific moral perspectives.

One of the more provocative manifestations of Maynard's conviction has been to argue that if doctors continue to provide treatment to patients which is not cost-effective (on his terms), this 'inefficient treatment' is 'unethical and should be construed by employers as prima facie evidence for dismissal.' (Maynard, 1996a, p.21)

Far from trying to modify QALYs to avoid the charge of ageism, Williams has recently proposed that age should be an explicit factor in making rationing decisions (Williams, 1997). This is both because age affects people's capacity to benefit from treatment, and also because older people are more likely to have had a 'fair innings', which might justify giving resources to younger people not yet in this position. He states clearly the belief that 'the values of the citizenry as a whole must override the values of a particular interest group within it' (p.822) He is not alone in proposing age as a rationing criterion; Callahan (1987) has also argued for limiting treatment for elderly people, and Menzel (1990) discusses our 'duty to die cheaply' (p.190).

(It is worth noting at this point that not all utilitarians support a QALY-based approach, precisely because it 'violates the principle that utilitarianism seeks to maximise the *autonomous preferences of individuals*...and substitutes rather the vicarious preferences of unaffected individuals' (Harvey, 1996, p.99). This difference of opinion is illustrated in the debate between Culyer (1997) and Harris (1997) as to whether the purpose of the NHS is to maximise the health of the whole community).

Many commentators who favour explicit decision-making are sceptical that some purely rational formula can be found which will make decisions painlessly on our behalf. Fuchs (1983, p.48) makes this point clearly: 'neither scientific data nor economic analyses are sufficient for resolving these policy dilemmas'; they can 'make explicit the distributional implications of any policy' but they cannot say what the policy should be. Klein *et al.* (1996), New and Le Grand (1996) and Lenaghan (1996), to name but a few, all argue that rationing health care is ultimately a series of political, moral and clinical judgements, for which no technical 'quick fix' can substitute.

This conclusion leads different writers in different directions. Doyal (1997) argues that explicit rationing at both macro and micro levels within the NHS is the only way to ensure that the main principle of the NHS (namely equal access to health care based on equal need) is adhered to. He suggests that 'there is too much secrecy in British public life already' (p.1118).

New (1997) and New and Le Grand (1996) favour national determination of the full range of NHS responsibilities; although some services might be excluded altogether, health authorities would be required to make available at least some of every service agreed to be an NHS responsibility. This addresses the 'what?' question, and ensures there is some degree of geographical equity; the issue of 'how much?', argue New and Le Grand, should be decided at health authority level, whilst the 'for whom?' question is a matter for individual clinical decision-making, because of the complexity and heterogeneity of medicine. Although doctors are recognised as 'the ultimate rationers' (p.71), their accountability can be strengthened; society can agree the principles which doctors should use in making their decisions (for example, extent of ill health and cost-effectiveness) and can institute improved monitoring of medical decision-making.

Klein (1997) rejects the idea of any national determination of NHS responsibilities, partly on the basis that if local health authorities still have discretion in how much of a service they will purchase, inconsistency and arbitrariness will remain and the problem of inequity will therefore not have been solved. However, his proposals are similar to those of New and Le Grand. He, too, recommends that micro rationing should be carried out implicitly by doctors, but with more explicit criteria for them to apply and improved collective professional accountability (Klein *et al.*, 1996). The final sentence of this text is important:

'The best we can hope for is to strive to improve the process by which we reach the decisions' (p.139).

This reflects the growing interest in procedural fairness, a theme which is examined particularly by Lenaghan (1996, 1997a) but which also runs through many current contributions to the debate, particularly those seeking to reconcile some elements of implicit and explicit decision-making at different levels of the organisation. Lenaghan (1996) argues that we need to consider developing fairer mechanisms for rationing because it is already becoming more explicit but in an *ad hoc* fashion. A rights-based approach would establish a national set of procedures and criteria (especially health care on the basis of need) to inform local decisions about how much to purchase and clinical decision-making between individuals. Patients would have 'rights to be heard, to consistent, relevant and unbiased decision-making, to be given reasons for decisions and to have a refusal of a service or a complaint independently reviewed' (p.94). There might also be national guidance on services which should normally be excluded from NHS provision (although this should not be a blanket exclusion).

An alternative way of describing this approach is to say that it would be explicit a) that rationing was taking place and b) by what criteria, even if individual decisions themselves were taken implicitly. If challenged, providers would be required to demonstrate that their decision had been taken in a correct and fair way, placing the emphasis on the process rather than directly on the content of the decision.

The procedural fairness approach lays some stress on the role of national government in both acknowledging and supporting explicit rationing, and taking a lead in defining the responsibilities of the NHS and the criteria by which decisions will be judged to be fair or otherwise. Lenaghan (1997b) and New (1997) both argue that this is essential to legitimise the process of rationing and to ensure that it takes place as equitably as possible.

Hunter (1997b) has been critical of this approach, on the grounds that rationing is too complex to be dealt with at national level, and that asking the government simply to take a lead without doing any rationing is a fruitless exercise. It would not help address the real problem of how to do it locally.

However, the approach suggested by Harrison and Hunter (1994) bears surprising similarities to the procedural fairness approach - another example of how people on apparently opposing sides of the rationing debate in fact use many of the same arguments. Unsurprisingly, given Hunter's views on the potential injustice of explicit rationing, they propose a very cautious approach, offering different 'scenarios' to help policy makers think about the way forward for rationing. One of these proposes local authorities as purchasers of health care, with freedom to choose their methods of rationing. 'Some might choose to establish specific local health care rights for individuals resident in the District, whereas others might prefer to establish broader objectives, within which waiting lists and clinical freedom would dispense rough justice' (p.67). However, they do add two explicit constraints to local authority freedoms under this scenario - there would be a prohibition on the use of purely social judgements (e.g. on the basis of lifestyle) in deciding entitlement to services, and there would be a requirement to pursue equity of outcome. The other scenario proposes an explicit, nationally defined minimum package of health care rights.

More recently, Harrison (1997) has proposed local government as an alternative to centralised rationing; purchasing responsibility would be transferred to local authorities, 'underpinned by strategic, centrally determined rules', which he anticipates would enable 'the logics of democracy and equity to be reconciled' (p.973). This, too, springs from a concern with procedural fairness, to be realised through democratic control.

Both Harrison and commentators on the ostensibly opposing side of the debate (Lenaghan, 1997b; New, 1997; Doyal, 1997) favour greater public involvement in decision-making, especially through citizens' juries, as an integral part of improving procedural fairness. Some legal writers, such as Newdick (1995) and Teff (1994), also favour greater public and patient participation, although Teff prefers the idea of a therapeutic alliance based on trust and honesty rather than the more adversarial formulation of 'patients' rights'.

Heginbotham (1992) and Klein (1992) both express caution about opening the floodgates of public debate about detailed priority-setting. Both suggest public involvement should be confined to decision-making at the broader policy level, with micro rationing remaining a matter primarily for clinical judgement, based on evidence of effectiveness and carried out within the constraints imposed by public policy decisions.

A more theoretical consideration of public involvement is offered by Dworkin (1994). He has proposed a 'prudent insurance' principle for trying to arrive at an equitable and affordable package of health care. He imagines a quasi-Rawlsian world in which five conditions hold: wealth is justly distributed; information on the costs and effectiveness of medical care is available to all; people make decisions rationally; parents place their children's interests on an equal footing with their own; and no-one has any information about social or genetic factors on which to predict their future health care needs. In this ideal insurance world, people would make rational decisions about which conditions or interventions they would insure themselves for - for example, most would probably not take out insurance to cover keeping them alive in a persistent vegetative state, or they might choose not to insure for further life-saving treatments when they are known to be in the last four months of life. The things which people would regard it as prudent to insure for are the things which should be covered by a basic comprehensive package of health care. He suggests the model could be used both by policy-makers seeking to define essential health care provision and for stimulating public discussion.

Daniels (1985) has also elaborated a concept of 'prudential or rational savings' (p.103), whereby it is assumed that 'prudent deliberators...would seek a health care and long term care system that protected their normal opportunity range at each stage of their lives' (p.99). Menzel (1990) formulates a rather harsher model of 'presumed prior consent' (p.23), which makes some controversial assumptions about the rights of newborns. The fact that they have no ability to consent to trade-offs between future risk against current savings brings into question their status as persons and gives others the right to make trade-offs on their behalf. Menzel's blunt approach highlights the danger of theoretical assumptions about what the public and individual people would choose - a danger which, it should be said, most of these authors themselves recognise.

The Government response

The government's response to the explicit rationing debate has consistently been evasive, because of the undoubted political sensitivities involved. The word 'rationing' has been avoided and replaced by 'priority-setting'; even 'priority-setting' has been firmly declared to be a local health authority responsibility or a matter for clinical judgement. However, the NHS Executive has shown signs of publicly accepting the need to make explicit choices. It has given some very broad national guidance, for instance in an Executive Letter which makes clear that health authorities will be set targets for reducing ineffective care and shifting investment into more effective treatments (NHSE, 1995). The White Paper 'The National Health Service: A Service with Ambitions' (Secretary of State for Health, 1996) reaffirmed the view that resources are limited and choices must be made, but that the assessment of need and the setting of priorities for resource allocation were primarily the responsibility of purchasers, within a broad framework established by central government. At provider level, clinicians should retain responsibility for deciding priorities between individual patients, but their decisions should be based on evidence of effectiveness.

Although the emphasis is on local responsibility, the White Paper does ask the NHS Executive to work with health authorities, NHS Trusts and the health care professions to establish a more systematic means of setting priorities at every level of the service. It states that no clinically effective treatments should be totally excluded from health authority funded services as a matter of principle. The availability of such treatments may be limited, but there should always be the possibility of making an exception in cases where clinical need can be demonstrated. Clinically ineffective treatments may, by implication, be totally excluded.

Although there is nationally funded and supported work to produce research evidence, such as the Effective Care Bulletins, the task of <u>interpreting</u> the evidence to decide what is effective or ineffective is still largely left to local decision-making. Indeed, if commissioners are not allowed to ban any treatments which carry some remote possibility of being effective, their only options may be rationing by limited numbers of cases or by allowing only 'exceptional' cases, both of which will place the burden back on clinicians facing individual patients (Harrison and Wistow, 1992; Klein *et al.*, 1996).

Perhaps the most significant recent development has been the production of the discussion document 'Priority Setting in the NHS' (Academy of Royal Colleges *et al.*, 1997). This is a report by the Priority Setting Working Party established following a workshop on priority setting in 1995. The Working Party was set up under the joint auspices of the Academy of Medical Royal Colleges, the British Medical Association, the National Association of Health Authorities and Trusts and (significantly) the NHS Executive, and all four bodies gave unanimous support to the report. The NHSE's support is significant because the report makes clear that it sees 'priority setting as a synonym for rationing' (p.6), a departure from the NHS Executive's normal preference for the term priority setting. It also recommends that the need to ration and to make trade-offs between values should be more explicitly acknowledged by all parties, including central government.

The Working Party does not recommend a national priority setting forum, although it suggests that this should be considered as an option as part of the public debate. It does argue for greater national consistency in the process of priority setting, if not the content, suggesting support for the idea of procedural fairness.

As noted in the previous chapter, the Labour Government has so far shown clear reluctance to get involved in the rationing debate. This is explored further in the final chapter.

The explicit rationing debate and the market

It should be noted that arguments supporting explicit rationing are generally not based on the idea that explicitness is inevitable, even if rationing is inevitable. Rather they are based on concerns about democracy, efficiency and equity, and the belief that we ought to be explicit in order to address these concerns. In other words, the assumption is that we have a choice about whether or not to be explicit. This has important consequences for the question of whether explicitness is a logical consequence of the market.

Having said this, there is also a widely held view (amongst many of the same authors) that the market is making rationing more explicit but in an unplanned way, and that this is one reason why we need to address it more systematically. For example, Lenaghan (1996) argues 'whereas inequity may have been an unwanted occurrence in the NHS of the past, it is now actually built into the system. The logic of the internal market, the purchaser-provider split, the freeing of trusts, the devolving of powers to individual health authorities all have led to geographical variations in access and provision of health care services' (p.11). The choice is thus to some extent made for us.

New and Le Grand (1996) suggest that 'developing new forms of rationing was not a specific government objective of the 1991 reforms. However, such forms did develop, largely as *ad hoc* responses to the reforms.' They describe the reforms as 'the catalyst for more significant change' in approaches to rationing (p.10).

Klein *et al.* (1996) imply that the greater explicitness inherent in the purchasing role was indeed intentional: 'decisions about resource allocation would therefore have to become more explicit as health authorities chose what package of health care to buy and what not to buy....That, at any rate, was the assumption.' They then go on to explore the 'apparent failure to follow the logic of the 1991 reforms', and analyse why it was never a realistic expectation (p.50).

The very nature of the disagreement about what rationing is and whether and how to be explicit provides some evidence of its high political and ethical costs, and suggests there is no easy logic between the establishment of a market and greater explicitness. The Conservative government's failure to acknowledge explicit rationing fully as an integral component of the system it put in place may have contributed to market failure - intervention in decisions supposedly devolved to purchasers when they became too awkward is a clear example of non-market decision-making running in parallel with the formal system.

On the other hand, there is a very strong movement in support of explicitness in priority setting, despite the reluctance in some quarters to call this rationing. It is important to note that this movement was underway before the market was established and is being experienced in other countries with different systems of funding and provision (see the following chapter). This suggests that it cannot be ascribed wholly to the creation of the market.

It is undoubtedly true, however, that the movement towards greater explicitness has gathered momentum since the advent of the purchaser provider split in this country. The volume of publications and conferences has increased, media reports of cases of explicit rationing continually hit the headlines, and discussion of explicit rationing has moved from being a minority interest to a mainstream preoccupation. Is this pure coincidence - a continuation of an independent trend? Do the media and health care professionals inaccurately blame the market as a convenient scapegoat, as part of wider criticism of Conservative reforms of the 1980s and 1990s? Or is there some genuine causal relationship at work? Has the market led to developments in explicit rationing which would not have happened before, or made them happen faster than they would otherwise have done? These questions are addressed in subsequent chapters.

Chapter Four

International experience of markets and explicit rationing

Before examining the evidence emerging from the case study health authorities, it is helpful to set UK experience within the wider international context. The introduction of a quasi-market in the NHS took place relatively recently, and there may be lessons to be learnt from other countries undertaking similar reforms. This chapter will therefore summarise the available evidence from other countries on the structure of healthcare systems and on the extent of different forms of rationing, and look for any evidence of interaction between them. Different possible combinations can be sought:

- countries which have explicit rationing but no market or quasi-market;
- countries which have a market-based system but no explicit rationing;
- countries which have both;
- countries which have neither.

Ham (1995) identifies three groups of policies which emerged internationally during the 1980s as responses to financial pressure:

- budgetary incentives for efficiency (e.g. co-payments and charges, fixed payments to providers)
- strengthening management (including doctors in management, audit, guidelines, quality management)
- convergence between markets and centralised planning and regulation - traditionally tax-funded systems have sought to import market forces, whilst market systems such as the US are experimenting with more regulation.

Alongside these three policy areas, he notes that continuing commitment to public funding has led to a fourth group of reforms, namely attempts to define the range of services. These can take the form of a list of core services, a list of excluded services, or the use of clinical guidelines to target services at particular groups and individuals.

For the purposes of this chapter, this typology is reorganised as follows:

1) Rationing of demand - e.g. by imposition of charges and co-payments (ability to pay)
2) Global rationing of supply - e.g. capping the budget (cash) or limiting facilities (in kind)
3) Explicit selective rationing of services - e.g. service exclusions. This can take place at three levels:
    - National level, by politicians
    - Local level, by politicians
    - Local level, by managers with political mandate
4) Structure of system - especially whether adopting quasi-market reforms
5) Nature of funding - especially whether tax-funded or not

This will enable a table to drawn up, showing the extent to which different kinds of rationing are related to different kinds of structure and funding. (The mechanisms of rationing demand and supply are many and complex - for instance, the World Health Organisation (1996) identifies 8 different forms of cost-sharing by patients, including extra billing, deductibles, co-payments, co-insurance and benefit maximums. For present purposes, a simplified approach is sufficient to point up differences in national strategy).

For comparison to be meaningful, it is sensible to look mainly at countries which are similar to the UK - in other words, industrialised western democracies. This is not to say developing countries or countries in Eastern Europe and the former Soviet Union, for example, have nothing in common with developed nations as far as health policy is concerned. The World Bank has commented that developing countries need to improve the targeting of scarce health care resources, particularly at cost-effective interventions, and should establish at least a minimum package of services available to all, including maternity care, family planning, control of tuberculosis and sexually transmitted diseases, and treatment for common diseases of childhood (World Bank, 1993). Defining a minimum package, should they choose to take the World Bank's advice, would be very much an explicit rationing activity, even though the context of the level of resources available may be radically different. On the other hand, it is precisely this different resource context which makes comparison invidious; the intricacies of rationing in a sophisticated quasi-market are worlds away from the struggle to establish even the most basic level of service in developing countries.

The former communist countries of Eastern Europe and the former Soviet Union are still in a state of transition. The chaotic nature of emerging market forces and of priority-setting mechanisms make comparison with western health services difficult, even though there may be distant parallels with the move in the UK from central planning to market processes.

If the search is confined to OECD nations, a problem emerges straight away in trying to find countries which do not have some form of market system, whether on the purchaser or the provider side, or both. Distinguishing a regulated but real market from a quasi-market is not always straightforward, especially under circumstances of convergence. The problem is complicated by the fact that different commentators may offer varying interpretations of how each system is organised - what one describes as compulsory social insurance another may describe as a system funded out of general taxation, for example. It is also difficult sometimes to define exactly what is public or private - social insurance schemes may be run by private companies, but may be so constrained by government regulations that they become 'quasi public agencies' (de Roo, 1995, p.46). For current purposes, only those social insurance schemes which are virtually universal and which are run by government, local or national, will be counted as tax-funded.

To try to establish the relevant facts, the situation in a number of EU/OECD countries has been analysed[1]. A summary is presented in the table below. This is followed by a more in-depth look at four countries particularly well-known to be experimenting with competition and/or explicit rationing. In each case, the information is linked back to the numbered typology already discussed. The chapter concludes with a discussion of the significance of these findings.

---

[1] A number of sources have been used to provide general descriptive data about the countries analysed. These are listed at the end of each section on the four countries analysed in depth; where a particularly significant point is raised, the reference is also given at the relevant point in the text. The sources used in compiling the summary table are:
Arvidsson, 1995; Bach, 1996; Bloor and Freemantle, 1996; Brommels, 1995; Cabases, 1995, 1997; de Roo, 1995; Economist Intelligence Unit (EIU), 1993-7; Gallo, 1997; Ham *et al.*, 1990; Health Services Management Centre, 1997; Iglehart, 1986, 1991; Klein *et al.*, 1996; Lønning, 1997; Nestman, 1996; NLN, 1990; NOU, 1987; OECD, 1994; Parston, 1994; Pfeiffer, 1996; Richard and Schönbach, 1996; Rosleff and Lister, 1995; Schwartz and Busse, 1996; Working Group on Prioritisation in Health Care, 1995; World Health Organisation, 1996.

4.1 SUMMARY OF INTERNATIONAL COMPARISONS

COUNTRY	DEMAND RATIONING	SUPPLY RATIONING	SELECTIVE RATIONING	QUASI-MARKET	TAX-FUNDED
Australia	Yes	Yes	Yes	No	Yes
Austria	Yes	No	No	No	No
Belgium	Yes	Yes	No	No	No
Canada	Yes	Yes	Yes	No	Yes
Denmark	Yes	Yes	Yes	No	Yes
Eire	Yes	Yes	No	No	Yes
Finland	Yes	Yes	Yes	Yes	Yes
France	Yes	Yes	No	No	No
Germany	Yes	Yes	No	No	No
Greece	Yes	No	No	No	No
Iceland	Yes	Yes	Yes	No	Yes
Italy	Yes	Yes	Yes	Yes	Yes
Japan	Yes	Yes	No	No	No
Luxembourg	Yes	No	No	No	No
Norway	Yes	Yes	Yes	No?	Yes
Portugal	Yes	Yes	No	No	Yes
Spain	Yes	Yes	Yes	Yes	Yes
Switzerland	Yes	Yes	No	No	No
US private insurance	Yes	Yes	No	No	No
US, Oregon - Medicaid	Yes	Yes	Yes	No	Yes
Netherlands	Yes	Yes	Yes	Yes?	Yes?
New Zealand	Yes	Yes	Yes	Yes	Yes
Sweden	Yes	Yes	Yes	Yes	Yes

Oregon, US

The State of Oregon is one of the best-known pioneers of explicit rationing, and the results have been extensively documented elsewhere. It remains the only place where a systematic policy of explicit rationing has actually been implemented. (For an up-to-date summary, see Klein *et al*, 1996; Ganiats and Kaplan, 1996).

The US is one of the few examples of a largely free market healthcare system, with a minimum safety net for the very poor and social insurance for the elderly. Although most purchasing is in fact done by third party payers (insurance companies, or Medicaid/ Medicare), these do not function in the same way as agency purchasers in a quasi-market, the main reason being that insurance companies are independent and have not traditionally had fixed budgets. The ability to pass costs back to the consumer (especially employers who pay premiums for their workers) in the form of increased premiums and higher co-payments has meant that there has been little incentive to carry out explicit rationing.

Only in more recent years, as employers have begun to resist price increases, has there been any move to impose fixed budgets on providers (for instance through the Health Maintenance Organisation system). This has been managed largely through efficiency savings and implicit rationing rather than the development of a list of excluded treatments or groups of patients, although it is worth noting that prior approval mechanisms operated by the insurance companies to vet each proposed admission may in practice have the same effect. The growth of arrangements in the US such as HMOs and preferred provider organisations is often taken as evidence of convergence between market-based systems and centrally planned and regulated systems. (Ham, 1995). There is now substantial pressure within the market towards cost control, which is in turn leading to demands for greater regulation and more planned rationing of services.

An already well-established rationing mechanism in the US is the exclusion of groups of patients from care - those who cannot afford insurance but who are not poor enough to qualify for Medicaid, as well as the substantial population of illegal immigrants. Although this is a recognised phenomenon, it has never been an intentional decision, but is seen merely as the outcome of market forces; it would be difficult therefore to define it as an explicit rationing mechanism, even though there may be those who offer an after-the-event rationalisation of it as a just outcome.

It was this gap in coverage, coupled with spiralling Medicaid costs, which prompted the State of Oregon to review its system. In early, crude attempts to limit Medicaid expenditure, a child was refused a bone-marrow transplant operation, and died as a result. The shock caused by this incident led to a fundamental re-examination of the principles of priority-setting, with the aim of defining a basic package of services to which all uninsured people would be entitled. A specially appointed commission was set up to undertake this work; after consulting as widely as possible with the local community about their values, and taking expert clinical and economic advice, the commission came up with a ranked list of treatment/condition pairings. The first list, drawn up in 1990, was never published and contained many errors and counter-intuitive results, such as ranking treatment for thumb-sucking above treatment for AIDS. A revised list, placing greater emphasis on the subjective judgements of members of the Commission and less emphasis on cost, was published in 1991 and contained 709 treatment/condition pairings. Federal approval for implementing the list was originally withheld on the grounds that it discriminated against people with disabilities. Two further revisions were undertaken, and the fourth list (with 688 pairings) was accepted in April 1993, and implemented in February 1994.

Although in 1991 the Oregon legislature decided to fund only 587 of the 709 pairings then on the list, it in fact had to vote extra resources to enable it to purchase this level of service, and the kinds of procedures excluded were similar to those excluded by some UK health authorities (i.e. treatment for very minor conditions, or treatments with little or no chance of a successful outcome). For instance, numbers 708 and 709 on the 1991 list are life support for extremely low birth weight babies (under 500g) of less than 23 weeks' gestation, and for anencephalic babies.

Oregon's pioneering work on rationing has clearly not resulted from the existence of a quasi-market, since the US does not have one; it results from the inflationary pressures arising from a free market in healthcare, where even state-funded elements of the service have offered unlimited coverage paid for at the going market rate. However, this is in itself instructive when trying to assess the likely outcome of trying to introduce more 'real market' forces into previously planned systems. Healthcare authorities in the US are seeking to control costs precisely by moving away from market forces towards greater intervention and regulation. This could therefore support the thesis that a move in the opposite direction could make resource pressures worse and thereby increase the likelihood of explicit rationing, as in Oregon. The Oregon example also demonstrates that actually implementing explicit rationing requires planned state intervention - it is significant that the only element of the system where such explicit rationing is happening is the tax-funded element.

Restricting the package available for Medicaid recipients has enabled previously uninsured people to be given coverage. Oregon is therefore trading off existing comprehensiveness in return for new universality. In other countries considered below, the question is somewhat different - universality and comprehensiveness are both already established, but can they be maintained? In practice, they may reach the same answer as Oregon - if they want to keep universal access, they may have to sacrifice comprehensive provision.

US private system as a whole	Oregon Medicaid
1) Demand rationing - yes, co-payments, exclusion by ability to pay	1) Demand rationing - yes, co-payments
2) Supply rationing - yes, fixed budgets	2) Supply rationing - yes, fixed budgets
3) Selective rationing - no	3) Selective rationing - yes
4) Quasi-market - no	4) Quasi-market - no
5) Tax-funded - no	5) Tax funded - yes

(Ganiats and Kaplan, 1996; Klein *et al.*, 1996; Ham, 1995)

Netherlands

Social insurance is compulsory in the Netherlands for employees below a certain income and for people dependent on social insurance. Others rely on private insurance or voluntary participation in a sick fund. There has been little in the way of co-payments for people insured with a sick fund, apart from having to pay the difference between the cost of a drug and its generic equivalent. Those privately insured have varying co-payment arrangements.

Rising premiums became a problem in the 1980s and led to increasing requirements for public subsidy, so the government began to investigate ways of containing costs more effectively. It has in the past attempted to impose some central planning and budgeting on the healthcare system, but the fact that most providers were private and independent made it difficult to enforce. The government-commissioned Dekker Report (Commissie Structuur en Financiering Gezondheidszorg, 1987) recommended that structural reform should focus on funding rather than provision, and that there should be greater reliance on market forces for effective cost-containment. (It is interesting to note that the Dekker report preceded 'Working for Patients' by some time).

A major aim of the report was to enable sick funds to compete with each other and with private insurers. This focus on competition amongst purchasers is unusual, although it is being tested in some social insurance systems such as Germany and Belgium and has similarities with the potential competition between health authorities and GP fundholders if the UK. The proposals were:

- to give everyone mandatory insurance rights, ending socio-economic differentiation of entitlements;
- to make this mandatory insurance cover all long term care and about 80% of existing sick fund coverage for acute care. The remainder would be funded by either voluntary additional insurance or by direct payments, whichever individuals prefer;
- to split premiums into two parts - one income-dependent, and the other a flat-rate ('nominal premium'). The income-related element would be collected by the central government revenue department into a central fund and then disbursed to insurers. This funding would deliberately fall short of total health care costs, so that insurers would have to meet the rest from the nominal premium, raised directly from subscribers. Allowing insurers the freedom to determine the level of the nominal premium would encourage them to be efficient and compete with each other by offering a lower premium;
- to alter the legal status of sick funds so that they could operate outside their previous geographical areas, offer private insurance, undercut officially fixed tariffs and withdraw contracts from providers they did not feel were offering acceptable quality or value for money. At the moment, they are legally obliged to offer contracts to all licensed providers.

In practice, several elements of the report have not been implemented. By the end of 1994, a Central Fund for income-dependent premiums had been established and the nominal premium had been introduced. However, the idea of a split between mandatory and voluntary insurance components is subject to continuing controversies, and all political parties have abandoned the idea of a single comprehensive mandatory scheme. Instead, the current government proposes a mandatory scheme for long term care (with fully income-dependent premium), another mandatory scheme for acute care, including a direct annual payment for the use of health services, and voluntary insurance for the rest.

Legal barriers to competition between insurers have been removed and regional monopolies abolished. Regulations to allow insurers to withdraw funding from unsatisfactory providers are being introduced gradually, but the system of funding providers through budgets has crucially not yet been replaced by competitive contract negotiations.

Dutch involvement in explicit rationing is at least partially connected with the structural reforms of the Dekker Report. The report of the Dunning Committee (Government Committee on Choices in Health Care, 1992) was intended to advise on which aspects of health services should form part of mandatory social insurance arrangements, given the government's aim of controlling the costs of public expenditure on insurance for dependent groups.

Van de Ven (1996) examines the complexity of trying to combine regulated competition and fixed budgets in healthcare; if the budget is exceeded, government intervention will be required to fix prices and volume, which he argues 'strongly reduces of even eliminates (the potential for) regulated competition.' However, the problem with allowing even a regulated market greater freedom to determine total expenditure is that this may interfere with the government's responsibility (in a comprehensive service) to ensure access for everyone. 'Access to care for sick and low-income people implies cross-subsidies from the healthy and high-income people...If more care becomes available to the rich, then also the decent minimum that has to be available for everyone, probably will be set at a higher level. Therefore the subsidies and also the income-related contributions for the compulsory health insurance will go up' (p.67). Thus the government's concern to ensure equitable access may require it to put an upper limit on public health care expenditure.

It is in this cost-limited context that the Dunning Committee recommended that treatments of low effectiveness and high cost should be excluded from the basic minimum package for compulsory health insurance. In order to be included in the basic package, treatments would have to pass through four 'sieves': 1) Is the care necessary to enhance participation in social life? 2) Is it effective? 3) Is it efficient? and 4) Is it something which could be left to individual responsibility? Only treatments which pass through all these filters would be provided. On the strength of this approach, some marginal services have already been excluded, including homeopathy, adult dental care and new technology until it has been properly evaluated. However, van de Ven (1995) reports that 'the Committee has applied its criteria to several forms of health services and concluded that it is not a simple matter, on the basis of these criteria, to leave complete services or parts of them out of the basic benefits package. The major reason for this is that effectiveness of care has to be considered in relation to the medical indication and the condition of the patient' (p.789) The upshot of this is that the Committee has not in fact come up with a clearly defined basic package, and has instead chosen to focus its efforts on the development and application of guidelines to ensure appropriateness of care at the individual level.

This summary poses two questions. Firstly, to what extent does the new system in the Netherlands really constitute a quasi-market? The failure to implement some substantial parts of the proposals, coupled with the problem identified by van de Ven of trying to maintain a fixed budget, suggest that the reforms are fairly limited. There is also the fact that the sick funds remain independent, rather than being state agencies as one would expect in a true quasi-market. However, the change in funding arrangements means they are now dependent on government for a substantial proportion of their budget and are subject to much more universal regulation than before. On balance, therefore, the Dutch system could be described as an evolving quasi-market that is not yet fully functional.

Secondly, to what extent have the recommendations of the Dunning Committee really resulted in explicit rationing? Again, the answer is that its scope is very limited.

The relationship between the market proposals and the explicit rationing proposals is complex. In one sense, the Dunning Committee has clearly been part of the market reforms, in trying to ensure the market does not result in inequitable access and coverage. However, it is important that the spirit of its enquiry was to make sure people were included in coverage, given a fixed budget, rather than deliberately to exclude. Furthermore, it was a planned central government initiative, rather than a natural outcome of market processes, and has been pursued in parallel with the market reforms.

1) Demand rationing - yes, co-payments and annual payment
2) Supply rationing - yes, fixed budgets and central planning, but limited success
3) Selective rationing - yes, but limited success at national level
4) Quasi-market - a qualified yes
5) Tax-funded - yes, mainly, but administered by sick funds

(Commissie Structuur en Financiering Gezondheidszorg, 1987; Government Committee on Choices in Health Care, 1992; de Roo, 1995; van de Ven, 1995, 1996; Klein *et al.*, 1996; WHO, 1996)

New Zealand

The election of the right-wing National Party to power in 1990 heralded a dramatic change in the structure of New Zealand's tax-funded national health service. Until that point, the main mechanisms for cost containment had been the use of fixed budgets, public sector pay freezes and extensive user co-payments, especially for drugs and for GP services, which are fee-for-service with partial reimbursement. (Co-payments were extended in 1992 to out-patient and in-patient services, although subsequently withdrawn again for the latter).

In line with its ideological preferences, the government decided upon rapid restructuring along market lines. From July 1993, the service had a purchaser/provider split extending to cabinet level, with the health minister overseeing the four regional health authorities (purchasers) and the minister of crown health enterprises overseeing providers. The crown health enterprises (CHEs) were amalgams of the former hospital boards and community services, and were limited liability companies. Their contracts were therefore commercially secret and legally binding, in marked contrast to the UK.

Ashton (1996) has been evaluating these reforms within Williamson's 'organisational failures' framework, and concludes that there is some evidence to support his thesis. She notes that the legally binding nature of contracts has meant every effort has been made to make them as complete as possible. 'As transaction cost theory predicts, this has clearly been a prolonged and costly process' (p.21). (The British Medical Journal (1996) has reported an estimate of a 40% increase in administration costs arising from the reforms). The amount of detail in service specifications depends partly on the kind of service being purchased; primary health services, for example, were described in much broader terms than surgical services. Although the surgical contracts were still effectively block contracts in 1994/95, they contained considerable detail on casemix and prices, and contract negotiations were protracted. Ashton suggests that purchasers are dissatisfied with the competitive structure and would prefer to move towards a more collaborative approach. However, providers may resist this as they feel vulnerable against monopsony purchasers and fear collaborative arrangements would perhaps give disproportionate advantage to purchasers.

Ashton does not report any evidence of explicit rationing resulting from the process of drawing up contracts within New Zealand's quasi-market. However, there has been nationally led exploration of the idea. As in the case of the Netherlands, this was undertaken in parallel with the market reforms by a national commission, and its original intention was to draw up a list of core services to be included in public health coverage. Hunter suggests that, unlike the Netherlands, the motivation was to promote private insurance and restrict public expenditure by finding services which could be excluded, rather than making sure the population had protected access to comprehensive basic coverage, despite limited public expenditure. The net effect may be the same, but the difference in perspective could be important.

It is not absolutely clear whether Hunter's assessment of the motivation is correct; the government's own white paper says 'in the past, rationing has been done informally and often without public scrutiny or control. Defining a set of "core health services" more explicitly will help ensure that the services the public believe to be the most important will be provided. It will also acknowledge more honestly that there are limits to the health services we can afford' (New Zealand Minister of Health, 1991, p.80) This would suggest that the government did see more positive reasons for identifying core services, although of course the reasoning given in public documents does not always reflect underlying political aims, if these may be unpopular. Cooper (1995) summarises the ambiguity thus: 'The core, in short, was to be both a statement of entitlement to the electorate and a way of capping the risk to the State' (p.804). In fact he also points out that the risk to the State could be increased, as 'rationing by means of a clearly defined core could make pressure for increased expenditure more difficult to resist' (p.804).

The Core Services Committee (now renamed the National Health Committee) quickly decided that its task was impossible. The Committee 'has not found any treatment or area of service within the current range of publicly funded services which can be completely excluded. The challenge is to determine the circumstances when people should have access to these services' (National Advisory Committee on Core Health and Disability Support Services, 1994, p.67). The Committee has therefore thrown its efforts into the development of detailed protocols for certain areas, and has also identified a list of broad purchasing priorities (services for children, mental health and substance abuse, integrated community care services, hospice care, emergency ambulance services and habilitation services). An example of the effect of protocols is given by Hadorn and Holmes (1997); a scoring system has been introduced for coronary by-pass operations, and patients whose score does not reach a certain level will not be booked for an operation. Clinicians have agreed to work with this threshold, although they regard it as too high, and believe it excludes many people they think would benefit from treatment.

The New Zealand experience suggests that quasi market reforms may prompt governments to pursue explicit rationing at national level, but that their chances of success are limited, at least as far as total exclusions are concerned. The potential for rationing by protocols, however, may be more far-reaching. It offers no proof either way whether the operation of the quasi-market will itself produce more explicit allocation decisions between purchaser and provider; it is still too early to draw firm conclusions. Furthermore, the election of a coalition government in Autumn 1996 has resulted in some significant changes to the quasi-market reforms: although the separation of provider responsibilities is being retained, the four regional health authorities are to be replaced by a single central funding body. At the same time, the crown health enterprises will be superseded by regional hospital and community services agencies. The emphasis is on replacing competition with collaboration and reducing transaction costs. It remains to be seen whether, under these circumstances, the purchaser provider split will have any continuing impact, or whether it will in fact represent a *de facto* return to hierarchical central planning. (Ham, 1997)

1) Demand rationing - yes, co-payments.
2) Supply rationing - yes, fixed budgets
3) Selective rationing - yes; no success in defining exclusions; growing rationing by protocol
4) Quasi-market - yes, but with significant changes recently announced.
5) Tax-funded - yes

(Hunter, 1996; Ashton, 1996; British Medical Journal, 1996; Cooper, 1995; Klein *et al.*, 1996; OECD, 1994; New Zealand Minister of Health, 1991; Core Services, 1995/96; Ham, 1997; Hadorn and Holmes, 1997)

Sweden

The Swedish health system is a comprehensive, tax-funded service. There has been extensive use of fixed budgets and planned levels of hospital capacity and clinical staffing to control costs, alongside patient co-payments for most services, including *per diem* in-patient care contributions.

During the 1980s, Swedish politics began to swing from their traditional social democratic position towards a more right-wing approach, which placed greater faith in competition and an already extant trend towards decentralisation to solve problems of rising public expenditure. Health care was already managed and financed at county level rather than by national government, but individual counties began to experiment with a purchaser provider split. About half now have some form of internal market. The locally-led nature of reform in Sweden makes generalisation difficult, but in both the case studies cited by Brommels (1995) the counties have set up several local health boards (comprised of local politicians, with support staff) to act as purchasers at sub-county level. In both cases, however, the local boards also manage primary care and in one case geriatric and psychiatric care, although hospital providers report directly to the county level.

Calltorp (1996) reports that Stockholm has made the most extensive developments of the purchaser role, but even here there have been practical problems of implementation. In particular, politicians, despite a real interest in purchasing, have allowed patient choice to override contracts, and as a result find themselves unable to tackle the pressing problem of hospital overcapacity. Brommels (1995) notes that it is in fact at county level where provider restructuring is being tackled, suggesting both that 'the internal market itself cannot do those "dirty jobs"' (P.104) and that it is at county level provider management that the real political power lies. Brommels is also sceptical about the extent to which purchasing has actually been implemented; in addition to the problem of controlling patient flows, he cites lack of information support as a major difficulty. Both of these problems have made it hard to link volume to price effectively. He concludes that 'the purchasing task is only gradually materialising, and the contracting process is thus far technical in nature and dominated by administrators' (P.101)

More recently, Whitehead *et al.* (1997) have noted that Stockholm County Council has begun to rethink its approach; as in New Zealand, the emphasis on competition is giving way to cooperation, with renewed interest in issues of equity and how to restore trust between purchasers and providers. In January 1996, the County Council set up a hospital board to oversee provision. This board reports directly to the central political board for health, which now co-ordinates all purchasing of hospital care over the nine district purchasers. The authors suggest this represents 'a considerable blurring of the boundary between purchaser and provider and an emphasis on setting up mechanisms for cooperation and priority setting' (p.938).

Calltorp (1995) notes that the introduction of the internal market in so many Swedish counties was a direct response to growing resource pressures, including bed overcapacity, technological advance, rising consumer expectations and numbers of elderly people. (Sweden has the highest percentage of people over 80 in the world). Interestingly he also says 'at the same time, public discussion on limits to health care also came to the surface' (p.793). This is perhaps the clearest example of explicit rationing emerging as an issue *simultaneously with* the introduction of a quasi-market, but in no way *resulting from* the quasi-market. This contrasts with the experience in New Zealand and the Netherlands where national work on core services was seen almost as part of regulating the new market reforms. This is particularly clear in Sweden because the market reforms have been developed and implemented by counties, whereas the work on rationing has been the initiative of central government, and both seem to stem independently from concern about lack of resources.

A National Priorities Commission was established and produced its final report in 1995 (Swedish Parliamentary Priorities Commission, 1995). It identifies three principles for setting priorities: human dignity, allocating resources on the basis of need (including solidarity with those vulnerable groups who may not be able to express their needs) and cost efficiency. Cost efficiency is only deemed to be an appropriate basis for comparing different treatments for the same disease; 'where different diseases are involved, fair comparison of the effects is impossible' (p.21). Defined age limits are deemed inappropriate, although judging whether the effects of old age in individuals reduce their ability to benefit is a legitimate part of clinical decision-making. Similarly, decisions about the care of premature babies should not be made according to inflexible weight limits, but should be taken according to the individual circumstances. The fact that a condition is self-inflicted, such as smoking, should not lead to negative discrimination, although again it may influence <u>clinical</u> decision-making. As Klein *et al.*(1996) point out, this leaves ample scope for implicit clinician-led rationing within the explicit national framework. It also allows scope for local priority-setting by purchasers.

The Commission also identified the following ranked list of priority areas to guide decision-making, and argued that currently insufficient priority is given to IB compared to categories II and III.

IA Acute care for life/permanent disability threatening conditions
IB Care for severe chronic conditions, palliative care and care for persons of reduced autonomy
II Prevention, rehabilitation
III Less severe acute or chronic conditions
IV Borderline cases
V Care for reasons other than disease or injury

The Commission's recommendations have yet to be implemented, so it is too early to tell what the outcome will be in practice. As in the Netherlands and New Zealand, the Commission's approach is likely to lead to the development of guidelines rather than lengthy lists of total exclusions. This would sit more easily with Sweden's firm commitment to equity; Calltorp (1995) argues that it is this commitment which requires a public debate about the issue, even if the outcome of the debate is to acknowledge that clinicians must make most of the decisions. Perhaps the most important feature is the political willingness to acknowledge that the welfare state has limits and that rationing in some form has to be countenanced.

1) Demand rationing - yes, co-payments
2) Supply rationing - yes, fixed budgets and planned capacity
3) Selective rationing - yes, in principle
4) Quasi-market - yes, but with significant changes underway
5) Tax-funded - yes

(Calltorp, 1995; Calltorp, 1996; McKee and Figueras, 1996; Brommels, 1995; Klein *et al.*, 1996; Swedish Parliamentary Priorities Commission, 1995; WHO, 1996; OECD, 1994; Whitehead *et al.*, 1997)

Discussion

International comparison cannot provide conclusive evidence about the relationship between explicit rationing and markets in healthcare, for a number of reasons:

- those countries which are experimenting with quasi-markets are at too early a stage to be certain about the effects of the change;
- those countries which are experimenting with explicit rationing are also at an early stage;
- in many cases, the rhetoric of both market experimentation and explicit rationing has not in fact been translated into reality - and, again, it is too soon to tell whether this state of affairs will persist or whether the proposals will finally be implemented;
- in some cases, notably New Zealand and Sweden, there is a degree of retreat from the quasi-market approach;
- it is impossible to determine whether countries which are at present not discussing explicit rationing will end up having to do so at a later date, regardless of organisational structure.

With these caveats in mind, international comparisons do give some clues which can be useful in analysing evidence emerging so far in the UK.

The summary table presented earlier produces some interesting results. All countries listed are operating some form of demand rationing, and virtually all ration supply as well. This is most often fixed budgeting, but also sometimes limits on capacity or staffing levels. Austria and Luxembourg are now experimenting with new activity-related budgets for hospitals using standardised costings for specific diagnostic groups, which should have some impact on supply. Greece has yet to tackle the issue.

More variation between countries emerges when explicit selective rationing, quasi-markets and tax funding are explored. Twelve countries are experimenting with selective rationing, either at national or local level, and all of them have a tax-funded health system. This applies even to Oregon, given that explicit rationing there is focused on the publicly funded Medicaid system only. It also applies to the new income-dependent health premium collected by the Dutch government. Only two countries with tax-funded systems are not yet looking at explicit rationing. Eire's tax-funded system only applies to 30-40% of the population, so has greater potential for rationing by ability to pay than universal systems. Portugal too seems to be following the route of encouraging private insurance to deal with the public funding problem. Australia has very limited explicit rationing, in the form of economic evaluation of new drugs before they are accepted for reimbursement - it is also pursuing the option of higher health taxes or mandatory private insurance for high earners.

Despite these exceptions, there does seem to be a very strong relation between tax funding and explicit rationing. What about quasi-market experimentation? Again, the five countries experimenting with quasi-market reforms (Finland, Italy, Netherlands, New Zealand and Sweden) are all tax-funded systems, and they are all examining explicit rationing, even though actual progress with both market reforms and rationing has been limited. In addition, Norway has some elements of a quasi-market in place, and Denmark and Spain are experimenting with a purchaser provider split in some local areas. However, there are countries with a tax-funded system, including Canada and Iceland, which are pursuing explicit rationing without any accompanying quasi-market reforms, and New Zealand and Sweden are rethinking their current approaches.

It could be argued that it is only tax-funded countries which can move towards quasi-market systems, because countries with insurance-based systems already have a real, if regulated, market. In fact the Netherlands does seems to provide an example of a previously insurance-based market system moving towards a more tax-based, quasi-market system. A trend of convergence towards planned, regulated markets in health care has been observed by a number of commentators, including Ham (1995) and Arvidsson (1995). It may be that other insurance-based systems will ultimately go as far as the Netherlands.

Common cause or coincidence?

It is noticeable that in all countries which are pursuing an active quasi-market reorganisation and also taking a serious interest in explicit rationing, the two strands have emerged more or less simultaneously, generally with a strong national lead. Superficially, therefore, it does not appear to be the case that the quasi-market experiment has been the cause of this interest in explicit rationing - although it is too early to say whether the quasi-market in each case will create additional pressure for more explicit rationing.

It is theoretically possible that it is pure coincidence the two strands have come together; however, this is not borne out by the way in which both have been openly characterised as responses to resource pressures. It seems more likely, therefore, that ultimately there is a common cause - namely the urgent need to control rising expenditure - which has given rise to both quasi-market and explicit rationing experiments.

The strong relationship between explicit rationing, quasi-market reforms and tax funding is surely significant. Countries which run social or private insurance schemes rely mainly on increasing premiums, or, if that fails, demand and supply rationing to cope with resource pressures. Increasing tax is a much more serious political step than increasing insurance premiums, because of greater public unwillingness to pay higher tax and because of the resulting electoral unpopularity. Thus tax-funded systems may run into deeper financial problems at an earlier stage than insurance-based systems. The fact that in Oregon the tax-funded service is not universal and is therefore not a source of social solidarity, may explain why it has got furthest with explicit rationing. The difficulty of persuading taxpayers to pay more for a service they themselves never use, such as Medicaid, is even greater than persuading people to pay more tax for a universal NHS.

Ham (1995) points out that continued commitment to public funding is a crucial factor in pursuing policies which seek to define a range of services; once this commitment is weakened and private funding assumes greater importance, as in Eire and Portugal, the pressure for explicit selective rationing is weakened too.

The actual level of healthcare expenditure does not seem to be as relevant as the source of funding. Klein *et al.*(1996) point out that countries which have set up special commissions on priority setting, such as New Zealand, Norway, Canada, Sweden and the Netherlands, all spend more on health than the UK, and furthermore vary substantially from each other as to how much they spend. The one thing these countries do have in common is that the level of health care spending is determined collectively.

The collective decision-making of a tax-funded system not only constrains expenditure; it also offers a ready mechanism for making national decisions about healthcare and ensuring they are implemented which more diffuse systems lack. Thus it both creates resource pressures but at the same time offers a clear chain of command for managing them through national priority-setting exercises which would be harder to contemplate in a system based on individual insurance.

This is not to say insurance-based systems will not also eventually get to a point where financial pressure is so great that they have to move towards explicit rationing, but it does help to explain why tax-based systems may get there first.

Independence or causation?

The suggestion that both quasi-markets and explicit rationing are the independent products of financial stringency could help to explain the position of other countries where one or other is taking place but not both. It is a perfectly reasonable response to financial pressure to try to improve structural efficiency and introduce competition, but to leave rationing to implicit mechanisms (at least for the time being) - it is also reasonable to opt for explicit rationing whilst leaving the structure untouched. What determines the choice may be a function of how serious the financial problem is perceived to be, what scope there is left in the system for simply increasing efficiency, political and public attitudes and historic differences in the current structure.

It was suggested earlier that the fact that quasi-markets and explicit rationing have so far usually emerged simultaneously gave superficial support to the view that they were independent. Certainly there is evidence of countries without quasi-markets turning to explicit rationing - but what are we to make of the fact that there is no country with a quasi-market which is not also investigating explicit rationing? Is this just chance, or does it indicate there is some additional relationship?

The relationship may not be causative in terms of quasi-markets happening and then explicit rationing in chronological sequence, but may be causative in the sense that governments experimenting with quasi-markets want to protect against possible inequities and therefore seek to define a basic minimum package. Even though this is intended to be inclusive rather than exclusive, the outcome may be to exclude services. The important distinction to make is that this is a political choice, rather than an inevitable logical process. It may also be important that most countries which have tried this approach have ended up backing away from hard and fast definitions of core services - perhaps just because it is too difficult, but perhaps also because they become aware of the irony that a measure intended to protect citizens' rights to equitable treatment ends up excluding them from some healthcare. It is significant that the only place which has not backed away from firm and detailed definitions is Oregon, where it was always clearly understood that the purpose was to exclude some treatments in return for universal coverage.

The above arguments do not necessarily refute the theory that quasi-markets will themselves increase the pressure for explicit rationing at a later date. Countries which have produced a strategy for explicit rationing may be merely pre-empting a need that will emerge anyway as their new quasi-markets take effect; their foresight may make it easier for them to cope with the extra pressure towards explicitness. There are two possible routes by which quasi-markets could make explicitness more likely:

1. They may force existing implicit rationing out into the open as the increased availability of information and the need to specify detailed contracts between purchaser and

provider formalise allocation decisions. Once in existence, contracts may provide a vehicle for the more explicit expression of new rationing decisions.

2. They may increase the total amount of rationing needed by ironically increasing the very costs they are supposed to control, because of the transaction cost problem, thus creating a cyclical link back to the ultimate common cause of resource pressures.

Unfortunately, it is still too early to be able to demonstrate this element of the equation; indeed, given the retreat from market reforms now happening in a number of places, it may never be possible to follow through the full effect of those reforms. However, it is worth noting a comment from Whitehead *et al.* (1997) in relation to Sweden that some aspects of the quasi-market reforms, notably fee for service payments, 'seemed the wrong incentives when cost control and structural changes in supply became political priorities in the 1990s' (p.938).

Summary

The following diagram shows how these different theories might link together.

explicit rationing
Resource ⇗ explicit rationing through contracts
pressures ⇘ ⇗
quasi-market reform ⇨ defining core services ⇨ explicit rationing
⇘
resource pressures ⇨ explicit rationing

Tax-funded systems are especially prone to the resource pressures which give rise to both explicit rationing and quasi-markets. Given the hierarchical nature of tax-funded systems, experimentation with market-style mechanisms is a natural development in the search for alternative ways of improving efficiency. Tax-based systems also offer a national collective decision-making framework which enables explicit rationing to be planned and implemented.

Quasi-markets are not a necessary condition for the emergence of explicit rationing, but may create additional impetus towards it, through some or all of the mechanisms described earlier.

Where does the UK sit within this framework? As a tax-funded system facing intense resource pressures, it is not surprising that there should have been experimentation with market reforms to improve efficiency. Klein *et al.* (1996) suggest these reforms stem from an interpretation that collective decision-making and lack of individual responsibility for paying for services inevitably lead to rationing (whatever the level of expenditure), so the only way to avoid it is to introduce a market system instead. Although they go on to expose the fallacy of assuming this would avoid rationing, their analysis of the government's motives sheds an interesting light on the absence of a national lead on explicit rationing unique amongst quasi-market experiments to date. It suggests the government was trying to avoid the tension between universality and comprehensiveness, rather than confronting it as other nations have tried to do.

In fact the original White paper 'Working for Patients' (Secretaries of State, 1989) stated that health authorities as purchasers would be required to define a list of core services to which it was essential that local access for their local population should be maintained. Initially this was to cover both self-governing trusts and directly managed units, and was intended to be a way of ensuring that providers could not make a unilateral decision to withdraw from providing some services. As such it was never really meant to be a rationing mechanism so much as a means of regulating the market; non-core services might not be available locally but it was never intended they should be excluded altogether, unlike in other countries. Very early on in implementing the reforms, however, the government retreated from this position, first of all (November 1989) by applying the core services definition only to self-governing trusts, then eventually abandoning the term altogether. The reasons why it did so were not made entirely clear, although one factor was the technical problem of one health authority trying to impose core service requirements for its residents on a unit directly managed by another district. One can only speculate whether ministers also thought it would constrict purchasers and providers too much, or whether they feared it might become a rationing mechanism against their original intentions.

What is clear is that there has been no national government lead on explicit rationing, either simultaneous with or after the introduction of the quasi-market. There is some national pressure towards explicit rationing emanating from academic and professional quarters, but this has been pursued largely in parallel with analysis of the internal market.

In any case, the Conservative government deliberately declined to sanction such debate in the White Paper 'The National Health Service: A Service with Ambitions' (Secretary of State for Health, 1996). Although the Secretary of State made important steps towards acknowledging the need for priority setting, he stressed that this must be at local, not national level, and must not countenance total exclusion of any service that may be of clinical benefit. Since most treatments could benefit some patients, however remote their chances, this meant that in practice total exclusions are not allowed. There are no signs to date that the Labour government will take a different line, although it remains to be seen how the proposed National Institute for Clinical Effectiveness will work in practice.

Given the lack of a national government lead on rationing so far, the UK offers the best opportunity internationally to analyse what impact the quasi-market has had on rationing; even so, the existence of other initiatives on explicit rationing at national level may be exercising an independent effect which must be allowed for.

Part Two

Fieldwork

Chapter Five

Methods

Secondary analysis of evidence presented to Health Select Committee

During 1994, the Parliamentary Health Select Committee undertook an investigation of purchasing and priority setting. Evidence was submitted by all fourteen regional health authorities and 49 district health authorities, amongst others (House of Commons, 1994a, 1994b). These responses have been analysed with particular reference to explicit rationing and the use of contracts in priority setting, and are then compared with the findings of the pilot fieldwork survey.

Pilot fieldwork survey

Also in 1994, a pilot fieldwork survey of 13 purchasers (8 health authorities and 5 GP fundholders) was carried out. The origins of this survey were that in 1994 I was asked to supervise a research project into 'purchasing levers' for Anglia and Oxford Regional Health Authority; as part of this broader project I was able to include my own specific questions on rationing, which participants knew were for my own research. The methods used were an initial questionnaire, followed by a semi-structured interview of about an hour's length. Those interviewed as representatives of the eight authorities comprised seven directors (of purchasing, commissioning, contracting or planning) and one contracts manager. The questionnaire and interview schedule were designed by me; the interviews were carried out by my assistant on the project, David Welsh, and I analysed the results. The results of the survey have been published separately by Anglia and Oxford (Locock, 1995a), but this did not include the questions on rationing. These have been published previously only in my M Phil thesis (see note at the end of this chapter).

Participants were asked the following questions:

1. Do you think the introduction of the purchasing function in the NHS has made it easier to establish new priorities for change and development?
2. If so, please describe how you have used your position as purchasers to establish new priorities.

3. Have you established a formal list of services you have decided not to purchase or to purchase only on an exceptional basis?
4. If so, what are they?
5. If there is no formal list, have you in practice refused to purchase certain services?
6. If so, what are they?
7. Have you made any decisions as purchasers to limit access to services for certain groups of people? (e.g. those over a certain age limit, smokers)
8. If so, please describe
9. If you have made any decisions not to purchase certain services or to limit access to them, do you think you would have been able to make such decisions before purchasing was introduced?
10. Has the development of the purchaser role made the rationing of services more explicit in your view?

Although the pilot survey included GP fundholders, it was decided to exclude them from subsequent fieldwork. The purpose of the research was to compare resource allocation decisions before and after the introduction of the quasi-market, and disentangle how much the market is responsible for greater explicitness. GPs did not have responsibility for financial resource allocation decisions before the market, so they cannot compare their position as fundholders with their previous position as ordinary GPs. Unlike health authorities, they have no point of reference against which to make such comparisons, and therefore cannot by definition say whether the market makes them ration more explicitly.

Of course GPs did previously have to live within the general budgetary constraints of the NHS, and they did make implicit rationing decisions in acting as gate-keepers to the service. The question of how or whether GP fundholders have rationed more explicitly is an appropriate topic for study but does not shed light on the hypotheses under investigation in this thesis. It is therefore only the findings related to health authority purchasers which are reported here.

Selection and design of the three new case studies

The approach adopted was that of purposive sampling. Time and resources available limited the choice of sites to three within reasonable reach of the author's home town. At the same time, the intention was to select authorities that fell within a middle range of population mix, with both urban and rural areas and without high concentrations of socio-economic deprivation.

Given the pressures under which health authorities are working, it was anticipated that there would be some difficulty in recruiting participants prepared to offer sufficient time and effort. An initial group of three chief executives was identified through personal contacts who might be willing to take part, as well as one whose authority had recently been involved in very high-profile rationing decisions. It was hoped that this authority could be compared with at least one other authority which did not have such a high national profile on the issue. In fact three authorities agreed to take part, including the high profile one.

All three had moderate to large urban areas within them, as well as rural areas. None were highly deprived, although all had pockets of deprivation within them. Two of the authorities included a university town with associated teaching hospitals; these were Avon in South and West Region, and Cambridge and Huntingdon in Anglia and Oxford Region. The third (which has requested anonymity) was in the home counties, in a Thames Region. Although the diversity of the NHS is such that it would be difficult to identify a 'representative' district, the three participating districts had no particularly unusual features, except for the fact that one had so recently hit the headlines with a very personalised case of explicit rationing. However, in terms of purchasing it is worth noting that there was a difference between them in the concentration of fundholding - in District Three, over three quarters of the population was covered by fundholding practices, whereas in Avon it was just over half the population and in Cambridge and Huntingdon just over a quarter.

Some basic characteristics of each district are as follows:

	Revenue 1997/98	Population (approx.)	No. of major providers (> £4m of contracts)
Avon	£472m	840,000	10
Cambridge and Huntingdon	£176m	450,000	4
District Three	£240m	490,000	8

The aim of the case studies was both to obtain the views of senior NHS managers about the factors affecting explicit rationing and to follow the negotiations for a complete contracting cycle carried out during 1996/97 (for contracts for 1997/98). The first round of interviews was conducted from July to September 1996. All interviews were taped and fully transcribed before analysis. In each district, interviews were arranged with the chief executive and the directors of finance, public health and commissioning/contracting, as well as a senior registrar in public health and the deputy director of commissioning in Cambridge and Huntingdon. In Avon, the deputy director of contracting was interviewed, as the director was ill. In District Three, the director of public health had not yet taken up post, so the acting director was interviewed. In the same district, there were two directors of commissioning, each with responsibility for a geographical patch. Both were interviewed.

The purpose of this round of interviews was to establish the current position in each district and to elicit interviewees' views of the factors affecting explicitness. Interviews were semi-structured, lasted between sixty and ninety minutes, and covered the following topics, which were notified to participants in advance:

- Work done locally on articulating values and defining the methods to be used in priority-setting decisions, and how useful this had been in practice
- Use made of research findings such as the Effectiveness Bulletins and mechanisms such as QALYs
- Who was involved in the priority-setting process
- Whether any treatments were limited or excluded, and if so on what grounds
- Whether interviewees themselves acknowledged this as rationing
- How explicit their rationing decisions were (e.g. explicit between managers or between managers and doctors, but still opaque to public)
- Whether their rationing decisions were included in contracts, or put in writing in some other form
- Whether they involved the public in their decision-making
- To what extent advice was taken from providers
- In whose interests interviewees thought it was to be more explicit, or to remain implicit
- What they perceived as the driving forces in favour of greater explicitness

Prompts: lack of money
patient's charter
market structure itself/contracting process
outcomes and effectiveness research
need for democratic involvement

- What constraints/obstacles had been encountered, and what the pressures were in favour of remaining implicit

Prompts: lack of information/research evidence
public and political opposition
inter-professional disagreement
ethical doubts

- The extent to which rationing decisions were put into practice and monitored effectively, either through the routine contracting process or separately
- Whether they had been able to identify any savings as a result; if so, whether it was worth the effort

The interviews concluded with a forward look at their expectations for the forthcoming contracting cycle, and more general views about the future of explicit rationing.

After this first round, it was concluded that directors of finance were not closely enough involved in the detail of rationing decisions to warrant interviewing them again, although this in itself was an interesting finding.

The second round of interviews took place in January and February 1997, with the directors of public health and commissioning in two districts, and one director of commissioning only in District Three (owing to scheduling difficulties). The main aim was to find out how contracting negotiations were progressing and whether any further explicit rationing proposals were being considered, and it was felt that at this stage public health and commissioning would be closest to the details of the process. In addition, interviewees were asked to compare current practice in allocating scarce resources with practice before the quasi-market was established and with the early years of the market. In two districts, this discussion was supplemented by comparing strategic and annual planning documents from previous years, to see if interviewees' recollections were supported by the written evidence. (There were some problems in obtaining appropriate comparable documentation, as all three districts had undergone several boundary changes in the last ten years. This meant the documents supplied referred to different geographical areas to the current health authorities. The changes had also resulted in office moves and changes in document storage, and one district, Cambridge and Huntingdon, was unable to supply anything).

The third round of interviews took place in June to July 1997. The original plan had been to interview the chief executive in each case, as well as the directors of public health and commissioning. In the event, however, one chief executive had to cancel. The third interviews had two main themes: firstly, to discover the final outcome of contracting negotiations for 1997/98, and secondly to elicit interviewees' understanding so far of the new Labour Government's proposals for abolishing the market, and their views on what effect they might have on explicit rationing.

Throughout the research, the participating health authorities have been very helpful in making available as many relevant documents as possible for analysis alongside the interviews. These have included papers concerning the values of the health authority, purchasing strategies and service reviews, draft and final annual purchasing plans for 1996/97 and 1997/98, resource assumptions and corporate contracts, as well as relevant samples of contracts drawn up with providers. A more detailed list is provided in the annexe to this chapter.

Inclusion of material from M Phil

The thesis incorporates material from an M Phil thesis already submitted at the University of Oxford, entitled 'Is the NHS market making health care rationing more explicit?' (Locock, 1995b). This material includes part of the literature reviews for the current thesis, although they have been substantially extended and updated, and the theoretical component has been considerably strengthened. It also includes the results of the pilot fieldwork survey and the analysis of evidence presented to the Health Select Committee. The LSE calendar [p.854] for 1995/6 states that '...a candidate shall not be precluded from incorporating work already submitted for a degree in this or any other university or institution in a thesis covering a wider field, provided that the candidate shall indicate on the entry form and also on the thesis any work which has been so incorporated'. I believe my work is consistent with this statement.

Annexe to Chapter Five - documents made available by case study health authorities

Avon

Proposals for change 97/8 (consultation document)
Budgets 97/8
Draft proposals for managing ECRs in 97/8
List of restricted treatments 97/8 (prepared in response to a regional office request for evidence to the minister of health)
Policy seminar paper on commissioning drug interventions 1997
Summary of service agreements 97/8
Extracts of service specifications 97/8
96/7 corporate contract
Budgets 96/7
Briefing paper for executive team on oncology, 1996
Service specs 96/7 (including list of restricted treatments)
Two draft statements of values, 1996
96/7 policy for management of ECRs
Contracts portfolio for 92/3
Contracts portfolio for 95/6
1986 Acute services review
Working for a healthier future (5-year plan, 1995-2000)

Cambridge and Huntingdon

Draft annual plan 97/8
Response to consultation on annual plan 97/8
Summary of service agreements 97/8
97/8 budget allocation summary
97/8 contracting brief
Acute trust contract offer, 97/8
Acute contract exclusions, 96/7
Indicative service levels, clinical effectiveness, part of 96/7 acute contracts
ECR policy 96/7

Involving the public in developing selection criteria for assisted conception treatments, 1995

Paper to board on citizens' juries, 1996

Final annual plan, 96/7

Corporate contract 96/7

Strategic framework for acute services, 95/96 to 2001/02

Public focus on rationing in the NHS: report on findings from focus groups, 1996

Purpose, roles, value, working practice and strategy, 1996

District Three

Funding plan and proposals for service developments 97/8

Health and healthcare plan 97/8

Purchasing plan 96/7

Contract stocktake 97/8

Contracts portfolio 96/7 (including contract exclusions, ECR referral policy, low priority treatments list, subfertility treatment policy)

Corporate contract and workplan, 96/7

Principles for making choices, 1995

Strategy for acute services 1989

County-wide acute services review 1993

Hospital closure consultation document 1993

Annual report and public health report 1991/2

Developing a strategy for involving local people in purchasing health care 1991

Chapter Six

Early experience of explicit rationing and the NHS quasi-market

This chapter summarises early experiences of explicit rationing and the internal market, as viewed by local actors. It begins with an analysis of regional health authorities' perceptions of experiences within their constituent district health authorities, and then examines district health authorities' own accounts. This draws on a secondary analysis of evidence presented by regional and district health authorities to the health select committee investigation of purchasing and priority setting in 1994 (House of Commons, 1994a, 1994b), and on the results of a pilot fieldwork survey of eight health authority purchasers, also carried out in 1994. The chapter concludes by comparing these findings with the findings of the NAHAT analyses of purchasing plans carried out over several years by Klein *et al.* at the University of Bath (Klein and Redmayne, 1992; Redmayne, Klein and Day, 1993; Redmayne, 1995; Redmayne, 1996).

Regional health authority perceptions - evidence to the health select committee

The questionnaire sent to regional health authorities by the health select committee asked a total of 23 questions about the experiences of their constituent districts, of which two were particularly relevant to the subject of explicit rationing and the market. These were :

Question 7 - What examples are there within your region of explicit policies by local purchasers to exclude certain services from NHS provision, or to discourage the provision of certain services?

Question 11 - What examples are there of contracts placed by local purchasers in your region which seek to ensure the purchase of effective care and discourage ineffective procedures?

Some of the evidence from regional health authorities inevitably overlaps with the district responses, and the districts they refer to are not always identified. This should be borne in mind when reading this analysis. Mersey region did not address either question directly and responded to the questionnaire in a different format than that used by the select committee. The other thirteen all gave direct answers, which are summarised in Table 6.1.

TABLE 6.1

REGIONAL EVIDENCE TO THE HEALTH SELECT COMMITTEE

Region	Local exclusion policies	Encouraging effective/discouraging ineffective care
Trent	Yes, in 5 DHAs - in some cases full exclusion, in others limited access only	Several protocols given as examples of both encouraging effective and discouraging ineffective care
E Anglia	Very few - 2 DHAs none, 4 DHAs in very limited areas	Protocols focus mainly on encouraging effective care
N Western	Few if any - no complete services excluded	Protocols may reduce or eliminate some procedures but not whole services
NW Thames	Very few and very limited - 2 examples given	A lot of work on protocols to limit inappropriate interventions, *to avoid need for rationing*
SE Thames	Regional policy (in line with national) is no exclusions - qualified restrictions on access acceptable if agreed by local people, however	Protocols mainly to encourage effective care
Northern	1 DHA excludes one procedure altogether. Others restrict by low volume contracts.	Protocols mainly on effective care, but beginning to look at ineffective procedures
NE Thames	Few examples - 1 current, 3 DHAs considering introducing	Contracts discourage ineffective care - but protocol for infertility focuses on effectiveness and requires extra resources

TABLE 6.1 (cont.)

Region	Local exclusion policies	Encouraging effective/discouraging ineffective care
Yorkshire	2 examples given	Protocol may introduce new infertility service. Aim to reduce D&C by 20% on effectiveness grounds. Other protocols for both effective and ineffective procedures
SW Thames	Regional policy follows Chief Medical Officer's line that purchasers should not exclude whole coverage of specific services - 3 examples given of DHAs defining limited access, by clinical appropriateness or by fixed budget	Active support for using information on ineffectiveness as a way to limit access
S Western	Examples are limited in scope (but clearly important to individuals affected). One example given for illustration. Usually allow for exceptions to be made	Effectiveness information may result in not purchasing a service at all, or in increased selectivity. Some contracts specify a maximum volume for some procedures
Wessex	Some DHAs still discussing, some already doing. One example of the latter given for illustration.	Most protocols focus on effective care, with a few on ineffective care
W Midlands	No Regional policy to exclude services, except for Regional guide-lines on aesthetic surgery. Defend clinicians' right to exclude individuals on clinical grounds (eg smokers), but not blanket policies.	Support for weeding out ineffective procedures expressed in general terms
Oxford	A few examples (one given for illustration) but more DHAs are considering policies	'Getting Research into Practice' project aims to discourage ineffective procedures.

The question of whether ceasing to provide ineffective care does or does not constitute rationing is an immediate problem in trying to analyse these regional responses. NW Thames clearly felt that ceasing ineffective care was an *alternative* to rationing; SW Thames, on the other hand, regarded using clinical appropriateness to limit access to services as acceptable rationing, but rationing nonetheless.

Further confusion arises from different interpretations of the term 'exclude certain services'. SE and SW Thames both took the Department of Health line that total exclusions of procedures that might have some clinical benefit was unacceptable; they argued that limiting access to certain procedures or to certain groups of people on clinical criteria was acceptable and did not constitute exclusion. Other regions took a less politically sensitive line, and described policies virtually identical to those in SE and SW Thames as exclusions, even if a limited number of people were in fact treated.

Whatever the intricacies of the language used to present what they were doing in an acceptable light, it is clear from this survey of regional health authorities that all over the country attempts were being made to draw up explicit specifications of what care would or would not be available. Several regions also stated that they believed more districts would soon be taking this approach and that it would be extended to a wider range of services, although it was currently fairly limited in scope. This was supported by the fact that, although most of the procedures listed as being rationed were the common targets such as fertility treatment, cosmetic procedures and sex change operations, there were already a few interesting additions to the list, including:

- non-clinical circumcision, or circumcision for children under 5 years of age
- treatment for clumsy children [sic]
- surgical treatment for impotence in men over 50
- antenatal triple test for Down's syndrome
- care at the interface between health and social services.

It is also possible that regions were not fully aware of the extent of rationing within their districts. For example, one named district was said by its regional health authority to have no policy of exclusions, yet this district was also one of the participants in my own survey and reported several exclusions.

Some other points of interest arise from the regional responses. Several regions mentioned the use of low volume contracts or fixed amounts of money to limit access to certain services such as IVF or cosmetic surgery. This is perhaps best described as a more explicit form of rationing by waiting list; it relies on a first come, first served principle, with no selection criteria applied to patients, but is targeted much more explicitly on single procedures than pre-market waiting lists and is determined by purchasers rather than clinicians. One district was reportedly operating a 'secondary waiting list' within its overall plastic surgery list, so that people wanting cosmetic surgery had to wait longer than anyone else, although they were still treated within the maximum times set out in the Patient's Charter. Another district which was said to exclude surgery for varicose veins either for cosmetic reasons or for discomfort nonetheless allowed GPs to continue to refer some patients at their discretion, a situation which sounds very similar to the traditional role of the GP as gate-keeper.

District health authority perceptions - evidence to the health select committee

Of the 49 district health authorities which gave evidence to the select committee, 5 were also participants in my own survey (see below). Their responses are included in the following analysis, as in some cases the information supplied is different or more detailed, but they will not be identified in order to preserve the confidentiality of their answers to my survey.

The relevant questions from the questionnaire for districts were:

Question 10 - Has your authority decided to exclude specific services from its contracts, e.g. tattoo removal, reversal of sterilisation? If so, which services have been excluded and why?

Question 11 - Has the authority made use of clinical guidelines or protocols designed to limit the provision of services to particular patients? If so, please give examples and reasons.

Question 14 - What contracts has your authority placed to ensure the purchase of effective care and to discourage ineffective procedures?

Question 15 - Has your authority made use of the Effective Care Bulletins produced by the Department of Health? If so, please give examples of which ones have been used and how this has changed the provision of services.

The sheer number of responses from districts makes presentation of each one individually unwieldy; the summary in Table 6.2 therefore takes the form of counting how many authorities fell into certain categories identified from the responses.

TABLE 6.2

DISTRICT HEALTH AUTHORITY RESPONSES TO HEALTH SELECT COMMITTEE

Arrangements operating	**Number of Districts**
Districts claiming no exclusions	21
Of these, the number in fact operating some form of exclusions	16
Districts reporting exclusions	28
Of these, the number which in fact mentioned that they allow some exceptions	19
Districts stating purpose of clinical guidelines should not be to limit access	27
Of these, the number in fact limiting access to services	23
Districts agreeing that purpose of guidelines may be to limit access	22

These categories require some explanation. It will be seen from the table that a high proportion of those claiming not to operate exclusions in their district did in fact operate some (according to their own evidence), and conversely that a high proportion of those stating they did operate exclusions in fact allowed exceptions to be made. In practice, what this means is that most of the districts did exercise some form of rationing or limitation of access to services, and that most of them also allowed exceptions to be made for a variety of reasons.

Only 5 of the districts claiming no exclusions at all did not go on to qualify this in some way. Of the districts which did report exclusions, 3 made no mention of exceptions being allowed for any of the procedures excluded, using expressions such as 'from 1st April 1993, the DHA has not funded IVF' or 'Gloucestershire Health does not make funds available for tattoo removal', with no qualifications. A further 6 reported that they allowed exceptions for some excluded procedures but not all. 2 districts which were currently operating a total ban on IVF treatments were reconsidering their position on the grounds that IVF was potentially effective for some women.

What made one district describe itself as having no exclusions whilst another with identical policies was ready to state that it did have exclusions seemed to be almost entirely a question of semantic emphasis, rather than substantial difference. Some districts chose to take the word 'exclusion' at face value, as meaning that absolutely no-one had access to a specific treatment, whereas others defined it more liberally as meaning that the treatment was normally excluded unless there were overriding reasons why an exception should be made, or that it was only available to selected groups of people. (These groups were usually those who would derive the greatest clinical benefit from it, but in some cases access was restricted purely by a fixed number of treatments being provided, on a first-come, first-served basis).

As with the regional evidence to the select committee, the question of whether excluding supposedly ineffective treatments amounted to rationing was not easily resolved. When asked about the use of clinical guidelines or protocols to limit the provision of services to particular patients, most districts stressed that their primary aim was to improve effectiveness and appropriateness, rather than to limit services. Even amongst those who did acknowledge that limiting services might be an acceptable aim, most emphasised that this was not the sole purpose, and that it was a by-product of more appropriate targeting of services on those who would derive the greatest benefit.

Only a few gave open recognition to the use of protocols to exclude treatments which might be effective but do not have high priority in the competition for limited resources. For example, Blackpool, Wyre and Fylde's evidence stated that:

> 'Clinical protocols apply to Preston residents for the provision of certain plastic surgery work. This is on the basis that demand far outstrips supply, and it is therefore considered inappropriate for social medicine to be undertaken from NHS funding.'

United Health (Grimsby and Scunthorpe Health Authority) said it had excluded services 'which are of a cosmetic nature or where a procedure is desirable but not necessary'. A few others made the distinction between excluding procedures on the grounds of 'low priority' (such as cosmetic operations) and excluding them because of poor effectiveness (such as sterilisation reversal).

Other authorities tended to couch all exclusions in terms of the lack of health gain they would bring. Gloucestershire Health Authority, for instance, stated that it did not make funds available for certain services 'on the basis that the inability to secure these services does not constitute a threat to the health of the individual or the population of Gloucestershire.'

The list of procedures being excluded (whether partially or totally) was in the main the familiar territory of cosmetic surgery, infertility treatments of various kinds, sex change and alternative/complementary therapies. In addition, a few districts mentioned the exclusion of radial keratomy (an operation for treating short sight), cochlear implants to correct deafness, dental implants and extraction of asymptomatic wisdom teeth. One authority had introduced eligibility criteria for a number of orthopaedic operations, including the provision of total joint replacement only for restriction of activity or persistent pain interfering with sleep and work. Another had decided that lower back pain was best treated by physiotherapy and psychological counselling, and had therefore restricted the availability of surgery for this condition. A third had excluded 'non-orthodox obstetrics', although no explanation of this was given.

The use of clinical guidelines and the Effectiveness Bulletins issued by the Department of Health had also led to a number of districts restricting the availability of operations for glue ear in children, dilatation and curettage in women under 40, and routine screening for osteoporosis and cholesterol levels. Although information on clinical effectiveness was frequently mentioned as the reason for restricting infertility services, several districts had also included social criteria in their protocols, such as whether the woman had a partner, and whether both partners were childless.

There is little further information to be added from the discussions that took place at select committee hearings attended by eleven of the participating districts. There was uncertainty and disagreement about the role of the public in helping to make rationing decisions, and about the potential for central government to issue national guidelines on what care should or should not be excluded. Some felt this would be helpful, particularly to overcome the dilemma of inequity between different districts, whereas others felt strongly that local needs and circumstances should be the deciding factor.

Several expressed concern that the information on effectiveness on which many of their rationing decisions relied was in fact weak and available only for a very few areas. In the words of North Cumbria's representative, 'a problem is that we may be trying to make objective decisions about things that simply cannot, with our present knowledge, be made into objective decisions.' The disagreement evident from the questionnaires as to whether restricting services on the basis of lack of effectiveness constituted rationing found no resolution in the Committee's discussions.

The select committee evidence is not very informative on the relation between the market and explicit rationing. Respondents were asked some questions about their purchasing practice, including those noted above concerning the use of effectiveness evidence in contracts, as well as questions about contract monitoring, performance monitoring and strategic shifts being achieved by purchasing. The underlying assumption of the entire review seems to have been that purchasing must be having some effect on the style of priority setting, and that this needed to be investigated. However, or perhaps precisely because of this underlying assumption, the questions were not structured in such a way as to explore how any causal relationship between purchasing and rationing might be working, and to what extent other causal factors might be involved. The answers given were correspondingly unenlightening, and tended to focus on good intentions and success stories. Blackpool, Wyre and Fylde's assessment of the impact of purchasing is typical: it has resulted in 'a shift towards primary care, the greater development of mental health services, a shift towards GP support and the furtherance of Health of the Nation targets, including changes in practice agreed with providers.'

District health authority perceptions - pilot fieldwork survey results

The following section contains an analysis of responses to part of a survey on approaches to purchasing carried out on behalf of the Anglia and Oxford Regional Health Authority in 1994. Eight health authorities and five GP fundholders took part. Only that part of the fieldwork which related to rationing by health authorities is analysed here. (See Chapter Five for further details on methodology).

Every health authority which participated in the survey was operating some kind of explicit rationing. The most common targets of explicit rationing were similar to those identified in the evidence to the select committee, namely assisted conception, cosmetic surgery and reversal of sterilisation, and less commonly gender reassignment, psychotherapy, insertion of grommets and surgery for obesity.

A summary of participants' responses is given in Table 6.3 for ease of reference, and a more detailed analysis follows, based on the questions used in the survey.

TABLE 6.3

HEALTH AUTHORITY RESPONSES TO PURCHASING LEVERS SURVEY

	Formal list of services not purchased	Refused in practice to purchase services	Limited access to certain groups	Able to make such decisions before purchasing introduced	Purchasing has made rationing more explicit
District 1	Yes	No	No	Yes	Yes
District 2	No	No	Yes	*	*
District 3	Yes	No	No	No	Yes
District 4	Yes	Yes	Yes	Yes	Yes
District 5	Yes	No	No	Yes	Yes
District 6	Yes	No	Yes	No	Yes
District 7	Yes	No	No	No	Yes
District 8	Yes	No	No	No	Yes

* = question not answered directly
(This district was one which did not complete a questionnaire, and these questions did not come up directly at interview).

Formal rationing

Only one authority had no formal list of services which it would not purchase, although it did limit access to certain groups of people (see later discussion). All the others had formal lists.

District 1 did not purchase assisted conception services, a decision taken on financial grounds. Since taking the decision, the authority had merged with another to form a new health commission; the second authority did purchase assisted conception, and it was unclear how this discrepancy would be resolved. District 1 had experienced some cases of GPs trying to by-pass the ban by referring direct to specialist centres, and was therefore tightening up its procedures for monitoring such referrals. A further difficulty was that some patients had been referred privately for assisted conception and started treatment, and then applied to transfer to the NHS for continuation of the treatment, which was proving difficult to refuse.

Apart from assisted conception there were no other absolute exclusions in District 1, although certain treatments such as insertion of grommets were designated for exceptional use only, on the grounds that they were of limited clinical value.

District 3 was formed by the merger of two smaller authorities and was still trying to reconcile the different policies it had inherited. In one of the previous authorities, a comparatively extensive list of excluded procedures had been developed, including some psychotherapy services, surgery for obesity and service costs of unapproved research projects and clinical trials, as well as the more usual fertility and cosmetic surgery services.

The list was adopted by the new merged authority, but there remained differences in implementation in the two halves of the authority. For example, exceptional requests for cosmetic surgery are decided through an informal discussion between clinicians and public health in one part of the authority, but in the other a scoring system operates.

District 4 reported an outright ban on cosmetic surgery procedures and also explicitly restricted purchasing of IVF and D&C (see section on limited access). It also operated a scoring system for oral/dental referrals; any cases falling below a certain score were referred back to a community dentist.

In District 5, explicit rationing was directed at cosmetic surgery and IVF.

District 6 had a formal protocol limiting access to IVF (see section on limited access below). Cosmetic surgery would only be approved if supported by an independent psychiatric report. Laparoscopic hernia repair was restricted in certain circumstances; the reason given was that one particular surgeon was 'clogging up his list' by attracting extra referrals for it, but 'there were not adequate clinical reasons to make it the only way of doing hernia repair.' This is an interesting example of intervention in an individual doctor's clinical practice.

In District 7, explicit rationing was aimed at 'the easy stuff', namely the insertion of grommets, IVF and cosmetic surgery. The fact that the authority had been receiving growth funding was recognised as a factor affecting its approach to rationing, in that it had not had to make many difficult choices so far. 'We can engineer change by targeting growth money to a large extent.'

District 8's interview was affected by poor taping quality, and there is therefore less information than for some of the other authorities. It had a formal list of treatments not purchased, covering gender assignment, sterilisation reversal and cosmetic surgery (which the respondent said constituted 60% of all plastic surgery activity nationally before the reforms).

Rationing in practice

The respondents in Districts 3 and 4 reported that on rare occasions extra contractual referrals had been refused on the basis of the clinical judgement of the director of public health and other clinical advice, but without any kind of formal rationing policy to back up the decision. The fact that other respondents did not mention this cannot necessarily be taken as proof that it was not happening in their authorities too, however.

Limiting access to certain groups

Three respondents reported formal policies limiting access to certain groups of people, and others felt it was a likely future development in their areas or suggested it was already happening at a clinical level.

Looking first at those which did have formal policies, District 2 described fairly complex restrictions on certain contentious areas, including IVF and cosmetic plastic surgery. On the subject of IVF, the respondent said:

> 'we have got a set of criteria about how long the couple have to have been together, indeed do they have to be a couple at all, about whether they have got any previous children on either side and all the rest of it. But if you fulfil the criteria then you join the waiting list and we fund about 30-40 treatments a year .... To my surprise in some ways, given we have a number of social criteria, the medical profession as a whole seems to recognise that it is as fair and equitable a way as we are going to come up with and we don't get the policy challenged any more.'

This represents a mixture of explicit rationing to certain groups of people who do not meet the criteria for eligibility and rationing by the traditional waiting list mechanism.

The respondent stressed that although he was comfortable with the idea of protocols defining who should be eligible for treatment he felt there should be an appeal or some sort of discretionary mechanism. He used the example of laser treatment for port wine stain birthmarks. The authority's protocol gave priority to children between the ages of two and eighteen with birthmarks on their face or neck, within a fixed budget for the procedure. This had been overridden to allow treatment of a teenage girl with a port wine stain on her thigh because she was suffering psychological damage from taunting at school.

He expressed concern at the idea that in some authorities 'essentially a couple of officers are deciding somewhere that a patient won't get treatment, and I do not think that is ethical, to be quite honest.' This represented to him a continuing kind of implicit rationing, but with managers rather than doctors making decisions, on financial grounds. It was for this reason that he preferred an overt policy. He did not appear to acknowledge that using managerial discretion to override that overt policy amounted to the very approach he was criticising; this may have been because he felt it was more evident in his case where responsibility for the decision lay, but he did accept that some of his colleagues saw him as 'slightly heretic' and were uncomfortable with the lack of clarity about how appeals should be made.

His defence of the need for overt, 'rational' policies was still set very clearly within a framework of human judgement and value-based decision-making. He felt using computer packages to draw up a quasi-scientific ranked list of priorities was 'retreating behind methodologies'; the question of whether the NHS should be comprehensive must remain primarily a moral debate, and part of his job as a public servant was to ensure the debate took place. In many cases, he felt a national decision-making process would have been better than local variation.

He felt more comfortable with more subjective judgements about rationing made by GPs in the light of their working experience. He described an occasion when GPs met with a local cardiologist and said in effect, 'will you please stop scraping up senile 75 year-old people off the pavements and bunging pacemakers in? It's a bloody nonsense. What I want is more district nursing.' The respondent felt this was a sensible stance, but not one that could have been built into a rational view of priorities.

However, he was uncertain what part, if any, the public should play in this debate. His particular concern was that it would be easy for majority groups to discriminate against the health needs of minorities, and suggested for instance that in an authority with a large concentration of elderly people, there would probably be little support for infertility services for younger, otherwise healthy people.

In District 4, a decision had been made not to purchase the gynaecological procedure of dilatation and curettage (D&C) for women under the age of 40. This was felt to be justified primarily on the grounds of lack of clinical effectiveness, although saving money was also an important aim.

The authority was preparing a formal protocol for rationing IVF, and a version of it was already operating informally. The eligibility criteria encompassed the woman's age and the number of previous IVF attempts - only three attempts per individual would be funded - and did not include any social criteria, such as the length and stability of the couple's relationship. These criteria were agreed by a public meeting of the health authority but had yet to be put in final written form. The protocol would also include a list of preferred providers to which referrals must be made, and this too was already working informally. GPs would not be allowed to refer a woman directly for IVF, but would have to refer her to a particular hospital infertility clinic. Doctors at the clinic would then decide whether to refer on to a specialist IVF centre. Again, the criteria were based on clinical effectiveness as well as the need to limit expenditure - the respondent described it as a waste of money to buy treatments with a low success rate, but emphasised that social criteria were excluded from the protocol because they had no bearing on clinical effectiveness.

District 6 also had a formal and quite detailed protocol rationing access to IVF, which included social as well as clinical criteria. Women seeking IVF must be under 38 and must be in a stable relationship of at least three years' duration. Both partners must be childless. The total amount available each year was cash-limited, within which treatment was available on a first-come first-served basis. GPs were not permitted to make direct referrals for IVF, but had to channel their referrals through one of two hospital contracts. No extra contractual referrals were allowed.

Prior to this protocol the authority had had no service within its boundaries, and had largely ignored the problem, funding the occasional extra contractual referral if put under sufficient pressure. It was then decided that 'we needed something a bit more formal so that people knew where they were. Either we were going to do something or we weren't. If we weren't, we would have to say so.' It was clear that there was more demand for the service than the authority could afford, so some selection criteria had to be established.

The respondent did not regard the criteria eventually selected as particularly fair; 'all criteria are going to be unfair to somebody, and if you are 38 and one month and you miss out, or if your now husband had a child by a partner 20 years ago, long since forgotten, it is not fair.' However, he felt that it was defensible for two main reasons. Firstly, because the authority had been receiving growth funding it could afford to invest in a new service without having to cut any other service, and it had been able to maintain relatively low waiting times for most of its services. This meant it could say to the public 'that nobody requiring a hip replacement will not get it as a result of this investment.' In addition, it had given a positive emphasis to the fact that it was investing in a new service for local women where previously none had existed, and publicly played down the restrictions on access.

Turning now to the authorities with no formal policies limiting access to certain treatments for specific groups of people, the respondent in District 1 believed clinical decisions of that nature were taken, with doctors refusing to operate on people who are grossly overweight or delaying treatment for smokers unless and until they give up smoking, on the grounds that it would not be effective if they continued to smoke. District 8 also stated that any rationing to groups of people on the basis of age or lifestyle, if it is taking place, is done purely by consultants and the authority has made no formal decisions in this area.

The respondent in District 3 thought individual purchasers trying to restrict services in their area to those over a certain age, for example, were on dangerous ground and very vulnerable to criticism from providers, especially provider clinicians. Although he believed continuing financial restrictions would eventually force purchasers to set clearer criteria about who should be eligible for, say, heart surgery or renal transplant, he felt this was an issue where purchasers needed to agree a joint line and support each other locally or even nationally, if possible.

Assuming that current trends towards greater explicitness would continue, the respondent in District 5 expected that rationing targeted at specific groups of people such as smokers could become a reality within three to four years. He thought the alternative of leaving everything to the value judgements of clinicians was 'not a bad thing in some respects', but that something more explicit would probably be needed in future. He, too, was concerned that some sort of collaboration between purchasers or a national policy was needed to avoid geographical variations in what was available. Some differences in the how much service was available would be acceptable, in response to local circumstances.

The effects of the market on rationing

All the authorities which answered this point (7 of the 8) agreed that the development of the purchasing role has made rationing more explicit, although they were not unanimous about whether they could have made such explicit decisions before the introduction of the market.

The respondent in District 1 felt that rationing was now becoming more overt because of the split of purchasers and providers into two separate organisations with different, sometimes conflicting interests. This he felt made it solely the responsibility of purchasers to make rationing decisions and to make clear to the public which items fell below the cut-off point on its list of priorities. Significantly, however, he pointed out that the local exclusion of assisted conception pre-dated the purchaser/provider split and had been in place for a number of years. It had been agreed jointly with providers and compliance had been good. His point was that it had been possible to take explicit rationing decisions before the market was introduced, but that the role of purchaser made it easier or at least more likely that such decision would be made and that they would be explicit not just within the NHS but to the public at large.

The respondent from District 3 said in the questionnaire that these kind of rationing decisions could not have been made before the introduction of the market, because 'I don't think the focus or information was available at that time.' At interview he elaborated that being required to set up contracts and to monitor extra contractual referrals had given purchasers a whole new set of information about what service activity was taking place - before that 'we had no idea what was going on, did we?' Thus although theoretically an old-style district planning department could have taken a decision not to fund sterilisation reversal, it had simply not found its way onto the agenda.

He believed the market had also brought discussions between purchasers and providers into sharper focus, by linking funding much more directly to activity. This meant purchasers now had greater leverage over providers to make sure that any rationing decisions they took were actually implemented.

The respondent from District 4 felt such decisions <u>could</u> have been taken before the introduction of the market but she was not sure if they <u>would</u> have been. She felt that the market had created a climate in which such decision-making processes were more likely. However, she added that so far this effect was operating 'only to an extremely limited degree - GPs and clinicians still act largely as responsible gate-keepers.' In other words, the great majority of rationing decisions were still being made implicitly, on the basis of clinical judgement.

In District 5, it was not solely the market that had brought about rationing of cosmetic surgery; it was a combination of responsibility for the contract being devolved to them from the regional health authority, and the introduction of the Patient's Charter. Plastic surgeons had used long waiting times as an implicit rationing mechanism, giving people with cosmetic surgery requirements lower priority than others. By imposing a maximum waiting time, the Patient's Charter exposed this situation, and the authority had to fund treatment for cosmetic cases in order to clear their waiting lists. This led to greater questioning of the kinds of operation being undertaken; since the implicit mechanism of long waiting times was no longer available, a more explicit alternative had to be found.

On the other hand, the authority's decision to ration IVF explicitly was a direct result of the introduction of the market. Because previously there had been no IVF provision within its boundaries, it had not been a financial issue, even though residents were able to obtain treatment in other districts. The setting up of the market meant the authority had to take on responsibility for funding all treatments for their residents, wherever they were treated. Following a detailed public health analysis of the costs of achieving a successful 'take home baby', it was decided not to fund any IVF, although it was recognised that this was a difficult and 'political' decision.

This respondent also felt that such decisions could have been made before the introduction of purchasing, but that the new system made them more likely and more explicit. Again, he pointed to the increased information made available through contracting as a crucial factor. Public health departments had also found the market a stimulus to undertake more detailed analyses of effectiveness. However, in a situation of static funding, new priorities identified by public health could only be funded by ceasing to purchase some other treatments and getting agreement on this remained very difficult.

In contrast, the respondent from District 7 believed these kinds of decisions could not have been made before the introduction of the market. He felt that the use of protocols was likely to be extended to other areas in the near future; this would involve looking at clinical outcomes and effectiveness. However, he was not convinced that this amounted to rationing, but was more a question of providing appropriate care and challenging traditional practices. He believed the market was making previously implicit rationing mechanisms more explicit; extra contractual referrals exposed areas where local services were inadequate, resulting in people seeking care elsewhere. The imposition of maximum waiting times was also given as an example of new explicitness being given to old rationing methods, although again this was more a result of the Patient's Charter than the market.

The key difference between the old system and the new market system was again identified as the availability of clearer information about the levels and types of activity going on, although it was noted that there are still flaws in the information process and that occasionally excluded procedures are performed before anyone realises. Increased knowledge, coupled with the firm linking of activity with money through contracts, has given purchasers much greater control and an explicit framework within which to organise change and make rationing decisions.

However, the existence of a clearer framework did not in itself make the hard decisions any easier. The authority recognised that growth money would probably dry up soon and that it would then come to a 'crunch point', when new services could only be set up if something else were rationed. For example, the authority was considering whether to reverse its ban on IVF and introduce some form of service, but 'the only way we could achieve that would be by pruning something else out.'

Finally, the respondent from District 8 also believed that purchasing was making rationing more explicit, and that the authority's current list of exclusions would have been impossible before the reforms because consultants would simply have refused to accept it and there would have been few ways of persuading them otherwise. Contracts now provided an effective lever to change behaviour - although despite this, the authority was apparently still unable to prevent one plastic surgeon from continuing to perform an excluded procedure.

However, although the market may have provided a mechanism to undertake explicit rationing, it was certainly not the sole cause. The respondent estimated that the Patient's Charter was about 70% responsible for the list of exclusions, because it had forced people to start prioritising.

Key points from pilot survey

- all the health authorities interviewed were carrying out some form of explicit rationing
- all but one had a formal policy stating which services would not normally be purchased (the 'what' question), although many of these policies allowed for clinical exceptions to be made
- in many cases, eligibility criteria or scoring systems were in use to determine who would have access to services (the 'for whom' question)
- sometimes a decision had been taken to limit the amount of a particular service available, either in financial or activity terms (the 'how much' question).
- although a few of the exclusions cited pre-dated the market, most were instituted after the reforms took place
- the point was made that in many cases explicit decisions were merely formalising implicit rationing that was already taking place
- the most commonly given reasons for decisions not to provide a particular service were financial pressures and evidence of low clinical effectiveness; it was queried by some whether decisions taken on the grounds of low effectiveness constituted rationing
- whilst some participants in the survey felt it might have been possible to take such decisions before the development of the purchasing function, most felt they would not have been taken in practice, and the general consensus was that the market was making rationing more explicit
- the purchaser provider split, greater availability of information and the ability to monitor and enforce rationing decisions through contracts were all seen as market-related factors contributing to greater explicitness
- in some cases, districts had been prompted to clarify conflicting policies following a merger with another district
- the Patient's Charter had also increased explicitness by closing off the option of rationing by waiting lists
- despite all this, implicit rationing through clinical decision-making was still predominant

The survey findings on the nature and extent of rationing being carried out by purchasers are consistent with the evidence presented to the select committee, and confirm that there is considerable blurring of the definition of 'exclusion'.

NAHAT analyses of purchasing plans

How do the findings presented so far compare with other analyses undertaken at the same stage of market development? Despite an increase during the early 1990s in the amount of theoretical work being done on explicit rationing, and growing media interest in individual high profile rationing cases, there was remarkably little empirical work undertaken. The early evaluation of the NHS reforms funded by the King's Fund (Robinson and Le Grand, 1994) provided the most comprehensive view of the development of the market, but did not directly address the question of explicit rationing and how significantly it was being affected by the market.

The most relevant empirical findings come from the series of analyses of the health authority annual purchasing plans commissioned over the last five years by the National Association of Health Authorities and Trusts (NAHAT) and carried out by Klein and colleagues at Bath University (Klein and Redmayne, 1992; Redmayne, Klein and Day, 1993; Redmayne, 1995; Redmayne, 1996). The lessons from these studies are drawn together in the authors' most recent book, 'Managing Scarcity' (Klein, Day and Redmayne, 1996). The most recent of these analyses concerns the purchasing plans for 1995/96, for which work would have started in late 1994, the same year as the select committee investigation and the pilot fieldwork survey.

The authors' assessment is that, in the first year studied (purchasing plans for 1992/93), health authorities were beginning to tackle the 'what' question and redefine the boundaries of the NHS's responsibilities, 'if only implicitly and unconsciously', by listing procedures which they would not purchase (Klein *et al.*, 1996, p.70). This 'explicit rationing by denial was very much at the margins of NHS activity' (p.69), comprising mainly cosmetic surgery, sterilisation reversal, assisted conception, gender reassignment and alternative medicine. The authors also point out that 'the criterion used seems to have been not so much whether the procedures were technically effective but whether they represented activities appropriate for a publicly funded health care system' (p.70), with a particular emphasis on self-inflicted problems such as tattoos and conditions where need is defined not by doctors but by the patient, such as cosmetic surgery. These findings are consistent with the evidence presented to the select committee and the findings of the pilot survey.

Over the next two years neither the number of health authorities reporting explicit rationing by denial nor the procedures concerned varied greatly, although some new procedures were added to the list, such as screening for prostate cancer and several instances of varicose veins. The authors suggest this means there was little increase in explicit rationing during this period, although there are some possible objections to accepting this suggestion. Firstly, it is not clear whether the purchasers identified each year as undertaking rationing were the same as those in previous years or in addition to them. Purchasing plans focus mainly on changes the authority proposes to make the following year, and it would be unusual for a purchasing plan to repeat descriptions of changes made the previous year. Thus each year it could be a new set of purchasers announcing that they are planning explicit rationing measures, which would indicate an increase in rationing activity, albeit not a very great increase.

The second problem is a related one: because purchasing plans are not fully comprehensive statements of services purchased, one cannot be sure that all authorities practising explicit rationing actually included it in their purchasing plans. The decision might already have been made part-way through the previous year at a meeting of the health authority members, and would therefore not necessarily find its way into a statement of next year's plans. Alternatively, the authority might simply have been reluctant to declare its rationing policies in such a public document as its purchasing plan, even though they were formalised and explicit within the Authority - this links back to the argument in Chapter Three that there is a continuum from implicit to explicit.

Nonetheless, the suggestion that explicit rationing remained a marginal activity is consistent with the findings already reported in this chapter, and it is unlikely that the objections discussed above would result in a serious under-estimate of the extent of explicit rationing.

Even if the extent and range of rationing did not change radically, Klein and colleagues note a marked shift in the manner of presentation, perhaps reflecting the reluctance of health authorities to get into an entrenched and very public position. Outright, explicit denial, they argue, risks provoking the anger not only of patients but also of doctors who see it as a challenge to their clinical autonomy. It could be added that it also risks provoking the wrath of the NHS Executive, which has become increasingly insistent that total exclusion of any potentially beneficial service is not acceptable. (See, for example, Secretary of State for Health, 1996).

They suggest that, in response to medical objections that the effectiveness of any given treatment is dependent on the individual circumstances, the NHS had moved away from 'limiting the NHS menu' to 'specifying the conditions of eligibility for treatment in co-operation with the medical profession. In effect, the criteria were re-medicalized' (Klein *et al.*, 1996, p.71). This reverted to a more traditional division of responsibilities between meso and micro levels of rationing, with doctors deciding how to turn broad allocation decisions into detailed implementation. However, there was a shift towards more collective medical decision-making based on effectiveness evidence and expressed through guidelines and protocols, rather than purely individual clinical judgements of what is appropriate and effective.

This certainly matches with the evidence already presented on the ambivalence surrounding the term 'exclusion', and the very widespread existence of exception clauses.

Klein *et al.* note a further but consistent shift in 1995/96 purchasing plans towards an increasing emphasis on effectiveness. The number of health authorities reporting exclusions (but with the all-important let-out clause of exceptions on clinical grounds) was increasing, as was the number of different procedures. However, some of the latter increase resulted from unpacking generic categories such as cosmetic surgery into specific procedures, and they remained largely marginal NHS activities. The emphasis on effectiveness was also leading to more discussion of reducing some procedures, rather than excluding them, and was making purchasers more cautious about new technologies such as beta interferon.

The authors conclude that this phase represented 'a partial retreat from explicit rationing by exclusion. The veil of clinical judgement had proved too useful to discard...Services and procedures would not be struck off the NHS menu but given low priority in resource allocation. And within restricted resource envelopes, clinicians would decide whom to treat, and how, according to their own criteria of appropriateness. Science, it seems, had come to the rescue of scarcity, offering a new legitimation of selectivity. Patients would be turned away not because resources were scarce but because treatment would not be appropriate in their case' (p.73).

It is important to re-iterate, however, the authors' point that clinician-led rationing is no longer the entirely individual and implicit enterprise it once was. The retreat from explicit rationing by exclusion is not so much to implicit rationing, as to explicit rationing by thresholds or by selection. Thresholds for treatment will be set on the basis of effectiveness or severity of the condition; the general criteria for determining whether a patient is above or below the threshold may be explicit, but their application will be informed by clinical discretion and the circumstances of the individual case. This finds an echo amongst many contributors to the current debate who favour the idea that doctors should exercise professional leadership in rationing, but within an explicit, collective framework. (e.g. New and Le Grand, 1996; Lenaghan, 1996).

Although Klein *et al.* (1996) thus demonstrate that explicit rationing (in some form or another) is a growing feature in the NHS, they suggest that the internal market might have been expected to produce much greater explicitness than it in fact has. This they describe as 'apparent failure to follow the logic of the 1991 reforms' (p.50), although they point out that the supposed 'logic' of the reforms may always have been unrealistic. Political and public unwillingness to countenance a challenge to the comprehensive, universal principles of the NHS, coupled with the practical problem of trying to reconcile hard and fast explicit rationing decisions with the need for clinical judgements about individual cases, were bound to constrain the effects of the quasi-market to some extent.

However, whilst it may be true that the market has not made rationing as explicit as might have been expected initially, it may still have made it more explicit than it would have been had the market never been established. The evidence from the pilot fieldwork survey suggests that practising managers at this stage felt the quasi-market was indeed an important influence. This is an issue which is examined in greater depth in the following chapters.

Chapter Seven

The experience of explicit rationing within the NHS quasi-market - evidence from first round case study interviews

This chapter analyses the results of the first round of interviews, undertaken between June and September 1996. (See Chapter Five for further detail on methods). They give a more recent account than that presented in Chapter Six and show how rationing has developed as resource constraints have tightened. In the early stages of the NHS reforms, NHS spending was increased to ease in the reforms. By the time these interviews were conducted, the new, much harsher limits to public spending were having their effect.

All of the interviewees described the 1996/97 contracting round as extremely tough, and none of them anticipated getting any growth funding for 1997/98 - on the contrary, they expected to have to make further reductions in service.

Values and criteria for priority setting

Before interviewees were questioned specifically about rationing and the market, they were asked about any work done locally on articulating values and criteria for priority setting, and whether this had been useful in practice. The aim was to establish whether there was a general climate of greater explicitness about how priority setting was to be achieved and on what principles it was to be based, but also whether this had been translated into practical decisions. It is quite possible for very clear statements of values to co-exist with implicit or ad hoc decision-making.

Avon

Avon had done a lot of work involving staff in trying to draw up an agreed statement of values, but the new authority from 1st April had decided to review what had been produced. Redrafting was still in progress, but a decision had been made not to consult staff again. It was decided the lead should come from the top - that authority members had to debate and agree their own values, which could then be shared with staff and their practical implications discussed.

The current draft at the time of the interview was as follows:

> 'Our key aim is to secure the best possible health for local people within the resources we have available by:
>
> - developing an understanding of people's health care needs in their particular local circumstances
> - seeking equity of access to health services for the whole community
> - focusing on those aspects of service which maximise health gain by demonstrating clear improvements in health status
> - targeting resources to individuals, groups or communities with particular needs
> - working closely with our residents and with all those involved in their health care
> - ensuring that services are delivered in an appropriate way to the service user, wherever possible in a primary care setting easily accessible to local residents
> - being publicly accountable and securing the best use of resources in terms of efficiency, quality and effectiveness

- valuing the rights of individuals and seeking methods for involving the community in health care decisions generally, whilst developing within individuals a sense of responsibility for their own health.'

The chief executive felt that until the value statement had been finalised it was of limited practical use. Nonetheless, she described clear examples where the authority's concern about geographical inequity and inaccessibility within its newly enlarged boundaries was prompting discussion about service change. She also stressed that a statement of values could never remove the ethical complexity of the issues faced, and that practical experience might mean further adaptation of the values.

The deputy director of contracting felt the values expressed in the statement were in the minds of contracting staff as they set priorities for negotiation, but that this was probably because those values were already deeply held. Writing them down had not made a huge difference. She also said contracting staff were very aware of and uncomfortable with the unavoidable trade-offs between competing values, and felt some of the decisions that they were called upon to defend did not match well with their own beliefs.

The director of finance took a similar view that they were already his values anyway, but that they were something to strive towards rather than actually achievable. Realistically they had to be subject to practical constraints.

The director of public health agreed that staff really did think about the values day to day, but identified a curiously British reluctance to talk about one's beliefs - 'it's like an Englishman trying to speak French.' One of the biggest problems, he felt, was not so much the conflict between the values themselves, but between them and various political imperatives from the Department of Health, such as reducing waiting times. However, conflict between values was also a problem, exemplified by the individualistic claims of expensive ECRs versus equity for the majority.

Cambridge and Huntingdon

Cambridge and Huntingdon set out its values in a document called 'Framework for Partnership' in 1994. These were taken from work already done by East Anglia Regional Health Authority, which was in turn based on work by Maxwell (1984), so their origins pre-date the market. They are known locally as the EEEAAR list:

- equity
- effectiveness
- efficiency
- accessibility
- appropriateness
- responsiveness

The chief executive explained that they were adopted at a time when Cambridge Health Authority, Huntingdon Health Authority and Cambridgeshire Family Health Services Authority were all trying to work as one health commission, whilst retaining separate identities and their own set of non-executives. As a result, decision-making was unwieldy, and the values were 'imposed', in his words. He did not feel it had been worth doing any more participative work with board members until the creation of the new health authorities on 1st April 1996. By the time of the interview, the director of public health was leading work on a new statement of values for the authority.

In the Commission's 'Strategic Framework for Acute Services 1995/96 to 2001/02', four principles were identified:

- the published values of the Health Commission (i.e. the EEEAAR list)
- sound evidence about what constitutes effective clinical practice
- responsiveness to the views of local residents and health professionals
- the requirement to sustain teaching and research.

(This covers broadly the same areas as Avon's statement of values. The two notable differences are that Avon does not make explicit reference to teaching and research, whilst Cambridge and Huntingdon does not single out individual rights and responsibilities, although these are to some extent implied by values such as appropriateness and responsiveness).

Inevitably, the very general and superficially incontrovertible nature of these principles meant there were mixed views about whether they had had any effect in practice, even though some further discussion had taken place with groups of staff about what the terms meant and how they could be applied.

The chief executive thought there were times when the values had been a prominent influence on decision-making and other times where there was 'evidence of not much more than a sort of after-the-event rationalisation'. This latter approach was also identified by the senior registrar in public health, who commented 'rather than the values driving the decisions, I think the decisions are made and then the values...inform it, but I don't think they drive it.' Although he believed some people in the organisation were well-informed about the statement of values and realistic about the tensions between values, he also thought there were probably many staff for whom they meant nothing or who had not heard of them.

The director of acute commissioning shared the view that a lot of things were done because 'they're politically flavour of the month'. When it came to rationing decisions about what <u>not</u> to do, however, he felt these were informed mainly by the question 'If we didn't do this, how much suffering would it cause?' As an example, he said they had not taken a very strict approach to the reduction of waiting times, seeing an 18 month wait for a varicose vein operation as less important in terms of human suffering than making sure the parents of a child with complex disability got immediate specialist support. This constitutes a judgement about equity, effectiveness and accessibility, albeit not fully spelt out.

The director of finance took a more positive view that, although there were problems translating values into practical effect, most people in the authority did adhere to them and tried to use them in their daily work.

District Three

District Three had done a considerable amount of work on values, culminating in the following four principles for making choices:

> The responsibility of the authority is to enable the local population to be as healthy as possible by use of the available resources.
>
> Choices and decisions on priorities should be made, to the greatest extent possible, on rational grounds.
>
> To justify their use, treatments (particularly, but not exclusively, new ones) should have the balance of evidence in their favour.
>
> Financial resources for NHS spending for local residents will always be finite (even if NHS funding is increased).

Although he had been one of the main proponents of establishing principles (including effectiveness) on which rationing decisions should be based, the acting director of public health argued the exercise had been 'totally useless' - it was helpful in 'keeping the profile [of rationing] up', but made no difference to practical decision-making. The pressures of the most recent contracting round had meant principles had taken a back seat, and propping up acute services had become the rationale for the choices made. One factor in this was 'the fear of the consequences of being tough...so in my more optimistic moments I think it's because we haven't had the will and the toughness to apply the principles, rather than that they're completely useless.' It was right in principle that decisions should be made on rational grounds, weighing up the benefits, risks and costs - it was just that in practice 'subjective opinion or prejudice' still tended to win the argument.

The director of finance confirmed that from a financial perspective the principles had made little difference. He had not been closely involved in their development, and felt they had little relevance to his day-to-day work.

The chief executive felt members of the public needed to be involved in reviewing and perhaps revising the principles over the next few months. He also discussed the difference between establishing broad values (such as equity or beneficence), about which he would expect most people to agree, and translating those into practical criteria for allocating resources, such as the person's age or whether they had economic dependants. In the absence of such criteria, statements of principles had only limited value. In his view, the rationing agenda was now shifting from denial of certain interventions towards an emphasis on clinical effectiveness and cost-effectiveness, and that this was the most promising route for finding practical criteria for priority setting.

Discussion

The amount of effort being put into the development of statements of values does support the idea that there is a general climate of greater explicitness about the principles on which priority setting is to be based. However, there is still much work to be done in exploring the underlying conflicts between the principles endorsed, translating these into practical criteria for decision-making, and in turn ensuring these criteria are actually applied. In the meantime, there is still room for implicit decision-making to co-exist with very clear statements of values.

Evidence of explicit rationing and contracting

Avon

Avon had taken legal advice on the question of excluding treatments and had been advised 'never to say never'. As a result, although they had a range of explicit limitations and virtual exclusions of services, they always allowed exceptions on clinical grounds. As they acknowledged, this placed some of the burden back on the clinician to manage rationing in practice, albeit within a very explicit framework.

Avon's service specifications included a list headed 'restricted purchasing of the following procedures', which included plastic surgery, sterilisation reversal, assisted conception and wart removal, as well as the more unusual ME, implantable defibrillators, bone marrow transplants, intervertebral surgery and epilepsy surgery. In some cases, quite detailed circumstances were described in which exceptions could be made. In others, the restriction took the form of a limited number of cases per year, without specifying how they were to be selected. The restrictions on IVF combined a limited number of cases with selection criteria which included the mother's age, number of previous children and whether the applicant was in a heterosexual relationship of at least two year's standing. For bone marrow transplants, there was a cash limit which was expected to cover an approximate number of cases.

Plastic surgery provided an interesting example of shifting rationing boundaries between the purchaser and the clinician. At the request of the plastic surgeons, indicative volume levels had originally been included against each of the restricted procedures. The contracting department had subsequently been approached by consultants to ask for an exception to be made for individual cases. The department's response was that it was up to consultants to decide priorities within the indicative volume. After three years, the consultants requested the removal of the indicative volume levels, preferring to manage the rationing themselves, within the overall volume of the contract. The director of public health confirmed that clinicians did not in general like the approach of restricted volumes for certain procedures, because they felt it demonstrated a lack of courage on the authority's part, and resulted in responsibility being unfairly loaded back onto them. 'And they've got a point, of course,' he added.

The deputy director of contracting was hesitant about the use of the word rationing to describe the plastic surgery restrictions, preferring the description 'a fairly carefully followed process about a limited resource and making the most use of it'. She expressed a strong organisational commitment to a population-wide view, which placed less emphasis on rationing for the individual and more on maximising health for the majority. She argued contracting or commissioning staff tended to feel greater corporate responsibility and less professional allegiance than public health staff, and that this could explain why the director of public health accepted the term 'rationing' more comfortably. Acknowledging rationing might somehow be disloyal - 'we may not personally agree with it, but we've taken the Queen's Shilling.'

In fact, the director of public health, when asked if staff locally used the term 'rationing' to describe exclusion of services said 'Yes. I think we do. Prioritisation, yes, I think we do.' When asked about this shift from 'rationing' to 'prioritisation', he added, 'Well, there used to be quite a big embargo against using the 'r' word, there really did, and it came right from the top through the regional offices, and so we couldn't ever use it in correspondence. It's lightening up now, I have to say - it's becoming more acceptable.'

The chief executive and director of finance agreed with the term rationing, although the director of finance qualified it by saying that not providing something which ought not to be an NHS responsibility, such as tattoo removal, did not seem like rationing to a purchaser, although it would to people outside the organisation.

Evidence on effectiveness was seen as an increasingly important influence on priority setting, and both 'improve effectiveness based on evidence' and 'maintain an active R&D programme' appeared as main headings in the authority's corporate contract for 1996/97. As the chief executive pointed out, however, it was easier to use such evidence to challenge new developments than to change mainstream services. She recognised that most provision continued to follow historical patterns and was not informed by research on effectiveness. What evidence there was about existing provision was not always used to limit or exclude - often it suggested that more of a particular service should be purchased. In a time of no financial growth, however, this meant something else would have to be rationed to achieve that reinvestment, either on the basis of low effectiveness or some other criterion such as low priority, or by more implicit means such as an across-the-board percentage cut. Thus effectiveness evidence could be both a direct and an indirect cause of rationing.

The director of public health also drew attention to the fact that, despite evidence on effectiveness, ensuring that services were properly targeted on those who could benefit most from them was another matter. He was also concerned that a very vocal minority might succeed in getting services of questionable benefit, because of the emotive nature of their appeals for funding.

He also felt strongly about government guidance that beta-interferon should be made available, despite research evidence that it was not very effective, which they could have used to support a decision to exclude it from the NHS. 'You don't want that kind of help from the government.' Avon had in fact decided not to make additional funding available for it, and was prepared to argue its case with the Department of Health if necessary. The director of public health blamed the pressures of the pharmaceutical market, which meant new drugs had to be promoted vigorously by their manufacturers.

Contract specifications in Avon were quite detailed, not only in listing exclusions but also in describing the range of services that would be provided. The director of public health estimated that about a third of the contracts contained exclusion clauses. The director of finance pointed out that most contracts were financially a simple block arrangement with each trust, specifically to avoid transaction costs - cost and volume arrangements accounted for less than 1% of total activity. However, this did not necessarily mean any loss of control over what is purchased - 'whilst we've got block contracts, we haven't got block activity.'

Although nobody commented on difficulties in monitoring and enforcing decisions to ration services, the director of finance did describe general difficulties in ensuring clinicians observed the terms of the contract. The director of public health said they tended to start specifying numbers of particular procedures only when there was an identified problem in that specialty, usually a waiting list. For the most part, contracts relied on discouragement to do certain procedures (such as grommets for chronic otitis media), rather than an absolute prohibition. He said 'we're quite a small organisation and we've found from bitter experience...not to look for trouble where you don't have to. So in many ways, if some specialties and services are OK and not clamouring, we're happy to accept that at face value, and get on with the specialties that are...I can happily leave much of the ENT work which I think genuinely is of low priority alone, despite the fact that it is of low priority.' He did feel contracts were really quite specific, but often reliant on co-operation rather than compulsion.

Cambridge and Huntingdon

Cambridge and Huntingdon is of course well known for one of the most public acts of explicit rationing to date, and one which has had a tremendous influence on the quantity and quality of debate on the issue - the decision to deny the leukaemic 'Child B' a second bone marrow transplant. This was also a clear example of effectiveness being used as the basis for a rationing decision, as the chances of success were perceived to be very low, even though the family wanted to pursue every possibility. It is worth noting, however, that this was a decision which arose not through the normal purchasing plan and routine contracting system, but as an extra contractual referral, a point which will be revisited later.

Unsurprisingly, then, staff in Cambridge and Huntingdon expressed willingness to take explicit rationing decisions. They were involved in national debate about the need for explicit rationing and for the government to acknowledge and support it. Staff were at ease using the word rationing - as the director of acute commissioning said, 'I think a lot of people try and avoid the "r" word and say it's prioritisation, but it just seems to be a matter of semantics. If you deny something to somebody that will be of benefit, however you manage it, however you define it, then it is rationing.'

Some services were completely excluded (homeopathy, osteopathy, chiropractic). Bearing in mind the Secretary of State's requirement that no service which is clinically effective should be excluded altogether, this was justified on the grounds that they were totally ineffective - in the words of one interviewee, 'to put it in a nutshell, homeopathy is to us pharmacological nonsense', although he acknowledged that people who believed in it would not accept that definition of effectiveness.

Some services which were experimental or of low effectiveness were listed in the contract with Addenbrooke's as 'not purchased by CHHA' [Cambridge and Huntingdon Health Authority], including small bowel transplant, laser treatment for soft palate (snoring) and sterilisation reversal. These were also excluded as ECRs to other providers. However, in interview the director of acute commissioning said that exceptions would be made on sterilisation reversal in particularly tragic and unforeseen circumstances - a social judgement of ability to benefit, rather than one based on technical effectiveness.

There was a lengthy list of plastic surgery interventions which were normally excluded, some of them accompanied by an explicit statement of when exceptions would be permissible (e.g. breast reduction where the body mass index is <26 and the weight of the breast is >500g, or tattoo removal where it is a barrier to employment). No exceptions were suggested for the correction of bat ears for people over 15 years, a judgement of low priority rather than low effectiveness. Gender reassignment, including surgery, was purchased but only 'within a limited cost envelope and following psychiatric tertiary referral'. Assisted conception was purchased, precisely on the basis that there was evidence of its effectiveness for some women, according to a set of eligibility criteria. Even so, the chief executive expressed some doubts whether age limits, although justified in terms of clinical effectiveness, were also used as a convenient 'hook to hang a rationing decision on', as a way of reducing the numbers to affordable levels. The publication of an explicit policy on assisted conception was necessitated by the merger of two health authorities, one of which did not purchase IVF, whilst the other did.

The Addenbrooke's contract contained a list headed 'Indicative service levels - clinical effectiveness'. This was an interesting combination of both interventions felt to be of dubious efficacy, where Cambridge and Huntingdon wanted to keep numbers down (such as varicose veins or tonsillectomy), and well-proven techniques such as hip replacements or cataract extractions, where the aim was to keep numbers up. The authority planned to monitor progress against these indicative levels specifically during 1996/97.

Finally, the director of finance/primary and community commissioning raised the issue of the use of eligibility criteria for continuing care as an example of explicit rationing. Unlike many other districts where the government's intervention was likely to mean reinvestment in long term health care, Cambridge and Huntingdon had historic over-provision of long term beds and was likely to end up reducing this in order to spend more on acute services. However, he felt that nobody within the organisation would regard this as rationing because it was retrenchment from something the health service should not have been doing anyway. The authority was also doing work on reducing some health visiting services, to improve targeting on those in particular need. The word 'targeting' was preferred to 'rationing', on the grounds that the overall level of resource was not going to be reduced.

As far as the contracts themselves were concerned, examples have already been given from the Addenbrooke's contract of quite detailed limitations and exclusions. Detailed listing of exclusions also applied to ECRs, where strict control of funding was exercised. Cambridge and Huntingdon was seeking to move from the current detailed annual contract towards longer term contracts with Addenbrooke's, based on an understanding of mutual reliance and partnership, subject to the constraint of annual budget setting. In practice, however, Addenbrooke's had yet to agree to sign the 1996/97 contract (in July 1996).

The deputy director of acute commissioning explained that the plastic surgery specification had been tightened up the previous year, when it was found that the previous protocol drawn up some years ago had had little or no impact in changing the pattern of provision. This suggests that monitoring and enforcing compliance can be a problem, even when a decision to ration a procedure had been made explicit. There were also continuing problems in finding out about what was being provided - laser treatment for snoring had originally been a research project which had been included in NHS provision without explicit agreement and was discovered by chance during a review of the ENT waiting list.

Despite Cambridge and Huntingdon's willingness to make explicit choices, the chief executive emphasised that most rationing was a product of historical levels of investment in different services, and that a traditional lack of provision for some vulnerable groups was probably the most important kind of exclusion - 'the hidden iceberg of rationing in the NHS,' as it is described by Klein *et al.* (1996, p.82). The chief executive cited the example of elderly people in the community being put to bed at 6.00 pm because district nurses were not available to do it later in the evening.

District Three

District Three operated a number of exclusions and limitations. Firstly, the purchasing plan for 1996/97 outlined three areas where, on the basis of low clinical effectiveness, it was planned to reduce services: these were grommets, D&C in women under 40 and in-patient care following a stroke. It was stated that this would be handled through 'working with clinicians rather than trying to impose standard protocols for care'.

Secondly, the contracts portfolio document for 96/7 contained several relevant sections, which will be described in turn. There was a six-page list of contract exclusions by provider, which covered a broad range of services such as bone marrow transplants, some expensive drugs, cochlear implants, eating disorder services and in-patient homeopathy, to name but a few. These were not necessarily excluded altogether; some of them were available as part of routine contracts with other providers, and some of them could be granted ECR approval at the hospitals where they were excluded from contracts.

There was also a list of 'special services', which explained where referrals could be made for these services. Some were mainstream but specialised services such as neonatal intensive care. Others were more controversial, such as bone densitometry, acupuncture or gastroplasty (stomach stapling) which in other authorities are often on the list of services not purchased, but which in District Three were available provided the referral was made to a specified provider. The list also included some services which were accessible only via a gatekeeper, including eating disorders and addiction services via local psychiatry, and pain relief via local anaesthetic services.

A separate policy on low priority treatments listed treatments which would not normally be purchased, although exceptions might be made via the ECR process. The list covered the familiar range of cosmetic surgery, sterilisation reversal, gender identity surgery, adult orthodontics and dental implants. The director of public health explained how this had been supported and even driven by GPs, including GP fundholders, seeking backing for what they felt were difficult but necessary rationing decisions.

Assisted conception techniques were also included, and were dealt with in more detail in the subfertility treatment purchasing policy. As in the case of Cambridge and Huntingdon, the need for an explicit policy was in part related to differing practice between merging health authorities. Public health did a considerable amount of work looking at research on cost-effectiveness, as a result of which it was decided that investigations and a detailed list of first-line drug/hormone therapies would be purchased, as well as donor insemination. Tubal surgery would be on an ECR basis only, and IVF, GIFT and IUI would not be purchased at all, except for couples already on the waiting list in April 1995. This reflected District Three's emphasis on cost-effectiveness, in that spending on drug/hormone therapy produces a much higher return in terms of healthy babies born per amount spent than IVF. In interviews it was explained that the decision had also been informed by a view amongst some executives and non-executives that unexplained infertility, once all possible investigations had been exhausted, was not strictly a medical condition but a social condition, and that therefore it should not be an NHS responsibility. The chief executive suggested he might challenge that approach and review the policy. He also indicated that in practice some cases of IVF are funded - sometimes because it is simply very difficult to say no to everything, and sometimes because GP fundholders have tried to circumvent the ban by proceeding so far with preparation for IVF that it becomes impossible to refuse to fund it.

District Three had decided it would fund beta interferon - in fact, the budget set aside was considerably underspent, following the introduction of a protocol agreed with local clinicians. The chief executive commented that national intervention on this issue had been helpful in prompting a debate with the Multiple Sclerosis Society on the need for careful targeting of the drug and for giving people with MS clear information about the potential unpleasant side effects. This meant the MS Society supported very cautious use of beta interferon.

Despite the lists of exclusions, the chief executive described the great majority of contracts as 'relatively broad brush', leaving a lot to the individual clinician to manage. However, the district was gradually disaggregating contracts between specialties and specifying more closely who would be treated. This was particularly true in developing specialties, such as renal medicine, which will be raised again later.

Future rationing plans

All three districts expected to have to make difficult, explicit choices in the coming contracting round.

Avon

Avon was expecting all its providers to come up with plans for releasing cash, in discussion with purchasers. The chief executive described the relationship with providers as one of 'armed neutrality' while they worked together on contentious proposals, including closing family planning clinics and reducing school nursing. Her preference was for greater explicitness; whilst recognising that clinical judgement would still have a major part to play in interpreting and applying purchasing policies at an individual level, this should be 'against a basis of clarity which we have the guts to provide'. The director of public health described how 10 or 12 main disease entities had been identified, for each of which they planned to list those interventions which are of key importance and those which are least important and might be the target for service reductions. He was also particularly concerned that both renal and oncology services were expanding rapidly, but did not foresee any explicit rationing in these areas in the immediate future.

Cambridge and Huntingdon

Cambridge and Huntingdon foresaw particular difficulties trying to keep elective surgery within the activity levels agreed with Addenbrooke's; although they did not expect to be limiting particular procedures, the provider was looking to them to give a lead on how to handle the problem, and it was anticipated that waiting times would rise across the board. At a more explicit level, they were also trying to reconcile two lists - one of new developments they felt were essential and would improve health, and the other of services they might cut to achieve that reinvestment. In the words of the chief executive, 'Now, the question is, how radical, brave, foolhardy are we going to be? Are we going to halve that list and know that potential health gain out there is lost, because we're spending it on things that frankly don't deliver that health gain, but politically we couldn't cut?' He argued that if the government were to allocate additional funding to the NHS in the run-up to the election, it would only delay the time when the inevitable need for explicit rationing would be acknowledged and action taken. The director of public health echoed this: 'I don't think we should duck this any more.'

District Three

In District Three, as in Avon, rapid growth in renal services was a pressing concern, and one which it was felt would have to be tackled more explicitly. The use of a financial cap was proving difficult to maintain, given local clinicians' increasing reluctance to make choices between patients to stay within budget. The authority was therefore working with the provider on a protocol which would set out eligibility criteria for being accepted onto the programme.

Discussion

Initial results from the three case study health authorities support the conclusions reached by Klein *et al.* (1996), namely that explicit rationing has increased, and that there is increasing use of the effectiveness criterion to justify the exclusion of services, but that explicit rationing remains marginal.

There was certainly evidence from the case studies that explicit rationing was taking place, and that all the health authorities believed more of it was inevitable. The kind of procedure-specific exclusions being applied and the detailed protocols about who would be accepted for treatment and where exceptions could be made are unlike anything seen at district level before the introduction of the market. Clinicians may in the past have had some similar guidelines - for example the use of strict eligibility criteria for renal dialysis in some treatment centres as early as the 1970s (Wing, 1983). However, as Klein (1995) points out, it was 'not ministers or civil servants' making these judgements, but 'the clinicians concerned' (p.78) What is new is the way that purchasers have become involved in agreeing or even leading the development of such protocols, and in using them as a way of controlling clinical activity.

A similar shift is apparent in attitudes to clinical effectiveness evidence. There is considerable evidence from the first round of interviews that purchasers now regard this as a central part of their territory, and believe ideally it should be one of the most important factors, if not the most important, in deciding how resources should be allocated.

In practice, however, health authorities themselves acknowledge that both explicit rationing based on relative social priorities and explicit rationing based on clinical effectiveness are still a relatively minor part of overall resource allocation decisions. Historical service patterns at a collective level and clinical judgement at an individual level continue to play a major role in rationing. There is also widespread concern that effectiveness is treated as a black and white solution to rationing, rather than a complex continuum. There are very few areas where it leads to unequivocal decisions for whole populations - for the most part, effective use of evidence still relies on clinical judgement in the light of individual patient characteristics.

What are the reasons for greater explicitness in rationing?

In the previous section, it was stated that the detail of procedure-specific exclusions and the exceptions to them are unlike anything seen at district level before the introduction of the market. This is by no means to say that the market is the cause; the timing could be purely coincidental, or both the market and the growth in explicit rationing could be the result of some third factor. As Ham *et al.* have described, the immediate catalyst for the review which led to the 1990 Act was a funding crisis within the NHS in the mid 1980s (Ham *et al.*, 1990), and the Government itself acknowledged the financial pressures which lay behind the review in the White Paper 'Working for Patients' (Secretaries of State for Health etc., 1989). Thus both the establishment of the internal market and the growth in explicit rationing could result from pressure on resources, even if the government hoped that the internal market might help avoid explicit rationing by achieving better value for the money.

Interviewees were asked a number of questions which aimed to elicit their views about the causes of explicit rationing, the impact of the market and, more specifically, the impact of contracting. The complexity of the situation is reflected in the uncertainty of the answers received and the way in which most interviewees developed their thinking in the course of the discussion. This uncertainty is in itself evidence of the multifactorial and circular nature of causation in the relationship between rationing, the market and other pressures.

In order to do justice to the complexity of each individual's arguments on this point, they are analysed individually.

Avon

Director of Finance

The director of finance identified lack of resources as the main driving force behind rationing. However, he noted that becoming more explicit had started at a time before the financial situation was as constrained as it now was, and had been prompted by providers wanting new developments. Even though there was no requirement to cut services to stay within budget, it was necessary to move money around with the trusts, away from things of little value towards the new developments. Purchasers had therefore been involved in identifying activities of low priority. He felt it was more appropriate for purchasers to do this than to leave it to implicit rationing amongst doctors, because of the need to maintain a focus on the needs of the population as a whole, and on cost-effectiveness (as opposed to effectiveness alone). Having said this, he did feel the market was increasing doctors' awareness of finite resources.

He felt the purchaser provider split was having an impact in terms of allowing purchasers to focus on health needs, rather than on providers. However, he had doubts about 'this business of charging all over the place for things' - Avon had kept mainly to block contracts specifically to avoid wasteful transaction costs, even though the activity was specified in some detail. He argued providers were realistic that simply having more complex contract formulae was not going to increase the total money available, although recent rises in emergency activity had led to more provider pressure for additional reimbursement.

The extent to which the activity levels specified in the contract were adhered to was questionable. The director of finance believed that sufficient leverage was exercised over providers through regular monitoring, but he also said they varied the levels of activity according to what the demand was anyway and virtually always overperformed. There seemed to be two factors at work here - one was that trust managers found it difficult to control clinical activity, but on the other hand they were genuinely (in his opinion) trying to work with purchasers to squeeze more out of the system. He said, 'certainly the way we have a relationship with the trusts, we're probably working more like one big health authority.' He suggested that, apart from the independent status of providers, the relationship between purchasers and providers was very similar to that between area health authorities and districts from 1974 to 1982. He did not believe abolition of the market would make much difference to the way that relationship worked and had always worked. Some form of contract would still be necessary, but not necessarily any different from what he felt was already a fairly flexible, informal and non-market contracting system now in place.

Finally he questioned whether not buying certain procedures, however explicit the process, was going to save resources - in particular, he felt a sense of shared responsibility with providers to cope with their fixed infrastructure costs. Moving resources around within one provider to buy more desirable developments might be feasible, but taking money out in large amounts was not.

Chief Executive

The chief executive was very conscious of geographical variations in access to services within the district, particularly in IVF, cardiac surgery and renal dialysis, which had become an issue because of the merger of districts with previously different purchasing policies. For the most part, she expected this to exert an indirect effect on explicit rationing; it was unlikely to be publicly acceptable for the district to say that, because people in one half of the district could not have IVF, nor could the other half, so a process of gradual 'evening up' was the only feasible route. To achieve this would require redirecting money from elsewhere, however, and this in turn meant more explicit decision-making elsewhere. Service variations were 'flushing out more explicitly the differences that there are and therefore in itself to a certain extent driving our thinking about where recycled money should go.'

Family planning clinic provision was being 'evened down' to match the level in the district with lower access, but this was coincidental - a decision had already been taken that reducing choice for women seeking family planning services was preferable to not having, say, sufficient renal or cardiac services. Demonstrating that this brought the service down to the same level as in the other half of the district was described as 'a *post hoc* commentary.'

As far as the list of excluded procedures was concerned, the chief executive's primary reason for these was 'our perception of clinical worth or value', in spite of the possibility of growing waiting lists. She added that she took the importance of lack of resources as read - in times of plenty, such procedures might be only '399th on the list' but might nonetheless be funded.

However, sometimes low priority services might not be funded even if the money were available. She explained that family planning staff were working on a proposal to save the money that closing the clinics would save, but in a different way which would allow them to remain open. Her planned response, if they produced such a proposal, was to be 'thank you but no - this is an area where we are not prepared to say that there is this degree of choice and duplication of service, when in other services there is none.'

Monitoring waiting times had been an important factor recently in tightening up referral criteria for varicose veins. Although waiting times were within Patient's Charter targets, monitoring had shown up a 'long tail' in general surgery, much of which was varicose veins. Exclusion of all but ulcerous veins had previously been agreed with clinical staff, but because GPs had kept referring they had continued placing them on the waiting list. The situation was 'brought to a head during this last year's contracting round' and clinical staff were given additional guidance on telling the GPs that they would not accept such referrals, backed up by the same message direct from the authority to GPs. The chief executive felt consultants needed reassurance that they could blame it on the authority rather than have to make the rationing decisions themselves.

The role of contracts and contract monitoring here was ambiguous - clinical staff had been resisting a previous agreement because they did not like enforcing it, but on the other hand monitoring was eventually effective in picking this up and making sure practice changes were achieved. In general, the chief executive thought there was fairly complex interaction between rationing and contracting, and that contracting had brought 'a degree of painful clarity and more precision than was ever the case before'. Sometimes difficult problems requiring explicit hard choices might come up during the year independently of the contracting cycle and contracts might then be simply a convenient place to record such decisions. Sometimes 'the intensive period of debate around January to March, around contracting time...is an appropriate and often inescapable time for it to come out into the open.'

Would growing resource pressures have caused explicit rationing to bubble up anyway, regardless of the contracting system? The chief executive could not imagine how they would have avoided it, and could certainly not envisage a return to the old system of paying without knowing much about what was being purchased. (However, this may be because NHS staff are now so used to contracts that they cannot imagine life without them, rather than because implicit rationing could genuinely not have coped with further resource pressures).

The chief executive was in no doubt that contracts had demanded a level of detailed information which there had never been before, and without which it would have been impossible to know what the authority was buying and decide what it was not going to buy. On the other hand, the quality of information in community and mental health services was still poor, despite the existence of contracts, making it difficult to 'really answer questions about what is being provided within the community part of budgets'. A particular structural problem she identified was the fact that three of the local trusts were in fact generic (acute, community and mental health services all in one trust). Although a separate contract for community services was negotiated with each of these trusts, 'the ability for fudging and little bits of financial flow' meant it was difficult for purchasers to keep control, and in effect the trusts acted like mini districts, doing their own implicit rationing to stay within global budgets. Indeed, whole district management teams had moved over to trusts in the first and second waves of trust creation, and 'they still substantially think of themselves as health authorities.' Here, perhaps, was an example of hierarchical organisation never having gone away, let alone creeping back in in the form of block agreements and partnerships.

Interestingly, this view has been supported by the chief executive of one purely community trust in Avon, who says of combined management of acute and community services 'we tried for 45 years and it didn't work. Every time the acute dinosaurs roared, the community primates were sacrificed to feed their insatiable appetite...I hear it's still happening now in the all-in game reserves' (Pughe-Morgan, 1996, p.20).

A side-effect of this structural problem was that the real district health authorities had been left stripped of expertise and with management costs so low that their ability to assemble and manipulate the necessary detailed information for strong negotiations was seriously impaired. 'Provider capture' was inevitable. It was only with the amalgamation of the three health authorities that a strong enough team could be put together to carry out effective contract negotiations.

The chief executive did not feel Labour had really thought through how comprehensive healthcare agreements would differ in practice from contracts. Abolishing the internal market might lead to some sensible reductions in information requirements such as the efficiency index, but the NHS would continue to need something broadly similar to existing contracts, if it were not to revert entirely to handing over a blank cheque. She argued that a return to an integrated organisation would not necessarily avoid explicit choices - she described the former budget negotiations between District HQ and directly managed units as very tense and often aggressive, with little sense of 'one big happy family', although she noted there may have been a greater cabinet-style closing of ranks after the negotiations had been settled than is the case now between purchasers and dissatisfied providers. Conversely, under the current system there were signs locally that providers were willing to work privately with purchasers on proposals for rationing services, and perhaps to consult jointly on these. Whether they would be willing to share responsibility in public was not yet clear, however.

Director of Public Health

The director of public health brought a new slant to the question of lack of resources: he identified three reasons why spending on some individuals had recently increased dramatically, thereby intensifying resource pressures for the majority of services and requiring explicit decisions one way or the other. These factors were a growing use of intensive care, ECRs for expensive and sometimes dubious treatments and new technology, especially expensive new drugs. For him, inequity in the distribution of resources was an important influence on the need for more explicit allocation mechanisms, and he was particularly concerned that a few individuals were making enormous demands on resources whilst others' needs were not adequately recognised.

ECRs for 'difficult to place' mental illness patients, child and adolescent psychiatry (especially eating disorders) and neuro-rehabilitation for people with behavioural disturbance following head injury have created a particular pressure. However, he was not convinced that the need to approve ECRs explicitly rather than simply making out-of-district referrals which came under another district's budget was the sole reason why these problems were surfacing now. 'It all coincides with the loss of the so-called back-ward, the so-called mental illness bins. You used to have pleasant-sounding named wards like the Ethel Johnson Ward for the Tragically Mad or whatever, and you find out that what it is is a real sink, where these incredibly difficult people are looked after in conditions of confinement...Of course, in our great desire not to tolerate or put up with that kind of thing, we have shut them all, and moved exactly that provision to the private sector, where it is to be had at a price.' Furthermore, professional sub-specialisation in mental health and the general 'contraction of the bed base' had led to much tighter definitions of which patients would or would not be accepted - 'there are now very great sort of lacunae between the various sub-specialties, where these patients just don't fit.'

When asked how rationing fitted into contracting and the wider process of commissioning, he said that at 'most steps along that process, we are implicitly, and in a consensus fashion, making choices'. Having to write things down in a contract had probably made it more explicit, 'but not to the great extent that we once thought it might'. As a small organisation (see chief executive's previous comments), they had learnt to focus their attention where it was most needed and to leave many service specifications broad. They also had a preference for dissuasion rather than prohibition in contracts, partly because contracts were in any case unenforceable.

He identified a moral and a technical component to rationing - in the former, one might decide that it was more equitable to shift money from one service and invest it in another. The technical component was how one actually implemented such a decision - 'and let me tell you, it is still as difficult in the health service to get things changed as before.' In fact, in some cases he thought contracting had made things worse, precisely because it required explicitness - 'you can't work with providers to achieve change that ordinarily they might have agreed with, but, because you're putting them on the spot and saying it's got to be a black and white thing, they kick up and say, "no, I'm not going to do it".' Nonetheless, this does to some extent support the idea that contracts make rationing more explicit, even if in practice it makes it less easy to do.

Contracts had in his view failed to explore the great bulk of provision, and explicit rationing remained marginal. They could not hope to specify every single thing done, and many possibly undesirable activities remained undiscovered. Even when problems were identified and changes in practice planned, contracts alone were often an inadequate vehicle for ensuring changes took place. In plastic surgery, where the authority had made an explicit decision to stop purchasing some procedures in order to meet Patient's Charter waiting time targets, consultants had undermined the decision by filling up their waiting lists with other cases. 'They weren't with us, they weren't in a kind of joint venture, which was to do with the general population, the greater good. It was all about "that's your problem, not mine".'

Nonetheless, the discipline of having to think constantly about waiting times had brought into much sharper focus what was really important and what was not; even if it ultimately only formalised an existing implicit position, it meant an explicit choice was made not to invest more money in that particular service to deal with waiting times.

The director of public health was undecided whether there was any ethical obligation to be explicit. He certainly felt that many public sector staff 'feel quite uncomfortable about doing things that really are down to our prejudices' and acknowledged a general pressure to give the public a clearer account of how and why decisions are made. Whether this should extend to getting the public involved in making rationing decisions was another matter; Avon was embarking on its own research with the University of Bristol to test the public's willingness and ability to be involved.

The possibility was discussed that health authority staff, like the public, might prefer a system where they could be involved in commenting on others' rationing decisions but not have to make the decisions themselves. In this case, it could be doctors who would have to explain how and why decisions had been made, but still retain responsibility for making them. The director of public health thought this might appeal to some authority staff, many of whom felt more comfortable with rules and a framework on which to lay responsibility for the decisions they had to take, rather than having to 'hold the smoking gun' themselves. However, he felt this was not realistic, partly because doctors trained in high tech medicine were increasingly uncomfortable about letting people die and 'don't like making rationing decisions any more', and partly because it was clearly a role for health authority staff to make such decisions.

One possibility for passing some rationing back to doctors lay in the handling of tertiary ECRs. Avon was keen to make paying for tertiary ECRs a provider responsibility as a way of controlling costs, or alternatively to make reductions to the contract price if such ECRs exceed a certain number. The director of public health commented, 'I would rather ration that kind of care than compromise basic health care for the population.'

Finally the director of public health explained his understanding of where the NHS's real financial problems lay: in the creeping increase in unnecessary investigations and use of new expensive drugs of dubious efficacy. He commented on the market's failure to improve the situation:

> 'It was meant to take care of all that, and manifestly it has failed, utterly, utterly failed, and in fact it's made things worse. We have left people who have nothing to do with funding decisions in charge of the consequences of funding, and of course they don't care, you know. And the market reforms have dislocated us who do care about funding decisions from being able to manage and manipulate the process. I could see the logic; I could see the logic, but it hasn't worked.'

Assistant Director of Contracting

Service variation was mentioned again in this interview as an influential factor, combined with financial pressures. Plastic surgery had risen at an unaffordable rate, and this had prompted comparison with the rates in other districts. It had been found that Bristol had an unusually high level of plastic surgery, although this was treated with some caution because of the existence of a sub-regional centre - the effect of this being that cases which might elsewhere be treated by a general surgeon or even a GP were referred instead to plastic surgery. However, it was decided that too much was being done and this had led to the development of the list of limited procedures. (This was the example where the consultants eventually decided they would like to move back to managing the rationing themselves, within broad activity levels).

The deputy director of contracting argued that contracting - and she was emphatic that she meant contracting, not commissioning or the purchaser provider split more generally - had led providers to market their services very aggressively, 'as if the market meant that there were freed resource somewhere, without recognising that it was still the same pot of money and it was just going to be divvied up differently.' This had accelerated clinical and public demand 'for new and better and more. And that has led to discussions which are about limiting that, and if it's not limiting the new and the better, it's about limiting some of the old to make way for it.' In her view, it was only in the last year that providers had begun to realise that there was no new pot of money. This was in direct contrast to the views of the director of finance that providers had been behaving realistically and responsibly until the last year.

Although she did feel the purchaser provider split had enabled purchasers to 'take our minds off the difficulty of implementing some of what we ask for', the virtual monopoly of local provision meant ultimately purchasers still had to have one eye on the effect of hard choices on local providers. This and the fact that the government was anxious to avoid the full logic of the market resulting in unsuccessful providers going bankrupt weakened purchasers' ability to make radical decisions.

She thought the Patient's Charter had been a very important factor, although its effect was now waning in the face of a rising tide of emergency workload, and government acceptance that this should take priority.

She discussed both public and clinical pressure to be explicit. The public 'want and expect explicitness, but they don't demand it' - where their influence was most strongly felt was retrospectively, in explaining the rationale behind individual decisions not to treat, or in consultation about specific service changes. Clinicians, on the other hand, 'repeatedly ask for explicit rationing decisions', sometimes because they felt they needed support from purchasers in turning down referrals they also thought were probably inappropriate, but sometimes because they were using it to try to push purchasers into making additional money available.

Making explicit rationing decisions was hard for purchasing managers and they often felt vulnerable - the deputy director described personally distressing negotiations where she had questioned the appropriateness of what she had been required to do, and even one case where she had had to stop purchasing a service which she had been responsible for setting up in a previous job as a provider. She was affected by the way some clinicians personalised their attack on purchasing decisions, for example by saying purchasing managers were endangering patients' lives. She also referred to the fact that 'the clinical directors are often men and the contract managers here are often women', which gave a particular slant to the power relations between the negotiating parties, as did the doctor versus non-doctor issue. However, she did not believe these kinds of ethical and personal difficulties were ultimately going to stop the process of explicit rationing, and she did not think there was any way of finding an easy technical fix to replace difficult judgements. 'If contracting was simply a technical process, all you'd need was a technician', whereas in fact contract managers were highly graded and trained.

Cambridge and Huntingdon

Cambridge and Huntingdon as an organisation was keen to draw a clear distinction between contracting and commissioning. Contracting, in the words of the director of public health, was 'a subset of commissioning' - the process of agreeing and writing down what the health authority would get for its money. Commissioning was the wider function of priority-setting, strategic planning and managing change with providers in the longer term.

Director of Finance/Primary and Community Care Commissioning

In discussing the development of eligibility criteria for continuing care, it was suggested to this interviewee that this could be an interesting example of explicit rationing which had come about independently of the market, because of government and public concern about inequality of access and erosion of NHS provision. Although he agreed that it was independent of the market, he also felt that it had become an issue locally before the government decided to make it a central initiative.

> 'I suppose it arose initially from a recognition that acute services in our patch were more thinly spread than most others. We were struggling with waiting times, Patient's Charter standards and...as we tried to examine why that should be the case in our patch, one of the things we examined our mix of expenditure...And when we were able to compare the broad analysis of our expenditure by programme areas with what was going on in other places, it was clear that we were spending less on acute services and more on continuing care, significantly so...We weren't getting best value for money from our investment in caring for the elderly in that way, as compared with the value we might get by investing in acute care.'

The idea that long term care should be excluded from NHS provision predates the market by a long way, and was a response to financial pressure. Indeed, the explosion of nursing home provision for former long stay NHS patients funded through the benefits system was what prompted the Griffiths report on community care (Griffiths, 1988), which in turn influenced the 1990 Act. Equally, Cambridge and Huntingdon's concern with inequity between specialties, and variation between their own expenditure pattern and that of other districts, is not specifically related to the market.

However, although there have always been attempts to compare expenditure between care groups, the advent of contracts with more detailed costing and activity analysis than ever before could be assumed to have made this kind of analysis easier and more informative.

The director of finance believed the NHS reforms had made allocation decisions more explicit, but this was because of the new focus on acquiring services for a population rather than managing providers. The negotiation of contracts was less significant. He felt contracts in their current form did not deal well with any kind of rationing, except the total exclusion of some marginal procedures. When it came to less clear-cut issues such as reducing the number of D&Cs performed on women under 40, this had to be pursued through persuasion and discussion with clinicians, allowing room for the operation of clinical judgement at individual patient level.

However, notes of such discussions were appended to the contract as an agreed statement of intent or 'planning agenda', perhaps less strictly defined than the contract but part of it nonetheless. He felt 'gentleman's agreement' was a fair description of this process, although the chief executive was subsequently to describe it as much firmer and more binding. He did also acknowledge that the very process of writing things down had led to greater explicitness, not least because there needed to be some record of what had been agreed against which monitoring could take place.

Other factors identified by the director of finance included lack of resources generally, and changing public attitudes, which meant that taxpayers wanted to know clearly what they were getting for their money and would not let public officials 'get away with' continued decision-making behind closed doors. He also described a strong ethical imperative to ensure that the interests of minority groups such as those with learning disabilities or mental illness were protected, and felt discrimination was much less likely to happen in an explicit system of resource allocation. There is certainly evidence that such groups have suffered from historic inequities in service provision, and that even explicit attempts to give them funding priority have been overwhelmed by political and professional pressures to subvert official priorities (Ham, 1992). On the other hand, there is a risk that explicit rationing may disadvantage such groups still further - the suggestion, for example, that fetuses and severely handicapped newborns do not have full personhood and may therefore be accorded lower priority (e.g. Tooley, 1972) is a case in point. The director of finance responded that, if this view of personhood were made explicit, he believed the public would not accept it.

Chief Executive

The chief executive's initial view on whether the purchaser/provider split or the process of contracting had caused greater explicitness was 'I don't think it's any of that, it's just running out of money.'

This lack of resources was often brought to a head by new technology, and he cited IVF (which Cambridge and Huntingdon purchased within a strict protocol) as an example. He thought many authorities had excluded it because they could not afford to keep investing more in it, 'and then all kinds of spurious, after-the-event rationalisation is used to justify that decision, one of which is to say that it's nothing to do with health.' In Cambridge and Huntingdon's particular case, merger of two authorities had been a factor in the development of the current policy, as one of the authorities did purchase IVF and the other did not, resulting in untenable inequity within the new district. The chief executive suggested that personal or family experience of infertility among health authority staff could be a more influential factor in whether an authority purchased IVF than evidence of effectiveness.

Because he saw lack of resources as the primary factor in explicit rationing, he had no doubt that it would be happening regardless of structure. However, he went on to say that, although he did not think it was dependent on the NHS reforms, they did appear to have made the process more open, especially because purchasers could think about the needs of the population first and foremost. The needs of provider institutions were less important to districts than they had been, and this was 'a cultural change for the good'.

He argued that greater explicitness within the NHS was part of a wider societal change in attitude, which meant the public felt it had a right to know more about what was going on. This in turn was influenced by the 'decline of deference' generally, and specifically deference to the professions. He made the same point in relation to the Patient's Charter - in his experience, GPs blamed the Charter for making patients more demanding and ready to complain, but he saw it rather as an attempt by the government to respond to changes in consumer attitudes already taking place. He equated explicitness with being open with the public, and regarded decisions which were explicit between, say, doctors and managers, as still implicit by his definition.

Although he accepted the imposition of maximum waiting times had forced implicit rationing of some procedures to become more explicit, he thought this was only marginal. Many other elective procedures continued to be purchased despite the fact that greater health benefit would be gained by spending that money elsewhere, because there was no political support for tackling the issue.

Although he felt the reforms had not made much difference to explicit rationing, he believed the fact that contracts were public documents and providers were signing up to a certain amount of work for a certain amount of money, and to changes in the way the service was delivered, did have some influence. He described previous attempts to change services, whether by introducing new procedures or reducing existing ones, as a wish list, which no-one had 'signed in blood'.

However, he argued that contracting was speeding up an existing movement towards increased explicitness which was being driven by other reasons. 'Things were moving in a direction really well before the NHS reforms - introducing management into the NHS, introducing the whole notion of setting priorities. I mean, in those days it wasn't called rationing because it was about priorities over new money, that was the thing - it wasn't about changing what we've got. And I think that's just got more and more intense.'

The role of extra contractual referrals was discussed in some detail. Because they are often highly expensive and often deal with controversial or rare treatments not catered for in mainstream contracts, they have been at the cutting edge of explicit rationing. Child B was a case in point. The chief executive was particularly concerned at government plans (subsequently deferred) to replace the system whereby each ECR had to be approved by the relevant health authority with automatic acceptance and payment. He felt the removal of this rationing mechanism was a recipe for disastrous loss of financial control.

Senior Registrar in Public Health

The views of the senior registrar are included because although, as he pointed out, 'I'm not one of the executives', he had been closely involved with rationing work at both local and national level.

Although he felt lack of new money had sharpened thinking about where less effective services could be curtailed in order to develop new services, he suggested the issues tackled so far, such as homeopathy, had not released significant savings. To make major savings would require major shifts in purchasing, beginning to eat into services that are effective to some extent, and he thought the resulting public and political outcry would be so great that it would not be feasible.

He was doubtful of the extent to which effectiveness evidence really influenced decisions, especially when it pointed to more rather than less investment. He had produced a detailed report on the evidence for greater investment in back pain - 'yet in terms of "does that make a health authority change the way it purchases?" the answer's "no" - certainly not here....That's because back pain is not as sexy as cancer and heart disease...The evidence is there and everyone round the country's talking about it, yet it's not making a lot of difference...Evidence does help, but there are a lot of political imperatives.'

He felt there was a move towards explicitness, but was sceptical that it was being put into practice to any real extent. The contracts, in his view, were a 'mechanism of expression' for the 'real decisions' which had been taking place in behind the scenes discussions between a few key people in the health authority and clinicians. Persuasion, influence, discussion - the broader activities of commissioning rather than contracting - were much more important in setting priorities than contracts themselves.

The fact that ultimately decisions were made by the chief executive and director of public health might, he suggested, be entirely right - someone has to take final responsibility for weighing up all the evidence and pressures on the service and making a judgement, and explicitness was only possible to a limited extent. He was concerned, however, that the 'rhetoric of explicitness' was not matched by any real transparency of process, and that this was essential for public accountability. Whilst each actual decision itself might be made behind closed doors, the authority ought to be able to show the public clear, written procedures for how decisions generally were made, who was involved and what factors were taken into consideration. He discussed a model of informed consent, with the authority as organisational doctor and the public as organisational patient - the patient might not want or be able to participate in every step of the doctor's decision-making about their case, but they would want to be reassured that he or she was acting according to certain principles and procedures, and had considered all relevant factors.

A number of other factors were identified, including media pressure and changing public attitudes in favour of more explicitness. Added to this was pressure from academic sources associated with the NHS and health care. Health authority staff increasingly felt a sense of obligation to be more explicit, and this was at least partly driven by the purchaser provider split.

Director of Public Health

In this interview, lack of resources was again cited as a primary factor. 'I think it is the *raison d'être* of health authorities to set priorities, to ration, to allocate resources - use whatever term you like...basically it's about the fact that there's a cake of a given size.' The resource management initiative, which predated the market, was identified as another major factor behind increased explicitness, because it had made the whole of the NHS, clinical staff included, more aware of cost. The other important influence, he felt, was the clinical audit movement, 'which has turned into the clinical effectiveness movement. Again, we don't need the purchaser provider split to get doctors and nurses to understand that it's not just their opinion that matters, but is there any scientific backing for what interventions they use?'

He did not believe the NHS was operating a real market in any sense, and described it as 'the worst of both worlds'. He felt the purchaser provider split had introduced perverse incentives; because providers were no longer 'part of the NHS family, working together within the framework of a finite cake', but at the same time not subject to genuine competitive pressure, it was in their interests to 'hype up technology' and generate unmanageable demand. Their lack of willingness to accept that there was only a finite sum of money was one of the reasons why difficult choices were having to be made - only a pure market, or reintegration under district control, would control the situation.

On the other hand, he argued that personally he did not think it right that purchasers should try to suppress the development of new technology - 'we should let providers do what they think is right medically. Then we should ration, and ration very hard. But no government will allow that.' His preferred situation would be one where national rationing would determine what was to be available as a core state funded service, and individuals would be free to purchase whatever extra services they wished. To the extent that the market had exposed the conflict between what the state could afford and what was medically possible more clearly than ever before, it was perhaps leading to greater explicitness, but he believed this would still be necessary without a market, as the mismatch between the possible and the affordable would remain. Thus the market affected the degree and speed of explicitness, rather than causing it.

Director and Deputy Director of Acute Commissioning

The director's view was that 'the mismatch between money and demand' was the main driving force behind greater explicitness in rationing. His deputy suggested that the contracting process had resulted in much better information being available about what their residents were and were not getting, and that this had itself prompted some explicit rationing decisions. The director's response was that availability of information in itself was not the reason for explicit rationing, and in any case, even if the market had never existed, other things would have happened to the NHS, particularly resource management and the Patient's Charter. He did not think implicit rationing was possible any more because the gap between resources and demand had become so much greater.

He then discussed whether the internal market had increased the pressure to be explicit, and like the director of public health pointed out that 'in the old days before the internal market, directly managed hospitals would be under the same pressure and be looking at ways to get round it', whereas now this constraint had gone. He also mentioned that, now purchasers were responsible for a resident population, the loophole of referring outside the district to a provider who would provide the relevant service had been closed off, which had made things more explicit. 'So I suppose I'm sort of arguing the reverse of what I originally said', he concluded. He also expressed concern that abolishing ECR controls would lead to GPs being able to by-pass local exclusions by referring elsewhere.

The deputy director reiterated the importance of information availability, although the director said lack of knowledge was still an obstacle to greater explicitness. 'We do have a much better idea of what's going on, but if you were to turn round and say, "well, how do we ration urology?"...then where you start is just a nightmare.' He felt detailed understanding of procedures was still relatively marginal, and all the easy or obvious ones had already been picked off.

The Patient's Charter was felt by the director to have made existing rationing more explicit, rather than actually increasing the total amount. On the other hand, one trust had made its own decision to start turning back referrals for a particular procedure, because it was compromising its ability to meet Patient's Charter targets. The case of experimental laser treatment for snoring (now excluded) had also only been discovered because of a review of ENT waiting times. However, a review of the ENT waiting list might not in itself have shown up such cases, were it not for the detailed requirements of the contract minimum data set now recorded for every episode.

Plastic surgery provided an interesting example of weaknesses in monitoring agreed rationing. A 'loose protocol' had been drawn up some years ago, but GPs were continuing to refer people for limited procedures and consultants did not feel able to refuse the referrals. It was only as a result of lengthening waiting times and rising emergency activity levels in plastic surgery in 1995 that this situation came to light, and the new, more explicit protocol was negotiated, including specific criteria for making exceptions.

The director identified an ethical obligation on health authorities to ration explicitly. He felt the gap between resources and demand was something 'we just can't run away from'. 'This is something most health authorities, and certainly central government, just want to duck and just hope it'll go away. I don't think it's fair on anybody, really. I think it's much more honest to say, "yes, there is a problem, we do have to ration, but we'll do it as humanely as possible".' He was concerned at attempts to disguise rationing as something else, such as patient empowerment, giving the example of interactive videos informing people about the side effects of prostate surgery as a way of dissuading them from having treatment.

District Three

Director of Finance

Although the director of finance argued that lack of money was the main driving force behind explicit rationing, the first example he discussed (grommets) was in his view driven more by the discovery of service variations - one local consultant was found to perform more than twice the national average. The savings identified for reducing this were 'not huge in financial terms'.

However, he identified an urgent need to develop explicit eligibility criteria for renal dialysis, where the financial problem was much greater. The kind of cosmetic surgery and low priority procedures rationed already were the easy ones, with little financial impact, but rationing mainstream services was going to be much harder.

Many of the decisions taken already had been stimulated by the Patient's Charter. Others had first come to the authority's attention as ECRs - these were often the developmental, alternative treatments and new drugs which were not covered under normal contracts. Like the director of public health in Avon, he identified lack of provision for mental illness patients as another factor in the burgeoning cost of ECRs, and suggested that the internal market had given the private sector a chance to develop expensive services which the NHS had not previously had or even thought it needed. This is further evidence of the market unleashing previously suppressed demand.

The market had also enabled providers to pass decisions back to the purchasers which providers had formerly had to make implicitly. Having said this, he added that consultants' increasing reluctance to take responsibility for rationing was probably a result of worsening financial pressures, and would have happened anyway regardless of structure. He concluded that the structure had probably speeded up the process, but it was impossible to be certain because so many other circumstances had changed over the last five years, particularly the development of new technology.

The effect of contracting more specifically was unclear. He felt rationing decisions and the contracting process were not always well integrated - sometimes things were picked up as part of negotiations, but just as often the discussions ran in parallel. One area where contracts had made a difference was in increasing the amount of detailed information available, particularly because providers needed accurate data to support invoices. He contrasted the current situation with the old-style approach to reducing waiting lists, which was to allocate more money, rather than analyse what procedures people were waiting for and whether they were necessary.

Enforcing rationing decisions through contracts was another matter - even if monitoring showed there was a lack of compliance, there was a limited amount one could then do to change the situation. Clinicians did not seem to find the fact that they were in a market relationship with the health authority a constraint on what they wanted to do, and if anything it made them feel free to 'push it to the hilt'. Peer pressure from clinical colleagues concerned about the inequitable use of the resources available might after all be a more effective route. The director of finance did feel, however, that contracts helped make clinicians more aware of each other's use of resources.

A further difficulty in using contracts as a way of enforcing rationing decisions was the presence of GP fundholders, who were not bound to follow the same decisions. Consultants were more reluctant to abide by exclusions in the contract from the district purchaser if they were continuing to provide that service to other patients.

The director of finance took a somewhat different view of public involvement to his colleagues. Whilst he felt it was entirely right and proper that decisions should be more explicit at purchaser level, there were great dangers in making them public. In his view, public scrutiny operated as a pressure towards implicitness.

In direct contrast with the views of Avon's chief executive, he did not feel that trusts, particularly generic trusts, were trying to do any of their own explicit rationing - the impetus came from purchasing. He believed having a generic trust made it easier to move money from acute services into community services, because it did not involve depriving the trust as a whole of some of its income. He did acknowledge, however, that there had been examples of generic trusts moving resources in the opposite direction to support acute care, and that this had been one factor in the change in national policy after the second wave towards generic trust applications.

A move away from trust status under a new government would not, he felt, make a great difference to current relationships with providers, which were already a process of continuous dialogue, but it probably would mean that 'the purist contracting way we do things will be watered down to a very large extent.'

Chief Executive

The chief executive felt evidence on effectiveness was becoming an increasingly important influence on rationing. 'Our discourse on priority setting, resource allocation and rationing has changed subtly but very firmly away from notions of rationing as denial to some people of certain interventions towards a much greater emphasis on clinical effectiveness, and on cost-effectiveness of treatments.' It was primarily for this reason that the authority did not operate any blanket exclusions except in IVF, because a demonstrable clinical need should always override the decision. He also agreed with his director of finance that fundholding made it impossible to hold the line on absolute exclusions.

Personal views and social judgements also played a strong part. They had been influential in discussions about whether IVF should be provided locally, and the chief executive said he thought previous experience in mental health provision amongst individual staff members had resulted in a more liberal purchasing policy towards various mental health interventions than in other districts. He also discussed the social judgements underlying policies on funding abortions.

He felt contracts had made rationing more explicit, although perhaps not as much as some people had thought it would. Many contracts were still relatively broad brush and left a lot to the individual clinician, but gradually 'as we have moved to disaggregate contracts between specialties...there is a move more and more to specify more closely who we will treat and on what basis.' Technological advances in some specialties, such as renal dialysis, were also forcing the pace because of the resource pressures they created.

The reason why many contracts remained broad brush was not that it was impossible to specify services in more detail. 'First of all, it wasn't felt perhaps necessary at the beginning. Secondly, there was opposition from consultants to specifying more clearly what they should do - the clinical freedom argument. It took some time for purchasers and providers to get a grip on that. Thirdly, the money has become tighter over time. At the beginning it may not have been as necessary as it now is to specify more clearly. Fourthly, I think, there were arguments that rationing...that there was a Holy Grail somewhere, if only we could find it, and we would be able to determine what should be treated and what shouldn't. We've realised that that's actually very difficult - that in many cases you can't say "never". And therefore you begin looking at what's possible....As we've gradually become more knowledgeable, as purchasing has developed, as we have recognised that effectiveness is the key, we've had to begin to think about developing protocols and guidelines for using the services.'

He concluded that lack of money and greater availability of research evidence were crucial driving forces and that contracting was 'simply the mechanism' for expressing the results, not in itself the cause of explicit rationing.

He added that the purchaser provider split, which had disrupted the historic funding to providers, and the competitive nature of the service were further factors. He illustrated this by describing how one hospital had begun to develop a particular specialist service against the policy of the district, which wanted to concentrate this specialism at another provider. The district was responding by developing a detailed protocol to specify safe levels of service and under what circumstances a patient should be transferred to the specialist provider.

The role of competition was limited, however; for the most part the system did not really operate as a market, so things would not change significantly under a Labour government. It was precisely the lack of a real market which made protocols and guidelines so important as an alternative way of controlling provider activity - 'I think the future does lie in terms of effectiveness issues, rather than in terms of buying where we get the best quality and price within the market.' The Labour idea of healthcare agreements was simply contracts by another name, and they would continue to be a way of enshrining agreed protocols. It was unclear how abolishing ECRs would work, however.

The chief executive anticipated that rationing would become more explicit, but gradually, rather than dramatically; although there were pressures to remain implicit, he did not think these would win the day. He felt one of the pressures in favour of implicitness came from the public, who would prefer not to have to face these decisions and would rather hold on to the belief that doctors can be trusted to make the best judgements on their behalf. However, the Patient's Charter, rising consumer expectations and debate over cases such as Child B meant that there was now increased awareness of the issue and a point of no return had probably been reached. Once people became aware doctors were making rationing decisions, they wanted these to be open and would challenge a decision not to provide active treatment. Some doctors might believe the point of no return had not yet been reached and that they could still maintain a purely implicit approach, but equally there were other doctors who would prefer not to have this responsibility any more.

Acting Director of Public Health

The acting director of public health was due to leave shortly after the interview to become director in another district. The new director had not yet taken up post.

He began by stating that much of the pressure to ration low priority treatments more explicitly had come from GPs, because many of them felt uneasy continuing to refer patients for such treatments, but unable to refuse. They felt 'it would strengthen our arm considerably if there is a policy.' GP fundholders were also involved in seeking an explicit policy, partly to give their own policies a wider framework of support, and partly to avoid losing patients to other practices if they wanted a treatment that the fundholding practice did not feel should be provided. This offered a somewhat different perspective to the chief executive and director of finance view that exclusions were made more difficult to operate by the presence of GP fundholders. The Patient's Charter had had some influence, 'but it felt more like an environmental factor - that that was part of what made GPs feel the way that they were feeling.'

The other main strand in work on low priority treatments was clinical effectiveness, which the authority wanted to make a driving force as a matter of principle. However, he said, 'I'm very worried that when we stand back from it and look at that, an awful lot of it has a very discriminatory flavour about it', thinking particularly of procedures such as tattoo removal, gender reassignment and reversal of sterilisation.

IVF had come up as an issue, as in Cambridge and Huntingdon, because of the merger of health authorities with different policies. Again, it had been approached from a rigorous effectiveness and cost-effectiveness approach; although there had been discussions about whether infertility was a social or medical condition, he did not feel this had influenced the final outcome. Even though IVF was excluded, other therapies for unexplained fertility, such as drugs, were not excluded, and the decision as to where to draw the line was based solely on 'cost per maternity'. Nonetheless, he felt there was an important point here for the future of rationing, 'as medical technology makes it possible to do things which are "unnatural"', and to alter conditions which 'it's much harder to define as diseases'.

Although he had been one of the main proponents of establishing principles (including effectiveness) on which rationing decisions should be based, he suggested that in practice decision-making was still largely dominated by 'subjective opinion or prejudice' and by political pressures such as the need to maintain emergency services.

His view on the processes of contracting was 'I don't think they've led to very much more explicitness. They've led to much more explicitness about levels of activity, and <u>some</u> sort of things around cost, but not really around the balance between choices...The naive concept of the quasi-market might have been that the purchasers would say what they want, and providers would think of ways of meeting those. And of course it doesn't happen that way at all.'

However, the ethos of competition had led to 'providers, especially clinicians, being more robust in their justifications of what they want to do.' This had had an impact on making choices, because it had thrust them back at purchasers and demanded a more explicit response. As other interviewees had suggested, 'the purchaser provider split has decoupled health authorities from the old methods of control that they used to use, which were basically control of supply', although occasionally providers did retain some sense of corporate responsibility. Competition had also made providers fearful that if they did not develop new services they would lose out to other providers, so there had been 'multiple developments of specialist services...Then the pressure is on to do something we might not have chosen to do.'

He did not believe contracts disaggregated by specialty made any difference - he thought in any case most authorities had specialty-specific activity, but 'nobody has contracts which are <u>rigid</u> by specialty.' Variations between contracts, or between specialty activity levels would still be handled by allowing underperformance in one to compensate for overperformance in another. However, he agreed that the difficulties in renal services were now such that the trust would no longer work within a general financial cap, and was seeking more explicit patient selection criteria from purchasers - in effect creating a specialty-based contract. This may be a case of the need to ration explicitly causing a contract to become more explicit, rather than the other way round.

In general, however, he felt that the need for flexibility, combined with prohibitive transaction costs, meant that most contracts would remain broader than this. This would tend to support the Williamson and Bartlett thesis that bounded rationality and uncertainty require something more adaptable than a rigid contract specification. He also suggested that where very specific contracting had been tried, the sum of the individual contracts negotiated always ended up being greater than the total purchasing budget available, so it was a dangerous route.

For him, the ultimate reason for explicit rationing was 'cost control versus all those pressures that we know about - elderly population, medical technology and all those sorts of things.' However, whilst cost control required choices to be made, these did not always have to be explicit but could be implicit. 'Some of the choices that are involved are so difficult that it is actually really quite hard to present them to people in ways that they can understand' - not least because of the complexity of effectiveness evidence and poor understanding of statistical risk. 'That's actually extremely difficult for people, individually or collectively, to make those choices - whether those people are doctors or not, actually. And there's some kind of negotiation processes that go on, where the implicitness is around those people who have that bit more knowledge, experience, training, and are more able to weigh those things up, and take that awfully heavy burden of doing that for other people.' He identified a different kind of transaction cost - that of trying to explain the full benefits and risks of a particular procedure to a patient, when it takes years of training for even the doctor to understand it.

Another advantage to implicitness was that 'we do face a lot of rabid values out there', amongst GPs as well as the public; whilst it might be paternalist to say it, it might be the only way to protect people with, say, learning disabilities from more extreme forms of rationing than they already experience. 'That's where I really become a benevolent dictator.'

Directors of Commissioning

District Three has two patch-based directors of commissioning, one of whom (DC2) had only taken up her post the week in which the interviews were carried out. Most of the comments below are therefore from the existing post-holder (DC1).

DC1 was sceptical about the ability of contracts to specify and control activity to the last detail, and that this might be in any case undesirable. She favoured using more professional self-regulation through the Royal Colleges and protocols to decide what care would and would not be provided, as opposed to the current approach of Royal Colleges of specifying ideal wish lists. At the moment, she said, the simplistic view of contracts was 'we specify; they do', whereas the reality was 'we specify, they think about it, and then carry on doing what they normally do.' Difficulties of monitoring and enforcing detailed contracts and their subversion by clinicians was a recurring theme during the interview. The other danger with a very detailed specification was that 'clinicians use it as a powerful lever to abrogate any responsibility that they have for making rationing decisions, and use it as a way of exercising leverage on us, for us to give them more money. So the specification can actually be used very negatively against purchasers.'

This did not necessarily mean in her view that the contracting process was bound to be inefficient, but that it could only be efficient if contracts were broad and flexible, and were backed up by changes in clinical practice led by the professions. 'In terms of the documents being at least manageable, then you have to assume that common sense and custom and practice will prevail. But actually putting that down in the document is particularly challenging.'

However, she too described how the problems with developments in a specialist service at one hospital which the authority did not support were resulting in precisely that kind of detailed and prescriptive contract being drawn up, if only 'to protect us from a medico-legal point of view'.

A further concern was that monitoring detailed contracts relied on professional staff having in place quality assurance mechanisms that sometimes turned out not to exist - 'there is the reluctance to use specifications because it does uncover weaknesses...in the current system.'

Both interviewees felt equity of access was a major motivating factor in making explicit rationing decisions, along with a strong sense of the unfairness of low priority treatments being continued at the expense of other more important interventions. DC1 acknowledged that, in fact, stopping low priority treatments rarely released much money to reinvest in the more important things - 'it's more the principle than the number of people'.

The Patient's Charter was felt to have operated in a number of ways - it had given purchasers 'added legitimacy' in challenging clinical practice, it had improved provider efficiency, and had removed the potential for consultants deliberately to maintain a long waiting list to create demand for private practice. However, DC1 noted that not providing certain treatments at all could also channel patients towards private care. The Patient's Charter could also exercise a negative influence; clinically necessary breast reductions (which took three to four hours to complete) might have to wait longer than they would otherwise have done, because consultants were reluctant to fill up their theatre sessions with one case and thereby slow down their throughput and lengthen their overall waiting times.

Both directors discussed the potential for sharing research evidence with patients to reduce the need for rationing, because some patients might then choose not to have the operation. On the other hand, they were concerned about the possible presentation of information in such a way as to persuade patients to choose the cheapest rather than the best option. An additional problem was that purchasers might simply not know what alternative treatments were available, and could not trust providers with vested interests to let them know. This was an area where GP fundholders could prove useful allies. They were undecided whether offering someone an alternative treatment that might be more beneficial but which they did not want counted as rationing - was need or demand the criterion against which rationing should be judged?

On the subject of personal ethical difficulties, both directors agreed explicit rationing was hard, especially at an individual patient level (as in ECRs), and that telling the public or providers could make them feel very vulnerable and unsure if they were doing the right thing. However, being open about it was also a defence and a necessary responsibility. DC2 said 'I think they're shared decisions, aren't they? Which makes it more bearable', and DC1 said the worst kind of decision was having to say 'no' to an ECR 'if I'm the last person left in the building'. This echoes the comment from Avon's director of public health that no-one wanted to be left 'holding the smoking gun', although the uncomfortable nature of the decisions was not actually preventing the directors of commissioning from taking them. They certainly did not believe it was right or even possible to get the public to make the decisions instead.

There was an ethical imperative for purchasers to mount an explicit challenge to some aspects of clinical practice - whilst lack of money might account for 'two thirds' of the reasons behind explicit rationing, DC1 felt it was an important way to stop patients having needless operations or being subjected to heroic intervention in the last days of life, if based on sound effectiveness evidence. It might also reduce clinical experimentation with resource-intensive procedures when it might be quicker and more efficient for patients as a whole 'if you just got the knife out and got on with it'.

The process of contracting might help by making available information that would otherwise not have been there - 'without the purchaser side of it would we ever have sat down and thought about what services were like and described them? I don't think that we would have.' 'I think that we may have done it in some areas, but not quite as comprehensively as I think we've been forced to.' It had accelerated a movement that had begun before the market to question what value would be obtained from appointing an additional member of staff - assessing investment in terms of service benefits rather than whole time equivalents.

Contracts were a means to start dialogue and also to record its results, but could not replace the delicate process of change management with clinicians; contracts had to be used 'as an agreement, rather than as a blunt instrument that we try and catch each other out with', otherwise they might be counter-productive. It was therefore unlikely that a change of government would mean any great change in contracts. What it might mean was a reduction in the transaction costs of 'invoices and lots of accountants and processing'.

Discussion

There are clearly identifiable recurring themes in interviewees' explanations of the driving forces behind greater explicitness in rationing. The most important of these is lack of resources, although so far the kind of rationing decisions being made are only scratching the surface of this problem. Nonetheless, the general climate of scarcity is extremely influential.

Lack of money on its own is not a sufficient explanation, however, as one possibility would be simply to increase the amount of explicit rationing taking place. Several other themes emerged during interviews - the possible interactions between these factors are discussed more fully in Chapter Nine. These themes included:

- The Patient's Charter
- Other government initiatives, such as eligibility criteria for continuing care
- Changing public attitudes
- Contracts (to a limited extent), especially the greater availability of information and the need to specify services in more detail than ever before
- The purchaser provider split more generally - the focus on the population, not on providers, has freed purchasers to make tougher decisions, and also made providers less likely to comply with implicit rationing
- Perceived inequity (often geographical, made more stark by HA mergers) and the need to protect disadvantaged groups
- An ethical obligation to be explicit, both because it is seen as a way of guaranteeing fairness and because it is a democratic responsibility
- Improved information for reasons other than contracting (e.g. resource management initiative, audit)
- The clinical effectiveness movement and improved availability of research findings.
- Managerialism and the culture of priority setting
- ECRs
- Academic fashion
- Professional sub-specialisation
- Lack of beds (particularly in mental health)
- Technological advance

What are the obstacles to explicit rationing, or pressures in favour of implicitness?

There was a remarkable degree of consistency between health authorities and between disciplines in analysing the obstacles to explicit rationing; their views are therefore discussed mainly in aggregate. Some of the obstacles have already been mentioned in the previous section, as interviewees explored the often contradictory effects of the factors they identified.

There was an absolutely clear agreement that **political reluctance** to be seen to be rationing explicitly was the most obvious difficulty. Views about whether it would be desirable to have some form of national rationing framework differed, but the unanimous view was that in reality it would never happen, because it would be politically unacceptable. This applied to all political parties.

The chief executive of Cambridge and Huntingdon commented on the planned abolition of the approval system for elective ECRs in this context. The ostensible reason for making ECRs automatic was to reduce transaction costs and bureaucracy, but the chief executive felt the reasons were in fact more overtly political - to 'prove that the internal market is responsive to clinical need, and they don't want managers to keep saying no to doctors.'

The director of acute commissioning in Cambridge and Huntingdon saw the ECR question somewhat differently: he agreed that the government did intend it to be 'a sop to GPs' to guarantee freedom of referral, and a sop to cutting bureaucracy and transaction costs, but saw the potential loss of financial control as 'cock-up rather than conspiracy', stemming from a lack of understanding of real practicalities.

The government requirement that no treatments should be subject to a total exclusion unless completely ineffective was also evidence of government unwillingness to allow the ideal of a comprehensive NHS to be openly challenged, even if rationing was accepted tacitly. Several interviewees commented on the government's decision to make beta interferon available on the NHS, despite evidence of low effectiveness which would have given them an opportunity to reject it. One respondent contrasted this with the government's willingness to introduce limited list prescribing before the market was established.

Interviewees felt **effectiveness evidence** did have the potential to encourage and to justify more explicit rationing decisions; on the other hand, taking effectiveness seriously also meant 'never say never'. There would always be individuals for whom a treatment of generally low effectiveness was worth trying, and this required flexibility for clinicians to exercise judgement. Allowing exceptions to be made meant allowing implicit rationing to maintain a significant role, albeit 'against a basis of clarity which we have the guts to provide', as the chief executive of Avon said.

Despite increased availability of effectiveness information, several interviewees pointed out that it still barely scratched the surface of the majority of healthcare provision, and that historical funding patterns continued to form the basis for most allocation decisions. Sometimes even where evidence was available it was ignored, because cutting a service would be politically and publicly unacceptable. Sometimes the evidence suggested that more of a particular service should be purchased, rather than offering a rationale for limiting it, but again, such evidence might be ignored because it would prove too difficult to cut other things to find the money. The urgent need to prop up emergency services in the last contracting round had also taken priority over other longer term projects.

Interviewees were very conscious that rationing was a difficult but inevitable responsibility, whether implicit or explicit. Personal and collective **ethical concerns** were already present, therefore, and did not generally act as an obstacle to explicitness; although psychological and emotional reluctance to be explicit was mentioned by several interviewees, it tended to be outweighed by a perceived ethical imperative precisely to be explicit. However, it was acknowledged that so far only procedures where there was a reasonable degree of consensus had been tackled, and where the effects of rationing were not generally life-threatening, so purchasers still felt on reasonably safe ethical ground. IVF was probably the most controversial of the commonly rationed procedures, and this was reflected in more ethical concerns as to whether the right decision had been made. There was general acknowledgement, too, that areas such as renal dialysis and oncology presented a huge ethical challenge, and that this was definitely delaying the introduction of explicit rationing. However, most seemed to feel that this delay would not be indefinite and that in the end the ethical problems would have to be overcome.

Two interviewees took opposing views about the benefits of explicitness for vulnerable groups such as those with learning disabilities. One suggested that it was the best way to protect their interests against those of the more glamorous acute services and ensure they were not surreptitiously denied appropriate services. This contrasted with the view that implicit, paternalist decision-making might be the only way to protect people with learning disabilities from more extreme forms of rationing than they already experience. However, this was more a theoretical argument in defence of implicitness rather than a belief that it would in practice act as an obstacle to explicitness.

The **views of the public** were seen in similar terms to ethical considerations - in some ways an obstacle to greater explicitness ('the pressure of the bureaucracy for a quiet life'), and one of the things that made explicitness most uncomfortable - but at the same time constituting a pressure towards greater explicitness. Many interviewees commented that they thought public attitudes had changed and that people would no longer tolerate secrecy, even if they did not like what they then found out.

Although contracting had increased the amount of detailed **information available**, interviewees reported that there were still large gaps in purchasers' knowledge of what was being provided and how worthwhile much of it was. Part of the problem was that to specify every eventuality would be a) very expensive and b) probably not possible, with the result that for many areas broad, flexible contracts would continue to be needed. (See further discussion of transaction costs below).

Furthermore, there was strong evidence of **difficulties in monitoring and enforcing compliance** with contracts - 'we specify, they think about it, and then carry on doing what they normally do' - and concern that providers would not always be open about what was being provided. There were several instances where lack of compliance had been discovered almost by chance. Once discovered, it frequently led to the contract being tightened up, with even more explicit specification of exclusions. However, several interviewees felt that no matter how specific contracts became, they could still not replace the need for co-operation, persuasion and influence needed to achieve changes in clinical practice. These were sometimes integrated with the contract negotiating process, but more often ran in parallel with it. One interviewee suggested very explicit contracts had actually made it more difficult to achieve such change, because they alienated clinicians who might otherwise have been more willing to compromise. Although most interviewees felt providers used the purchaser provider split as a way of forcing rationing back onto purchasers, there was limited evidence from Avon that clinicians felt so constrained by the result that they decided to re-establish more implicit mechanisms.

Most interviewees felt the market did not work as a market because of local monopolies (amongst other factors), which enabled providers effectively to ignore contracts if they so chose. Despite the feeling that purchasers had been liberated from worries about provider interests, some continued to express a sense of shared responsibility for providers' fixed infrastructure costs, suggesting that monopoly provision also restricted their ability to make explicit rationing decisions, let alone enforce them.

Two districts were actively pursuing the idea of longer term contracts, and in practice all the districts regarded contracts as expressions of longer term commitment to providers, albeit with a continuing need for annual re-evaluation. There was a general view that they were already seeking to negotiate the kind of contracts described by Labour as healthcare agreements.

The existence of **GP fundholding** was felt by many to be an obstacle for health authority purchasers in enforcing explicit rationing decisions, because it was hard to stick to the decision if other patients locally were exempt. Even in cases where GP fundholders were not officially able to buy the procedure which was excluded, they were able sometimes to circumvent this unofficially.

Other structural problems included the **existence of generic trusts** and the sometimes linked problem of **inadequate staff numbers and expertise at district level**. Generic trusts were felt to reinforce the difficulty of getting adequate information out of providers, and to allow providers to retain greater control over continued implicit rationing between community and acute services, particularly where a previous district management team had transferred to the trust virtually intact. The knock-on effect was to leave the health authority drained of skill, local knowledge and sheer numbers of people.

Views on the extent of **transaction costs** were somewhat mixed. The chief executive of Cambridge and Huntingdon said, 'I just do not believe there has been a massive increase in transaction costs between health authorities and trusts. Where there has been a massive increase in transaction costs has been all the transaction costs associated with GP fundholding...they occur in the GP practice, but they also occur in the trusts, because the trusts have to manage a whole series of tiny contracts.' His views were based partly on local experience of health authority mergers, which had led to reductions in staffing costs. However, this was comparing the costs of two health authorities already set up as purchasers as against one authority, rather than comparing the cost of a purchasing agency with an old-style district before the market.

He did acknowledge that the purchaser provider split and the contracting process did entail some extra costs but added 'whether the transaction costs that that generates are outweighed by the benefits is an interesting question. I genuinely don't know the answer to that.' He was emphatic that transaction costs could not be judged on their own, without some assessment of whether the system was achieving better value for money as a result, so did not see high transaction costs in themselves as prohibitive. This was particularly the case with ECRs, where he thought investing in a few staff to scrutinise requests for approval rigorously could save the authority much more than it cost.

He did identify managerial pressure to get a quick decision as a pressure in favour of implicitness, because 'explicitness means openness of discussion, it means consultation periods, it means public meetings.' This could be interpreted as unaffordable transaction costs acting as an obstacle to explicitness.

The chief executive of Avon echoed the view that GP fundholding was the main culprit in increasing transaction costs and that its abolition would lead to a substantial reduction in staffing in trusts and general practice. In health authorities, by contrast, what purchasing staff 'spend most of their time doing in most of the year is actually working with the trusts to ensure that the services are what we want them to be - that they're of a decent quality, and that we are putting the time in together planning the new things...the old service planning agenda, if you like. The fact that then, for three to four months of the year, they nearly kill themselves working about 80% overtime for a period of time, does mean that you would not halve the number of contracting staff [if contracting were abolished]...Of that team of about twenty people, if you didn't have as precise a contracting arrangement, you'd probably take out a couple, quite honestly.' A few information posts might also go, and simplifying some of the information requirements 'would take a degree of the silliness and the messiness and the playing arithmetic games out, and that would be entirely sensible.' Essentially, however, she supported the view that transaction costs were not that significant and gave good value for money.

Her director of finance pointed out that most contracts were financially a simple block arrangement with each trust, specifically to avoid 'this business of charging all over the place for things' - cost and volume arrangements accounted for less than 1% of total activity - even though within the block of money activity could be specified in some detail.

Although the director of public health in Avon agreed that transaction costs were a good reason for sticking to block contracts in financial terms, he was less sanguine about whether it was possible or even desirable to specify activity in great detail nonetheless. 'I can envisage it going on exactly the same as if we were health authorities and the senior people in the health authority were setting service level agreements with their operational people *[i.e. pre or post market]*. And is that because contracts can just never get that explicit? I mean, you know, listing absolutely everything that's done, because it would just be too much work and a waste of time?'

Other interviewees, particularly in District Three, expressed the view more strongly that the purchaser provider split and the process of writing contracts had indeed entailed extra costs, particularly in financial administration, and that the value obtained was sometimes doubtful. The acting director of public health believed further increases in transaction costs, coupled with the need to retain flexibility, would prevent further development of explicit contracts on a large scale, even though it would happen in a few areas where particular resource pressures were forcing the pace.

Discussion

At one level, some interviewees argued that transaction costs were not an obstacle to explicit rationing because although they undoubtedly existed they were not in fact very high, at least in health authorities. Yet at another level there was recognition that they could be a lot higher if contracts were specified in greater detail across the board, and that a deliberate decision had been made to avoid increasing transaction costs in this way, because the gains would not justify the expense.

It would appear that there was a trade-off at work here, which was not necessarily fully acknowledged. The perception that the transaction costs of running a market and writing contracts were not very high in fact reflected a view that an appropriate balance had been struck between the benefits and costs of greater explicitness. There is a point at which the costs of obtaining information, negotiating and monitoring does become an obstacle to making contracts more specific - interviewees took differing views as to whether this point had already been crossed or had been skilfully avoided.

Whether this obstacle would prove to be insurmountable was another matter. Most interviewees believed explicit rationing would continue to spread, albeit in selected areas rather than across the board. Again, this probably demonstrates the view that in these areas there were still gains to be had from greater explicitness that would outweigh the transaction costs, whereas in other areas the costs could not be justified.

Most did not believe there was a significantly cheaper way of achieving the same result, because they could not envisage a return to a hierarchical and integrated organisation. They did not believe the Labour Party's proposals would mean a return to hierarchy; although they might make marginal reductions in transaction costs, the essence of the contracting system would have to remain. One interviewee pointed out that even before the purchaser provider split, the NHS had been moving towards more sophisticated cost and management accountancy and this would have incurred its own transaction costs.

Chapter Eight

Back from the brink? The effects of a pre-election budget

Introduction

This chapter presents the results of the second and third round of interviews in the three case study health authorities. (See also Chapter Five on methods). The second round of interviews was carried out in January and February 1997. It was planned to interview the directors of public health and the directors of contracting/commissioning, as they would be closest to the detail of contract negotiations. Owing to timetabling difficulties, in District Three it was possible to interview only the director of commissioning.

At the first round of interviews, it was striking that the directors of finance expressed a feeling of distance from the dilemmas of rationing, and did not feel they were the best people to interview. Their sense of remoteness from the debate about rationing is in itself an important finding, particularly in terms of the extent to which greater explicitness is finance-driven, or otherwise. At any rate, it certainly does not seem to be finance director-driven.

The third round of interviews took place in June and July 1997; the directors of public health and contracting were re-interviewed, and the chief executive in two of the three districts. The third was unable to keep the appointment.

At the first interviews in the summer of 1996, there was a general consensus that the financial position for 1997/98 was going to be extremely tight, and that further developments in explicit rationing would be inevitable in order to deal with this. A prophetic note was sounded by one chief executive, however: 'Something that I think would be a backward step would be for a government in the dying days before an election to shove a whole lot of money into the NHS, because it would just literally be buying time, because next year we would be at the point where in fact we're at this year. So no amount of money will do anything other than act as a time delay.'

There is always disagreement within the NHS and between the NHS and the government about exactly what percentage of real growth a particular budget settlement represents, after allowing for inflation, efficiency savings and keeping pace with technological and demographic change. Different authorities use terms such as 'deficits', 'unavoidable commitments', 'essential developments' or 'priorities for development' in different ways, and it is not easy to be sure like is being compared with like. Nonetheless, the settlement for 1997/98 did seem to make a radical difference to the authorities being studied, although the effect was by no means uniform. In fact it resulted in considerable divergence between the three authorities, and between their plans for explicit rationing.

All three authorities received more than the 1.89% allocated nationally. Avon received 2.2% growth, Cambridge and Huntingdon 2.5% and District Three 2%, according to their own documentation. However, both Cambridge and Huntingdon and District Three reported that serious deficits experienced during 1996/97 were a first call on growth money, which reduced the amount available to them considerably. This was in addition to ECR shortfalls, which Avon also reported.

The original aim of following the case studies through a whole contracting cycle had been to pick out one or two concrete examples of new explicit rationing decisions in each authority and to explore how they were made and whether they were in fact implemented. It became apparent in January that the changed financial position meant this would not be possible as planned.

Given the divergence between authorities, the experience in each one is analysed separately.

Contract negotiations for 1997/98

Avon

In Avon, the director of contracting reported that their original expectation of a £2m additional allocation had been transformed into an actual addition of £8m, a fourfold increase. In fact by the time of the third interviews, the extra allocation had turned out to be £8.4m (2.2% real growth), plus an additional £0.6m from reductions in regional levies. Some of this was due to a greater than expected increase in general growth funding and some of it was due to a change to their advantage in the way that the 'market forces factor' was applied to allocations. There was some speculation about the extent to which this was a result of political lobbying by local conservative MPs in marginal constituencies, worried about losing their seats at the election if the local health service was seen to be having funding difficulties.

As a result of this allocation increase, the district's senior managers felt they could defer major explicit rationing decisions within existing services that might otherwise have been necessary, although in the longer term they believed their allocation would be reduced again as further adjustments to weighted capitation were made. They were therefore still pursuing less immediate plans for explicit service reductions, especially in oncology services, and in 1997/98 would continue to challenge some existing provision (such as D&Cs in women under 40) on the grounds of low cost-effectiveness, but this was likely to be marginal to their annual purchasing plan. The rationing issues were 'still there, and they're happening all the time, but what we haven't had is the urgent need to suddenly impose a new layer.' The district had no deficit to make up, but all the available growth money went on essential developments, particularly in coping with growing acute service pressures. The point was also made that waiting times had to be allowed to lengthen in order to stay within the allocation, an example of the continued use of more implicit mechanisms.

By the time the third interviews took place, district managers were optimistic that further adjustments to weighted capitation would not affect them as adversely as they had anticipated, so the financial position looked more stable.

The director of public health anticipated continued 'unofficial suppression of demand', particularly in the field of new drug technology. This was an area which interviewees in other districts also identified as a problem. Unofficial suppression took the form of the authority telling the clinicians that additional money for new drugs would simply not be made available. In several cases this had been accepted and the clinicians were not prescribing the drug in question. Although this agreement was unofficial, it may well have reflected a belief on the part of the clinicians that the authority would formalise the decision if unofficial mechanisms were not accepted, as indeed it had done very explicitly in the case of beta interferon the previous year.

The authority took differing approaches for different drugs. In the case of beta interferon, it had refused to make additional funding available, but accepted that it might be effective for selected individuals and that, if the neurologists wanted to prescribe it from within their existing budgets, they could do so. In other cases where concerns about effectiveness were stronger, such as Riluzole for motor neurone disease, 'we went further, and said "not only are we not going to fund it, but we don't think clinicians in the area should use it either." But that wasn't public in the sense that we advertised that at a public health authority meeting, or took the decision in public. It's public in that that's the content of letters which have gone to trusts and clinicians.' Neither beta interferon nor Riluzole were mentioned in contracts. This was partly influenced by continuing uncertainty as to whether a drug constitutes a treatment in its own right, which should be specified in the contract, or whether it is merely an in-put to treatment, in the same way as staffing levels or medical supplies, which would be assumed as part of the price and not specified separately.

The use of recombinant Factor VIII for haemophiliacs was also the target of explicit decision-making. Recombinant Factor VIII is a synthetic, genetically engineered clotting agent, which can be used instead of the traditional product derived from human blood. As this was the source of HIV infection which affected so many haemophiliacs, they are understandably keen to use the new recombinant version for absolute safety. The authority, on the other hand, took the view that new production methods for old-style

Factor VIII meant it no longer posed any infection risk. In all other aspects it performs as well as recombinant Factor VIII and is a fraction of the price[1].

The director of public health felt that, despite 'vituperative' opposition to the authority's decision, it would be stated explicitly in contracts:

> 'What we're increasingly doing, I have to say, is being explicit about it all right, but not agreeing with them. We don't reach agreement - there's no point at which they say, "all right, fair cop", and there's no point at which we say, "all right, we give in". We just simply say, "no, we are not making that money available". The trusts can't hold up contracts for that, so we have this kind of "not agreed" position.'

In fact by the time of the third interviews, the authority had not formally agreed its decision, so the issue had not been concluded in the contracts for 1997/98.

Both the director of public health and the director of contracting agreed that refusing to purchase new services in this way did constitute rationing, and yet at the same time they both stated that the financial allocation for 1997/98 meant they did not have to do much explicit rationing. The public perception of what rationing means is generally making cuts in <u>existing</u> services, and health authority staff respond to this perception in the way they present their decisions, even though in their own minds they are clear that there is no rational difference between cuts and not purchasing new services. As noted in Chapter Three, differences in perceptions and use of terminology are a recurrent feature of the rationing debate.

In some ways it is not surprising that decisions not to purchase new technology sometimes fail to find their way into contracts, since the purpose of a contract is to state what <u>will</u> be provided. Whilst it may be expected that contracts would state the exclusion of something that has formed part of the contract in the past and now does not, it may not be reasonable

---

[1] It is interesting to note that media coverage of a law suit against three health authorities in the north west of England to try to force them to pay for recombinant Factor VIII for four haemophiliac children presented the issue as one of geographical inequity. The families too stressed that they thought it was unfair that their sons should not be treated with it when health authorities in the south routinely paid for recombinant Factor VIII. The authority in this case study <u>was</u> in the south and was choosing not to buy the treatment despite a year of unforeseen plenty in financial terms.
The question of infection risk has recently become an issue again, this time in relation to new variant CJD.

to expect contracts to describe services that have never been provided and are not going to be provided.

All the interviewees in Avon expressed particular interest in the use of thresholds as a way of rationing - rather than blanket exclusions (and perhaps partly to cope with the fact that these go against ministerial guidance), the aim is to cap demand by setting thresholds for treatment using explicit criteria. The financial position meant that they could delay debating one such move, which would have been to reduce the frequency of breast-screening. On the other hand, continuing resource pressures in oncology meant they would be developing more explicit protocols, covering issues such as restricting access to chemotherapy for palliative care and shortening the standard period of radiotherapy from six weeks to a month, whilst at the same time providing additional funding for oncology. In continence services, on the other hand, user complaints about a previously determined norm of three incontinence pads per day had led to a relaxation of the policy: ‘the thresholds are still reasonably there, but it gives more discretion around what does constitute an exceptional case.’

The use of thresholds is also a response to the complexity of information about effectiveness and the need to target services at those most likely to benefit from them. There was a sense that most of the easy targets for rationing had already been picked off and that any further rationing was going to get into more expensive and more contentious areas. The very complexity of the issues, however, and the ethical dilemmas of rationing potentially life-saving or life-prolonging treatment made it dangerous territory for purchasers. The relationship between purchasers and clinicians when discussing limits to treatment was described as 'manoeuvring' or 'a bit of a dance with the renal physicians'. In some cases, clinicians were known to have their own guidelines for withholding treatment, for example in neonatology, but the extent to which purchasers could move in to enforce such guidelines was acknowledged to be a contested area. On the other hand, there was concern that unless purchasers got more closely involved, the guidelines would not be held to and would therefore not be effective as rationing mechanisms. In many cases, the interviewees felt guidelines might not be developed at all without purchaser pressure.

As with decisions not to purchase new drug treatments, the inclusion of decisions about thresholds in contracts was variable. In response to a request from the regional office, Avon had undertaken a 'stocktake' of all its services subject to exclusions or restrictions. Although the director of public health commented that it had revealed 'a surprisingly large amount of limiting material in our contracts', it did not cover most of the rationing by threshold, which was largely agreed through other channels. The stocktake exercise, coupled with a review of a patient complaint about plastic surgery exclusions, had led the authority to look again at its decision-making process. The chief executive commented on the need 'to have a more established route through which to bring those things to the authority'. Although rationing decisions were not in any way intended to be secret, there were inconsistencies in the process of formal agreement through the health authority and formal expression in public documents which needed to be addressed. In many ways, concerns about procedural fairness and equity were as important as financial pressures in being more explicit.

Although the longer term rationing agenda was still being pursued, particularly through thresholds, the lifting of the immediate pressure to take decisions in such difficult areas was greeted with relief. In the words of the director of public health, 'we want to avoid that as much as we can, because we get such bad publicity and because we rarely bring it off satisfactorily.' Later he added 'we really don't want to undermine the public's confidence in the NHS unless we really have to. Some of us think we probably will really have to, but the moment is not now.' He reported having heard that the Treasury took the view the NHS could only take one or two years of really hard financial pressure at a time, after which it was necessary to take the pressure off for a breathing space to prevent total collapse. This, he said, reinforced his view that NHS crises were cyclical rather than a linear process. At the same time, he and the chief executive both acknowledged that financial crises could sometimes serve a useful purpose in creating a climate for radical change, and perhaps for greater openness and information sharing on the part of providers. Without that background of crisis, decisions continued to be largely incremental and financial difficulties were handled through general budget trimming.

The director of contracting suggested that they had fairly deliberately provoked political intervention to stop them from making further unpalatable explicit rationing choices, precisely by making some very high profile choices last year, including cutting family planning (see previous chapter) and reducing palliative radiotherapy by referring patients back to their GPs for morphine instead.

> 'I think there are examples where we've been seen to be tying ourselves to the stake and pouring the petrol on. And of course the strength of our position, I think, is also that this authority adamantly refuses to overspend in any circumstance. That's the one thing they're absolutely clear on. And of course that's becoming quite unusual now for health authorities. And in a way I suppose it's quite a powerful purist position, I think, to refuse to overspend and then to burn yourself at the stake over all these kind of rationing decisions. It puts quite a lot of pressure, I think, on the regional office and the centre, because nobody can condemn you for refusing to overspend. And it would be a much easier way out than what we keep doing, which is pouring on more petrol and chucking a match on.'

It would be hard to find a clearer exposition of brinkmanship.

The extent to which the most difficult rationing decisions can or should be pursued through contract negotiations was questioned. With reference to oncology, the director of contracting felt that contracts suffered from a poor information base and that 'contracting certainly lends itself to simplistic ways of counting oncology work', which did not capture the specialist and complex nature of the service. 'I think it is easy to get diverted by the much more simple areas where you can say "this operation's not of very much value and therefore let's cut it out" or "we'll only do twenty". These are much more fine and complicated issues.' The district's approach therefore was to work on a separate briefing paper and to pursue negotiations between public health, GPs and the clinicians concerned. 'I think one of the big dangers of the system we've got is allowing things to happen through contracting, really...We're [the contracting team] generally there as the organisers and facilitators and implementers, and I think it's very dangerous to allow contracting to get exposed and into such a way that it's actually trying to take those decisions. I don't see how contract managers can, because they're not clinical people - I don't see how they should, either.'

The director of public health commented on the lack of knowledge at purchaser level about the effect of their decisions on provision; 'it's quite possible that the contracted for limitations have no effect whatsoever, and the ones that we don't specify have the most massive effects.'

One major source of financial difficulties in Avon was tertiary ECRs, particularly in mental health. Here the district was planning to make it a pre-condition of all contract negotiations that secondary care providers should take the budget for tertiary ECRs and accept the financial responsibility for controlling them. This is an example of moving back from the very stark purchaser-led decision-making process associated with ECRs towards a more implicit clinician-led process. The director of contracting was uncertain whether they would be successful in negotiating this with trusts; the director of public health agreed that hitherto providers have always resisted such proposals, but felt the fact that they were in a growth year might enable the district to persuade providers to accept it in exchange for some other developments, or alternatively by being told they would not get developments unless they took on tertiary ECRs as well. 'What gets things changed is the judicious application of blackmail and bribes - the blackmail being "if you don't play ball with this particular piece of policy, we will withdraw something or other which you need to get things done". The bribe is "you know, we were going to invest *x* in you. I don't think we will now". And we've always done that....It is essential that you have some money to spend in a year. It's the dynamic that you can wrap a great deal of change around.'

He also commented that it was the expectation of new investment overall that was encouraging the oncologists to co-operate with the development of thresholds for some aspects of cancer: 'they're quite confident that their time has come, that this is an era in the NHS when oncologists are going to have their ten years in the sun.'

In practice, by the time contract negotiations had been concluded there had been varying progress in including ECRs in contracts. Providers were anxious about being left holding the entire risk, so some risk sharing agreements had been negotiated. In one case, the authority had made a decision not to fund an additional neurologist. However, the neurology department had been given to understand that, if they accepted budgetary responsibility for managing their own tertiary ECRs, where there had recently been substantial overspending, they could use any savings to fund a new neurologist themselves. Thus an apparently explicit decision was in fact left to the discretion of providers, as long as they remained within budget.

There was some discussion with the directors of public health and contracting about how the same rationing issue can be presented differently in different authorities. Explicit restrictions on renal services in a neighbouring district were described as 'so marginal that they're not really what other people might recognise as restrictions. They're what the clinicians do every day.' Avon had recently been criticised on television by a local oncologist for not funding trials of Taxol, a new anti-cancer drug, whilst a neighbouring district (where the oncologist also worked part-time) was said to be funding it. 'It wasn't - they didn't know what they were funding. It was just being used on their patients because there was enough money left in their contract.'

The issue of cataract removal for the second eye, which was an issue in Cambridge and Huntingdon (see below), was also discussed. The director of public health suspected that not many such operations were performed in Avon - 'it just may be that that's happened by custom and practice...an unofficial bit of rationing has taken place, and we don't need to say "don't do second eyes".' The director of contracting stated that in fact there had been discussions with the ophthalmologists about whether to restrict second eye cataract removal. 'They all have slightly different policies from each other...and slightly different views on the clinical priority that they should give to it.' As a result, it had been decided that it would not be appropriate to make it an official exclusion.

Cambridge and Huntingdon

In Cambridge and Huntingdon, the general growth allocation had also increased, from an expected maximum of £1.5m to £3.875m (2.51%). However, at least £1m of this was expected to disappear straight away into inflation, the allocation for which was felt to be unrealistically low. Furthermore, there was a list of service pressures (including recovering from a £1.5m deficit incurred in 1996/97) and priorities for development which were thought to require additional funding in 1997/98 and this list outstripped the remaining growth funding, so explicit rationing was still very much on the agenda.

Rather than impose rationing decisions on providers, the health authority took the approach of developing a menu of 29 possible rationing or 'disinvestment' areas, although some of these could in fact be achieved by service rationalisation rather than reduction. Some (such as replacing complex pacemakers with simpler ones) would result in reduced quality but the same volume of service. Included on the list were cataract surgery for the second eye, male and female sterilisation, health visiting and school nursing, varicose veins, prostate surgery, sleep apnoea, psychiatric rehabilitation and community midwifery, to name but a few. The authority was careful to insist that none of these would be subject to a blanket exclusion; clinically necessary exceptions would be allowed, in line with the recent White Paper 'The National Health Service: A Service with Ambitions' (Secretary of State for Health, 1996). However, it was also made clear in the Draft Annual Plan that 'the Authority recognises that there will be a health loss associated with each and every one...but considers that the relatively greater health gain resulting from investment in the priorities for development must take precedence.'

Providers were told only a fixed amount of money was available, and if, in order to live within that amount of money, they had to make reductions in any of the services included on the menu, the authority would endorse that decision. This was on the assumption that providers could not find any further efficiency savings to meet the shortfall. If, on the other hand, they could come up with their own alternatives for saving the money and those alternatives were acceptable to the authority, then this would be welcomed. The director of public health was particularly insistent that this approach did not mean the authority was making a decision to stop providing sterilisations, for example. Rather it was offering a list of suggestions to the trusts to allow clinicians to decide the most appropriate areas for disinvestment, an approach which bears strong similarities to the developments in explicit rationing described by Klein *et al.* (1996) which were discussed at the end of Chapter Six. Given this authority's past history of very high-profile rationing decisions, and the controversial inclusion on the list of health visiting and sterilisations, the possibility that this was intended as a way of off-loading responsibility for hard choices is not an immediately plausible explanation. Indeed, the director of public health said at one point 'we would have the guts to say "no" to *x*, except that we've been told by the Secretary of State we're not allowed to say "no" to *x*.' Furthermore, he stressed the authority's willingness to take final public responsibility for the choices made at provider level.

> 'The authority puts its hand up and says, "I've taken 20% off the ENT budget or ophthalmology, and so you're not going to get all the services that you want, and it is our fault, because we can't afford it," but the decision as to whether they do a cataract rather than a squint, Mrs X rather than Mr Jones, has surely got to be a clinical decision. And implicit at that, I suspect.'

He regarded this emphasis on affordability as being an even more explicit basis for rationing than previous approaches, which had confused issues of effectiveness with issues of affordability, as in the case of beta interferon. However, he recognised that it would be 'worrying for politicians'.

As a general principle, he felt the sharing of responsibility for rationing at the different levels could be described as an '80/20 situation - or even 90/10'. 'Deciding how much we put into ENT or gynaecology...is 90% our responsibility. We take the can for the million pounds available for ENT. But when it comes to how you spend that million pounds on Mrs Jones and Mrs Smith...that is 90% your [the clinicians'] decision...Although the two are not totally separate...it is separating out rationing at the meso level and the micro level.' The rationing suggestions for 1997/98 were thus presumably meant to fall within the 10% of micro rationing where purchaser involvement was appropriate, although it would in fact seem more interventionist than that. This can be partly explained by conflicting views within trusts identified by both interviewees. 'You get this tension between "well, don't try and manage clinical decision-making in detail" on the one hand, but "if you want us to reduce the service you want, you have to tell us in detail which services you're not going to purchase" and the two don't add up.' Trust chief executives in particular were reported to be unhappy about rationing decisions being passed back to them.

Faced with this situation, the authority took the line that 'flagging up some broad areas for disinvestment is about as far as we can go without meddling in the details of running a hospital...The mix between them and the degree to which they disinvest in each of these individual lines is down to them to plan.' However, he already felt confident in January that some of the proposals would definitely happen, some would probably not and was uncertain about the outcome in only in a few cases.

As it turned out, the response from providers and others consulted, such as GPs, was regarded as a disappointing demonstration that providers did not believe the rationing proposals would really be implemented. Despite the strong emphasis from the authority that there was a fixed amount of money available to each provider, negotiations for more money had continued until quite late in the process. As a result, few rationing alternatives had been put forward, and the authority's original list was implemented almost in its entirety.

However, it had been agreed with the ophthalmologists to withdraw the proposal to ration cataract surgery for the second eye, and to give the clinicians a savings target which they would manage themselves. This was felt to be a move in the right direction by the director of public health.

The decision to exclude male and female sterilisations had been implemented, but had created such concern amongst the gynaecologists that they were now prepared to consider cutting other procedures in order to reinstate sterilisations. 'It has taught us that maybe to get their attention, instead of threatening to cut, you've got to cut first and then negotiate.' Part of the reasons for the gynaecologists' concern seemed to be that, although they were able to make exceptions within the terms of the authority's decision, there was uncertainty whether exceptions were to be made only when pregnancy posed a real health risk (for example when hormonal changes in pregnancy could accelerate some forms of cancer), or when there was a broader 'genuine social need'. The gynaecologists were uncomfortable with this position.

The Draft Annual Plan also included a list of ten areas which were originally considered for inclusion on the menu but were rejected following discussion by members of the authority, for a variety of political, personal or scientific reasons. These include assisted conception, termination of pregnancy, impotence treatment, music and art therapy, patient advocacy and gender dysphoria. Proposals from providers for service reductions in these areas would therefore not be accepted.

The director of public health reported that there had been considerable disagreement about terminations amongst authority members. The reasoning why terminations had been treated differently from sterilisations had not been clearly articulated, but was felt to have been influenced by the fact that in the case of termination a fetus was already present and that therefore there was an immediate need, whereas sterilisation was preventive and therefore less immediate. There was also a perception that those seeking termination were likely to be less able to pay for themselves than those seeking sterilisation.

The director of commissioning commented on the irony that IVF, which was dropped from the list of rationing suggestions, was not as effective as sterilisation but added 'neither's ventilating old people or ventilating babies, but...it's all about values in the end...There is no scientific way of making these decisions', despite the fact that evidence on effectiveness had been the theoretical basis for a number of the decisions discussed. In the case of impotence, the argument had been 'why target one particular dysfunction when there's lots of others that we treat quite happily?' The decision to drop gender surgery from the list was taken purely on the grounds that so little was purchased anyway that to reduce it further would risk contravening the requirement not to ban anything altogether.

In fact, consultation on the rationing menu had resulted in a reconsideration of assisted conception and gender reassignment, and both these procedures were ultimately re-included on the list of services to be rationed.

As well as changes in the authority's thinking about rationing, the decision to offer providers a menu from which to choose can also be attributed to changing views about how best to manage negotiations and about the purpose and form of contracts. The director of commissioning in Cambridge and Huntingdon commented on the irrelevance of distinctions between contracting and other mechanisms

> 'It's all the same really...there is a sum of money - what are we going to get for that? I mean, that's the contract negotiation. It's basically that the contract negotiation isn't about money - it's about service. Which is what it should be about, really, because otherwise you spend lots and lots of time arguing about notional amounts of money, without getting down to what the service issues [are]. It's easy to get obsessed with money, you know - there's a huge NHS infrastructure devoted to having debates about money, but it's relating money to the actual service that's the important thing.'

This perspective reflects the general view in this health authority that the contract was simply the expression of agreements reached through the wider commissioning process, and not in itself a driving force. 'Whether it's in the contract or not is not the important thing, it's whether the people that matter within trusts are signed up to it or not.' 'Most if it's in people's heads, that's the important thing.'

The director of commissioning expressed considerable disillusion with the use of activity specifications in contracts, and felt negotiations over activity levels were futile given the constraints of a fixed budget.

> 'What we're trying to do here is get away from that, and say 'well, look, it doesn't really matter what the activity is - there is a certain amount of money, and that's all there is to it...You can keep within that cash limit in a number of ways. You can either pull back on your activity, by things like the sterilisations and things, or you can cut back on your infrastructure. Really we don't mind which you do. It is open to you either way. But at the end of the day there's a certain amount of money and that's it, so there's not much point saying "well, the activity's going out the roof". And there's no point. I mean, it's just a pointless conversation, because unless one or other of us goes into deficit there is no other way round that.'

This demonstrated precisely the frustration of trying to reconcile a supposedly demand-led market with a fixed global budget - a recurrent feature in interviews throughout this research has been the perception that providers behave as though there is no limit on money and if they increase their activity levels they should simply be reimbursed accordingly. As the director of commissioning said, 'we've got an internal reimbursement system, but we don't have the money to reimburse everything everybody wants reimbursing.'

Does this shift in approach mean that, after all, Cambridge and Huntingdon was off-loading the responsibility for hard choices onto providers? It is conceivable that a district which had had such harsh experience of media reaction to open rationing might want to share some of the burden - although this was not how they themselves saw it. The director of commissioning responded in two ways. Firstly, he argued that:

> ‘activity doesn't make the process more explicit, because nobody controls activity...Health authorities may say that “oh yes, we'll buy twenty less of HRG 64”...but it's not like that. It's not like a fundholder who can decide that “we're going to do eight hip replacements, and Mrs Smith and Mrs Jones and Mrs Brown will have them”...We're managing a system where we don't directly manage the activity. Now, we can put things in that will hopefully have an effect on activity, like we can have a rationing policy on sterilisations which will have an effect on activity. We can give a steer that we want them to cut down on grommets and tonsillectomies...But the translation from that into activity is guesswork. It's not scientific.’

In other words, the apparent explicitness of clear activity ceilings in contracts was spurious; therefore omitting such activity specifications from contracts did not represent a real reduction in explicitness.

Secondly, he argued that they would be just as explicit in contracts about what they wanted for their money, but that it would be expressed in terms of quality standards and service targets.

> 'If you take it to the extreme, you could say "well, what we want is access for emergency admissions, twelve month or fifteen month waiting targets, we want these particular infrastructure improvements...This is the amount of money that's on the table. Get on with it. And what we want from you - what we want in terms of contract monitoring is what your waiting times are, the number of days you've restricted

> emergency admissions, trolley waits, when you've put in this forensic development...We don't want to know about activity - activity doesn't matter. If you can manage, if you can deliver those service targets, it doesn't really matter if the activity's 80% of what we think it's going to be or if it's 120%...The key thing is the service targets.'

In fact the director of commissioning recognised that this would be an extreme position, and that an understanding of activity levels would always be necessary - 'because how else do you measure whether hospital x is working harder than hospital y, who's got the more pressure? You need some way of counting it.' It was more a question of developing a different emphasis and being realistic about the limitations of contracts. The contracts for 1997/98 did not contain detailed targets for reductions in activity in the areas included on the list of rationed procedures, but they did set out an agreement that there would be reductions. It had also been agreed with providers to reduce the total activity assumptions specified in contracts in recognition of the budgetary constraints imposed upon them, although there were doubts whether providers would actually keep to those reduced levels. By the time of the third interview, overspending was already becoming apparent in some trusts.

This change of emphasis from detailed activity specifications back towards global budgets could be seen as an example of a move back to hierarchy, in response to the unworkable nature of the quasi-market. Alternatively, it could be seen as a refinement of the market relationship, to focus on more appropriate outcomes than crude activity levels.

In any event, it marks an increasing perception that the proper role of purchasers was to set an explicit framework, within which implicit decision-making could appropriately continue placing trust and responsibility with clinicians. This was not so much a return to implicit rationing, as a reassertion of its importance and the relatively marginal role of explicit rationing. The director of commissioning commented, 'we haven't gone from implicit to explicit, we've gone from implicit to implicit plus explicit.'

It was planned to begin discussions with providers about next year's contracts much earlier than the previous year, reinforcing the principle of a fixed budget, within which they would have to manage. Reflecting the authority's position that responsibility for micro rationing lay with clinicians, it was not planned that the authority should itself come up with a detailed rationing menu again. 'It's really trying to throw it back to them, and say "well, hang on, we did all the running last year - this list of things came entirely from us. It's not a realistic way to manage the NHS. There's 150 of us, and there's thousands of you out there, working in the trusts, and you've got to take some ownership of the problem and work out how you're going to cope".' The director of commissioning was in no doubt that the purchaser provider split had led to 'unprecedented cost inflation' as providers had struggled to increase their share of an ultimately cash limited system; the approach now being adopted in Cambridge and Huntingdon sought 'to confront them with the stark reality of the foolishness of that approach.'

He was emphatic that there was no single approach to controlling costs that could provide the answer. 'People keep casting around for the Holy Grail, and I don't think there is one.' Certainly explicit rationing alone could not tackle the problem, unless really radical exclusions of whole areas of service were undertaken, but these would be politically unacceptable. The current fashion for reconfiguration was also insufficient.

> 'If you use any policy instrument as your sole way of managing the NHS then you're doomed to failure. It's more complex. It requires some explicit rationing, it requires some implicit rationing. It needs pursuing efficiency, it needs rigid cost control, and it needs all of these things in order to function.'

District Three

District Three was experiencing a substantial deficit in 1996/97, largely as a result of fundholder overspends. Some providers had also been running up deficits. This had affected the financial outlook for 1997/98; even though the additional growth allocation of £3.85m (2%) was slightly better than anticipated, they did not expect any of it to be available to fund new developments. Approximately half of the allocation would go to meet a shortfall on inflation, and the rest would be taken up by covering deficits, meeting cost pressures in the budgets for ECRs and continuing care placements and some unavoidable prior commitments to develop mental health services and resettlement of people with a learning disability. This position assumed that the district would be successful in negotiating down the deficit within trusts to some extent; it was hoped this would be achieved 'just by sheer haggling and horse-trading, rather than explicit rationing'. If necessary, some of the unavoidable developments could be slipped to start later in the year, building up a full year cost problem for 1998/99.

There was therefore still a strong feeling that resources remained very tight; one interviewee likened the settlement to 'somebody helping you pay some of your credit card bill' but not solving the longer term mismatch between resources and demands.

Given the expectation of an approximate break-even, the director of commissioning reported that the authority was 'not actually looking at any positive disinvestment this coming year'. This meant that there would be no explicit rationing decisions which involved taking money out of contracts. However, 'there are things that trusts might offer us as a development, which we wouldn't want to accept' - and again the pressures of new drug technology was mentioned - 'and there are odd things such as sleep studies, that...we currently purchase on an ECR basis, and we may start saying "thank you but no thank you".' A reduction in sleep studies would probably be implemented by refining the existing clinical protocol to 'raise the threshold' for referral. She anticipated that this would be a temporary measure for the next year, with a view to phasing it out altogether in subsequent years, on the grounds of ineffectiveness. This was 'not mega-bucks in the scheme of things, but it's probably one more thing we can try.' In the event, 1997/98 contracts focused on ensuring all referrals for sleep studies were channelled through one tertiary referral route.

In addition there were some procedures within contracts which would continue to be reduced, such as grommets and D&Cs in women under 40. These were mentioned in the purchasing plan for 1996/97, but were proving difficult to implement. The director of commissioning said 'we're chasing up those to make sure we know what's gone through, what's been paid for and really being very assertive with the trusts to actually show tangible reductions in our contract volumes when we've said we won't buy something.' The estimated potential savings from these two changes of £350,000 a year would be financially much more significant than sleep studies. However, the director of commissioning said the figure was ambitious and was based on being able to withdraw full costs, which might not be realistic. In practice providers were told the savings would not be taken away from them, but should be used to offset the overall deficit they were asking the purchaser to make good; no explicit targets for activity reductions were included in contracts. This was not 'positive disinvestment' because the resources stayed with the provider; however, it was a redirection of resources to support higher priorities. The director of commissioning felt 'we have a very delicate path to tread, in terms of policies and procedures that, by their very nature, are going to cap demand', and that even this kind of less visible approach could face considerable opposition from trusts.

More substantial conflict was reported between purchaser and provider over renal services. The district was seeking to develop protocols to restrict demand in both general nephrology and dialysis, whilst at the same time putting in more resources. At the time of the second interview (mid-February 1997), the trust had not yet agreed to sign the contract for 1996/97, the current financial year. Independent clinical advice had been sought on issues such as the balance between peritoneal dialysis and haemodialysis (the former being considerably cheaper). The district was also concerned about the appropriateness of using haemodialysis as part of palliative care in people who were dying of some other condition and developed kidney problems as well.

However, there was still clinical opposition to carrying out rationing decisions on the authority's behalf. Negotiations were being pursued both through and in parallel with contract discussions and the director of commissioning commented, 'It just goes to show, as a purchaser, when you come up against resistance from a trust, it really does get to the heart of what influence you have as a purchaser, without sort of resorting to "well, we'll stop paying you completely"...we've got in many respects very, very crude tools to use, which at the end of the day we probably wouldn't use, because that's inappropriate and unprofessional as well.'

A particular problem in the case of renal services was the perceived lack of support from trust management. 'Trusts don't necessarily feel an obligation to work within the contracted volumes, and you will get the chief executive siding with the clinicians to demonstrate why they need more money, rather than managing their clinicians to force through an agenda of "Are we doing too much? Is it clinically appropriate? Is it best practice? Is it acceptable practice?" So we're not getting the internal support, I believe, from the trusts, that certainly was originally envisaged.' This offered an example of how the purchaser provider split has made previously implicit rationing decisions more explicit. The director of commissioning said 'it only works if we're all part of the same conspiracy', but the conspiracy was breaking down and trust managers were prepared to support their clinicians in refusing to conspire.

By the third interview, close examination of local renal services had provided some evidence that their unit costs were considerably higher than in other districts. The implication that the district might move services elsewhere had put the local trust under pressure to bring costs down. The director of commissioning remarked 'I think probably we're still trying to squeeze the last drop out of any contracting or management intervention, without having to go down the road of more explicit rationing.' This seemed to be a counter-example where market competition, rather than driving rationing into the open, was making it possible to use a more implicit approach by reducing prices. A focus on high unit costs could also be effective within one trust, by encouraging peer pressure from colleagues who felt their own service developments were being jeopardised by another department's high spending.

As elsewhere, the question of expensive new drugs and whether it was legitimate to exclude them from normal contracts had been a major problem. The approach being taken in District Three was to work with local clinicians (and users) to develop protocols for determining access, a 'sensible half-way house' between a blanket 'no' and a blanket 'yes'. In some cases this might mean the purchaser would require evidence that the patient had been involved in a discussion of alternatives to the new drug, such as palliation. The director of public health reported that clinicians were finding this approach helpful, as it gave them continued clinical discretion but protected them from accusations of wasteful or inappropriate use of new drugs. The chief executive and director of public health were both strongly pursuing the line that drug-exclusive contracts were unacceptable.

District Three was also seeking to negotiate some financial responsibility with secondary care providers for tertiary ECRs, particularly in mental health. They planned to restrict access to private hospital beds where they felt these were being inappropriately used, by refusing to accept the responsibility to pay; if secondary care providers continued to use them, they were likely to have funding withdrawn from their contracts to pay for the ECR. The director of commissioning commented, however, that the chances of successfully negotiating this kind of arrangement with providers were not good, given the fact that most were 'exceptionally risk-averse'. By the third interview, she reported that the principle had in fact been accepted, although there was still some ‘horsetrading around how much money we give them to manage the tertiary ECRs’.

In the longer term, she thought providers would be more positive about plans to work with them on ‘repatriating’ referrals which had traditionally gone elsewhere, either as ECRs or through contracts. This would mean giving local providers the money normally spent on such referrals to enable them to build up local services. In some cases, this might also help purchasers regain more explicit control over provider activity. She described the position in one out-of-district provider where patients referred for routine secondary care were often referred on within the system and purchasers were required to pay invoices for subsequent treatment many years after the original referral.

At the third interview, the chief executive reported that it had been necessary to ‘rein back on activity to get within budget’ and conclude contract negotiations. An important element of this had been to ‘put a ceiling on the cash that any of our in-area or local providers will get from this authority, whether they get it on block contract from the authority, from a Total Purchasing Pilot or from a GP fundholder....Fixed budgets - we’ve already got back to that.’ This was partly a response to trusts seeking to ‘play the system and bring the business in and force us to run an overspend’. Although this was to some extent a return to implicit rationing by providers within their overall budgets, it was accompanied by some explicit reductions in contract activity levels, and an explicit understanding that this would result in longer waiting times.

Furthermore, the chief executive described plans to incorporate more explicit priority setting mechanisms within waiting lists, so that there would be clearly specified criteria for assessing individual patients' level of need. This was expected to be similar to mechanisms being tried in New Zealand (Hadorn and Holmes, 1997) and would result in patients who met the priority criteria being given a booked admission date. Patients not meeting the criteria would not be put on a waiting list. He felt many consultants 'would argue that waiting lists are a good form of rationing - that waiting lists in many ways are not unethical', particularly if combined with a more equitable way of deciding who should be placed on the list.

Although GPs (including fundholders) were largely supporting the fixed budget approach, there was some concern that they might try to circumvent it by referring patients elsewhere on an ECR basis. However, it was suggested that fundholders, who constitute a high proportion of the district's GPs, were actually experiencing quite strong budgetary constraints themselves, so it was not expected that referrals away from main providers would become a very serious problem.

Even though the 1997/98 contracting round had not required many very explicit rationing decisions, it was not anticipated that this would be sustainable in the longer term. 'We're not quite there yet, but I think this time next year will be very difficult.' The very fact that the new government had made more money available than anticipated might contribute to the problem - 'cost pressures emerge, simply because there's money there to meet them'. All the interviewees believed that further explicit rationing decisions would be required for 1998/99. Although the scope for further exclusions, as opposed to rationing by thresholds, was limited, there might be some new additions to the list of items not normally purchased, particularly new drugs of unproven or low effectiveness, such as Taxol. It was also suggested that chiropody might follow the example set by dentistry and be excluded from the NHS or only offered to certain age groups.

The director of public health expressed concern that 'you need to be very careful that you're not allowing cheap, inappropriate, ineffective treatments for common conditions, at the expense of something that's more appropriate and effective, but where the individual cost is a little higher.' The director of commissioning agreed that there were a number of high volume low cost areas where activity reductions might be appropriate, but suggested that in practice gains were likely to be greater in reducing a small number of very expensive procedures. In addition, 'we've got to be conscious of our management capacity to actually bring about some of these things. And therefore I think that does drive us down the low volume high cost route.' Both interviewees were conscious that trying to tackle the whole agenda at once would be very costly and probably impractical.

There remained a commitment to some major service changes: 'we wouldn't in any way back away from what we're trying to do with the acute services, irrespective of what the settlement gave us. We might not have to cut so deep, but it wouldn't stop us from making a good start....Because the settlement's been slightly more generous, some of that money can actually be used positively between us and social services and the trust to bring about some more radical change.'

The use of thresholds was expected to become much more widespread. The director of commissioning commented on the practical difficulties for purchasers of monitoring compliance with threshold criteria - 'if we wanted to look at every case note there would be thousands' - and suggested that medical peer pressure might actually be a more effective means of ensuring compliance. Nonetheless, the development of treatment thresholds and eligibility criteria was seen as a more promising and realistic way forward than exclusions, and it would be expected that purchasers would take an active role in ensuring protocols were in place, and sometimes in writing them.

As in other districts, there was a growing sense that health authorities had a responsibility to set an explicit framework within which clinicians could work:

> 'We're currently asking GPs and also hospital consultants about things that they feel they would prefer not to do, or would prefer not to refer. And I think interestingly this is, I suppose, crystallising the role of the health authority, in terms of we would be where the buck stops - if such and such a service was no longer purchased, that we would take full responsibility for that decision being made. And that consultants and GPs in particular would say it is not the policy of the health authority to purchase this.'

Discussion

What can be learned from the experience of these three districts in negotiating contracts for 1997/98? Inevitably the answer is complex. To some extent, it re-emphasises the primacy of shortage of resources as a driving force behind greater explicitness. It has already been pointed out that many explicit rationing proposals do not in fact yield huge savings, so they may not solve the financial problems to which they are a response. Perhaps more important is the organisational atmosphere generated by financial problems, which makes managers feel they must bring hard choices out into the open. It may also be a gradual process, beginning with the more obvious targets but progressing into more difficult areas, where greater savings may be obtainable, if resource pressures continue to grow.

If, on the other hand, there is a temporary financial let-up, managers may be only too glad to step back from the brink, even though they know it may not last for long.

There seem to be differences in how committed different managers and authorities are to the principle of explicit allocation decisions. Some expressed considerable doubts about the rightness of explicit decision-making, leaving aside the view that it was probably inevitable. On the other hand, there are those, like the chief executive quoted at the beginning of this chapter, who feel impatient to emerge from the trenches and tackle the issue head on.

> 'Something that I think would be a backward step would be for a government in the dying days before an election to shove a whole lot of money into the NHS, because it would just literally be buying time, because next year we would be at the point where in fact we're at this year. So no amount of money will do anything other than act as a time delay.'

To him a generous allocation would mean an opportunity to go beyond the point of no return had been missed. He did not want to feel an invisible political hand pulling them back from the brink.

These differences of opinion suggest managerial attitude may be an important factor in how far authorities are prepared to go. Nevertheless, the chief executive's comment suggests that even the most committed would feel obliged not to pursue extensive explicit rationing if given a pre-election bonanza. As it happens, he was chief executive of Cambridge and Huntingdon, where there remained a large gap between the allocation and funding requirements,. However, it is intriguing that Cambridge and Huntingdon reported the largest percentage growth increase of the three districts, and was yet the district pursuing the most explicit rationing. Some of this is undoubtedly because of the need to cover existing deficits, which was not such an issue in Avon; however, deficits were a serious issue in District Three, and yet they had not pursued the same overt approach as Cambridge and Huntingdon.

There is of course an important difference between explicit rationing on the basis of good evidence that a treatment is not very cost-effective, and rationing highly effective treatments. As already mentioned, Avon will still be looking at reducing areas where there is consensus on low cost-effectiveness, such as D&Cs in women under 40. However, this is far removed from Cambridge and Huntingdon's inclusion of effective procedures such as cataract surgery and sterilisation on its rationing menu. The district is entirely open that its proposals for disinvestment all involve 'health loss'.

Evidence on effectiveness continues to provide a <u>rationale</u> for many explicit rationing decisions, although whether it in fact <u>causes</u> them to be made is another matter. Certainly it is not always the decisive factor, and many interviewees commented on the value judgements involved. Moreover, its effects are still felt largely in marginal areas, although this may change as pressure for more substantial savings directs attention to more complex clinical areas where effectiveness evidence can be used to set thresholds for treatment.

Explicit decisions do not always find their way into contracts. Some major decisions are taken mid-year in response to financial crisis and some are pursued through parallel routes. Contracts may catch up with this later by formally recording the decisions made, but are not necessarily the mechanism by which agreement has been reached. Nonetheless, there was broad agreement amongst the interviewees that levels of explicitness had increased dramatically since the introduction of the quasi-market, and this is explored further in the next section.

However, implicit rationing mechanisms continue to play an important role, and there is evidence of districts returning to or re-emphasising implicit mechanisms in response to the perceived failure of more explicit methods to achieve the desired results, and in response to doubts about its appropriateness at purchaser level. The re-inclusion of financial responsibility for tertiary ECRs in secondary providers' contracts is one example of this.

More striking is the evolution in both District Three and Cambridge and Huntingdon towards contracts based on fixed budgets, which lay greater emphasis on the role of clinicians in making allocation decisions, albeit within explicit guidance. This does seem to provide support for the theory that managing a fixed budget within a quasi-market produces so many irreconcilable tensions that there will be a drift back to less visible ways of capping demand. It could be argued that it is unclear whether this will actually be achievable now that the 'conspiracy' of implicit rationing has been exposed by the purchaser provider split. However, it is important to note the consensus amongst interviewees that in making these changes they did not intend to abrogate responsibility for hard decisions. There was a clear view that they retained responsibility for making meso-level decisions (about how much should be spent in different areas); for taking a lead on the development of more explicit criteria for access to services and suggesting appropriate areas where reductions could be made; and finally for supporting micro rationing decisions made in their name by clinicians and 'taking the blame' in public. There was no desire to conspire, but rather to ensure a proper balance between implicit and explicit approaches.

A more important doubt about the likely success of this strategy is the extent to which clinicians - and trust managers - will be willing to accept it.

Reflections on previous approaches to managing resource pressures

In addition to investigating the current contracting cycle, the second round of interviews was also used to ask participants to reflect on previous approaches to managing resource pressures, both before the market was introduced and in the early days of the market. They were asked to compare what they would have done then with what they were doing now, and to consider the reasons for any differences in approach.

Documentary analysis

To supplement this discussion, each district was asked if it could supply examples of strategic and annual planning documents from past years and particularly from 1989/90 and 1992/93. The purpose of this was to see whether the interviewees' recollections of how things had changed was supported by written evidence. It was anticipated that there might be problems with this, however, as all three districts had undergone a variety of boundary changes and mergers during the past ten years, resulting also in office moves and changes in document storage. In the event, Cambridge and Huntingdon was unable to supply any such documentation.

Avon and District Three were able to produce quite a few documents from their previous constituent districts, from which some interesting contrasts emerged. For example, both produced an acute services review, in Avon dating from as far back as 1986 and in District Three dating from 1989. This document from District Three provided a classic example of pre-market facilities-based planning; the reasons given for the review were almost entirely to do with the viability of the local district general hospital, decommissioning an older hospital with unsuitable accommodation and providing an adequate base for medical training purposes. Although the review referred to 'Working for Patients' (Secretaries of State, etc., 1989) it was simply to use the idea of locally available 'core services' as a justification for expanding the local hospital. There was strikingly little mention of health needs, apart from one paragraph on standardised mortality ratios, half a page on demography and a third of a page on the social characteristics of the local population. This contrasted with an entire chapter devoted to estate issues. Discussion about service levels was mainly in terms of staffing levels and traditionally calculated bed targets.

District Three also made available a further acute services review for the whole county from 1993, the main purpose of which was to discuss moving contracts away from expensive providers in central London towards more local providers who could provide high quality care at lower cost. As might be expected, this placed considerably greater emphasis on health needs and activity levels, using detailed ICD and OPCS codes, rather than sites and physical capacity. This was founded on a distinct separation between purchasers' interests, on behalf of the population, and providers' interests in running their own hospitals. It showed a sophisticated understanding of referral patterns for both elective and emergency work and of the need to make any proposals sensitive to the different reasons underlying such referral patterns. For example, it might be appropriate to cease buying lithotripsy from London, because alternative provision was available locally, but the Lane Fox Unit at St Thomas' provides a unique specialist service which they should continue to purchase. Information from contract specifications and ECRs clearly formed the basis for this level of understanding, and yet the document referred to the 'relative insensitivity of the analysis' - perhaps illustrating that the more you know, the more you realise how much you do not know.

The impact of this shift away from describing facilities in terms of staffing and sites towards detailed activity breakdowns is that it paves the way for explicit rationing of services. Saying that a certain hospital or ward will close does of course imply that there will be some reduction in service, but it is not specified who will be refused treatment or for what conditions. It is also possible that activity levels will in fact remain unchanged and will ~~be~~ simply be absorbed elsewhere in the system. However, once the description is couched in terms of clinical activity, it becomes much harder to conceal the effect of any changes on that activity, and indeed activity becomes the common currency for expressing change.

In Avon, documentary evidence of the move from facilities and staffing-based planning to service activity and quality was not as clear-cut. The Acute Services Review from 1986 was astonishingly detailed in service terms for its date, for example giving a breakdown of urology activity by ICD codes which included the number of people treated for malignant neoplasms of the kidney and urinary organs, diseases of the urinary system, diseases of the prostate and hydrocele, to name but a few. Having said this, most of the proposals for change were expressed in terms of site rationalisation, number of clinic and theatre sessions, additional consultant posts and bed numbers.

One of the few examples of a clear decision not to provide a particular treatment was IVF at the local teaching hospital. The region decided not to make ear-marked funding available 'because its priority compared with other services did not indicate that earlier priorities for service development should be over-ridden.' Of the three districts which make up the present single authority, two decided not to fund it themselves, and one decided it would fund current levels of provision only for its own residents (at a time when district of residence was not normally a criterion in planning who should get treatment). For those not covered, the system would continue as it had been up till then - namely as 'a semi-private non-profit-making service involving donations by couples to the University'. As mentioned in Chapter Three, IVF was one of the few examples of pre-market explicit rationing, but there was always the possibility that women could get accepted for NHS treatment at a hospital in another district.

Despite the detail on activity contained in the 1986 document, it was clear from Avon's contracts portfolio for 1992/93 that things had moved on. The contracts contained very little on staffing or facilities issues, and considerable information on activity broken down by procedure, as well as comprehensive descriptions of the range of services provided by each specialty and a statement of special issues and developments in each one. There were more explicit references to rationed procedures, such as laser treatment for birthmarks, but not many. It is notable, for instance, that there was no reference to restricting cosmetic surgery, D&Cs, sterilisation reversals, varicose veins or grommets. Indeed, it was stated that varicose veins, hernias, tonsillectomies and grommet insertions should be performed to a <u>minimum</u> level, rather than a maximum being set. More emphasis was placed on increasing the proportion of these operations performed as day cases.

This is substantially different to the service specifications in 1995/96 contracts, which in plastic surgery listed eight individual cosmetic procedures which were not normally to be undertaken, together with an expected number of exceptions for each procedure and the grounds on which exceptions could be made (such as psychological distress or severe difficulties in finding employment). However, this is an interesting example, as it is in this district that the plastic surgeons requested that the indicative volume levels be removed from the contract for 1996/97, as they were finding it too restrictive and wanted to return to more implicit clinical decision-making to determine how many of each of the restricted procedures should be allowed.

The documentary evidence accords generally with the views expressed at interview discussed below, namely that in general terms resource allocation had shifted from a facilities and staffing approach to an approach focused on activity and quality targets. There was also consensus that rationing had become more explicit. However, the reasons for this were not seen as solely or even primarily to do with the market.

Interview analysis

Avon

One factor identified by the director of public health in Avon was changing clinical attitudes at provider level. This did not always work towards greater explicitness, as demonstrated by the evolving attitudes of the plastic surgeons. However, in most cases there was an increasing tendency for clinical staff to decline responsibility for implicit rationing, perhaps in response to growing resource pressures. Before the market was introduced, there had been a comparatively low rate of take-up for the renal replacement programme, but the clinical staff at the time had genuinely felt they were not turning anyone away who felt they could benefit from treatment. The director of public health, by contrast, believed they had been rationing subconsciously by operating a high eligibility threshold. Since then, new consultant staff had been appointed who wanted to treat more people and were reluctant to apply financial thresholds which they felt did not match appropriate clinical thresholds. In fact the director of contracting suggested that so far the renal physicians were expressing this reluctance much more strongly to other purchasers than to their host purchaser, 'although they do threaten, you know, they do get a bit menacing from time to time.' She added that GPs, whilst generally supportive of the need for the authority to take hard decisions, could occasionally force things out into the open that they felt were unacceptable. It was a GP who had called in the media the previous year when palliative radiotherapy was reduced.

Both the director of public health and the director of contracting believed there was a considerable amount of rationing going on before the market, but it was largely in terms of limiting supply by closing wards and reducing staffing, as the whole focus was on providers. The director of public health said, 'It was much more permissible to think in terms of the good of a hospital, of an institution, rather than the population - and I'm saying that as a public health person. So very often we would take a stance which would be at best equivocal, in terms of its effect on the population, but that was taken specifically to further the aims of X Hospital or Y Hospital...Now that still is important these days, because we want to ensure the health of our institutions and we don't want them to collapse. But on the other hand we're much less concerned about that - much, much less.'

His recollection of how resource pressures would have been handled before the introduction of the market was that a 'cost-savings programme' would have been set up, which would have meant 'we'd have weeded the gardens a bit less often, all that sort of stuff, or the hotel side of it being raided.' By the time the reforms came, he argued that most of the potential for this kind of infrastructure efficiency had been exhausted, and that to some extent therefore it was coincidence that the introduction of the market had come at the same time as health authorities had turned their attention much more towards clinical services in looking for efficiency.

However, the introduction of the market itself had (at least initially) marked a watershed in the way decisions were made and expressed.

> 'That was less because there was pressure on the budgets and increased demand...but because we were very conscious that that was what purchasing and providing was meant to do. It was meant to polarise, it was meant to create a body in the health service whose job was to say what got treated and what didn't get treated, and to do it out loud, and another body that was supposed to get on with the work it was commissioned to do. But I think we pretty quickly settled down into realising that we couldn't do that overnight. Partly it's because it's one thing to say that's what we should be doing, and another thing to get it done. It's very unpopular. And also we weren't in such terrible trouble in the beginning. We've only really had one bad year, I suspect - about 1992/93, 1993/94, it's one of those early years - when things did look absolutely calamitous. It was when medical emergency admissions had one of their what I now recognise to be cyclical hike-ups, and...hospitals were closed to admissions regularly. We had all sorts of trouble just literally keeping the basics going. So, although I think we wanted to be explicit and that was where we wanted to practice, we rapidly settled down and realised how very difficult that is.'

This casts a somewhat different light on the role of extreme resource pressures. Here, the director of public health seems to be arguing that it actually <u>stopped</u> purchasers from taking explicit decisions that they felt driven towards by the market, because the situation had got so bad they resorted to capping supply by closing to emergencies. In his view, all the international evidence seemed to suggest that extensive explicit rationing was not sustainable. He was also 'convinced that the market notion is a failed notion...It's unthinkable that we let an institution go phut because it's not doing what we want it to do, and they [providers] know that.' Ultimately, therefore, the need to keep the system going would force purchasers back into taking a more implicit, facilities-based approach to managing demand and allocating resources.

The director of contracting in Avon held similarly ambivalent views. As she pointed out, 'there have always been processes within the health service for deciding priorities, and they're perhaps slightly different now than they used to be...We used to have a bidding process where all the departments could put in proposals for the district's plan, and then there'd be a big power struggle, and some things would win and some things would lose.' She agreed that this had usually been expressed in terms of facilities and staffing, although she gave examples of how the public health department had taken an explicit approach to rationing clinical services (such as neonatology) before the market. That approach remained confidential between public health and hospital doctors, however; the difference now was that 'the whole health authority has tried to move towards those ways of thinking, it's much more widespread currency. And of course there's a gulf between ourselves and the trusts as a result of that, because we think and like to express ourselves in certain ways, and they still tend to think in in-put terms.' The development of the purchasing function had meant 'a real conversion' in the way district-level staff began to think about the detail of service provision; in the early days of the market, there were weekly health needs meetings, at which 'public health would try and educate us about health needs'.

She had mixed views about the difference made by contracting - by and large she felt it simply <u>reflected</u> changes already driven by policy (such as a decision to reduce the amount of mental health care available to people with minor psychiatric disorders in order to concentrate on the severely disturbed). In addition, 'the process of contract-setting does tend to get explicit when there have already been problems identified. So...all that explicit stuff in the plastic surgery specification came after months of discussion between public health and ourselves and the plastic surgeons. So by the time it was at the stage of being reflected in the contract, the conflict was over. Usually the contract - the process of writing a spec - isn't used as the kind of battering ram to force the issue. The issue's out and being discussed and dealt with.'

Nonetheless, the development of information and monitoring systems set up to support contracting had been an essential pre-requisite for actually knowing what the plastic surgeons were doing, and 'allows you to implement policies that you couldn't implement before'. (In practice even decisions specified in contracts are not always implemented - the director of public health said he knew varicose vein operations were still being done by local providers, despite a general ban in contracts).

The purchaser provider split in general and contracts in particular had also had the effect of making the consequences of decisions more explicit than they had been before the market. An important factor here was fundholding, although the same argument can apply to neighbouring health authorities: the fragmentation of decision-making amongst several purchasers in the same area exposed rationing in a way that would not have happened when a single district determined policy in all its local providers. 'If you actually sit in the hospital as a clinician or as a patient, you can often see that the person from the fundholding practice or the next door health authority is getting a level of service which you are not getting.'

Although the general effect of the purchaser provider split had been to make providers feel less responsible for resource allocation than they had before the market, there were exceptions. The director of contracting reported a conversation with a respected trust manager during contract negotiations two or three years ago - the timescale may be important, as it was in the earlier days of the market.

> '"Fuzzy is good", he said. "This is the way the health service has always been run - it's been fuzzy - and we've got along because we've allowed things to be fuzzy. You don't want to make things too clear." And I thought that was very interesting, because it allows just that little bit of leeway...I thought it was quite enlightened. And I think with some of our trusts at least, quite a few of them have still got the "let's not rock the boat" view about things - "we're all in this together, we're trying to do the best thing for a population that we serve and for patients, whether they come from further afield, and that we're not going to do that by having kind of high confrontation levels and things being over-explicit and all the rest of it".'

She noted that some of this was probably to do with local circumstances which meant that most of the previous district-level staff had moved over to trusts. 'The bad side of it is that they don't accept that we've got the right to do things, but the good side of it is that some of them feel a quite broad responsibility.'

This seems to be an example of an organisation dealing with complexity and uncertainty 'in an adaptive, sequential fashion...Rather than specifying the decision tree exhaustively in advance,...events are permitted to unfold and attention is restricted to only the actual rather than all possible outcomes' (Williamson, 1975, p.25). Thus it could be seen as further evidence of an unofficial drift back to hierarchy (or never having left formally left it), although it is clearly not the way decisions are always made, even in this district.

Cambridge and Huntingdon

Both the director of public health and the director of commissioning in Cambridge and Huntingdon supported the general idea that many of the easy rationalisations, supply reductions and savings on hotel services had already been exhausted by the time the market was introduced. Both believed there were still inefficiencies to be found in trusts, but they were now harder to find and were more likely to be 'in the clinical sphere, rather than in the sphere of porters and caterers and heating and lighting'. Thus to some extent the move towards explicit rationing of services was coincidental with the introduction of the market.

At one point, the director of commissioning suggested that financial pressures were no worse now than in, say, the winter of 1987/88; on the other hand he raised again the issue of providers having no incentive to 'dampen down on their operating costs', and therefore inflating resource pressures.

Both interviewees expressed reservations about the wisdom of abandoning facilities-based planning and the idea of a budget which covered all your operating costs, and said the increasing focus on activity was in many ways undesirable. This shift in focus was ascribed to the process of contracting. The director of public health was critical of the way in which trusts sometimes declare they have run out of income and will not perform elective work until the next financial year. 'So for a whole month they've got their plant lying idle, they've got their surgeons lying idle. In fact the genuine marginal cost is probably very small...and yet they feel obliged to charge £400 or £800 for that procedure.' In effect, the focus on activity creates an artificial need for rationing.

They also questioned whether the internal market could ever really function as a market because of the reality of a fixed budget, and because of political reluctance to allow trusts to go out of business. Anglia Harbours, he said, was 'one trust among 600 over five years - in business you'd expect about a third of trusts to have gone out of business by now, if there was a real market...it's peanuts, really.' In a situation where trusts could not expect to get significant resources redirected to them from other trusts because of the risk of forcing them out of business, they were confined to bidding for growth money. 'And having this huge contracting infrastructure to have debates about that much money is crazy, really crazy...We have all the disadvantages of a market, in terms of the contractual relationship, in terms of the transaction costs, but none of the advantages which are really forcing people to make radical changes because of market pressure.'

The director of public health, although vehemently supporting the claim that the market was nothing like a real market, took a slightly different view of its potential to achieve change.

> 'I think the market in itself has probably done very, very little. I think it's the longer terms trends - resource management, clinical audit, explicitness. I think what the market has done is it's been much easier to create change on two sides. I think that change management has been made easier for the trusts when they've been able to blame some third party, whilst if you were both the resource allocator and the operational manager, it's really very difficult to take a million pounds out of service x in order to invest that million in service y. Because the people in service x are your own employees and it's much easier to maintain the status quo. But if some third party forces you to do it...'

District Three

The above view was to some extent challenged by the director of commissioning in District Three when she described trust chief executives 'siding with clinicians to demonstrate why they need more money, rather than managing their clinicians to force through an agenda' of change. This was an example of a provider not taking the opportunity to make change and blame it on the purchaser, but resisting change strongly. She contrasted this with pre-market planning and rationing, which she felt was collaborative, implicit and largely focused on capping supply.

She too argued that although financial pressures had existed before the market, 'we were never acutely aware of deficits. We knew there were problems with the trusts [sic] and closing wards, but I think I would say there was probably enough flexibility in the system to actually accommodate a lot of that, so nobody minded too much.' The ability to run long waiting lists had been one factor in that flexibility, and much higher bed capacity was another. 'What we're seeing now is a combination of trusts squeezing everything out of the system in terms of reduced length of stay, increased day cases and cutting down on their estate so they're working in much tighter areas.' This was chiefly a result of perpetual efficiency targets rather than the market *per se*.

The cost of expensive new drugs was again raised as one of the main pressures on resources, but at the same time purchasers' awareness of drug costs had grown as a result of more explicit contract negotiations. In pre-market days there had been occasional debates about individual preparations, such as erythropoietin and Factor VIII, but these were seen as the exception.

The director of commissioning felt contracts had resulted in a complete shift of approach at provider level to managing financial pressures, which resulted in many issues (such as costly drugs) being thrust up the line to purchasers. 'The trusts are just getting more and more adept at charging differently, presenting information differently.' Purchasers were 'running hard to keep up with 76 contracts', but could only do so much to keep control of the situation. 'The rest of the agenda is just so big that if you take your mind off one of those items that's going to pop up as a problem, next year. And you're forever doing the old plate-spinning, and it does get quite wearing.' This illustrated how the introduction of the market polarised responsibility for managing resources and undermined collaboration between districts and providers.

Although the director of commissioning was certain contracting had created extra work, she was not convinced that this led to higher transaction costs. 'I would say in the last three months of the year, it is time that could be better spent...A fair proportion of that is actually done because people work much longer working weeks than they used to, so in terms of manpower we cope with it, but if you look at the number of hours those people are working, that's where the pressure is.'

Expectations of change under the Labour Government

By the time of the third interviews, there had been a change of government, and the opportunity was taken to discuss the potential implications of this change more extensively, although the issue had been touched on at previous interviews. It should be borne in mind that at this point no White Paper had been published, so there was no definitive account of what Labour planned to do. However, interviewees were able to comment on proposals to introduce GP commissioning throughout the NHS and the principle of abolishing the market, and what effect this might have on rationing.

Avon

The director of contracting in Avon believed GP commissioning would not simply be the market by another name, although it would probably only be 'slightly different'. She thought there was a danger that it could prove to be an 'emasculated' version of GP fundholding, with fewer budgetary incentives to make change and negotiate hard with providers. If providers were given a more guaranteed income, this would also make them less prepared to change.

On the other hand, she perceived some potential advantages in reviewing

> 'the levels at which the contracting approach makes sense. So in community services it might be mainly sensible to think of things at the practice level, and for ordinary secondary care it may be sensible to think of things at the district levels, and for the specialised stuff, the tertiary stuff, to get rid of ECRs and revert to block funding in the way that A&E departments are block funded. Which would require some formula to adjust the growth on some basis, but then you would be returning to a situation where you did what you did in the old days - providing the service until the money ran out and then you stopped. And you stopped for everybody and for most things.....Tertiary services have run away with the money in recent years, because of the incentives in the system. So it might be quite tempting to try and cap them, and that might then lead to a different approach to rationing.'

Although she did not believe there would be a wholesale return to implicit rationing of this kind, she thought the government would find this a pragmatic option in some areas, such as tertiary services.

She was uncertain what the government's approach would be to tackling inequity between districts with regard to what was rationed and what was not, even though she anticipated that this would be a major concern for the government. 'It's easy for Labour to have a go at the fundholders and issues of equity and inequity, but they just don't know what to say about the situation between districts and so they're not thinking about it.'

The chief executive agreed that the government would find it difficult to sustain the inequity of the 'total lottery approach to decision-making on priorities', but that they would also face huge political difficulties in trying to tackle it. Whatever they did, she did not see it making much difference to the amount of explicit rationing, even if it were to become more consistent nationally.

Both she and the director of public health agreed that competition had been completely unsuccessful as a way of controlling costs and that some other mechanism would be needed, but still within an internal market framework. They anticipated that funding would move from a capitation basis to budgets for providers, based on expected activity levels. The chief executive thought the government might usefully opt for greater openness of accounting by trusts, to enable commissioners to be better informed about where real financial pressures were being experienced, and where trusts were 'crying wolf'. Although these developments would make it 'feel very different for trusts', it was unlikely that there would be a return to direct management. There would still be a need for a commissioning function, although the chief executive was in no doubt that reducing the transaction costs incurred by GP fundholding in particular was sensible.

The director of public health was uncertain what effect, if any, the proposed changes would have on explicit rationing. Superficially, Labour's emphasis on open government and freedom of information might suggest an interest in more explicit rationing, but this was politically difficult territory. It was also practically difficult to realise the ideal of democratic legitimacy - 'even the most complex democracy doesn't get anywhere near it, and yet we knock ourselves senseless trying to achieve it.'

He personally felt 'terribly torn about whether to be explicit or implicit', and thought ultimately responsibility would have to be shared between managers and clinicians; certainly an approach which did not involve clinicians and implicit rationing was untenable. He argued that a retreat from even existing levels of explicit rationing was possible, because

> 'exclusions don't work....I suppose the way it would happen is they'd just decay into kind of misuse over time. So for example you'd have quite a tight specification for varicose veins: it's got to be associated with ulceration, pain and distress of a certain point on the scale, and all the rest of it, which the providers stick to. And then they find that the general surgery contract, for all sorts of other reasons, is doing quite well, they have a bit of space, they have a research fellow who needs to learn the trade as well. So they start to - off their own bat - lower the threshold, and before you know where you are...And it would not be noticeable - we wouldn't know. We might see that there was an increase in varicose veins, but we might think, "well, what the hell?"'

This perhaps raises a question mark over the extent to which even an explicit framework for implicit rationing is sustainable in the long term.

Cambridge and Huntingdon

The director of commissioning in Cambridge and Huntingdon remarked that abolishing the market yet keeping commissioning was

> 'almost the worst of both worlds - we'll keep the institutions and the expense of the internal market, but we'll remove any ability to actually use it creatively to lever change...I do think fundholding was a policy mistake, but they're leaving fundholding, but they're taking away any incentives on fundholders to be imaginative. Because a lot of the fundholders have delivered good change.'

Both interviewees in Cambridge and Huntingdon were in favour of some kind of return to provider-based budgets, perhaps routed through host health authorities but based on expected levels of activity at that provider rather than on capitation for the local population. Patients from other districts could either be funded as part of this budget, or through limited funding transfers between health authorities. Providers would then 'have to cope with all comers within their allocation'.

It was not anticipated that there would be a return to direct management, as 'one of the good things that has been a success with the purchaser provider split is exactly the separation of the operational management of trusts from the strategic organisation', and this was a benefit worth retaining. Nonetheless, the director of commissioning expressed some interest in restoring direct management, to tackle trust managers' 'freedom to mismanage the system - collectively the NHS is in debt and people have just ploughed ahead with unaffordable schemes.'

The director of public health suggested that the best way to implement GP commissioning would be to leave health authorities responsible for the planning and funding of secondary care, whilst GP localities held budgets for primary and community services. If the government opted for the 'bureaucracy of every GP commissioning secondary services', the transaction costs would be 'crazy', and worse than the already 'huge transaction costs' incurred by GP fundholding.

The preference in Cambridge and Huntingdon for fixed provider budgets was in line with their view that meso-level rationing by health authorities should remain explicit, but 'you have got to retreat from it at the micro level, because I don't think it's bearable', as the director of public health said. Fixed budgets would re-create a structure where implicit rationing by clinicians was seen as legitimate and desirable. They did not believe pursuing GP commissioning would achieve the same result.

District Three

The very high proportion of fundholding in District Three was felt to influence the way local GPs felt about possible changes: 'a lot of our fundholders are saying that fundholding's run its course. They're interested in unified budgets, they're interested in developing total purchasing pilot arrangements, but fundholding *per se* isn't actually terribly attractive any more.'

The chief executive felt they would only be interested in GP commissioning if it included fully devolved budgets, not just indicative budgets. However, this would raise concerns about accountability and maintaining financial control, and would certainly not avoid the need for rationing.

> 'In fact it'll make it worse in some ways, because you've got smaller risk pools, more localities having to make separate decisions....I can't see why GP commissioning will suddenly make the financial situation better.'

Although it could have many advantages in terms of bringing decision-making closer to local needs and creating closer links between consultants and GPs, the management costs associated with it could be substantial. He did not believe that it would result in a return to more implicit rationing; in any GP commissioning group there were always likely to be one or two practices who would be unhappy with any rationing decision agreed privately within the group, and would want to bring it to public attention.

The director of commissioning agreed with this assessment, and did not think GP commissioning could replicate the current position within fundholding, where GPs were able to make implicit individual judgements about cases they were also responsible for funding. The collective nature of GP commissioning was likely to make decisions more explicit than fundholding decisions. It could also reinforce the role of health authorities and public health staff in developing clear population-wide policies, with greater public consultation and more consistent use of effectiveness evidence.

She welcomed the principle of reducing the transaction costs of fundholding, and the action already taken by the government to reduce the administrative costs of checking every invoice by suggesting GPs use a sampling approach. However, she agreed with the chief executive that GP commissioning brought its own risk of higher management costs, and would require considerable organisational development support.

As far as providers were concerned, she was doubtful that a move to three-year contracts would have a significant impact on transaction costs, however desirable it might be. It might have a greater impact 'if trusts are instructed not to negotiate for high prices'. At a more general level, she was hopeful that the government's approach would 'send very positive messages down to trusts' not to continue setting up new services in the hope of generating more income. A stronger emphasis on mergers and partnerships between trusts would also 'cut down some of the unnecessary competition and the vying for scarce resources'.

The director of public health expressed some optimism about three-year contracts. She argued it was possible

> 'to abolish the aspects of markets that are associated with transactions for transactions' sake, and if you move from having an annual contracting round to having a three to five year comprehensive care agreement, you can refocus activity from the market bit of the relationship to the health and health care bit of the relationship.'

On the other hand, she believed people had overestimated the impact of the market, and that 'differences and tensions between the health authorities and the hospitals' had always existed, although 'admittedly there isn't the line management relationship that there was previously'. She did not envisage a return to direct management.

The director of public health had concerns both about the management costs associated with fundholding and the lack of accountability for the use of fundholder savings. She commented 'all the potential gains of fundholding can be made by strong GP commissioning, by strong GP leadership. It doesn't need fundholding to do them.' She felt that different parts of the market had been subject to 'different regulations and different constraints', and that competition had therefore never been fully realised. 'If you remove the perverse incentives to good practice that developed in the market, then you have the basis for longer term collaboration, but something where you have healthy tension, but without the threat. Because a lot of the market threats were never real threats.'

The purchaser and provider split was thus a positive influence, which could be maintained and improved if other aspects of the market were abolished. The market had 'simplified, and arguably oversimplified, a number of complex things in order to help people tackle them', but ultimately the need for a longer term, complex, collaborative process had to be recognised.

Discussion

As is evident from the previous section where interviewees compared the current situation with the pre-market NHS, there was substantial agreement amongst all interviewees that the NHS market did not really function as would be expected of a real market. The lack of a real threat that providers would go out of business meant there was little genuinely competitive behaviour; competition manifested itself simply as spiralling bids for more resources against a background of a fixed budget for the NHS. As the director of public health in Cambridge and Huntingdon commented, 'the combination of the purchaser provider split at a time of tight resources is explosive - they are two completely opposing directions.'

This accords with the views expressed by Whitehead *et al.* (1997) in relation to Sweden's reconsideration of its market reforms, that reimbursement-style funding arrangements created

> 'the wrong incentives when cost control and structural changes in supply became political priorities in the 1990s. These incentives were designed to solve "yesterday's" problems of decreasing productivity and access. But when the reforms were implemented the underlying problems - decreasing tax revenues and rising unemployment in society - were completely different and the reform solutions were counter-productive' (p.938).

This scepticism about the NHS market has two important implications for understanding interviewees' thoughts about changes under Labour. Firstly, the proposal to 'abolish the market' was seen as something of an overstatement or a misunderstanding, since there had never really been a market to abolish. Nonetheless, the principle of trying to resolve some of the problems was generally welcomed.

However, the second implication is that they were doubtful how far Labour's proposals would result in practical improvement, because it appeared that some key problematic aspects of the NHS market would remain in place, despite the intention to abolish the market. The maintenance of a split between commissioning and providing and the creation of GP commissioning are examples of this - 'the concept of commissioning requires a purchaser and a provider', as the director of public health in Cambridge and Huntingdon pointed out, 'so how can you get rid of the market and still have commissioning?'

This is not to say that interviewees felt market forces were necessarily of no value; their concern was rather that both the current situation and perhaps future plans represented a half-way house between market forces and bureaucracy, and that neither was allowed to work properly.

There was some uncertainty about the likely impact of proposed changes on rationing; some interviewees thought it might enable a return to implicit rationing to some extent, but others disagreed. The growing interest in using fixed budgets, combining implicit clinical judgements with an explicit purchaser-led framework, was already having an impact in advance of any changes to the internal market. Paradoxically, GP commissioning might actually make it harder to pursue this approach.

There was agreement that political reluctance to acknowledge rationing more openly at government level would persist, although it was just possible that Labour's interest in equity and in open government could overcome this reluctance.

Part Three

Discussion

Chapter Nine

The relationship between explicit rationing, the market and other factors

'Remember, it is not truth which matters but perceptions. And the perception, now widespread, is that the NHS is in serious difficulty once again.' (Hunter, 1997a, p.24)

The aim of this chapter is to discuss the findings of the thesis so far, with an analysis of the different factors affecting rationing. The results of this analysis are then compared against the original three hypotheses.

Discussion of the research findings

What kinds of rationing are occurring?

There is clear evidence from the case studies, and other sources, that explicit rationing is growing. This is in two senses: rationing which was already happening implicitly is becoming more explicit, and new cases of explicit rationing are occurring, particularly in the case of new technologies or drugs. An example of the former is cosmetic surgery, which used to be delayed by long waiting lists; an example of the latter is the drug beta-interferon.

It is less easy to demonstrate new cases of explicit rationing of services already provided but never previously rationed at all. It is virtually impossible to find an existing service which was not previously subject to some form of implicit rationing, however well-concealed. For example, the decision by District Three to limit IVF led to a review of other sub-fertility services, with the result that tubal surgery was to be funded on an ECR basis only. Yet it is probable that tubal surgery was already being rationed informally by individual clinicians, given what was already known about its low effectiveness. It may be that once such a service becomes the target of explicit restrictions, however, it is more strictly rationed than it was by implicit mechanisms.

In Chapter Two, three kinds of rationing questions were identified: *what* should be provided, *how much* and *for whom*? There is evidence that explicit responses are being made to all these questions.

Health authorities are deciding what should be provided, even in the absence of any national lead, and despite the government's rejection of blanket exclusions of any potentially effective treatment. Whilst authorities are generally careful to allow clinical exceptions, some services are virtually excluded altogether. The possibility that exceptions may be made is not always clear from written policies (e.g. IVF in District Three). In other cases, the authority may refuse specific funding but allow clinicians to make their own exceptions within existing resources (e.g. beta interferon in Avon). In the case of Riluzole, Avon had suggested even this possibility should not be allowed. These virtual exclusions are sometimes justified by stating that the services not purchased are ineffective and therefore do not contravene the ban. Sometimes the justification is that it is a social need rather than a health care need - again, IVF is an example. There are also services such as homeopathy, osteopathy and acupuncture which never have been part of mainstream NHS provision; this provides some justification for arguing they are not an NHS responsibility, and lack of evidence on effectiveness helps to support the decision to exclude them.

How much of a service is purchased is clearly stated in many contracts. The level of detail varies substantially - at one level, a contract giving a maximum volume of general medical episodes is a form of explicit rationing, although it still leaves generous scope for implicit 'micro rationing' at individual level (Klein *et al.*, 1996, p.8). This total may be split between sub-specialties, groups of procedures, or individual procedures. Maximum numbers of individual procedures may still allow for doctors to decide how they will be allocated, whether on a first-come first-served basis, on the basis of criteria agreed amongst clinical colleagues but not shared with purchasers, or on the basis of each clinician's personal judgement. Alternatively, purchasers may use contracts to specify the criteria they would like to see applied, which begins to take them into the territory of explicit 'for whom' rationing (see below). Sometimes how much of a service is to be bought may be expressed in financial rather than activity terms; although this may seem a more implicit way of rationing, it can be very explicit where the funding will buy only a small number of cases. This approach is found for example in gender reassignment and bone marrow transplants.

For whom: The suggestion that rationing by exclusion may be giving way to rationing by thresholds has been discussed in previous chapters. Whether this is clinician-led or purchaser-led, the net result is the development of eligibility criteria or guidelines which give explicit support to clinicians in determining who will receive the limited amount of treatment available - in other words, who meets the threshold for treatment. Again, these criteria may be based on social considerations or on evidence of clinical effectiveness, or both. Examples of the former include the marital and parental status of couples seeking IVF, and difficulty in obtaining employment as a criterion for offering tattoo removal. Examples of the latter include offering IVF, D&Cs and cancer screening only to certain age groups and the use of indicative blood cholesterol levels for prescribing lipid lowering drugs. There is also a half-way category of selection on the basis of severity of the condition, or severity of its impact on the person's life. This applies to some plastic surgery interventions, such as birth mark removal - the operation is equally effective whether the birthmark is on the face or neck, or elsewhere on the body, but the degree of psychological distress caused by the condition is judged to be less if it is not permanently visible.

There is international evidence that other countries, such as New Zealand, are also moving away from exclusions towards an emphasis on guidelines and criteria for treatment.

The result of this shift of emphasis is a renewed reliance on professionals' clinical judgement, but within a more explicit framework than before. Most interviewees felt this was a responsible way forward and struck an appropriate balance between public accountability and the need for flexibility in the face of complex individual cases.

Factors affecting explicit rationing

The following sections examine the evidence for the impact of a number of different factors. It will be evident from the text that some are clearly forces in favour of explicitness, and a few are equally clearly forces against. However, there are a number where the effect is unclear or ambivalent.

'Money is indeed the most important thing in the world' (Shaw)[1]

It is perhaps a truism to say that the fundamental backdrop to explicit rationing is **scarcity of resources.** A particularly strong influence on scarcity and rising expenditure seems to be technological advance - this featured more regularly in interviews than the ageing population or consumer demand. Interviewees were especially concerned about the impact

---

[1] See bibliography for sources of section headings

of new drugs and developments in services such as renal therapy and oncology.

Evidence from the 1997/98 contracting negotiations initially seemed to demonstrate how the temporary relief from scarcity afforded by a more generous budget allocation could give purchasers an opportunity to step back from explicit rationing, whereas those still facing serious deficits feel compelled to make more explicit choices.

Scarcity is a necessary but not sufficient condition for explicit rationing - necessary because it is unlikely doctors would want to withhold any potentially effective treatments if they could all be afforded, and nor would the public want them withheld. It is not sufficient because there has always been such scarcity and it has hitherto been dealt with mainly implicitly. Theoretically, one response to a worsening financial position could be simply to increase the extent of implicit rationing.

This may in practice no longer be possible, for three reasons:

1. There may be a limit to how much can be dealt with implicitly. The tighter money gets, the more likely it is that the implicit bubble will burst - if implicit rationing grows beyond a certain point, it may become too big to hide. Several interviewees thought it might be coincidence that there had been greater explicitness since the market, the real reason being that by the end of the 1980s they had run out of easy efficiency savings and implicit ways of capping supply.

2. Pandora's box has been opened - it becomes difficult to sustain implicit rationing once people know it is happening. Perhaps we are at one of those points identified by Calabresi and Bobbitt (1978) where the conflict between values has come to the surface and we have to make a clear shift in our allocation processes to reassure ourselves (no doubt falsely) that the problem is being addressed. As one interviewee suggested, academic fashion has played a part in exposing implicit rationing and perhaps taking society beyond the point of no return.
3. The Patient's Charter has blocked off one of the primary implicit rationing routes, namely very long waiting lists. Even if we could still pull back from the brink of explicit rationing, would there be nowhere to go?

How convincing are these arguments, particularly in the light of the final interviews? A number of problems can be identified.

Evidence from international comparisons suggests that most countries with a tax-funded health service have already reached a point where resource pressures can no longer be contained by purely implicit measures. However, it is interesting to note that healthcare expenditure in these countries varies substantially, despite the convergence on more explicit forms of rationing. It would seem that it is the pressure to spend more that is crucial, in conjunction with a perception that the taxable limits have been reached, rather than the level of expenditure *per se*.

Most explicit rationing to date has been marginal and has not generated major savings. Often it has simply formalised an existing position, such as excluding cosmetic surgery procedures for which the waiting list was so long they were unlikely ever to be performed. If lack of resources is the problem, as virtually all interviewees agreed it was, explicit rationing (at least by exclusion) does not seem to offer a solution any more than implicit mechanisms. This is borne out by the experience in countries which have tried to formulate a national list of excluded services and have had to retreat from the complexities of available evidence in most conditions.

One practical counter-argument is that, whilst explicit rationing may not be making much impact on existing expenditure, it does at least pre-empt the need to spend even more money to solve a waiting list problem or meet a growing demand for some new alternative therapy, such as expensive drugs.

At a different level, it could be argued that it does not matter whether explicit rationing is successful at dealing with scarcity. What matters is that it is clearly a response to scarcity. The general climate of financial difficulty affects the way health authorities behave; making explicit choices may give them a sense that they are dealing with the problem, even if the results are ultimately limited.

It may be that explicit rationing will become the answer to the problem in the longer term, and not just a reaction to it. There is some evidence from interviews that health authorities in this country are gradually moving into more major and controversial areas to look for financial savings.

Of crucial importance, however, is the way in which they approach this task. After the first round of interviews, lack of resources seemed to be a clear pressure in favour of explicitness. As the case studies progressed, however, a more complex picture emerged. Purchasers' reactions to worsening resource pressures have gone through distinct phases - having initially felt driven by them to greater explicitness, particularly in rationing by exclusion, they are now concerned that this is not an effective strategy. Those same resource pressures that prompted the growth in explicitness are now encouraging renewed emphasis on fixed budgets at provider level as a more successful way to manage scarcity. It could be argued there is a limit to what can be dealt with explicitly, rather than a limit to what can be dealt with implicitly.

However, this renewed interest in implicit mechanisms does not automatically mean a retreat from explicit rationing; there was considerable agreement amongst interviewees that explicit decisions would not go away altogether. Rather than a retreat from explicitness, this new phase represents a search for a better balance between implicit and explicit means of allocating resources, and the pursuit of more sophisticated approaches than straightforward exclusion. Waiting lists provide one example of this, where waiting times are being allowed to lengthen as a way of coping with financial difficulties but at the same time there are moves to develop more explicit and equitable criteria for admission to the list. There may also be greater explicitness to the public that rationing is happening and a clearer, more consistent decision-making process may be developed, whilst at the same time the actual content of individual decisions remains a matter for clinical judgement.

In summary, then, the effect of resource pressures is still towards explicitness, but towards a different kind of explicitness from the early emphasis on exclusion. *Simultaneously*, it is also an influence in favour of more implicit means of capping supply to stay within budget.

'And mighty proud I am... that I am able to have a spare bed for my friends' (Pepys)

Closely linked to the scarcity of resources argument is the loss of flexibility within hospitals; years of efficiency targets and underfunding of inflation have resulted in **reductions in the number of beds** and in lengths of stay, with increased occupancy and throughput. This capping of supply has made it more difficult to cope with unexpected fluctuations in workload. It has also made clinicians and managers more conscious of having to be selective in deciding who gets access to the reduced number of beds, particularly when coupled with the requirement to reduce long waiting times.

Mental health services have been particularly affected by bed losses. Old-style locked wards, where severely disturbed patients were confined, have been closed on humanitarian grounds, but 'we have moved exactly that provision to the private sector, where it is to be had at a price.' This accounts for a substantial element of the pressure experienced in mental health ECRs. Professional sub-specialisation in mental health and other fields has also led to much tighter definitions of which patients will or will not be accepted on particular wards.

'I do perceive here a divided duty' (Shakespeare)

The **purchaser/provider split** (leaving aside for the moment the actual process of contracting) has had two major effects: it has freed purchasers from the responsibility of managing providers, and in doing so has freed them to make hard choices which are in the interests of the population, but may conflict with the interests of providers. Such choices may also conflict with the interests of existing individual patients. The corollary to this is that providers no longer feel collectively responsible for helping to manage scarce resources and to collaborate in implicit means of control. On the contrary, the interests of individual trusts are in stimulating greater demand and allowing technology to develop, in competition with other trusts and at odds with purchasers' requirement to keep costs down. This in turn worsens the lack of resources already discussed. Thus the quasi-market may make explicit rationing more likely simply because it weakens informal cost control. This is not to say that resource pressures are solely the result of the quasi-market - they are clearly not - but to point out that it may reinforce and hasten them. This is in addition to the burden of transaction costs discussed later.

Of course, there have always been tensions between health authorities and hospitals; it has always been in the interests of hospitals and individual specialties to make a strong case for additional funding, and to give a forceful account of what may happen to patient care if the financial crisis is not resolved. However, organisational separation has both made these tensions more apparent and required them to be settled by negotiation rather than command; without direct management, it is less easy to suppress providers' dissatisfaction with the outcome of negotiations. It has resulted in a more open division within the managerial class, with provider managers supporting more explicitly the resource maximising strategies of their clinicians against the budget holders at district level.

It is also true that the fixed budget for the NHS as a whole and for each health authority has never gone away, but what changed as a result of the quasi-market is the presentation of the budget at provider level. The very direct linking of funding and activity embodied in **contracts** has encouraged providers to think less in terms of a fixed budget and more in terms of fee-for-service reimbursement, which Whitehead *et al.* (1997) have noted sits uneasily with the requirement to control costs. Theoretically the overall fixed budget may still be there, but contracts and the rhetoric of 'money following the patient' have given providers the illusion that they individually are not subject to it, and created even stronger incentives to increase activity as a way of obtaining additional funding. If funding is not forthcoming, both doctors and managers at trust level take the view that it is up to purchasers to tell them what to cut.

It is worth exploring in more detail how this differs from pre-market tensions. In the past, hospitals may have used extra activity as an argument in annual budget negotiations, but equally often discussions were couched in terms of needing more staff to deal with existing levels of workload at a more acceptable level of quality, or having to close beds to stay within budget. Bids for new theatre sessions might be made in order to alleviate waiting lists - in other words the extra workload would not be undertaken until extra funding had been promised. Now the sequence is often reversed - the extra activity is presented as a *fait accompli*, which places additional pressure on purchasers to fund it.

Of course, this was sometimes used as a tactic even in pre-market days, but arguably it stood less chance of success under a fixed budget system than a contract system. The measurement of activity was often inaccurate and was generally not broken down in detail to different sub-specialties, making it difficult to present a convincing case. Furthermore, the link between activity and price or cost was still very weak, whereas it was easy to present the cost of staffing requirements, or opening more beds or theatre sessions. Increases in activity were generally absorbed within the existing budget through increased efficiency, until something snapped and a step-wise increase in funding took place. The relation between funding and activity was certainly not linear.

Indeed, one of the major reasons for the 1991 reforms was precisely that the old system of fixed budgets allowed no flexibility to give additional financial rewards to high performing hospitals, and even made it potentially in providers' interests to avoid extra workload - safe in the knowledge that they were unlikely to lose any of their budget (especially if any underspendings were quickly spent on equipment and furnishings before the end of the financial year).

As Klein (1995) points out, the government had been steadily moving from an emphasis on inputs (beds, staffing) to an emphasis on outputs (activity), to try to get away from the relentless pressure for more money which the focus on bed and staffing numbers encouraged. (He also notes that their success in achieving this shift was limited, the debate between those favouring inputs and those preferring outputs being a 'dialogue of the deaf', p.178). The 1991 reforms represented the logical conclusion of this process - all funding was to be linked to activity, and only those producing more would get more money. Inefficient hospitals would theoretically lose out.

Evidence from the case studies suggests there is growing concern that this primary focus on activity has stimulated an unmanageable increase in workload as trusts have competed on this basis for a larger slice of the cake. Activity-based contracts and confidential accounting have also made it difficult for health authorities to establish whether the extra activity is actually costing providers as much as they claim, or whether it could in fact be absorbed within existing resources. Rationing decisions have also become activity-based, rather than facility or staffing-based. These concerns lie behind health authorities' increasing disillusion with activity as the basis for negotiations, and their renewed emphasis on providers' shared responsibility to manage limited resources; both trends are also reflected in the new White Paper (Secretary of State for Health, 1997).

It is also worth noting, as does the White Paper, that fundholding has 'artificially separated responsibility for emergency and planned care'(paragraph 5.5). This, too, has made it more difficult to manage emergency pressures implicitly by transferring resources from elective services as needed; these resources are now ring-fenced and providers have an incentive to maintain elective workload to retain essential income.

A problem with the lack of corporate feeling between purchasers and providers is that whilst it may force purchasers into taking explicit decisions, it may simultaneously undermine their implementation. There were examples in the interviews of consultants ignoring explicit exclusions or feeling unable to enforce them. Sometimes this comes full circle when the monitoring process reveals this divergence between contract and reality, and a further tightening of referral criteria is imposed or agreed. A key factor is whether clinicians agree that certain procedures ought not to be available but find it hard to say no, or whether they disagree fundamentally with the exclusion.

Although most interviewees concluded that the specification of activity inherent in contracts and the contract monitoring process had exerted some influence on explicit rationing, most felt contracts were not themselves the major reason for it. Some explicit decisions were taken mid-year in response to financial crisis and some were pursued in parallel with contract negotiations. Contracts might catch up with this later by formally recording the decisions made, but were not necessarily the mechanism by which agreement was reached.

ECRs have been at the cutting edge of explicit rationing, because they were often expensive and often dealt with controversial or rare treatments not covered by mainstream contracts. Although the initials stand for extra contractual referrals, they were officially individual cost per case contracts, and formed an integral part of the contracting system. The individual nature of ECRs made their refusal especially sensitive. The Conservative Government was planning to abolish the requirement for providers to seek approval from the relevant purchaser for elective ECRs; the stated reason was to cut transaction costs, but those interviewees who discussed it was driven by political reluctance to allow purchasers to challenge GP freedom of referral so obviously and painfully. In fact the threatened loss of financial control - which might have proved more expensive than the transaction costs of running the system - seems to have prompted the government to defer abolishing ECRs for a further year, in effect leaving it for the new Labour government to deal with. (See next chapter).

'Knowledge itself is power' (Bacon)

To ration explicitly requires more **detailed information** than was available in the 1980s. **Contract specifications** have been an undoubted influence - district planners, however well informed, were unlikely to know in detail which procedures were available in each specialty and who was getting them. There were exceptions, for instance when new specialist services were being developed - such as new cardiac units being set up to do a certain number of CABGs and angioplasties - or when there were specific waiting list initiatives on cataracts or hip replacements. However, even in these cases actually obtaining the information was often difficult. District planners certainly had no knowledge of their residents being treated outside the district, apart from occasional regional specialty cases or some long term mental illness and learning disability patients. Discovering where all local residents were being treated, however, is more a feature of the purchaser provider split generally than contracting specifically.

There may be some difference here between public health specialists and lay managers; one director of public health felt that contracting had made little difference to the level of information available, but acknowledged that having a medical background may have been important in obtaining information from hospital clinicians.

'We had evidence and no doubt' (Eliot)

Contracting is not the only source of additional information; **the Resource Management Initiative (RMI), clinical audit,** and the **increase in evidence from research into effectiveness and outcomes** have all played a non-market part, although RMI has not been as successful as the Department of Health had hoped (Cross, 1996). The requirement to write contracts may have speeded up the development of information from these other sources; it is impossible now to tell whether growing resource pressures would have meant that the process would have had to speed up anyway, if contracts had not been there to provide momentum. That they did provide momentum is beyond doubt - the lead-up to the introduction of the market was characterised by frantic activity to put the necessary IT in place. Since then, improvements in IT have enabled further sophistication in contracts.

The relationship between rationing, contracts and effectiveness evidence is equally convoluted. Effectiveness evidence has certainly provided the justification for many of the most commonly rationed procedures, but whether it is the cause of explicit rationing is another question. Certainly it is not always the decisive factor.

Part of the explosion in the availability of effectiveness information, which has been encouraged by the NHSE, can be explained as a response to the needs of purchasers, desperate for sources of independent advice to help them fulfil their new role. However, effectiveness research has not been purely driven by the existence of the market, and has been gathering pace for some time. Some of the most well-known examples such as grommets and D&Cs in women under 40 pre-date the market altogether. Having to negotiate contracts with providers, especially provider clinicians, has given purchasers additional impetus in seeking and utilising such evidence, and contracts have provided a vehicle for expressing a requirement that providers should use evidence-based medicine.

It should also be borne in mind that not all effectiveness evidence points towards limiting or excluding services - some of it may contribute to resource pressures by suggesting that more or better services should be purchased. Interviewees were unanimous in pointing to the complexity and uncertainty of much evidence and the need to assess effectiveness for individual patients. This means it is difficult to justify total exclusions in most cases, which in turn means allowing clinicians the flexibility to exercise their judgement - in effect confirming the place of implicit rationing, albeit within an explicit framework.

Despite the growth in effectiveness evidence, there is general agreement that it barely scratches the surface and that historical funding patterns continue to form the basis of for most allocation decisions. Sometimes even where there is clear evidence either to cut or purchase more of something, it may be ignored because the consequences would be politically too difficult.

The Patient's Charter and the setting of maximum waiting times have exerted an independent influence on explicit rationing. This is an interesting case of a rationing mechanism that is in itself essentially implicit (namely waiting lists) being made more explicit and also itself producing more explicitness. Although the Patient's Charter reflects a general trend towards consumerism, and has an important influence on the quality standards set in contracts, it was launched as a separate initiative independently of the NHS quasi-market reforms. Again, however, the existence of contracts and the accompanying availability of information may have fostered a greater degree of explicitness in implementing and monitoring the Patient's Charter than would otherwise have been possible.

The explicit rationing prompted by the Patient's Charter has been mostly on a one-off rather than a continuing basis: once the tattoo removals and sterilisation reversals have been excluded, it is unlikely that further low priority treatments previously rationed implicitly will be unearthed. A possible continuing influence may be in diverting resources away from other aspects of care, particularly emergencies and community services, which may then be subject to new explicit rationing - alternatively this pressure may be handled implicitly, through across-the-board cuts or simple unavailability. However, there were signs during 1996/97 that the government would not allow either to happen, especially in the run-up to an election. Lengthening waiting times were officially tolerated again in order to protect emergency services, and the Treasury made available emergency funding of £25m during 1996/97, as an advance on the settlement for 1997/98 (Cervi, 1996).

However, relaxation of Patient's Charter targets does not alter the conclusion that meeting the original targets did indeed force some rationing decisions out into the open. The development of criteria for deciding who will be admitted to a waiting list (on the basis of need) suggests the Patient's Charter is continuing to produce more explicit allocation mechanisms, to work in conjunction with more traditional implicit mechanisms.

'The old order changeth, yielding place to new' (Tennyson)

The effect of **managerialism and the creation of a culture of priority setting** was identified by some interviewees as an important influence which had its origins before the market. The whole thrust of the Griffiths reforms, which introduced general management to the NHS, was to create a clear and authoritative decision-making structure; managers were encouraged to challenge traditional clinical practice and to set clear priorities for resource allocation. Commentators from a legal background, such as Teff (1994) and Newdick (1995), give particular emphasis to the rise of managerial power, and argue it is a more influential factor in the exposure of rationing than contracts, even if, as Newdick suggests, contracts do to some extent 'force the parties to be more explicit about the services which they wish to provide and, by implication, those which they do not' (p.64). By legal standards, NHS contracts are not contracts at all, but simply another manifestation of hierarchy. Longley (1993) argues 'the actual degree of contractual freedom of purchasers and providers is ultimately determined centrally by executive decision, not by the market. The true nature of the contract mechanism in the health service therefore is not an undertaking of any commercial risk, but merely another stratagem for administrative planning' (p.48).

The managerial challenge to clinical practice was accompanied by developments such as the Resource Management Initiative, which aimed to improve knowledge of costs and outcomes and to increase awareness amongst clinical staff of the demands of managing within a fixed budget. Over the last decade, therefore, the whole NHS has become acclimatised to talk of priorities and resources, which has paved the way for explicit rationing.

Although managerial culture has thus promoted the language of priority-setting and the tools to carry it out, it has also acted as an obstacle to acknowledging it as rationing. One reason for this is that managers may bow to political pressure to keep it quiet and disguise its true nature; the increased centralisation of accountability which has accompanied the shift to managerialism has undoubtedly strengthened political control. It may also have something to do with managers' professional self-interest. May (1997) comments on the difficulty of obtaining independent, objective information about local health needs, and blames this on the current structure of the NHS:

> 'Until recently, for example, annual public health reports gave a reasonably clear warts-and-all picture of local problems. Now these have been replaced by purchasers' glossy documents. Complete with a politically correct sprinkling of multi-ethnic photographs, they link local problems in financial terms with local purchasing plans. If there's no money to tackle an issue, it won't be there. The prevailing culture means managers are reluctant to acknowledge anything exists which might be construed as "management failure"' (p.18).

The purchaser provider split may be partly to blame, but May also identifies a more general point about management culture. This amounts to implicit rationing by suppressing information about unmet need.

Managerial attitudes seem to exercise a somewhat ambivalent influence, and doubtless vary from individual to individual. It is worth remembering, for example, the view expressed by one manager that 'fuzzy is good', stressing the positive role ambiguity and implicitness can play in managing a complex and uncertain environment.

The **views of the public** were seen in some ways an obstacle to greater explicitness: 'the pressure of the bureaucracy for a quiet life', and sometimes the need to make a quick decision unhampered by the lengthy demands of consultation, meant that there was an incentive for managers to avoid bringing difficult issues out into the open. At the same time many interviewees felt public attitudes had changed and that they would no longer tolerate secrecy.

From the 1997/98 contracting round, it would seem that some managers may welcome the chance of a generous budget to avoid explicit rationing for the time being, however much they feel it is right and/or inevitable in the longer term. On the other hand, other managers are keen to maintain the momentum towards explicitness. Developments in individual districts may therefore be affected not only by the level of the budget each year but also by how committed local managers are to the principle and practice of explicitness, and their assessment of the best way of achieving it.

A Conservative Government is an organised hypocrisy (Disraeli)

There was an absolutely clear agreement that **political reluctance** to be seen to be rationing explicitly was a major obstacle. The insistence by the government that no treatments should be subject to a total exclusion unless absolutely ineffective demonstrates its unwillingness to allow the ideal of a comprehensive NHS to be openly challenged, even if rationing is accepted tacitly.

Views about whether it would be desirable to have some form of national rationing framework differed, but the unanimous view was that in reality it would never happen, because it would be politically unacceptable. This applied to all political parties, not just to the Conservatives.

There have been some signs of movement on the political front: the House of Commons Health Committee (1995) was exceptionally forthright about the need for rationing (although largely couched in terms of priority setting), and as described in Chapter Three the NHSE recently endorsed a report which includes the words 'we therefore use the term *priority setting* as a synonym for *rationing*' (Academy of Medical Royal Colleges *et al.*, 1997, p.6). It is significant that neither of these developments involved ministers directly, who may prefer to take the line that priority setting may be necessary but it is not for them to do it.

However, since the election there has been little sign of any political willingness to address the issue. The new Secretary of State and Ministers declined an invitation to attend a conference on rationing in July 1997 organised by the Rationing Agenda Group, BMA and King's Fund, amongst others. The Labour backbencher who did attend argued that rationing should not be considered until better public health and preventive measures have been instituted, and money redirected from the wasteful transaction costs of the internal market, the Private Finance Initiative and prescription fraud. She reiterated the Conservative emphasis on priority setting (which was required) rather than rationing (which was not), and on local rather than national decision-making. (Starkey, 1997)

More recently, there were suggestions (e.g. Chadda, 1997) that Labour might recommend a national priority setting commission as part of the new White Paper (Secretary of State for Health, 1997). In the event, however, the potential for the proposed National Institute for Clinical Excellence to take a lead on rationing has been considerably down-played, a question which is discussed further in the next chapter.

'Getting and spending, we lay waste our powers' (Wordsworth)

This section deals with the question of Williamson's **organisational failures framework** (Williamson, 1975) and the impact of **transaction costs**.

Although contracting necessitated a quantum leap in the amount of information available, there are still large gaps in purchasers' knowledge of what is provided and how worthwhile it is. Specifying every detail would be both very expensive and probably not possible, however much were spent, supporting Williamson's argument that bounded rationality and uncertainty require something more vague and adaptable than a rigid contract specification.

Furthermore, there is evidence of **difficulties in monitoring and enforcing compliance** with contracts and a lack of trust between purchasers and providers. Once lack of compliance is discovered, it may lead to the contract being tightened up, with even more explicit specification of exclusions. However, several interviewees felt that no matter how specific contracts became, they could not replace the need for co-operation, persuasion and influence to achieve changes in clinical practice. Very explicit contracts may even make it more difficult to achieve change, because they risk hardening positions where otherwise compromise might have been possible.

This problem of compliance fits well with Williamson's description of the results of opportunism combined with small numbers exchange. Most interviewees felt there was no real market, because local monopolies enabled providers effectively to ignore contracts if they so chose. Despite the feeling that purchasers had been liberated from worries about provider interests, their continuing sense of shared responsibility for providers' fixed infrastructure costs suggests that monopoly provision also restricts their ability to make explicit rationing decisions, let alone enforce them.

Although most interviewees felt providers used the purchaser provider split to force rationing back onto purchasers, there was evidence of a return to more implicit mechanisms in response to the perceived failure of more explicit methods. Examples include the re-inclusion of financial responsibility for tertiary ECRs within secondary providers' budgets, and the decision by plastic surgeons in Avon to move away from explicit targets for each procedure, which they themselves had originally requested.

Two districts were actively pursuing the idea of longer term contracts as a response to these concerns, and in practice all the districts regarded their contracts as expressions of longer term commitment. Their views that little would change if contracts were replaced with healthcare agreements under Labour supports the suggestion that they are already seeking ways round the market and looking to partnership or even hierarchy rather than arm's length contracts to solve their problems. Developments in Cambridge and District Three in particular seem to provide support for the theory that managing a fixed budget within a quasi-market produces so many irreconcilable tensions that there will be a drift back to less visible ways of capping demand, albeit within an explicit framework.

Interviewees' comments on the relative unimportance of contract negotiations in making allocation decisions and the desirability of longer term commitment is supported by the findings of the ESRC research programme on contracts and competition. Williams and Flynn (1997) conclude:

> 'While there may be a cosmetic element in the elimination of the word "contract", much of the work reported here, with its emphasis on the need for long-term collaboration with limited contestability, would support the strategic and organizational sense of the move towards some kind of service-level agreement with a longer time-scale. Some may see this as contracting in another guise, but the work reported here indicates that much of the important negotiation between purchasers and providers takes place outside the contracting process' (p.158).

On the other hand, plans to work in partnership with providers may be wishful thinking, given the way the purchaser provider split has decoupled providers from shared responsibility for managing scarce resources, and encouraged them to stimulate demand and develop new services. This pressure may have made explicit rationing all the more necessary, but simultaneously made it harder to achieve in practice. Any reforms which maintain a purchaser provider split will still have to confront this issue - the extent to which the new White Paper in fact maintains such a split is discussed in the final chapter.

Structural problems reported by interviewees included the **existence of generic trusts** and the sometimes linked problem of **inadequate staff numbers and expertise at district level**. Generic trusts reinforce the difficulty of getting adequate information out of providers. They also tend to retain greater control over continued implicit rationing between community and acute services, particularly where a previous district management team has gone over to the trust virtually intact. Here, perhaps, is an example of hierarchical organisation having persisted despite the quasi-market. The knock-on effect is to leave the health authority drained of skill, local knowledge and sheer numbers of people. However, the question of numbers of purchasing staff is not only a result of trust configurations - it is affected by levels of remuneration, perceived attractiveness (or otherwise) of doing the job and by government imposition of management cost cutting, in response to concerns about transaction costs.

Views on the extent of **transaction costs** were somewhat mixed. Most interviewees agreed that the major cause of high transaction costs was GP fundholding, where contracting in its fullest sense was taken furthest, albeit for a restricted range of services. There have undoubtedly been higher management and information technology costs in fundholding practices, and providers have also employed extra staff to deal with fundholding. It has been pointed out in a recent review of management costs in health authorities and GP fundholding practices led by Jenny Griffiths that the existence of two parallel systems of purchasing leads to duplication - health authorities cannot necessarily shed functions (and costs) in direct proportion to the increasing number of fundholders (Millar, 1997). Interviewees' comments that GP fundholding is expensive in its own right are to be expected, given that they all come from health authority backgrounds, but it is perhaps the persistence of parallel systems that is the problem.

There was less agreement about transaction costs incurred within health authorities. Although they all acknowledged some extra costs had arisen, the picture was complicated by management savings arising from mergers and recent cost-cutting exercises. Furthermore, it was argued transaction costs were not necessarily in themselves prohibitive, if the system was achieving better value for money as a result, for instance in preventing an increase in expensive ECRs or in negotiating better contracts. Some interviewees did see transaction costs as a more serious problem, although there was a general view that much of the extra workload had been absorbed by existing staff and that savings from abolishing the market might not be as substantial as suggested. Accountancy and information were mentioned as the most obvious areas for savings.

One director of finance pointed out that most contracts were financially a simple block arrangement with each trust, specifically to avoid 'this business of charging all over the place for things', although he argued it was still possible to be very specific about activity within block contracts. Others were less convinced - the director of public health in the same district felt 'it would just be too much work and a waste of time' to have detailed activity targets in all contracts, as opposed to a few acknowledged problem contracts.

This suggests a trade-off was being made between the benefits and costs of greater explicitness, which was not necessarily fully acknowledged. There is a point at which the costs of obtaining information, negotiating and monitoring does become an obstacle to making contracts more specific - interviewees took differing views as to whether this point had already been crossed or had been skilfully avoided. The difficulties in monitoring and enforcing contracts described above, whilst stemming partly from provider and clinician behaviour, would suggest purchasers are already foregoing some potential transaction costs; it does not represent good value for money in their eyes to pursue enforcement to the last detail.

Despite these difficulties, most interviewees believed explicit rationing would continue to spread, but in selected areas rather than across the board, and through the use of thresholds rather than exclusions. The answer to the question of whether transaction costs in a complex quasi-market prevent contracts from making clear rationing decisions is probably not a simple yes or no, but lies somewhere along a continuum, where the benefits of greater explicitness are traded off against the costs involved. Most interviewees thought transaction costs had so far been kept reasonably low, but this has not prevented definite increases in explicit contracting and rationing. One could perhaps argue that the NHS is settling into a half-way house, pursuing the full logic of the market where it suits a particular purchaser or provider to do so because of a specific problem, but behaving in a more informal way where possible - or perhaps where a problem is so intractable that market relationships are seen as unhelpful.

Most did not believe there was a significantly cheaper way of achieving the same result, because they could not envisage a return to a hierarchical and integrated organisation, even if this were desirable. The growing emphasis on fixed budgets suggests health authorities are trying to move in that direction, but the rationale is mainly to keep expenditure under control overall, rather than to limit transaction costs. Most felt Labour's proposals might make marginal reductions in transaction costs, but thought the essence of the contracting system would remain, and that GP commissioning would create new costs. It was also pointed out that even before the purchaser provider split, the NHS had been moving towards more sophisticated cost and management accountancy and this would have incurred its own transaction costs.

Conclusive objective information about the impact of the market on transaction costs is hard to come by, not least because the basis for measuring management costs has been repeatedly changed. Before the 1997 election, the King's Fund reviewed the available evidence, and concluded 'our best estimate on the increase in the total administration costs...is from about 9% of total revenue expenditure in 1988/89 to around 12% in 1994/95' (King's Fund, 1997, p.1). What proportion of this is market-related is not stated.

Let us be moral (Dickens)

There are a number of **ethical and professional concerns** which have an impact on explicit rationing, both at a collective and at an individual level.

Most interviewees did comment on the ethical dilemmas of making rationing decisions, whether explicit or implicit. At the same time, there was a consensus that explicitness was an ethical requirement for purchasers, at least in some circumstances.

It has become fashionable to characterise implicit rationing as secretive, undemocratic and paternalist, whilst explicitness is seen as honest, participative and accountable. The current wave of academic discussion of explicit rationing has been a powerful force here.

Given this background, it is hardly surprising that health authority managers feel an ethical obligation to be explicit. Some added their own reasons, including the need to protect the disadvantaged - the theory being that it would be easier to discriminate against them in secret than in public. (In fact the argument may also work the other way - explicitly quantifying the value of different people's lives and health may give lower priority to people with severe disabilities, as Maclean (1993) suggests. One interviewee thought implicit decision-making gave vulnerable people better protection from popular prejudice).

However, the belief that explicitness was necessary and morally right was not seen by interviewees as an absolute. As one interviewee said, 'I don't think it's bearable' for purchasers to try to make all decisions explicit; a distinction is needed between meso level rationing, where health authorities have a responsibility to be open about what they are doing and why, and micro-level rationing, which may be guided by explicit principles but must continue to embrace implicit clinical judgements.

This growing interest in formalising the value of a mixed approach is also reflected in academic literature. Coast (1997) is developing the concept of the utility of ignorance, which seeks to express from an economics perspective how the public may benefit from not knowing too much about rationing. A number of commentators advocate a mixture of implicit and explicit methods, depending on the situation and the level at which the decision is being taken (New and Le Grand, 1996; Klein, 1996; Mechanic, 1995) and there is increasing interest in the idea of procedural fairness (i.e. an explicit decision-making process rather than explicit decisions in every individual case). (Lenaghan, 1996).

Concerns about inequity in the NHS are a powerful motivating force for managers, although their record - and that of the NHS as a whole - in actually tackling inequalities is not particularly strong. Needless to say, inequity was a problem long before the market was introduced, but in some respects the market has made it worse or made it more evident. The purchaser provider split has made differences in level of provision between purchasers (both health authorities and GP fundholders) more evident. More explicit contracts and contract exclusions have also made people aware of what they are not getting; again, the increased availability of information is crucial in detecting inequalities. The removal of the potential to obtain treatment elsewhere has also worsened inequity, in the sense that there is now an absolute difference in access for people resident in different areas, which cannot be ameliorated by even the most assertive patient. The only way to obtain services available from another purchaser would be to move house, an option only open to the more affluent and well-informed.

At a different level, the reduction in direct management responsibilities for health authorities has led to mergers, which have exposed differences in purchasing policy. Managers rightly feel such variation is unacceptable within the boundaries of one district. Sometimes the response will be to bring services up the level in the best-provided district, but in other cases a choice will be made to restrict services to the lower level.

Several interviewees mentioned the personal experience of managers as a factor in making policy decisions, particularly in the case of infertility services. It was suggested that managers' views about whether this was a health service responsibility or not was affected by whether they themselves or a close relative or friend had ever experienced infertility. It is also clear that health authority members bring their personal values into the debate about which services should be excluded, and that this may in the end be more influential than effectiveness evidence, costs or numbers of people affected.

The NHS is made up of a variety of different groups with divergent interests and concerns; how these are played out affects resource allocation decisions as much as formal structure, if not more so. The focus of this research has been health authority managers, who may have very different values, interests and incentives to other groups, which affect their response to the question of rationing. These divergent interests may always have been present, but may have been magnified or exposed by the purchaser provider split. The purchaser provider split may also create new incentives and constraints. An analysis of what is at stake for different groups in the NHS may offer some help in understanding the impulse to be explicit - can we determine whose interests are best served by explicitness, and whose are best served by continued implicit decision-making?

Politicians have a clear electoral interest in minimising the tax burden, and minimising public awareness of rationing. They are therefore unlikely to feel comfortable with any explicit rationing, at whatever level of the organisation it takes place, and would certainly not wish to take this responsibility themselves. Faced with growing pressure towards greater explicitness, their best hope is to ensure responsibility is diffused to local health authorities, from whose perhaps unpalatable decisions they can then distance themselves.

At the opposite end of the spectrum, clinicians' interests could be served by either implicitness or explicitness, depending on the circumstances. If they themselves have to take responsibility for rationing, as a way of preserving their clinical freedom, it is likely they will prefer implicitness, to protect themselves both from the emotional costs of what they are doing and from the anger of the public and individual patients. However, as rationing becomes more extensive and forces them to deny more and more treatments, they are less likely to want to retain responsibility. The conflict with their professional, ethical imperative to do the best possible for individual patients becomes increasingly direct and painful. In this situation, they might support more explicit decision-making if responsibility for it can be transferred to purchasers, relieving them of the burden and allowing them to lay the blame elsewhere.

Provider managers' interests are similar - there are advantages for them in terms of maximising their resources and minimising responsibility if explicit rationing can be passed to purchasers. In a hierarchical situation, where responsibility is shared, providers are more likely to prefer implicit decisions. This offers some challenge to the suggestion that a quasi-market will fail to produce clear decisions; in fact it makes it easier for providers to force purchasers into making explicit choices.

For the actors considered so far, explicitness is only an advantage if responsibility can be passed to someone else. At first sight, this might be true of purchasers, too - just as providers are pushing responsibility up to them, so they might wish to pass it further up to the national/ political level. Indeed, there is evidence from the current debate that they are trying to do this to some extent. At the 'Rationing in the NHS: Time to Get Real' conference in July 1997, only seven people out of an audience of approximately 250 people voted against the motion that the government should take a lead in rationing. However, it was apparent from the conference debate, and from the case study interviews, that the emphasis is on leadership, rather than on transferring responsibility wholesale. Purchasers do not want the government to take all the decisions; as one interviewee explained, the support purchasers seek is similar to that sought by clinicians from purchasers. If purchasers have set an explicit framework, the clinician can feel '"oh well, at least I'm not copping all the responsibility myself", and it enables them to continue their own implicit rationing'. Similarly purchasers 'feel uncomfortable about it, we do it all the time, we want somebody up there to bear the responsibility alongside us'. The difference is that purchasers are seeking support to continue their explicit rationing.

It seems purchasers are unique in believing not only that explicit rationing is right and desirable, but also that they themselves should be doing it, albeit with more overt support from government and with less emphasis on exclusion. There are several possible reasons.

Firstly, it may be because purchasers know the government is more powerful than they are and would never agree to take total responsibility for explicit rationing, so they may as well say this was not what they wanted anyway. It could also be to avoid relinquishing too much responsibility, which could lead to questions over the value of their role.

Secondly, a number of interviewees pointed to the use of explicit decision-making as a resource maximising strategy, by making clear to the government and the public how tight resources are and thus increasing pressure for more. It is also resource maximising in the sense that it enables money be to diverted away from treatments of low effectiveness or low priority to higher priorities, or to meeting savings targets.

However, as the concerns about equity discussed earlier indicate, this is not the whole story. Purchasers' professional interest also lies in making a reality of their role as representatives of the local population, distinct from provider organisations. This entails several different strands: legitimising and to some extent democratising their decision-making processes; challenging traditional provider practices and specifying more clearly what they should be doing; challenging historical patterns of resource distribution between specialties, and between hospital and community services. Their whole *raison d'être* is precisely to allocate scarce resources; from a public choice perspective, their jobs depend on taking responsibility for rationing. Individual and bureaucratic pressures for the quiet life may yield to these professional interests, all of which tend towards explicitness.

Finding legitimacy for decisions includes an element of individual self-defence, too. One interviewee explained, 'us quango workers would probably feel more comfortable if we could just get on with it, but actually I'm not sure that's the case. We feel quite uncomfortable about doing things that really are down to our prejudices. And we do occasionally pull ourselves up short...particularly when people get through into the lager, you know, and they get through on a telephone to me and I think how am I going to talk to this...? I mean, this bloke'll quite often say "well, who said you could do this? I mean you're just a bloody civil servant, you wretch...Why haven't you made it clear?" And he's right - people feel that.' Being open is hard because it invites criticism from the public and the media, but it relieves the burden of bearing responsibility in private and questioning the morality of one's actions.

As noted earlier, the belief that more explicitness is ethically desirable does not mean purchasers believe all resource allocation decisions should be explicit. There is still considerable emphasis on the need for clinical decision-making - perhaps partly in self-defence, but also in genuine recognition of the inappropriateness of purchasing usurping clinical judgement. The search for an acceptable balance between the two continues.

Summary of factors affecting explicitness

From the above discussion, the factors affecting levels of explicitness in rationing can be summarised in the following groups:

Factors for
Purchaser/provider split
Reduction in bed numbers
ECRs
Patient's Charter
Academic fashion

Factors against
Political reluctance
Generic trusts
Inadequate staffing/levels of expertise at district

Unclear/ambivalent
Lack of resources
Contracting process
Information availability
Effectiveness evidence
Views of the public
Transaction costs
Managerialism/culture of priority setting
Ethical and professional concerns

How well do the research findings match with the three hypotheses?

Hypothesis one - rationing in the NHS will become more explicit as a result of the introduction of the internal market.

There is clear evidence that rationing has become more explicit and that a good part of this is related to the NHS market, both the purchaser provider split more generally and also the process of negotiating, writing and monitoring contracts. Against this, explicit rationing remains marginal, many contracts remain vague and implicit rationing still plays a major role in decision-making. This is not in itself enough to disprove the hypothesis; the development of explicit rationing may be a gradual process - if the internal market were to continue, explicit rationing might spread to the more difficult areas mentioned by some interviewees. However, it is important to note interviewees' comments on the perceived folly and virtual impossibility of being totally explicit in all contracts, and their renewed interest in fixed budgets at trust level which place some of the onus back on providers to make rationing decisions.

Although the internal market has played a role in increasing explicitness, it is not the only factor. In some cases, it does seem to have been the cause of greater explicitness. In other cases it may have simply speeded up the effect of other factors such as the use of effectiveness evidence. The use of contracts certainly affects the way explicit rationing decisions are expressed, but most interviewees felt contract negotiations were a vehicle for making such decisions, rather than the reason for them.

Hypothesis one does have some validity, but it does not tell the whole story.

Hypothesis two - the market will fail to produce clear contracts and unambiguous allocations, because of transactional, political and ethical costs

The behaviour of purchasers and providers offers some support to Williamson's 'organisational failures' framework. As already stated, many contracts remain vague and implicit allocation methods remain important. There are some signs of drifting back to hierarchy, and in some cases (e.g. generic trusts) of never having left it. However, there is still considerable tension between purchasers and providers, and the fact that they are pursuing disparate aims is blamed for much of the pressure on resources. This could be evidence that Williamson was right about the problems of 'self-interest seeking with guile' where opportunism and small numbers exchange interact. A return to hierarchy or a new emphasis on partnership might provide a solution to this, although relying on participants in the market to reach that position by themselves without external reform may be over-optimistic.

Transaction costs are cited as a reason for not going into greater detail; although a number of purchasers did not see them as a significant obstacle, this is because they are already taking action to minimise them. Block financial arrangements may co-exist with detailed activity specifications and exclusions, so keeping transaction costs low does not automatically preclude explicit rationing. However, it may limit its extent.

There has been clear political reluctance to acknowledge explicit rationing or to get involved in it. Again, however, this is not an insurmountable obstacle. The absence of a government-led priorities commission will not prevent local health authorities from taking their own action, nor will it prevent academic and professional bodies from stepping in to take a national lead. It remains to be seen whether the new National Institute for Clinical Excellence will lead ministers to take a more explicit approach in future.

Ethical obstacles to explicitness were discussed by a number of interviewees, but in general there was considerable consensus that the ethical imperative is in favour of explicit rationing in at least some circumstances, no matter how difficult and uncomfortable it is. Explicitness has become well-established in health authority purchasing decisions, and marks a clear difference from pre-market planning. It is possible that ethical costs are subject to the same kind of trade-off as transaction costs - purchasers may be avoiding the highest ethical costs by hesitating over the most controversial areas and re-emphasising the importance of clinical judgement. It is as yet too early to tell how a mixed implicit/explicit approach will work in practice, and how far purchasers will feel explicitness should go.

The evidence thus supports hypothesis two to some extent, but not fully. The fact that explicit rationing has continued to grow would suggest that the organisation is not so much failing as finding its own level of functioning, trading off transaction (and ethical) costs against the benefits of greater explicitness, and still seeking to find the appropriate boundary between meso and micro decision-making

Hypothesis three - other factors, particularly rising expenditure, outcomes research and the Patient's Charter may prove so strong that explicit rationing will continue in spite of the costs, and regardless of the structural form the NHS takes.

There is no doubt that other factors are as important, if not more so, than contracting and the purchaser provider split. Of these other factors, simple scarcity of resources has probably been the most influential, although its impact now seems to be shifting towards more implicit means of capping supply.

All the interviewees thought it was inevitable that explicit rationing would continue and increase over the next few years, regardless of structure. However, this was partly based on their perception that the current structure will remain fundamentally intact under Labour. It also reflects a changing view of the kind of explicit decisions which will be made in future.

More fundamentally, once the 'conspiracy' of implicit rationing has been exposed, it is uncertain whether it could ever be re-instated, even if the market were fully abolished. One possibility is that a substantial injection of additional funds could enable a retreat from rationing (both explicit and implicit) that would give time for memories to fade and for implicit mechanisms to reassert themselves if financial pressure increased again. Another possibility is that existing exclusions could remain in place but gradually become corrupted by an ever-widening definition of what constitutes exceptional circumstances. At the same time, further developments in explicit rationing could fail to materialise.

There are two problems with hypothesis three as it is now formulated. Firstly, it assumes hypothesis two is correct and that the driving forces behind explicit rationing are the other factors listed, not the quasi-market . As already stated, there is good evidence to believe that the market has been a vital factor, alongside other influences. If it is abandoned, those other factors may ensure explicit rationing continues, but this should not obscure the part the market has played in creating the present situation.

Secondly, it is not sensitive to the distinctions between different kinds of explicit rationing.

It seems unlikely, on the basis of the evidence presented, that explicit rationing by exclusion has much left to offer. A more likely scenario is an evolution towards a different kind of explicit rationing, based on thresholds and eligibility criteria for selecting which patients will be treated, which reconfirms equity and need as the principles guiding the allocation of NHS resources. The application of these principles and criteria to individual cases will continue to require the involvement of doctors making implicit judgements.

Conclusion

Each of the three hypotheses is supported to some extent by the evidence, but none provides a perfect fit. The NHS is a complex organisation, which operates within a complex political and ethical context. It is not surprising, therefore, that theories do not always match neatly with reality.

Any new theory will be qualified by the fact that it is impossible to say with certainty what would have happened without the introduction of the quasi-market, or what would have happened had it been allowed to develop without further national reform. Nonetheless, it is possible to reformulate elements of the three original hypotheses into an alternative theory of the relationship between explicit rationing, the quasi-market and other factors.

An alternative theory

Despite transaction costs, and despite political reluctance, the internal market generally and contracting specifically have contributed to explicit rationing in their own right. They seem also to have speeded up the operation of other factors, and may in some cases have been a catalyst, activating chains of events that would not otherwise have happened. Equally, contracting on its own might not have been so influential were it not for other pressures happening at the same time, and it is impossible to know now whether these pressures would have resulted in explicit rationing in the absence of the internal market.

It seems likely that explicit rationing will continue to increase, but there is strong evidence to suggest that it is changing course and perhaps becoming less dominant in health authorities' thinking about how to manage resource allocation. This is partly because the quasi-market is not seen to be working well at controlling expenditure. Transaction costs form part of the reason for this perception; however, a more significant factor is the divergence of interest between purchaser and provider fostered by the quasi-market, and the incentives for providers to increase activity. The shift towards a more formally recognised mix of implicit and explicit approaches recognises that some alternative to formal market relations may offer a more efficient allocative system, but this is a wider issue than simply transaction costs.

This in turn leads to the conclusion that it is too simplistic to say explicit rationing will continue to increase regardless of the structure of the NHS. At one level this may be true, but structure may have a strong bearing on the *kind* of explicit rationing which is practised. A more hierarchical structure, or one based on partnership rather than contract, may be more conducive to the use of an explicit framework of principles and criteria, which guides decisions about who should have access to what kinds of treatment, but within which there is flexibility to make implicit judgements in response to complex individual cases.

Chapter Ten

The implications of the 1997 White Paper

Introduction

'An end to rationing - we want to see an end to rationing' (Milburn, 1997)

With these words, the Minister for Health ended his interview on 'Newsnight' on the day 'The New NHS' White Paper was published (Secretary of State for Health, 1997). There could be no clearer reaffirmation of political reluctance to acknowledge rationing - indeed, the refusal to accept even implicit rationing took the political debate several steps back. The potential for the National Institute for Clinical Excellence to offer a national lead on rationing was denied.

How realistic is it to seek an end to rationing? Before the 1997 general election, the Institute for Fiscal Studies criticised the assertion by both major political parties that their low public spending plans would have no detrimental effect on the NHS.

> 'At a time when the public's demand for healthcare is inevitably rising, this will have serious implications for how well the NHS will be able to continue in its role as a comprehensive universal provider of free healthcare' (Dilnot and Johnson, 1997, p.15)

> 'Either the current shape of the welfare state will be maintained and the next government will have to spend and tax more than planned, or the welfare state will shrink and provide fewer services to fewer people' (Dilnot and Johnson, 1997, press release).

The availability of resources will remain a crucial factor. In a tax-funded system facing major resource constraints, some form of rationing is inevitable. The extent to which this is implicit or explicit is determined by a variety of factors as discussed in the previous chapter. Opinion from interviewees suggests that the tighter resource constraints become, the more likely it is that implicit mechanisms alone will no longer suffice - or at least will no longer be thought to suffice by those having to cope with them. Evidence from international experience suggests tax-funded systems have been at the forefront of explicit rationing, whether or not they have introduced quasi-market reforms. Social or private insurance systems may also get to this point eventually, but there is greater leeway in raising premiums than in raising taxes, so tax-funded systems will experience resource constraints sooner.

However, having caused an initial upsurge of interest in exclusions and detailed limitations on eligibility for treatment, those same resource pressures are now stimulating revived interest in fixed budgets framed by explicit but perhaps less interventionist principles. Again, this is supported by international evidence.

Given the complex nature of the relationship presented between explicit rationing, the market and other factors, what is the likely effect on rationing of Labour's proposals to abolish the market? It should be noted that this question is independent of whether the market has been one of the causes in the recent growth in explicit rationing. In a variety of ways already described the market has contributed to explicit rationing, both in its own right and by speeding up the effect of other influences. However, abolishing the market would not necessarily reverse the position.

To examine the potential effect of Labour's proposals, it is first necessary to outline what these proposals are, focusing particularly on those most likely to affect rationing.

The Content of the White Paper[1]

At national level, there are four important developments: one is the **National Institute for Clinical Excellence** (NICE), which will produce and disseminate 'clinical guidelines based on relevant evidence of clinical and cost-effectiveness' (7.11), in a programme of work agreed and funded by the Department of Health. The second initiative is to develop **National Service Frameworks**, drawing together the evidence on particular conditions to 'establish principles for the pattern and level of services required' (7.9). The aim is to reduce national variation in service provision. A new **Commission for Health Improvement** will monitor quality at local level, and will have powers to intervene where serious problems occur (7.13). It may also develop systematic reviews to ensure NICE

---

[1] References are to paragraph numbers.

guidelines and National Service Frameworks are properly implemented. Finally the **NHS Direct** telephone helpline will offer 24-hour information direct to patients (1.11); Pencheon (1998) suggests this may be 'part of the solution to managing ever-increasing demand' by enabling 'graduated access to the right care at the right time in the right way by the right person' (p.215).

The White Paper states that the **market will be abolished**; this is expected to reduce transaction costs substantially. However, **'the separation between the planning of hospital care and its provision'** (2.6) will be retained. Although the word 'purchasing' is carefully avoided and replaced by 'commissioning', it is arguable that this simply constitutes a renamed purchaser provider split. Contracting is explicitly rejected as a failed aspect of the market, which will be 'replaced' by longer term service agreements (9.14). Again, the extent to which these will differ from contracts as they have evolved in recent years is questionable.

Responsibility for negotiating service agreements will gradually transfer from health authorities to **'primary care groups' (PCGs)**, which are expected to cover all GPs in the country in groups of practices covering up to 100,000 patients (5.16). Fundholding will also eventually be replaced by the PCGs. PCGs will be able to develop to different levels, from advising the health authority (which negotiates agreements on its behalf) or holding a devolved budget, to independent budget holding bodies (Primary Care Trusts), accountable to the health authority. In some cases they may also take on responsibility for providing all community health services to their population (5.11).

PCGs will combine the traditionally separate budgets for prescribing, general medical services and hospital and community services in one unified and flexible budget. This funding will still be population-based, and they will retain the right to withdraw services commissioned from a provider who persistently fails to meet required standards and transfer them to an alternative provider.

Health authorities, as well as their residual negotiating role on behalf of PCGs, will have a major strategic planning function, and will be responsible for drawing up a local three-year **Health Improvement Programme (HIP)**, in consultation with local health and social care organisations (4.7). Both PCGs and trusts will have a duty to operate according to the HIP. Health authorities may organise authority-wide commissioning of certain more specialised services, in discussion with PCGs and trusts. Regional offices may perform a similar function for highly specialised services such as bone marrow transplants and medium secure psychiatric care (7.23).

Trusts will retain control over operational management matters, but there are proposals for stronger control of costs, with the publication of **national reference costs** (3.11). A new **national performance framework** will monitor a number of quality indicators, including whether trusts are ensuring 'fair access' to their services and are providing effective and appropriate services, with a reduction in 'inappropriate treatments' (8.5). Quality monitoring will be further supported by a new annual **national user survey** (8.10). This more quality-based approach to performance management will mean 'there will no longer be a narrow obsession with counting activity for the sake of it', which has had 'a perverse impact on NHS performance' and 'rewarded [trusts] for hospitalising patients even where more appropriate treatments may have been given in the community' (8.3, 8.4). Trusts will also be monitored on the development of **'clinical governance'**, an important part of which is ensuring clinical practice is evidence-based and consistent, and that adverse events, service variations and poor practice are identified and dealt with (6.12).

Significantly trusts will have a new **statutory duty of partnership**, rather than competition, requiring them to contribute to and operate in accordance with the HIP (6.6). Regional offices will have powers to intervene on behalf of the Secretary of State where monitoring of trusts reveals persistent problems or failure to abide by the HIP (7.18).

'Partnership will be dependent on sharing of information with other NHS organisations. The days of the NHS Trust acting alone without regard for others are over'(6.7). The White Paper notes that 'market-style incentives drove NHS Trusts to compete to expand their "business" irrespective of whether this reflected local NHS priorities.' Although some tried to 'overcome the limitations of the market' and work in partnership, 'most found themselves driven by these inappropriate incentives' (6.2, 6.3).

The new service agreements, backed by all local GPs, together with authority-wide or regional level commissioning of more specialised services, are expected to cover virtually all referrals. **ECRs will therefore be abolished**, 'and replaced by simplified arrangements that minimise bureaucracy and eliminate incentives to "play the market".' Instead of invoicing, adjustments will be made to PCG and health authority allocations to reflect the small number of referrals not covered by agreements. Further guidance on this is scheduled for the summer of 1998. The White Paper stresses the importance of aligning clinical and financial responsibility, 'coupling the freedom to refer with the ability to fund' (9.17).

The Implications of the White Paper

How significant is the change of language from 'purchasing' and 'contracting' to 'commissioning'? Commissioning does not carry quite the same market assumptions as the other two words, but is still compatible with a market framework. Oxford English Dictionary definitions of 'commission' offer some important clues, including: 'empower, give authority to, entrust with an office or duty'; 'authority, especially delegated authority to act in a specific capacity or manner'; 'authority to act as agent for another in trade'; 'command, instruction' ; 'charge or matter entrusted to another to perform; an order for the execution of particular work'. These definitions offer an interesting mix of hierarchical command on the one hand and trust on the other hand. In a market context, 'commission' often carries overtones of command, for example from a patron to a craftsman, rather than an exchange negotiated on an equal footing. Yet the relationship involves considerable delegation and reliance on the expert's skills to produce the required outcome.

It is no surely no accident that the word 'commissioning' conveys the same ambivalence between hierarchy and trust as does the White Paper, as it seeks an appropriate alternative to market relationships for the NHS. The concept of hierarchy may not seem immediately relevant, given the White Paper's assertion that 'there will be no return to the old centralised command and control systems of the 1970s' (2.1)[2]. Certainly there will be no return to direct management of providers by health authorities, and relationships between commissioning bodies and providers will be based on negotiated partnerships rather than command. (Many would argue that in practice the NHS has always been based on negotiation rather than command to some extent, both between levels or parts of the

---

[2] Parston and McMahon (1998) note that the Scottish White Paper (Secretary of State for Scotland, 1998) leaves commissioning with health boards and does offer the possibility to 'use this reinstated hierarchy of command to slip back into operational management' (p.213).

organisation and between professional groupings, although such negotiations have not always been cordial).

However, it is dangerous to overlook the fact that the NHS is indeed 'the organisation'; whilst commissioners and providers may not be in a strictly hierarchical relationship with each other, there may be an informal version of hierarchy at work between them. Furthermore they are both part of a bureaucratic whole, and are both very clearly in a hierarchical relationship with the NHSE.

Indeed, in a situation where commissioning and providing remain separate, the national hierarchy may be all the more important in providing the necessary co-ordination to override that separation. The White Paper certainly places strong emphasis on performance management, with greater powers at all levels to intervene if partnership fails to deliver.

Nonetheless, the emphasis on partnership is equally strong and sets the tone for the new relationship between health authorities, primary care groups and providers. This fits well with the growing interest in network theory, which suggests there is 'a third way', as the White Paper suggests (2.2), between markets and hierarchies. The theoretical debate is summarised in Thompson *et al.* (1991). Williamson himself suggests intermediate levels of uncertainty or asset specificity may lead to intermediate organisational forms, such as quasi-vertical integration, with stable long-term 'relational contracts'(Williamson, 1985, p.73). However, Bradach and Eccles (1991) argue it is oversimplistic to view markets, hierarchies and networks (or relational contracting) as mutually exclusive, and take issue with Williamson's attempts to fit networks into his model as a discrete alternative form. Rather than seeing price, authority and trust - 'which map roughly on to market, hierarchy and relational contracting' (p.279) - as opposed to each other, they emphasise 'how these control mechanisms are *combined* in empirical situations...specifically we examine *Plural Forms*, where distinct and different control mechanisms in the same organizational structure are operated simultaneously by a company to perform the same function' (p.278).

The proposals in the White Paper may constitute just such a plural form, where some form of market continues to exist, but where relationships between players are governed by partnership and trust rather than competition, and supported by hierarchy where necessary. Whilst the balance between the elements may have shifted more towards a combination of partnership and hierarchy, the claim that the market has been abolished seems premature.

As described in Chapter Eight, most interviewees thought the political rhetoric of abolition was an overstatement or a misunderstanding, since there had never really been a proper market to abolish. Nonetheless, the quasi-market as it stood had a number of serious flaws which needed to be addressed. However, as one interviewee said, ‘how can you get rid of the market and still have commissioning?’

In maintaining a split between commissioning and providing, and developing Primary Care Group commissioning, the White Paper leaves in place one of the key causes of explicit rationing identified by interviewees. There will still be divergent interests between commissioners and providers, which will have to be settled through negotiation rather than *Diktat*, and recorded in service agreements and the HIP. The shift from activity towards quality as the currency of agreements may not prevent providers from insisting commissioners should tell them explicitly what else to cut in order to maintain the desired levels of quality. There is also a remaining threat that commissioners may move their services elsewhere if providers do not meet their requirements, although this is as a last resort rather than an active encouragement of competitive business-seeking.

It seems likely, therefore, that Labour's proposals for the NHS will retain many of the pressures towards explicitness inherent in the quasi-market, as well as leaving it to contend with severe public expenditure control. It remains to be seen whether the inevitable rivalry for resources between providers will continue to be played out in the very public and explicit way encouraged by the purchaser provider split or whether a new spirit of co-operation will make a real difference to the way hard decisions are taken. It is also unclear whether and how the abolition of the ECR system will make any difference; in theory, at least, it could remove one of the more exposed areas of decision-making arising from the market.

One unknown quantity is the effect of placing GPs at the head of the commissioning process throughout the NHS. Results of the pilot fieldwork survey (Locock, 1995) suggested there was very little explicit rationing amongst GPs at that stage (summer

1994); Glennerster *et al.* (1994) and Coulter and Bradlow (1993) also reported finding little if any evidence that fund-holding was causing GPs to under-refer or under-treat for financial reasons. This may be simply because of initially advantageous funding incentives for GP fundholders. Alternatively, it may result from deeper reluctance on the part of fundholders to acknowledge rationing decisions openly, as this would place them in an extremely difficult position in face-to-face contact with individual patients.

At a more structural level, one of the very reasons why the purchaser provider split made decision-making more explicit is because it separated clinicians more sharply from the resource allocation process. Lay purchasers cannot make the same kind of implicit, informal and individualised rationing decisions that doctors can make, and have been obliged to resort to more formalised general statements of what can or cannot be purchased. GP fundholding places resource allocation under clinical control (albeit at primary rather than secondary care level) and has thus perhaps not forced decisions out into the open in the same way. GP commissioning, by realigning clinical and financial responsibility, and reintegrating the funding of elective and emergency services, may offer scope for a return to more implicit rationing .Whether or not they hold budgets could prove to be a crucial factor - if PCGs do not themselves have responsibility for managing the finances, they may feel more inclined to lay the blame for rationing decisions openly on the health authority.

However, the research cited above suggesting GP fundholders do not do much explicit rationing is now somewhat dated and the situation has moved on, no doubt partly as a result of a general tightening of finances affecting all types of purchaser. There is now some evidence that GP fundholders have begun to think more actively about the need to ration (see, for example, Crisp *et al.*, 1996), even if this is still not widespread and is largely at the level of statements of principle rather than action. Total purchasing pilots may have accelerated awareness of rationing issues, although there is insufficient empirical evidence to be certain. Even if PCGs could revert to more implicit mechanisms, there is no guarantee that they would or even that they would want to.

The government may be attracted to PCGs as 'a way of devolving responsibility and blame for unpopular rationing decisions' (Ham, 1998, p.212); however, this very fragmentation may increase the pressure on resources by creating smaller risk pools. Fragmentation may also be incompatible with the current trend towards fixed budgets for providers and the associated increased reliance on implicit mechanisms, because it will be more difficult to co-ordinate agreement and compliance amongst all purchasers. The transaction costs associated with PCGs are as yet unpredictable, but could be substantial.

Despite the political reluctance to describe the National Institute for Clinical Excellence (NICE) as a priority setting or rationing body, it is one of a number of initiatives in the White Paper aimed at reducing variation, eliminating less effective services, identifying effective services and targeting them at those individuals who will benefit most, and making the consideration of cost-effectiveness a primary concern. Re-emphasising equality of access is also a major theme. As discussed in previous chapters, the use of effectiveness and cost-effectiveness evidence in determining who should have priority in getting particular treatments is part of the rationing debate, and cannot legitimately be seen as a separate activity. The only exception to this is the elimination of procedures which are completely ineffective in all cases, and there are very few of these.

This important theme reflects two related developments already happening before the White Paper. One is the recognition that the complexity of interpreting effectiveness evidence for individual patients requires clinical judgement and cannot be specified in advance. The other is the recognition that ensuring consistent and effective treatment is targeted at those most in need cannot rely solely on the discretion of individual doctors, and needs to be guided by more explicit frameworks and criteria.

These developments were gathering pace within the NHS market; whether this was *despite* the constraints of the market structure, or *because* the market itself was already being unofficially abandoned in some respects, is to some extent irrelevant. What is more important is that the new combined model of partnership with strong monitoring and powers of intervention lends itself well to this mixed model of rationing, which places trust in clinical judgement to interpret the explicit criteria established at meso level.

This is the theory - how different groups will react in practice to the duty of partnership is a different and unpredictable matter. The preference expressed by purchasing interviewees for a return to fixed budgets at trust level and greater reliance on providers to make some of their own rationing decisions will probably remain attractive to health authorities in their new strategic role, and to the new commissioning groups - as Boyce and Lamont (1998) note, commissioning skills are thin on the ground, and PCGs will not be able to establish elaborate and detailed decision-making structures. The fixed budget approach would certainly be popular with politicians.

As already noted, however, it may not be so attractive to providers. One might speculate that a more genuinely integrated organisation will improve the likelihood of their compliance, in return for greater involvement in the development of the HIP and greater financial stability. Equally, however, trusts might feel they are net losers in the new structure, having lost much of their independence and ability to compete, and gained a more onerous and interventionist performance management system. If so, they may remain unwilling to relieve commissioners of the burden of making the hard choices about what is unaffordable.

One line of thought is that the White Paper will result in increased provider dominance, because, it is argued, greater stability for trusts can only be achieved at the cost of weakened and fragmented purchasing power and loss of contestability. (Boyce and Lamont, 1998; Light, 1998). Whether providers will use their dominance to forcing rationing decisions explicitly back to commissioners, or to re-appropriate decision-making to themselves, only time will tell.

How far they will be satisfied if commissioners, and NICE, are prepared to take responsibility for developing explicit criteria to guide provider decision-making is equally uncertain. It is likely that health authorities at least, and perhaps the new PCGs, will retain a sense of ethical obligation to develop a more explicit rationale for resource allocation decisions, even if this has to be at the level of principles and has to rely on doctors to implement it. Indeed, health authorities liberated not only from operational management but now also from the pragmatic pressures of negotiating contracts may feel even more able to take explicit decisions in the collective interest, through the HIP. If providers can be convinced that partnership will give them greater involvement and restore the emphasis on clinical judgement, whilst at the same time not leaving them to bear the whole burden of rationing, there may indeed be a way forward.

The complexity of the NHS and the plurality of interests contained within it make it highly unpredictable. It is hard enough to analyse with certainty what has already happened and why, let alone to forecast the effect of further changes, given all the different variables. As Dixon and Mays (1997) note:

> 'at best these reforms could give the service a real chance to manage scarcity better - through effective managed care. At worst they could just be the internal market with its motor removed, while perennial problems which undermine support for the NHS - haphazard rationing, financial deficits, the "winter crises", and lengthening waiting time - go unaddressed' (p.1640).

Whatever the reactions of different players to the new interests and incentives created by the White Paper, the use of effectiveness evidence to set explicit guidelines and thresholds seems set to continue irrespective of the structure of the NHS. If it does, it will allow explicit rationing to develop on a much broader scale than ever before, but in a direction which restores legitimacy to implicit judgement.

There is of course no certainty that any of these different approaches to resource allocation will provide a more efficient or more ethical solution than any of its rivals. Here we return to the relevance of Calabresi and Bobbitt's theory that cycles of different allocation processes offer a way of preserving conflicting values and limiting the destructive impact of openly confronting tragic choices - the NHS also needs to go through cycles to preserve a sense that we are doing all we can to act in a morally acceptable way. This applies both to explicit versus implicit approaches and market versus hierarchical structures. The White Paper may simply form part of this restless search for an unattainable goal.

> *'Why do approaches to tragic allocations change? Such changes are not mindlessly made; they have, in fact, represented quite rational responses preceded by discussions as rational as discussions termed rational usually are. The criticisms of the pre-existing system have described in generally accurate detail its fundamental flaws and have invoked the basic values which that system degrades. But the defenders of the pre-existing system are just as rational. They usually are penetrating in their recognition of the flaws inherent in the proposed reform. And when the reform is accepted and has become the vested method, it is eventually*

> *seen to display the very shortcomings which its critics had predicted (and to degrade those values which they sought to protect). Are these* mistakes? *If they are not, why do we move restlessly from one system which proves inadequate to another?*
>
> *'The answer is, we have come to think, that a society may limit the destructive impact of tragic choices by choosing to mix approaches over time. Endangered values are reaffirmed. The ultimate cost to other values is not immediately borne. Change itself brings two dividends, though all too often of an illusory kind we have associated with subterfuges. First, a reconceptualization of the problem arouses hope that its final price will not be exacted; the certainties of the discarded method are replaced. Second, the society is acting, and action has some palliative benefit since it too implies that necessity can somehow be evaded if only we try harder, plan better than those we followed, avoid their mistakes, and so forth. More important, because more honest, the deep knowledge that change will come again carries with it the hope that values currently degraded will not for all that be abandoned' (Calabresi and Bobbitt, 1978, p.196-7).*

Or, in the words of one interviewee:

'People keep casting around for the Holy Grail, and I don't think there is one.'

Is the search therefore fruitless? Or does the solution lie, not in any one method of resource allocation, but precisely in the *search* for a method, as Calabresi and Bobbitt imply? Perhaps change itself is the Holy Grail.

Bibliography[1]

Aaron, H. and Schwartz, W. (1984) **The Painful Prescription: Rationing Hospital Care**, Washington: Brookings Institution

Academy of Medical Royal Colleges, British Medical Association, National Association of Health Authorities and Trusts, NHS Executive (1997) **Priority Setting in the NHS: A Discussion Document**, Tunbridge Wells: South East Institute of Public Health

Appleby, J., Smith, P., Ranade, W. and Little, V. (1994) 'Monitoring Managed Competition' in Robinson, R. and Le Grand, J. (eds.) **Evaluating the NHS Reforms**, London: King's Fund Institute

Appleyard, B. (1992) 'High Cost of Dying', **The Times**, 8 July 1992 (p.14)

Arvidsson, G. (1995) 'Regulation of Planned Markets in Health Care', in Saltman, R.B. and von Otter, C. (eds.) **Implementing Planned Markets in Health Care: Balancing Social and Economic Responsibility**, Buckingham: Open University Press

Ashton, T. (1996) **Contracting for Health Services in New Zealand: Early Experiences**, Paper presented at the International Health Economics Association Inaugural Conference, Vancouver, 19-23 May 1996

Audit Commission (1996) **What the Doctor Ordered: A Study of GP Fundholders in England and Wales**, London: HMSO

Bach, S. '*Les Misérables*: What Drove French Doctors onto the Streets?', **Health Service Journal,** 29th February 1996, pp.28-29

Bagust, A. (1994) 'In Line for a New Look at Queueing', **Health Service Journal**, 30 June 1994 (p.17)

---

[1] Page references for direct quotations are given in the main body of the text

Barr, N. (1993) **The Economics of the Welfare State** (2nd edition), London: Weidenfeld and Nicholson

Bartlett, W. (1991) 'Quasi-markets and contracts: a markets and hierarchies perspective on NHS reforms', **Public Money and Management**, Autumn: 53-61

Baumol, W.J., Panzer, J. and Willig, R. (1982) **Contestable Markets**, New York: Harcourt Brace Jovanovich

Besley, T., Hall, J. and Preston, I. (1996) **Private Health Insurance and the State of the NHS**, London: Institute for Fiscal Studies

Bion, J (1995) 'Rationing Intensive Care', **British Medical Journal**, 310: 682-3

Blaxter, M. and Paterson, E. (1982) **Mothers and Daughters - a three-generational study of health attitudes and behaviour,** London: Heinemann

Bloor, K. and Freemantle, N. (1996) 'Lessons from international experience in controlling pharmaceutical expenditure III: regulating industry', **British Medical Journal**, 313: 33-5

Boucher Report (1957) **Survey of services available to the chronic sick and elderly 1954-55,** Reports on Public Health and Medical Subjects no.98 London: HMSO

Boyce, J. and Lamont, T. (1998) 'The new health authorities: moving forward, moving back?', **British Medical Journal**, 316: 215

Bradach, J.L. and Eccles, R.G. (1991) 'Price, Authority and Trust: From ideal types to Plural Forms' in Thompson, G., Frances, J., Levacic, R. and Mitchell, J. (eds.) **Markets, Hierarchies and Networks: The Coordination of Social Life**, London: Sage

British Medical Journal (1994) Comment in Minerva column, 308:1178

British Medical Journal (1996) 'New Zealand's voters focus on health', **British Medical Journal** 313:901

Brommels, M. (1995) 'Contracting and Political Boards in Planned Markets', in Saltman, R.B. and von Otter, C. (eds.) **Implementing Planned Markets in Health Care: Balancing Social and Economic Responsibility**, Buckingham: Open University Press

Byrne, P. (ed.) (1988) **Health, Rights and Resources**, King's College studies 1987-88, London: King Edward Hospital Fund for London

Cabases, J.M. (1995) 'The Guaranteed Health Care Entitlement in Spain' Paper given at the 'Hard Choices in Health Care' Conference, London, 21 Novemeber 1995.

Cabases, J.M. (1997) 'Guaranteed Health Care Entitlement in Spain', in Lenaghan, J. (ed.) **Hard Choices in Health Care**, London: BMJ Publishing Group

Calabresi, G. and Bobbitt, P. (1978) **Tragic Choices**, New York: Norton

Callahan, D. (1987) **Setting Limits: Medical Goals in an Ageing Society**, New York: Simon and Schuster

Calltorp, J. (1995) 'Sweden: No Easy Choices'. in Maxwell, R. (ed.) **Rationing Health Care**, British Medical Bulletin, 51 no.4, Edinburgh: Churchill Livingstone

Calltorp, J. (1996) 'Swedish Experience with Fixed Regional Budgets' in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Carr-Hill, R. (1991) 'Allocating Resources to Health Care: Is the QALY a Technical Solution to a Political Problem?', **International Journal of Health Services**, 21(2):351-63

Carr-Hill, R., Hardman, G., Martin, S., Peacock, S., Sheldon, T. and Smith, P. (1994) **A Formula for Distributing NHS Revenues Based on Small Area Use of Hospital Beds**, York: York University

Central Statistical Office (1994) **Social Trends 24,** London: HMSO

Cervi, B. (1996) 'HAs offered £25m winter lifeline', **Health Service Journal** (5th December 1996, p.7)

Chadda, D. (1997) 'White paper overhaul will join together GPs and trusts', **Health Service Journal** (20 November 1997, p.3)

Clarke, M., Chesworth, J. and Harrison, S. (1993) **Information and National Health Service Purchasing Authorities: the HIPPS System**, Leeds: University of Leeds Nuffield Institute for Health

Coast, J. (1997) 'Rationing within the NHS should be explicit - the case against', **British Medical Journal** 314: 1118-22

Commissie Structuur en Financiering Gezondheidszorg (1987) **Bereidheid tot Verandering**, Den Haag: DOP

Cooper, M.H. (1975) **Rationing Health Care**, London: Croom Helm

Cooper, M.H. (1995) 'Core Services and the New Zealand Health Reforms', in Maxwell, R.J. (ed.) **Rationing Health Care**, British Medical Bulletin, 51 no.4, Edinburgh: Churchill Livingstone

Coulter, A. and Bradlow, J. (1993) 'Effect of NHS Reforms on General Practitioners' Referral Patterns', **British Medical Journal** 306: 433-7

Crisp, R., Hope, T. and Ebbs, D. (1996) 'The Asbury Draft Policy on Ethical Use of Resources', **British Medical Journal** 312:1528-31

Cross, M. (1996) 'When the Chips are Down', **Health Service Journal**, 7 November 1996 (p.14)

Culyer, A.J. (1985) **Economics**, Oxford: Blackwell

Culyer, A.J. (1997) 'Maximising the Health of the Whole Community - the case for', **British Medical Journal,** 314: 667-9

Daniels, N. (1985) **Just Health Care**, Cambridge: Cambridge University Press

Department of Health (1991) **The Patient's Charter**, London: HMSO

Department of Health (Expert Advisory Group on Cancer) (1994) **A Policy Framework for Commissioning Cancer Services**, London: HMSO

Department of Health (1995) **Burdens of Paperwork in NHS Trusts and Health Authorities: Terms of Reference**, London: Department of Health

Department of Health and Social Security (1976) **Sharing Resources for Health in England: Report of the Resource Allocation Working Party**, London: HMSO

de Roo, A.A. (1995) 'Contracting and Solidarity: Market-Oriented Changes in Dutch Health Insurance Schemes', in Saltman, R.B. and von Otter, C. (eds.) **Implementing Planned Markets in Health Care: Balancing Social and Economic Responsibility**, Buckingham: Open University Press

Dilnot, A. and Johnson, P. (eds.) (1997) **Election Briefing 1997, Commentary 60**, London: Institute for Fiscal Studies

Dixon, J. and Mays, N. (1997) 'New Labour, new NHS? The White Paper spells evolution not revolution', **British Medical Journal**, 315: 1639-40

Dixon, J. and Welch, H.G. (1991) 'Priority-setting: lessons from Oregon', **Lancet**, 337:891-4

Doyal, L. (1993) 'The Role of the Public in Health Care Rationing', **Critical Public Health**, 4(1): 49-54

Doyal, L. (1997) 'Rationing within the NHS should be explicit - the case for', **British Medical Journal,** 314: 1114-8

Dunleavy, P. (1991) **Democracy, Bureaucracy and Public Choice**, Hemel Hempstead: Harvester Wheatsheaf

Dworkin, R. (1994) 'Will Clinton's Plan Be Fair?', **New York Review**, 13 January (pp.20-25)

Economist Intelligence Unit (EIU) (1993-1997) **Country Reports**:

- **Australia 1996-97, 1997-8**
- **Austria 1994-95**
- **Belgium Luxembourg 1995-96**
- **Canada 1993-94**
- **Denmark Iceland 1994-95**
- **Eire 1995-96**
- **Finland 1994-95**
- **France 1996-7**
- **Germany 1995-96**
- **Italy 1995-96**
- **Japan 1995-96**
- **Norway 1994-95**
- **Portugal 1994-95**
- **Spain 1996-97**
- **Switzerland 1994-95**
- **Netherlands 1995-96**
- **New Zealand 1996-97**

Ellis, J., Mulligan, I., Rowe, J. and Sackett, D.L. (1995) 'In-patient general medicine is evidence-based', **Lancet**, 346: 407-10

Enkin, M., Keirse, M.J.N.C. and Chalmers, I. (1989) **A Guide to Effective Care in Pregnancy and Childbirth**, Oxford: Oxford University Press

Ermisch, J. (1990) **Fewer Babies, Longer Lives,** York: Joseph Rowntree Foundation

Etzioni, A. (1991) 'Health Care Rationing: a Critical Evaluation', **Health Affairs**, 10(2): 88-95

Flynn, R. and Williams, G. (1997) **Contracting for Health: Quasi-Markets and the National Health Service**, Oxford: Oxford University Press

Friedman, M.(1962) **Capitalism and Freedom**, Chicago: University of Chicago Press

Fuchs, V.R. (1983) **How We Live**, Harvard: Harvard University Press

Gallo, P. (1997) Personal communication

Ganiats, T.G. and Kaplan, R.M. (1996) 'Priority Setting: the Oregon Example', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Glennerster, H. (1975) **Social Service Budgets and Social Policy: British and American Experience**, London: Allen and Unwin

Glennerster, H. (1992) **Paying for Welfare: the 1990s**, Hemel Hempstead: Harvester Wheatsheaf

Glennerster, H. and Le Grand, J. (1994) **The Development of Quasi-markets in Welfare Provision**, STICERD Welfare State Programme Discussion Paper WSP/102, London: London School of Economics

Glennerster, H., Matsaganis, M. and Owens, P. with Hancock, S. (1994) **Implementing GP Fundholding,** Buckingham: Open University Press

Government Committee on Choices in Health Care (1992) **Choices in Health Care,** Rijswijk, Netherlands: Ministry of Welfare, Health and Cultural Affairs

Griffiths, R. (1988) **Community Care: agenda for action**, London: HMSO

Hadorn, D. and Holmes, A. (1997) ' The New Zealand Priority Criteria Project. Part 1: Overview and Part 2: Coronary Artery Bypass Graft Surgery', **British Medical Journal** 314:131-4 (part one) and 314:135-8 (part two)

Ham, C. (1982) **Health Policy in Britain,** (1st edition) London: Macmillan

Ham, C. (1992) **Health Policy in Britain**, (3rd edition) London: Macmillan

Ham, C. (1995) 'Synthesis: What Can We Learn from International Exeprience?' in Maxwell, R. (ed.) **Rationing Health Care**, British Medical Bulletin, 51 no.4, Edinburgh: Churchill Livingstone

Ham, C. (1996) 'Contestability: a Middle Path for Health Care', **British Medical Journal**, 312: 70-71

Ham, C. (1997) 'Reforming the New Zealand Health Reforms: Big bang gives way to incrementalism as competition is abandoned', **British Medical Journal**, 314: 1844-5

Ham, C. (1998) 'Financing the NHS', **British Medical Journal**, 316: 212-3

Ham, C., Robinson, R. and Benzeval, M. (1990) **Health Check: Health Care Reforms in an International Context,** London: King's Fund Institute

Hancock, C. (1997) 'Rationing - not the Spice of life', **Health Service Journal**, 26 June 1997 (p.24)

Harley, M. (1995) 'Signs of the Times', **Health Service Journal**, 17 August 1995 (pp. 24-25)

Harris, J. (1985) **The Value of Life: an Introduction to Medical Ethics**, London: Routledge and Kegan Paul

Harris, J. (1986) 'The Survival Lottery' in Singer, P. (ed.) **Applied Ethics**, Cambridge: Cambridge University Press

Harris, J. (1987) 'QALYfying the Value of Life', **Journal of Medical Ethics**, 13: 117-123

Harris, J. (1988) 'eQALYty', in Byrne, P. (ed.) **Health, Rights and Resources**, King's College studies 1987-88, London: King Edward Hospital Fund for London

Harris, J. (1997) 'Maximising the Health of the Whole Community - the case against', **British Medical Journal** 314: 669-72

Harrison, S. (1997) 'Central government should have a greater role in rationing decisions - the case against', **British Medical Journal** 314: 970-3

Harrison, S. and Hunter, D.J. (1994) **Rationing Health Care**, London: Institute for Public Policy Research

Harrison, S. and Wistow, G. (1992) 'The Purchaser/Provider Split in English health Care: Towards Explicit Rationing?', **Policy and Politics**, 20 (2)

Harvey, I. (1996) 'Philosophical Perspectives on Priority Setting' in Coast, J., Donovan, J. and Frankel, S. (eds.) **Priority Setting: The Health Care Debate**, Chichester: Wiley

Hayek, F.A. (1944) **The Road to Serfdom**, London: Routledge and Kegan Paul

Hayek, F.A. (1976) **Law, Legislation and Liberty, Vol.2, 'The mirage of social justice'**, London: Routledge and Kegan Paul

Health Service Journal (1994a) News item, 23 June 1994 (p.6)

Health Service Journal (1994b), 'Waiting list at all-time high', news item, **Health Service Journal**, 18 August 1994 (p.6)

Health Service Journal (1996) News item, 14 March 1996 (p.8)

Health Services Management Centre (1997) **Briefing on Priority Setting no.1**, Birmingham: Health Services Management Centre, University of Birmingham

Heginbotham, C. (1992) 'Rationing', **British Medical Journal**, 304: 496-9

House of Commons, Health Committee (1994a) **Priority setting in the NHS: purchasing. HC 227. Memoranda received from Regional Health Authorities** Vol.1 London: HMSO

House of Commons, Health Committee (1994b) **Priority setting in the NHS: purchasing. HC 288-I. Minutes of evidence and appendices** Vol.1 London: HMSO

Hunter, D.J. (1993a) 'Rationing and Health Gain', **Critical Public Health**, 4(1):27-33

Hunter, D.J. (1993b) **Rationing Dilemmas in Health Care**, Birmingham: NAHAT

Hunter, D.J. (1995) 'Rationing Health Care: The Political Perspective' in Maxwell, R. J. (ed.) **Rationing Health Care**, British Medical Bulletin, 51 no. 4, Edinburgh: Churchill Livingstone

Hunter, D.J. (1997a) 'Prepare for a future shock', **Health Service Journal,** 27 March 1997, p.24

Hunter, D.J. (1997b) 'Formulating a Rational Debate', **Health Service Journal,** 31 July 1997, p.18

Hunter, H. (1996) 'New Zealand Chops', **Health Service Journal** 9 May 1996, pp.20-23

Iglehart, J. (1986) 'Canada's Health Care System', **New England Journal of Medicine** 315: 202-8, 778-84, 1623-8

Iglehart, J. (1991) 'Germany's Health Care System', **New England Journal of Medicine** 324: 503-8, 1750-6

Institute of Health Services Management (1991) Lecture on ethics and contracting given to a meeting of the Oxford Regional branch of the Institute of Health Services Management, 1 May 1991.

Johnson, P. and Falkingham, J. (1992) **Ageing and Economic Welfare**, London: Sage

Joshi, H. (ed.) (1989) **The Changing Population of Britain,** Oxford: Blackwell

King's Fund (1997) **Press Briefing: Management Costs**, London: King's Fund

Klein, R. (1989) **The Politics of the NHS** (2nd edition), London: Longman

Klein, R. (1992) 'Warning Signals from Oregon', **British Medical Journal**, 304: 1457-8

Klein, R. (1995) **The New Politics of the NHS** (3rd edition), London: Longman

Klein. R (1997) 'Defining a package of healthcare services the NHS is responsible for - the case against', **British Medical Journal** 314: 506-9

Klein, R. and Redmayne, S. (1992) **Patterns of Priorities: a Study of the Purchasing and Rationing Policies of Health Authorities**. NAHAT Research Paper no.7. Birmingham: National Association of Health Authorities and Trusts

Klein, R., Day, P. and Redmayne, S. (1996) **Managing Scarcity: Priority Setting and Rationing in the National Health Service**, Buckingham: Open University Press

Laidlaw, D.A., Bloom, P.A., Hughes, A.O., Sparrow, J.M. and Marmion, V.J. (1994) 'The sight test fee: effect on ophthalmology referrals and rate of glaucoma detection', **British Medical Journal** 309: 634-6

Le Grand, J. (1982) **The Strategy of Equality: Redistribution and the Social Services**, London: Allen and Unwin

Le Grand, J. (1991) **Equity and Choice: an Essay in Applied Philosophy**, London and New York: Harper Collins

Lenaghan, J. (1996) **Rationing and Rights in Health Care**, London:Institute for Public Policy Research

Lenaghan, J. (ed.) (1997a) **Hard Choices in Health Care**, London: BMJ Publishing Group

Lenaghan, J. (1997b) 'Central government should have a greater role in rationing decisions - the case for', **British Medical Journal** 314: 967-70

Levenson, R. (1996) 'The Cinderella Syndrome', **Health Service Journal**, 21 March 1996 (pp.28-9)

Lévi-Strauss, C. (1969) **The elementary structures of kinship**, (Les structures élémentaires de la parenté), (2nd edition) London: Eyre and Spottiswoode

Light, D.W. (1998) 'Is NHS purchasing serious? An American perspective', **British Medical Journal**, 316: 217-20

Locock, L. (1995a) **Pulling the Right Levers: How Health Authorities and GP Fundholders Aim to Get the Best from Providers**, Oxford: Anglia and Oxford Regional Health Authority

Locock, L. (1995b) **Is the NHS Market Making Rationing More Explicit?** M Phil thesis

Longley, D. (1993) **Public Law and Health Service Accountability**, Buckingham: Open University Press

Lønning, I. (1997) 'Public Involvement in Rationing - an Overseas Experience'. Paper presented at the 'Rationing in the NHS: time to get real' conference held in London on 11 July 1997

Lubitz, J. and Riley, G. (1993) 'Trends in Medicare Payments in the Last Year of Life', **New England Journal of Medicine**, 328: 1092-6

McGuire, A., Fenn, P. and Mayhew, K. (1994) **Providing Health Care**, Oxford: Oxford University Press

McKee, M. and Clarke, A. (1995) 'Guidelines, Enthusiasms, Uncertainty and the Limits of Purchasing' **British Medical Journal**, 310: 101-3

McKee, M. and Figueras, J. (1996) 'Setting Priorities: Can Britain Learn from Sweden?', **British Medical Journal**, 312:691-4

Maclean, A. (1993) **The Elimination of Morality: Reflections on Utilitarianism and Bioethics**, London: Routledge and Kegan Paul

Mahon, A., Wilkin, D. and Whitehouse, C. (1994) 'Choice of Hospital for Elective Surgery Referrals: GPs' and Patients' Views' in Robinson, R. and Le Grand, J. (eds.) **Evaluating the NHS Reforms**, London: King's Fund Institute

Malek, M. (ed.) (1994) **Setting Priorities in Health Care**, Chichester: Wiley

Manning, K. (1996) Speech to NAHAT conference, reported in **Health Service Journal**, 1 February 1996 (p.16)

Maxwell, R. (1984) 'Quality Assessment in Health', **British Medical Journal** 288:1470-2

Maxwell, R.J. (ed.) (1995) **Rationing Health Care**, British Medical Bulletin, 51 no.4, Edinburgh: Churchill Livingstone

May, A. (1997) 'Talking Turkey about PFI', **Health Service Journal** (9 January 97, p.18)

Maynard, A. (1994) 'Prioritising Health Care - Dreams and Reality' in Malek, M. (ed.) **Setting Priorities in Health Care**, Chichester: Wiley

Maynard, A. (1996a) 'Lean, Mean Rationing Machine', **Health Service Journal**, 1 February 1996 (p.21)

Maynard, A. (1996b) 'Rationing Health Care', **British Medical Journal** 313: 1499

Mechanic, D. (1995) 'Dilemmas in Rationing Health Care Services: the Case for Implicit Rationing', **British Medical Journal**, 310: 1655-9

Menzel, T. (1990) **Strong Medicine: the Ethical Rationing of Health Care**, New York: Oxford University Press

Milburn, A. (1997) Interview about the White Paper on 'Newsnight', 9th December 1997

Millar, B. (1997) 'Nine-to-five', **Health Service Journal,** 9 January 1997, pp 12-13

Miller, F. and Miller, G. (1986) 'The Painful Prescription: a Procrustean Perspective', **New England Journal of Medicine**, 314: 1383-6

Mueller, D.C. (1989) **Public Choice II**, Cambridge: Cambridge University Press

Mullen, P. (1995) **Is Health Care Rationing Really Necessary?** Discussion Paper 36, Birmingham: Health Services Management Centre, University of Birmingham

National Advisory Committee on Core Health and Disability Support Services (1994) **Core Services for 1995/96**, Wellington: National Advisory Committee on Core Health and Disability Support Services

Nestman, L.J. (1996) 'Federal and Provincial Roles in Canadian Health Care Budgets', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

New, B. and Le Grand, J. (1996) **Rationing in the NHS: Principles and Pragmatism**, London: King's Fund

New, B. (1997) 'Defining a package of healthcare services the NHS is responsible for - the case for', **British Medical Journal** 314: 503-5

New Zealand Minister of Health (1991) **Your Health and the Public Health (Green and White Paper): A Statement of Government Health Policy**, Wellington: Minister of Health

Newdick, C. (1995) **Who Should We Treat? Law, Patients and Resources in the NHS**, Oxford: Oxford University Press

NHS Centre for Reviews and Dissemination (1995), **Relationship between Volume and Quality of Health Care**, CRD report no.2, York: York University

NHSE (1995) **Improving the Effectiveness of Clinical Services**, EL(95)105

Niskanen, W. (1971) **Bureaucracy and Representative Government**, Chicago: Aldine Atherton

NLN (1990) **Nordic Statistics on Medicines 1987-89** NLN publication no.30

NOU (1987) **Retningelinjer for Prioriteringer innen Norsk Helsevesen (Guidelines for Priority Setting in the Norwegian Health Service)**, Oslo: Universiteitsforlaget

OECD (1994) **The Reform of Health Care Systems. A Review of Seventeen OECD Countries**, Health Policy Studies no.5, Paris: OECD

OECD (1995) **New Directions in Health Care Policy**, Health Policy Studies Series no.7, Paris: OECD

Office of Health Economics (1985) **Measurement of Health**, London: Office of Health Economics

Parker, R.A. (1967) 'Social Administration and Scarcity: the Problem of Rationing', **Social Work**, April 1967

Parker, R.A. (1975) 'Social Administration and Scarcity' in Butterworth, E. and Holtman, R. (eds.) **Social Welfare in Modern Britain**, London: Fontana

Parston, G. (1994) 'Power with Responsibility', **Health Service Journal**, 27 October 1994, p.25

Parston, G. and McMahon, L. (1998) 'A third way? England - yes; Scotland - maybe', **British Medical Journal**, 316: 213

Pencheon, D. (1998) 'NHS Direct: managing demand', **British Medical Journal**, 316: 215-6

Pfeiffer, K.P. (1996) 'The Possible Effects of a New Hospital Financing System in Austria', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Pughe-Morgan, P. (1996) 'Frankly, it's time to take the gloves off', letter to the **Health Service Journal**, 14th November 1996 (p.20)

Rawles, J. (1989) 'Castigating QALYs', **Journal of Medical Ethics**, 15: 143-7

Redmayne, S. (1995) **Reshaping the NHS: Strategies, Priorities and Resource Allocation**. NAHAT Research Paper no.16 Birmingham: National Association of Health Authorities and Trusts

Redmayne, S. (1996) **Small Steps, Big Goals: Purchasing Policies in the NHS**, Birmingham: National Association of Health Authorities and Trusts

Redmayne, S., Klein, R. and Day, P. (1993) **Sharing Out Resources : Purchasing and Priority Setting in the NHS**. NAHAT Research Paper no.11. Birmingham: NAHAT

Richard, S. and Schönbach, K.H. (1996) 'German Sickness Funds under Fixed Budgets', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Rawles, J. (1989) 'Castigating QALYs', **Journal of Medical Ethics**, 15: 143-7

Robbins, L.C. (1932) **An Essay on the Nature and Significance of Economic Science**, London: Macmillan

Roberts, C. *et al.* (1995) 'Rationing is a Desperate Measure', **Health Service Journal**, 12 January 1995 (p.15)

Robinson, R. and Le Grand, J. (eds.) (1994) **Evaluating the NHS Reforms**, London: King's Fund Institute

Robinson, R. and New, B. (1992) 'Health Economics and Health Economists in the NHS', **British Medical Journal** 305: 1361

Rosleff, F. and Lister, G. (1995) **European Health Care Trends: Towards Managed Care in Europe**, Coopers and Lybrand report in 'Solutions for Business' series

Sackett, D.L., Rosenberg, W.M.C., Muir Gray, J.A., Haynes, R.B. and Richardson, W.S. (1996) 'Evidence-based Medicine: what it is and what it isn't', **British Medical Journal**, 312: 71-2

Saltman, R.B. and von Otter, C. (eds.) (1995) **Implementing Planned Markets in Health Care**, Buckingham: Open University Press

Schwartz, F.W. and Busse, R. (1996) 'Fixed Budgets in the Ambulatory Care Sector: the German Experience', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) (1996) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Secretaries of State for Health, Wales, Northern Ireland and Scotland (1989) **Working for Patients**, London: HMSO

Secretary of State for Health (1996) **The National Health Service: A Service With Ambitions**, London: HMSO

Secretary of State for Health (1997) **The New NHS**, London: Department of Health

Secretary of State for Scotland (1997) **Designed to Care**, Edinburgh: The Stationery Office

Sheldon, A. (1995) 'Keep Taking the Pill', **Health Service Journal**, 16 November 1995 (pp.14-15)

Singer, P. (ed.) (1986) **Applied Ethics**, Cambridge: Cambridge University Press

Smith, P. and Peacock, S. (1995) **The Resource allocation Consequences of the New NHS Needs Formula**, Discussion paper 134, Centre for Health Economics, York: York University

Starkey, P. (1997) Speech opposing the motion 'This house believes that government has the obligation to take the lead in health care rationing' at the 'Rationing in the NHS: time to get real' debate held in London on 10 July 1997

Swedish Parliamentary Priorities Commission (1995) **Priorities in Health Care: Ethics, Economy, Implementation (Vårdens svåra val: slutbetänkande av Prioriteringsutredningen)**, Stockholm: Ministry of Health and Social Affairs, SOU 1995:5

Taylor-Gooby, P. (1991) **Social Change, Social Welfare and Social Science**, Hemel Hempstead: Harvester Wheatsheaf

Teff, H. (1994) **Reasonable Care: Legal Perspectives on the Doctor-Patient Relationship**, Oxford: Clarendon Press

Thane, P. (1989) 'Old Age: Burden of Benefit?' in Joshi, H. (ed.) **The Changing Population of Britain,** Oxford: Blackwell

Thompson, G., Frances, J., Levacic, R. and Mitchell, J. (1991) (eds.) **Markets, Hierarchies and Networks: The Coordination of Social Life**, London: Sage

Tooley, M. (1972) 'Abortion and infanticide', **Philosophy and Public Affairs** 2: 37-65

Townsend, P. and Davidson, N. (1982) **Inequalities in Health: The Black Report**, Harmondsworth: Penguin

Van de Ven, W.P.M.M. (1995) 'Choices in Health Care: a Contribution from the Netherlands', in Maxwell, R. (ed.) **Rationing Health Care**, British Medical Bulletin, 51 no.4, Edinburgh: Churchill Livingstone

Van de Ven, W.P.M.M. (1996) 'Regulation, Competition and Equity: With or Without a Fixed Budget?', in Schwartz, F.W., Glennerster, H. and Saltman, R.B. (eds.) **Fixing Health Budgets: Experience from Europe and North America**, Chichester: Wiley

Weijer, C. and Elliott, C. (1995) 'Pulling the Plug on Futility', **British Medical Journal**, 310: 683-4

Whitehead, M., Gustafsson, R. Å. and Diderichsen, F. (1997) 'Why is Sweden rethinking its NHS style reforms?', **British Medical Journal**, 315: 935-9

Williams, A. (1985) 'Economics of coronary artery by-pass grafting', **British Medical Journal**, 291:326-9

Williams, A. (1997) 'Rationing Health Care by Age - the Case For', **British Medical Journal**, 314: 820-2

Williams, M.H. and Frankel, S.J. (1993) 'The Myth of Infinite Demand', **Critical Public Health**, 4(1): 13-18

Williamson, O.E. (1975) **Markets and Hierarchies: analysis and anti-trust implications**, New York: Free Press

Williamson, O.E. (1985) **The Economic Institutions of Capitalism**, New York: Free Press

Wing, A.J. (1983) 'Why Don't the British Treat More Patients with Kidney Failure?', **British Medical Journal**, 287: 1157-8

Wordsworth, S., Donaldson, C. and Scott, A. (1996) **Can We Afford the NHS?** London: Institute for Public Policy Research

Working Group on Prioritisation in Health Care (1995) **From Values to Choices: Report of the Working Group on Prioritisation in Health Care (Arvoista valintoihin Sosiaala-ja terveydenhuollon priorisointiryhmän raportti)**, Helsinki: National Research and Development Centre for Welfare and Health (STAKES)

World Bank (1993) **World Development Report 1993**, New York: Oxford University Press

World Health Organisation Regional Office for Europe (1996) **European Health Care Reforms: Analysis of Current Strategies**, Copenhagen: WHO Regional Office for Europe

Sources for quotations used as sub-headings in Chapter Nine

Francis Bacon, *'Knowledge itself is power' ('nam et ipsa scientia potestas est')*, from **Religious Meditations (Of Heresies)**

Charles Dickens, *'Let us be moral. Let us contemplate existence'*, Mr Pecksniff in **Martin Chuzzlewit**

Benjamin Disraeli, *'A Conservative Government is an organised hypocrisy'*, House of Commons speech 17 March 1845 (period of opposition within party to Prime Minister Robert Peel)

T.S. Eliot, *'Were we led all that way for*
Birth or death? There was a
Birth, certainly. We had evidence
And no doubt.' **Journey of the Magi**

John Milton, *'They also servewho only stand and wait'*, **Sonnet xvi, On His Blindness**

Samuel Pepys, *'And mighty proud I am (and ought to be thankful to God Almighty) that I am able to have a spare bed for my friends'*, **Diary**, 8th August 1666

William Shakespeare, *'I do perceive here a divided duty'*, Desdemona in **Othello**, Act I scene iii (l.181), torn between duty to her father and her husband

George Bernard Shaw, *'Money is indeed the most important thing in the world; and all sound and successful personal and national morality should have this fact for its basis'*, **The Irrational Knot** (preface)

Alfred Tennyson, *'And slowly answer'd Arthur from the barge:*
"The old order changeth, yielding place to new"'.
The Passing of Arthur (l.407)

William Wordsworth, *'The world is too much with us; late and soon,*
Getting and spending, we lay waste our powers.
Little we see in Nature that is ours'.
The World is Too Much With Us