

# THINKING ABOUT RATIONING

Rudolf Klein

Jo Maybin

---

The King's Fund seeks to understand how the health system in England can be improved. Using that insight, we help to shape policy, transform services and bring about behaviour change. Our work includes research, analysis, leadership development and service improvement. We also offer a wide range of resources to help everyone working in health to share knowledge, learning and ideas.

---

© The King's Fund 2012

First published 2012 by The King's Fund

Charity registration number: 1126980

All rights reserved, including the right of reproduction in whole or in part in any form

ISBN: 978 1 85717 636 0

A catalogue record for this publication is available from the British Library

Available from:

The King's Fund  
11-13 Cavendish Square

London W1G 0AN

Tel: 020 7307 2591

Fax: 020 7307 2801

Email: [publications@kingsfund.org.uk](mailto:publications@kingsfund.org.uk)

[www.kingsfund.org.uk/publications](http://www.kingsfund.org.uk/publications)

Edited by Jane Edmonds

Typeset by Grasshopper Design Company

Printed in the UK by The King's Fund

---

# Contents

---

<b>About the authors</b>	iv
<b>Acknowledgements</b>	v
<b>Executive summary</b>	vi
<b>Introduction</b>	1
<b>Values, criteria and actors</b>	4
<b>Competing criteria and values</b>	6
The technocratic model	10
The commissioner's tool box	13
<b>From theory to practice: how commissioners ration</b>	15
Increasing waiting times	15
Limiting volumes of treatments	16
Rationing by selection: eligibility thresholds	17
Rationing by deflection: risks to NHS continuing care?	20
Exclusion lists	21
Exceptional funding requests	25
<b>The legal framework for policy decisions</b>	26
Legal requirements on commissioners' decision-making processes	27
<b>Decoding provider rationing</b>	30
The secret garden of clinical rationing	33
<b>Concluding discussion</b>	35
Is there a case for postcode rationing?	36
Accountability for reasonableness: but to whom?	40
Issues for future debate	42
<b>References</b>	44

---

## About the authors

---

**Rudolf Klein**, Emeritus Professor of Social Policy at the University of Bath, is a visiting fellow at The King's Fund. He is also a visiting professor at the London School of Economics and at the London School of Hygiene & Tropical Medicine. Born in Prague, he was educated in the UK and has an MA in History from the University of Oxford. He spent the first half of his career as a journalist with the *Evening Standard* and *The Observer*. On switching careers, he became Professor of Social Policy at the University of Bath in 1978. His awards include an Honorary Doctorate of Arts (Oxford Brookes University) and a CBE. In 2006 he was elected Fellow of the British Academy and of the Academy of Medicine.

Rudolf has published many articles in journals ranging from the *British Journal of Political Science* to the *British Medical Journal*. He is author of *The New Politics of the NHS*, the sixth edition of which was published in 2010, as well as of books on rationing, accountability, patient complaints and performance measurement.

**Jo Maybin** is a fellow in health policy at The King's Fund. Jo joined the Fund in 2006 having previously worked as a social affairs analyst for BBC News. In 2010 she co-edited a major review of NHS performance under the last Labour government, and in 2011 she led an analysis of the impact of the coalition government's NHS reform programme on accountability in the health service.

Jo has degrees in social and political sciences from Cambridge University and the LSE and is on sabbatical leave until autumn 2012 to complete her PhD at the University of Edinburgh. Her doctoral research examines how civil servants in England's Department of Health use knowledge in their work.

---

# Acknowledgements

---

In preparing this paper we benefited from helpful conversations with clinicians and NHS managers including Muaaze Ahmad, Peter Brambleby, Caroline Clarke, Ros Crooks, Kim Fleming, Tim Hiles, Celia Ingham-Clark, Ann Jacklin, Stephen Powis and Daron Smith. We are grateful to Amy Ford, University of Manchester, and to Sophie Coronini for sharing with us research in progress. We would also like to thank our reviewers for their insightful comments: Albert Weale of University College London, Suzanne Robinson of the University of Birmingham, and our ever-helpful colleagues at The King's Fund: John Appleby, Anna Dixon and Chris Ham. Any errors or omissions are our own.

---

# Executive summary

---

- Rationing has been a fact of life in the NHS in England since its launch. It is certain to become an ever more dominant issue in the hard times ahead. Accordingly, the aim of this discussion paper is to provide a guide through the intellectual maze of rationing: to analyse and clarify the puzzles and controversies that beset the issue, and to identify what is and isn't known about how it works out in practice.
- Rationing takes many forms. It could be a headline-catching refusal by commissioners to fund particular treatments (rationing by denial), or to fund them only for particular patients (rationing by selection). Alternatively, patients may be put off a service by obstacles to access (rationing by deterrence), or shunted off to another institution (rationing by deflection). Rationing by dilution refers to a situation where a service may continue to be offered but its quality declines as cuts are made to staff numbers, equipment and so on. This form of rationing may be the least visible, but it may also be the most pervasive.
- Decisions that determine who will eventually get what are taken at all levels within the system: central government determines the overall budget for the NHS, commissioners and providers decide between competing priorities when allocating funding, and clinicians decide how to allocate their limited time and resources.
- The criteria used in making such decisions vary with the level at which they are made. In the making of decisions about allocating resources between different services and interventions (priority-setting), utilitarian principles tend to be used: the aim is to maximise the impact for the whole population. In making decisions about which patients should be treated and how (rationing), more individualist or social, ethical criteria are invoked. The tension between population-based and individual-orientated criteria lies at the heart of much of the controversy in this area.
- Given that conflicting values are involved, there is now much emphasis on establishing fair processes by which decisions ought to be taken (rather than on the criteria for decisions themselves). The Accountability for Reasonableness Framework specifies that decisions should be public, based on relevant reasons, and open to challenge or revision. They should also be subject to regulation to ensure that the three other conditions are met. This framework is echoed in much of the legal and policy guidance to commissioners.

- The National Institute for Health and Clinical Excellence (NICE) represents an attempt both to depoliticise decisions and to base them on expert evidence. However, its reliance on Quality Adjusted Life Years (QALYs) continues to be conceptually and methodologically controversial, making NICE vulnerable to political pressure. NICE acknowledges that social value judgements have to inform its decisions, but it is not clear how they do so.
- The NHS Constitution enshrines the right of patients ‘to drugs and treatments that have been recommended by NICE’. But while this limits the freedom of commissioners – primary care trusts (PCTs) in the past, clinical commissioning groups in future – these still have a battery of strategies available to them. Among them are: introducing minimum waiting times for treatment (rationing by delay), shifting responsibility for continuing care funding onto local authorities and individual patients (rationing by deflection), increasing the range of treatments for which patients must meet certain commissioner-set eligibility criteria in order for GPs to be allowed to refer them to a specialist (rationing by selection), and increasing the lists of treatments that the commissioner decides to exclude from NHS funding on the grounds of their limited cost-effectiveness or cosmetic nature (rationing by denial).
- PCTs have restricted funding – whether by outright exclusion or by raising the treatment threshold – for some 250 procedures. But despite the headline-grabbing nature of these decisions, the Audit Commission estimated that if all commissioners were as rigorous as the top 25 per cent in terms of limiting access, there could be a national saving of £441 million – a drop in the ocean of the estimated £15–20 billion of efficiency savings needed by the NHS.
- No exclusion can be absolute. Commissioners must allow for ‘exceptions’: ie, they must provide patients with an opportunity to argue that their case deserves special consideration. Some 26,000 such applications for exceptional funding were made in the most recent year for which statistics are available. The outcome of such applications varied greatly: some PCTs approved all applications, others rejected all.
- It is not only the outcomes of the decision-making process that differ; so, too, do the processes by which they are reached. There are no clear and binding guidelines about how ‘exceptionality’ ought to be interpreted or what criteria should be used in decision-making. Commissioners’ decision-making processes on funding are, however, subject to a range of legal requirements. The NHS Constitution sets out that their decisions must be ‘rational’ and meet some limited requirements around transparency, and because commissioners are public bodies, their decisions may be subject to appeal through judicial review if the decision-making process is found to be improper or outside of their legal powers.

- While these requirements should ensure that priority-setting and rationing decisions by commissioners are explicit and in the public domain, the picture is much more opaque in the case of providers. We know least about precisely those areas – hospitals – where some of the most critical decisions about the allocation of resources to patients are made. Knowing that there are staff cuts and budget pressures does not in itself tell us about their impact on patients.
- Outright denial of particular services or treatments within hospitals still seems to be relatively rare, though pressure to reduce follow-up appointments, consultant-to-consultant referrals and orders for diagnostic tests considered to be of marginal benefit is on the rise. In all this, managerial strategy is to create frameworks within which such decisions are taken by clinicians. Seemingly minor changes at the edges of services may, however, have a cumulative effect. They are also precisely the kinds of changes that may result in rationing by dilution: a general diminution in the quality of care offered to patients.
- Most importantly, we know very little about ‘bedside rationing’ – the effects of decisions by clinicians determining who gets what. International studies have found that most strategies employed by doctors were implicit, but factors influencing their decisions included a patient’s age (those over 85 were less likely to be referred for tests), the patient’s ability to exercise pressure, and their relative contribution to society.
- We conclude by examining some of the critical issues that should feature on the future agenda of debate about priority-setting and rationing. The first is whether ‘postcode rationing’ is acceptable. Variations in decisions by commissioners about what services, interventions or drugs to provide – or not – are widely seen and condemned as incompatible with the whole notion of a national health care system. It is argued that geography should not determine the care that people get.
- But geography does determine – in many respects – what people get. Large, difficult-to-explain and impossible-to-justify variations in access to care have characterised the NHS since the start. Given the historical legacy of differences in provision, postcode prioritising by commissioners to rebalance local health economies is both inevitable and right.
- There is also reason for scepticism about a much-touted solution to the problem of postcode prioritising and rationing: introducing a national menu. Instead of allowing individual commissioners to decide on what not to provide, it is argued, there should be a national list. But as the experience of PCTs suggests, the savings tend to be marginal, while the bureaucratic costs of dealing with ‘exceptional’ cases can be large. Moreover, commissioners

would almost certainly have to be selective in their use of the national menu to reflect local circumstances, so uniformity would still remain an illusory policy goal.

- When translating priorities about resources into rationing decisions about individual patients, however, a different argument applies. Here there is a strong case for having a national set of criteria to determine whether any given patient should count as an 'exception'. But even this may not be able to prevent seemingly anomalous outcomes, given the heterogeneity of patients, the uncertainty of medicine and the fallibility of doctors.
- The introduction of clinical commissioning groups accurately reflects public trust in doctors (rather than managers or politicians) to take rationing decisions and act as the patient's agent. However, this trust may be undermined if the public suspects that decisions are being taken for financial rather than clinical reasons – the double-agent dilemma.
- While much is made of quantitative differences in access to care, much less attention has been paid to qualitative differences. Yet there is much evidence of poor care being provided in poor environments to older people in particular (though not exclusively). Although the relationship between the level of resources and the quality of care is not straightforward, this would seem to be an example of rationing by dilution.
- The case of older people points to an apparent inconsistency in the principles used in prioritising and rationing. In the case of drugs for terminal cancer care, NICE (pushed by the government) decided that special value should be attached to the last few months of life. Should not the same extra value be attached to the last few months of life for older people in hospital? Why should drugs, as distinct from care, be privileged?
- Debates about priorities and rationing place great emphasis on accountability, but exactly who should call commissioners and providers to account is less clear. The courts, clearly, have a role. But they cannot routinely scrutinise decisions, far less outcomes. Here there could be an important role for HealthWatch, particularly if its link with the Care Quality Commission could be used to strengthen the analytic ability of local agencies. The challenge, for both HealthWatch and local authority scrutiny committees, will be to sound the alarm when efficiency savings become a euphemism for rationing by dilution.



---

# Introduction

---

Decisions about how to allocate and use resources are taken 24 hours a day, and at every level, in every health care system. Governments strive to limit the seemingly unappeasable appetite for more funds for health care; clinicians have to decide how much time to spend with individual patients or how many tests to order. Not all rationing is necessarily funding related: hard choices have to be made, too, about how to allocate kidneys and other organs for transplants. The starting point of this paper is, therefore, that decisions about what to provide – and more importantly what not to provide – are inevitable in any health care system. The result is rationing – a word that once encapsulated the notion of an entitlement to a fair share of resources in conditions of scarcity but that now is often referred to with pursed lips to imply a deficit between expectations and what is available: a deprivation of possible benefits.

In the case of the NHS, rationing is a fact of life, as it has been ever since its foundation and as it is in health care systems throughout the world (Coulter and Ham 2000). The form may vary between systems and over time. If waiting lists symbolised rationing in the NHS for many decades, lack of health care insurance coverage characterised rationing in the United States. And while waiting lists are no longer the dominating symbol of the NHS's inability to match supply to demand, other forms of rationing, such as denying funding for some treatments, are increasingly commanding the headlines. Nor is rationing evidence of original sin in publicly funded systems like the NHS. Private health care insurance policies carry lists of exclusions specifying what is not covered: lists that would have ministerial heads crowning lamp-posts in Whitehall if governments sought to introduce anything similar in the NHS.

But if there is nothing new about the fact of rationing in its many forms, the context in England has changed and is continuing to do so. Most obviously, the NHS faces a bleak financial future. The expectation that the NHS can achieve £20 billion of efficiency savings, as distinct from cuts to services, represents the triumph of hope over experience: while politicians invoke the scope for greater efficiency, they are often reluctant to support the changes required to bring it about. Further, the boundary between efficiency savings and service cuts is porous and blurred: while efficiency savings mean (in theory) delivering the same service at less cost without any loss of quality, they can (in practice) mean cutting both quantity and quality. The most insidious, if also least visible, forms of rationing may well be the by-product of, or fall-out from, other policies as distinct from explicit decisions.

There are other changes likely to affect both the practice of rationing and how public debate about it is framed. Successive governments have increasingly used the rhetoric of localism, culminating in the launch of clinical commissioning groups and the involvement of local authorities in health policy-making. The aim is to enhance the ‘democratic’ legitimacy of local decision-making. But would increased legitimacy serve to justify continued – or even accentuated – geographic variations in what the NHS offers? Is postcode rationing acceptable in what is supposed to be a *national* health system? And, again, there is the rhetoric of ‘no decision about me without me’, which carries the risk that choice between treatment options may come to fuel rising patient expectations about the menu that should be available. Personal budgets may create new tensions: what if the patients use the funds to purchase snake oil?

A changing and financially meaner environment will also give new urgency to issues that have long been debated but not resolved, and which perhaps can never be resolved. What criteria should be used in deciding who should get what? And how should such decisions be arrived at? How can we best accommodate conflicting values? The National Institute for Health and Clinical Excellence (NICE) represents a globally admired attempt to make rationing acceptable by basing it on clinical and economic expertise, yet its decisions and methodology continue to be questioned. Can rationing, then, be insulated from politics? Should rationing be explicit or implicit? Is the way decisions are taken as important as – perhaps even more important than – the criteria and methodology used? We shall explore these and other questions further in what follows and return to some of them in our conclusions.

Before doing so, however, there is a prior issue. The assumption shaping the argument so far has been that rationing is an ineluctable feature of all health care systems, reflecting an inevitable imbalance between resources and demands. Should not this assumption be challenged? The evidence is clear: existing resources could be used much more efficiently and effectively. Demands are not cast in concrete: self-care can reduce them, as can the successful promotion of healthy lifestyles. The largest single element in health expenditure is the wage bill, the outcome of political bargaining about the level of rewards between governments, the professions and others. So in theory there is an argument to be made that rationing reflects the poor use made of existing resources and a lack of political will. In practice, however, the notion that rationing would be redundant if only waste were eliminated turns out to be unrealistic (Blustein and Marmor 1992): ‘waste’ is an elusive concept given uncertainty about what works for whom and changing definitions of good medical practice over time. Moreover, the argument ignores the pressure on budgets created by the ever-expanding menu of expensive pharmaceutical and other interventions – particularly evident in

the case of cancer drugs (Sullivan *et al* 2011) – which, as we shall see, features prominently in current controversies about rationing.

The aim of this discussion paper is to give the reader an understanding of the issues involved in rationing in the evolving context of the contemporary NHS in England. Our paper is not intended as a primer: we provide no advice about how rationing decisions ought to be made. Nor is it meant to be a comprehensive report on how rationing is practised in the NHS. We have scanned the available literature and talked to practitioners (*see Acknowledgements*) in order to provide examples and illustrations of what is happening but with no pretensions to provide anything like a full picture. Rather our intention is to provide a guide through the intellectual maze of rationing, to analyse and, wherever possible, clarify the puzzles and controversies encountered and to identify what we do not know as well as set out what we do know.

For some readers this study may be a first introduction to the rationing debate. Accordingly, it does not assume knowledge but sets out (albeit briefly and schematically) the basic conceptual, methodological and institutional elements. We hope, however, that those familiar with the topic will – with judicious skipping – find enough to stimulate their interest or provoke their disagreement.

---

# Values, criteria and actors

---

Debate about rationing is characterised by controversy. Disagreement starts with the word itself. Rationing in a health context is conventionally defined as the denial of potentially beneficial interventions or treatment. According to this definition, refusing to prescribe snake oil is not rationing. But what if even snake oil has some beneficial effect – if only psychological – for some patients? A much-invoked phrase when interventions are struck off the NHS menu (as we shall see) is that these offer only ‘marginal benefit’. But the concept of ‘marginality’ remains elastic in its application. Further, denial is only one – though by far the most visible – form of rationing. There are other forms (*see* box below), of which rationing by dilution is the least visible but arguably the most pervasive and likely to become more so in hard times.

## NHS rationing strategies

**Rationing by denial** This is the most headline-catching form of rationing. Specific forms of intervention are excluded from the NHS services on offer, on the grounds of lack of effectiveness, high cost or a combination of the two.

**Rationing by selection** Service providers select those patients who are most likely to benefit from interventions or raise the threshold of eligibility for treatment.

**Rationing by delay** The traditional form of rationing in the NHS, designed to control access to the system and match demand to supply by making patients wait.

**Rationing by deterrence** If patients are not put off by queues, there are other ways of raising barriers to, and the costs of, entry into the health care system. Receptionists may be unhelpful, information leaflets may be unavailable, access may be difficult.

**Rationing by deflection** All else failing, patients may be shunted off to another institution, agency or programme. ‘Difficult cases’ may be referred to another hospital or specialist.

**Rationing by dilution** Services or programmes continue to be offered, but there are fewer nurses on the ward, doctors order fewer tests, the palatability of hospital food plunges, and the quality of care and treatment declines.

Conventionally, the notion of rationing is conflated with that of priority-setting. This is to miss an important distinction (Klein *et al* 1996). Priority-setting describes decisions about the allocation of resources between the competing claims of different services, different patient groups or different elements of care. Rationing, strictly speaking, describes the effect of those decisions on individual patients. Giving priority to service A when allocating resources does not tell us anything about whether patients in services B, C or D are deprived of potentially beneficial interventions or suffer a loss in the quality of care.

The distinction between priority-setting and rationing is also helpful as a reminder that decisions that eventually determine who gets what – the resources devoted to individual patients – are taken at all levels of the NHS. Central government decides on the priority to be given to the NHS in England as against the competing claims of other public services and the economy. The Department of Health then designs the formula used to distribute the budget to individual commissioners (Bevan 2009). Individual commissioners – primary care trusts (PCTs) hitherto, clinical commissioning groups in future – decide between competing priorities when allocating resources to specific services. So do providers. Finally, there comes the point when the cascade of decisions about priorities from the national to the local level impinges on the individual patient: when clinicians, whether in primary or community care or in hospitals, decide how to allocate their limited time and resources – what is known as ‘bedside rationing’.

There is, then, in effect a hierarchy of decision-making about priorities in resource allocation in the NHS where the end point is the decision about what the patient gets – rationing in the strict sense. Similarly, there is a hierarchy in the criteria used in decision-making. The further decision-making is from the individual patient, the more invoked is the criterion of using resources to maximise the health of the population as a whole. The nearer decisions get to the patient, the more the traditional ethical criteria are invoked: the focus switches from populations to individuals. And it is precisely at the point where the two perspectives come together – when decisions based on seeking to maximise the health of the population have to be applied to individuals – that there is most friction. Further, cutting across all levels in the hierarchy of decision-making, there is another issue. This is how decisions should be taken: what processes give decisions legitimacy and thus make them publicly acceptable?

The next section briefly discusses the competing criteria applied to resource allocation. Subsequent sections illustrate how, and whether, different institutional actors apply them in practice, and the conclusion discusses some of the major tensions identified.

---

# Competing criteria and values

---

The focus of the population health-maximising principle of resource allocation – rooted in utilitarianism – is on comparing the increments of benefit yielded by different interventions. Its tool, cost-utility or cost-effectiveness analysis, is designed to maximise the total impact of any given input of resources and to take account of opportunity costs. Here the crucial question is not whether resources allocated to a particular individual would or would not improve her or his condition, but whether the same resources would produce an even bigger increment in welfare – ie, be more cost-effective – if applied elsewhere or devoted to another intervention. Population gains trump individual gains. It is a highly egalitarian principle since it is based on the Benthamite formula that everyone should count for one and no one for more than one. It also embodies, it has been argued (Williams 1992), a moral imperative, inasmuch as to ignore cost-effectiveness is to inflict a loss of potential benefits on others. In summary, and crudely, the aim is to ensure the biggest bang for the buck.

Counting the bucks is, of course, easy. Measuring the bangs is another matter. The challenge is to find a common currency for comparing the benefits of different interventions in a heterogeneous population. Here the standard numerator is the Quality Adjusted Life Year (QALY). As we shall see in the next section, this is not without problems. More fundamentally still, the notion that maximising population health can or should be the exclusive criterion when allocating or rationing resources has few supporters. The utilitarian approach may produce an efficient distribution of resources, but is it necessarily fair? What if we value equity as well as efficiency? In that case, we may wish to weight any formula for allocating resources toward those interventions that favour the most disadvantaged. Or we may think about equity across the lifetime of individuals, and argue (Williams 1997) that preference should be given to the young since older people have already enjoyed a ‘fair innings’. Conversely, we may place a higher value on the final weeks or months of life.

Again, cost-effectiveness analysis may well demonstrate that investing in cheap routine procedures or preventive measures yields higher benefits, as measured by QALYs, than engaging in expensive life-saving interventions. So should all intensive care units be shut down and the resources switched to more cost-effective interventions? Surely not, for we value the NHS’s commitment to doing the utmost to save our lives should we be struck down by some catastrophic event. The sense of security that the NHS gives us has a high value, even though we may not know how to put a cash figure on it.

Indeed, it is at precisely this point, when population-based criteria for allocating resources (priority-setting) have to be translated into what the results mean for individual patients (rationing), that a new set of criteria or ethical principles comes into play. Here two principles compete. The first is that resources are allocated according to the capacity to benefit: the doctor's attention or the scarce kidney goes to the patient with the best prospect of surviving longest. The other is the rule of rescue: the patient with the highest risk of death has first claim on the available care, no matter what the costs.

Social criteria for allocating resources to individuals may also be invoked. Here the notion of 'desert' or 'merit' comes into play. Some people may be deemed to merit preference because of their contribution to society – whether, for example, as outstanding musicians or soldiers wounded in the service of their country. Or the notion may be stretched to include a mother with children dependent on her continued capacity to look after them. Conversely, it is sometimes argued that those who are responsible for their own condition – failing to exert self-discipline by over-eating, say – should go to the back of any queue. However, this is contested territory: 'merit' or 'desert' as an allocative principle is challenged by those who believe that only medical need should count in a national health care system.

The complexity of decisions about whom to treat and how can be further illustrated by a case that attracted much attention in the 1990s and raises issues that are still very much alive. This was the case of Jaymee Bowen, analysed in Ham and Pickard's 1998 study (Ham and Pickard 1998). In 1995 the 10-year-old was diagnosed as having a recurrence of acute myeloid leukaemia, having received a bone marrow transplant for the condition nine months earlier. The clinicians involved all advised against further chemotherapy and a second transplant, calculating that the chance for the success of intensive treatment working was around 1 per cent. Given the pain and discomfort involved, and the low chance of success, they advised against heroic treatment. However, Jaymee's father sought other opinions and applied to the local health authority for agreement to refer his daughter to another hospital where the clinicians took a more optimistic view. The authority refused. A long legal and media battle followed. Eventually Jaymee was treated in the private sector, an anonymous donor providing the funds. An experimental treatment gave Jaymee some respite – she was even able to return to school for some time – but 10 months later she died.

The rule of rescue might have suggested that everything possible should be done, however slim the chances of success and however brief the extension of life. But the case brought into play three founding principles of medical ethics: non-maleficence, beneficence and autonomy (Gillon 1994) – that is, the duty of doctors to do more benefit than harm, while respecting the autonomy of

patients. The clinicians who decided against further chemotherapy and a second transplant decided that the harm involved would outweigh any benefits. Another set of clinicians took a different view based on their assessment of the case and the odds on a favourable outcome: a reminder of the uncertainty involved in clinical decision-making, with the result that medical opinion can speak with many voices. In turn, the principle of autonomy raised the question of what weight should be given to the views of the main actor in the drama, the patient: Jaymee's father saw the health authority's refusal to sanction further treatment as a breach of the principle of autonomy, in that it ran against his daughter's expressed wishes. Thus the case not only raises the question of who should speak for a minor; it also underlines the wider issue of who, if not the patient, can put a value on extending life if only by a few months or on the quality of the life thus extended – an issue sharply raised by current controversies about paying for expensive drugs in end-of-life care (*see* pp 12–13).

The case is also a reminder that cost constraints are not necessarily the only or dominant factor in decisions about whom to treat and how. So, for example, the dilemmas of choice are often most difficult and tragic when it comes to deciding who should receive scarce organs for transplant (Calabresi and Bobbitt 1978). In the Jaymee Bowen case, population-based criteria were certainly involved: the health authority took the view that substantial expenditure with small chances of success would not be an effective use of resources given the needs of other patients. But the main justification for the decision was that the proposed treatment would not be in the best interests of the individual concerned. In short, the case is an example of the challenge faced by purchasers of health care – health authorities in the past, clinical commissioning groups in the future – when seeking to accommodate competing principles.

Given conflicting criteria and values, increasing emphasis has in recent years been put on the process of decision-making. If we cannot agree on the criteria, can we at least agree on a fair way of reaching decisions? Enter the Accountability for Reasonableness Framework (Daniels and Sabin 2002) (*see* box opposite), which sets out the four conditions required for the making of fair decisions. The framework's strength lies in the fact that it does not require agreement with the final decision – which would be to demand the impossible – but only requires agreement that relevant factors and evidence have been taken into account while irrelevant considerations (such as favouring red-headed patients) have been disregarded. It is a formula for establishing the rules of the game, not necessarily for producing a consensus about the outcome.

### The Accountability for Reasonableness Framework

- **Publicity** Both the decisions about the allocation of resources and the grounds for reaching them must be made public.
- **Relevance** The reasons for reaching decisions must be ones that fair-minded people would agree are relevant in the particular context.
- **Challenge and revision** There must be opportunities for challenging decisions, mechanisms for resolving disputes and transparent systems for revising decisions if more evidence becomes available.
- **Regulation** There must be public regulation of the decision-making process to ensure that it meets the demands of the first three conditions.

The ‘accountability for reasonableness’ approach is reflected in much of public policy, as in the requirement for transparency in decision-making by PCTs. Further, the conditions overlap with, and reinforce, some of the criteria used by the courts when reviewing the process of decision-making about priorities (*see* pp 26–9). For example, it is not self-evident what should be the ‘relevant factors’ in any given situation, so there is still scope for debate.

Further, there is a long-standing controversy: should decision-making be explicit or implicit? In the case of decisions about priority-setting, there is now general agreement that decisions about priorities should be explicit and that the transparency condition should be met. But it is far from evident that rationing decisions in individual cases should invariably also be explicit. The ‘no decision about me without me’ rhetoric would suggest that every time a doctor decides that budgetary constraints compel restraint in prescribing or ordering another test, this would have to be discussed and agreed with the patient. However, as Mechanic (1997) has argued, implicit rationing by clinicians has considerable strengths: ‘its discretion, flexibility, and ability to take account of emotions, aspirations and preferences’. The case for explicitness does not take account of the uncertainty inherent in many clinical encounters. So would it suffice if doctors could, if challenged, account for the reasonableness of their decisions retrospectively? Nor do the rules of procedural fairness tell us who should be involved in the decision-making process: should patients or the public be represented? This remains a very much alive issue.

## The technocratic model

If any institution comes near to embodying the population health-maximising approach to resource allocation, it is the National Institute for Health and Clinical Excellence. NICE has a variety of functions, including the preparation of guidelines for clinicians, and its remit has changed since it was set up in 1999 and continues to do so. But its central – certainly most visible and often most controversial – role is to evaluate health technology, ie, pharmaceuticals and surgical interventions. The recommendations that flow from these ‘technology appraisals’ are mandatory on NHS funders, and the NHS Constitution enshrines the right of patients ‘to drugs and treatments that have been recommended by NICE’. However, the converse does not follow: purchasers are free to fund a treatment that is not recommended by NICE. More recently, NICE has put more emphasis on the scope for disinvestments. For example, it issued guidelines to reduce antibiotic prescribing for respiratory tract infection in primary care.

NICE’s role will change somewhat as from January 2014 when a new system of value-based pricing for drugs is due to be introduced. This is planned to link the price the NHS pays to the value that a new medicine delivers to patients. However, ‘The NHS will be required to fund drugs already recommended by NICE, as well as drug treatments subject to the value-based pricing regime. This means patients will continue to have a legal right to clinically appropriate, cost-effective drugs and treatments as set out in the NHS Constitution’ (Secretary of State for Health 2011).

NICE’s methodology is based on expert panels taking decisions based on the evidence about the effectiveness of new drugs or other interventions. On the basis of the evidence, the cost per QALY (*see* box opposite) is calculated and the recommendation follows. The NICE decision may be to recommend or reject the use of a particular new drug in the NHS or to limit its use to a specified sub-group of patients. It is an apparently dispassionate, scientific process. Yet NICE’s decisions are often controversial and contested. In part this is inevitable. Not only may the evidence base be challenged; the results of clinical trials do not necessarily predict accurately the impact in practice on a heterogeneous patient population, many of whom may suffer from co-morbidities. More crucially still, decisions may look very different when applied to the case of the individual patient where different criteria are invoked, as already argued. And, as we shall see, this is precisely the point in the chain of decision-making about resource allocation in the NHS – the point where priority-setting turns into rationing – when decisions are challenged.

## What is a QALY and how is it used?

A Quality Adjusted Life Year is the unit of measurement used to compare the cost-effectiveness of different treatments for the same condition or across treatments for different conditions. QALYs have two dimensions. The first is the length of life – months or years – that the patient can expect following treatment. The second is the quality of that life. The quality is measured on a scale ranging from 0 (death) to 1 (perfect health). The scale takes into account mobility, pain or discomfort, anxiety or depression and the ability to pursue the usual activities of daily living.

This allows new treatments to be compared with standard existing care. So (to take an example used by NICE) a notional patient with a serious, life-threatening condition may have a life expectancy of one year with a quality score of 0.4 given existing care, but if he or she receives a new drug the life expectancy will rise to 15 months with a quality score of 0.6. So while standard care yields a gain of 0.4 QALYs (one extra year at 0.4), the new drug yields a gain of 0.75 QALYs (15 months at 0.6). So the new treatment produces 0.35 additional QALYs.

The final stage in the process of using this methodology in decisions about new drugs or other interventions is to calculate the cost of achieving the QALY gain. Assuming that the cost of new drug is £10,000 as against standard treatment costs of £3,000, the difference of £7,000 is then divided by the QALYs gained to calculate the cost per QALY: £20,000 in this imaginary example.

But there are also some fundamental problems about the methodology itself. Having calculated the cost per QALY, what follows? How does NICE define the limits of acceptable spending? In practice NICE has operated on the basis that a cost per QALY of less than £20,000 is acceptable, that a cost per QALY of between £20,000 and £30,000 needs justification, and that if the cost exceeds £30,000, it's thumbs down. NICE interprets these thresholds with some flexibility. Many of its decisions are conditional – for example, the use of a new drug may be limited to particular groups of patients. It takes into account the burden of disease and the availability (or not) of alternative treatments (Devlin and Parkin 2004). The difficulty remains that, as NICE's chairman has acknowledged, its threshold lacks any empirical basis (Rawlins and Culyer 2004). In short, the limits – apparently set after a trawl of economists – are arbitrary, based on neither theory nor evidence.

So it can be argued that the figures should be much lower, in line with the judgements made by commissioners in the NHS (Appleby *et al* 2007). Why should NHS commissioners be forced to spend money on expensive NICE-recommended drugs, say, when they can buy better-value QALYs at a lower price by investing in other services? Do NICE decisions crowd out better investment opportunities? Which should have higher priority – national or local decisions? Or, come to that, why has NICE not up-rated its threshold figures – which have seemingly remained the same since 1999 – in line with inflation (House of Commons Health Committee 2007)?

Nor has the concept of the QALY – the foundation of NICE’s approach – escaped criticism. The QALY ratings for specific states of health on the 0 to 1 scale are derived from a study of public views. But it is not self-evident that there is a match between public and patient views about how to rate a particular condition; indeed, there is evidence to suggest that patients take a rather different view (Dolan *et al* 2009). Similarly, it has been argued that the use of QALYs discriminates against the disabled. Given the same life expectancy after treatment, a disabled person would score less than a non-disabled person. In short, the use of QALYs implies a judgement on what life is worth to a person – a judgement that may not be shared by the person concerned (Oliver and Sorensen 2005).

The notion that there is a neat, technocratic fix for taking decisions about resource allocation – as epitomised by NICE – therefore turns out to be a mirage. Indeed, NICE itself has recognised this in a variety of ways. It accepts that social value judgements are inevitably involved in the decision-making process (NICE 2009). For example, NICE considers that it ‘has a duty to take into account the impact of its guidance on health inequalities’. However, it is not apparent how other criteria beside cost-effectiveness influence the recommendations of NICE panels: we simply do not know. NICE has also set up a Citizens’ Council to explore issues such as whether the severity of a disease should be taken into account when making decisions: a recognition in effect, that technocratic expertise is not a sufficient source of legitimacy and that ways have to be found to test the public acceptability of NICE’s methodology. It is far from clear, however, that the existence of a Citizens’ Council is sufficient to bridge any legitimacy gap (Syrrett 2003).

The fragility of some of the assumptions shaping its methodology, and doubts about whether it has a strong constituency of public support, combine to make NICE vulnerable to political and media pressure. The hope that its invention would insulate decisions about the introduction of drugs in the NHS from ministerial intervention proved illusory. So, for example, in 2008 NICE introduced new criteria, and higher thresholds, for drugs for small patient groups

with short life expectancies (Raftery 2009). This followed a campaign by patient groups, in part funded by the pharmaceutical industry, representing terminal cancer sufferers. In effect, the change put a premium on the last weeks or months of life, threatening to displace more cost-effective treatments (Maynard and Bloor 2009). The episode was a reminder also that diseases differ in their emotional resonance when it comes to resource allocation; soon afterwards the Conservative Party committed itself to creating a special fund for cancer drugs, a pledge since implemented.

NICE's social-value guidelines also stress the importance of process: the way decisions are taken. They include a set of procedural principles such as transparency, inclusiveness (giving professional bodies, patients, pharmaceutical companies and others an opportunity to participate in the decision-making process) and challenge (giving an opportunity to comment and appeal). In stressing these principles, NICE reflects the demands of procedural justice as set out in the Accountability for Reasonableness Framework (*see* box, p 9).

If the NICE decision process turns out to be more complex and more flexible than appears at first sight, the need to accommodate a variety of criteria or values is even greater when we move to the next point in the hierarchy: priority-setting by PCTs. PCTs are, of course, on their way to the dustbin of history – but in what follows we assume that GP commissioners may start their lives by building on the PCT legacy.

## The commissioner's tool box

Any commissioner of health care in search of tools for taking decisions about priorities can consult a rich library of acronyms: PBMA (programme budgeting and marginal analysis), MCDA (multiple criteria decision analysis) and others (for a review, see Yorkshire and Humber Public Health Observatory 2010). Here there is no intention to offer a comprehensive guide to the various decision-making techniques that are on offer. Instead, by taking two examples from the opposite ends of a spectrum running from an emphasis on costs and benefits to the accommodation of multiple criteria, we can illustrate some of the issues involved.

At the end of the spectrum nearest to the NICE approach is PBMA (Donaldson 2008). The first step in this model is to set out how existing resources are used under various programme headings – for example, cancer or mental health services (Brambleby *et al* 2008). Next, then, changes in the pattern are considered, comparing the gains (or losses) in health benefits that can be achieved by changes at the margins. Unlike NICE, PBMA is not dependent on the use of QALYs: at a programme level the necessary information may not be available. But the aim

is the same: ‘to maximise benefits to the community’. And so is the emphasis on cost-effectiveness. However, as one of the key exponents of the PBMA approach explains, ‘an ethical rather than economic, framework, may be more appropriate to take account of criteria relating to equity and fairness when deciding how to allocate healthcare resources’ (Donaldson 2008). So those involved in any PBMA exercise have to decide what criteria – beside cost-effectiveness – to apply when making a final judgement.

At the other end of the spectrum there is the Modified Portsmouth Scorecard, used by many PCTs, where funding proposals are scored against a number of weighted criteria – a variant of MCDA. The criteria include strength of evidence (scoring up to 40 points), magnitude of benefit (up to 40 points), patient acceptability (up to 40 points), whether it is a national requirement or target (up to 40 points), whether it is addressing health inequalities (up to 20 points), whether it is the only treatment (up to 20 points) and the strength of local feeling (up to 10 points). This model, in contrast to PBMA and some others, is not driven principally by economic, health-maximising aspirations, and its dependency on qualitative judgements appears to be even greater than that of the other models. But it does have the advantage that it can be applied to micro developments rather than programmes, thus coming closer to rationing decisions.

The other models on offer are variations on these themes. So, for example, decision conferencing (Airoldi and Bevan 2010) puts much emphasis on the social dynamics of the process: the iterative way in which information about value for money is presented and discussed. However, while we know quite a bit about how many PCTs use the different decision-making tools when setting their priorities (Robinson *et al* 2011), we have no systematic knowledge about the relationship between the use of the different methodologies, the composition of the panels involved in the process and the outcomes. And we know even less about how decisions about priorities translate into policies that affect individual clinicians and patients directly, ie, rationing.

---

# From theory to practice: how commissioners ration

---

How do rationing decisions get made in practice by commissioners? How do they limit what's available, and on what basis do they make such decisions? Commissioners currently adopt some combination of five approaches:

- slowing down treatment rates by increasing waiting times (*rationing by delay*)
- introducing eligibility thresholds for treatments in order to limit the number of referrals (*rationing by selection*)
- reviewing spending on continuing care and diverting demand to social services (*rationing by deflection*)
- capping the number of treatments hospitals can provide (*rationing by denial*)
- developing 'exclusion lists' of treatments that will no longer be funded (*rationing by denial*).

However, as we shall see, commissioners must also allow for the possibility that individual patients may represent 'exceptions' to any of their general policies on funding, and have a process in place for considering, and ruling on, such cases.

## Increasing waiting times

In the past, waiting times reflected lack of capacity (as well as poor management and a low value being placed on patients' time); now waiting times often reflect a lack of ability on the part of commissioners to pay providers for treatments. Commissioners are stopping providers from treating patients by introducing minimum waiting times. This approach can effectively defer payments to a new financial year, and potentially reduce expenditure overall since, as one PCT put it in response to a survey, 'experience suggests that if patients wait longer then some will remove themselves from the list or will no longer require treatment when it is finally offered' (Co-operation and Competition Panel 2011). So this approach is one of deterrence as well as delay. Commissioners are bound by the 18-week maximum referral-to-treatment time target in the Handbook to the NHS Constitution, but they can allow waiting times to rise so long as they do not exceed that point, a practice that a report by the NHS Co-operation and Competition Panel in 2011 judged to be 'reasonably widespread'.

Headlines in November 2011 proclaimed that the Secretary of State was to ‘ban’ minimum waits (Wilson 2011). In fact, the restriction, which all commissioners must comply with by March 2012, is on minimum waiting times ‘that do not take account of health care needs of individual patients’ (Department of Health 2011a). The objection is not to minimum waits *per se*, but to blanket policies that are insensitive to the particular needs of any individual patient. In fact, this echoes a ruling by the European Court of Justice in 2006, which found that a commissioner offering treatment within a national waiting time target does not in itself mean the treatment is being provided without ‘undue delay’ (a trigger for commissioners having to fund a patient’s treatment abroad within the EU). The court judged that undue delay may still have occurred if the waiting time ‘appears to exceed in the individual case concerned an acceptable period having regard to an objective medical assessment of all the circumstances of the situation and the clinical needs of the person concerned’ (Watts v Bedfordshire PCT, Case C-372/04).

These new national instructions to commissioners thus focus on how commissioners reach a decision about waiting times (eg, whether an individual’s particular clinical circumstances are taken into account) rather than on what that decision ought to be, a theme that we will return to later.

## Limiting volumes of treatments

More radical than delaying treatment is to limit the numbers of treatments for which any provider will be funded. The implementation of the policies of patient choice and Payment by Results for many aspects of routine, planned hospital care over the past half decade has meant a shift away from block contracts, making it more difficult for commissioners to control the volume of care carried out by any one provider. But some areas of care – mental health services and most community services, for example – are not yet subject to payment by results. And even for those types of care that are, the standard acute contract used by commissioners includes an ‘activity plan’ agreed with the provider, and allows that ‘financial adjustments’ are made to the contract if the provider persists in performing more treatments than had been agreed.

PCTs have taken different approaches to using this type of planning, but some are reported to have used activity plans to introduce limits beyond which they withhold or challenge payments for treatments (Co-operation and Competition Panel 2011). Other PCTs have gone further and effectively suspended the operation of payment by results and put in place block-volume, capped contracts in order to manage their costs; this was the case for some of the acute trusts whose staff we spoke to in preparing this paper.

Commissioners also limit activity by imposing restrictions on the number of follow-up appointments that patients are given after their surgery or main consultation. Such strategies may be starting to bite: analysis of hospital trust data in 2011 found that the ratio of follow-up appointments to surgical procedures fell by 4.3 per cent in the last year and 5.9 per cent over the last two years (Pulse 2011). What is difficult to establish is whether this is rationing in the sense of depriving patients of beneficial treatment. There does not appear to be a clear consensus on how many appointments are necessary or appropriate.

Furthermore, how do such restrictions filter through to clinical practice? Does management pressure change the informal threshold for referral with which consultants and junior doctors operate in practice when deciding whether to refer a patient on for further treatment? Are patients seen towards the end of a quarter or a financial year less likely to be treated because the hospital is approaching the upper limit for its activity?

The strength of the Department of Health's reaction to these practices (particularly in a context where they are committed to holding back from issuing edicts to the service) suggests that the Department is anxious about the effects of such restrictions on patients, and public perceptions of the services. The Department published a strong warning on volume-capping practices, stating that 'activity schedules in the NHS Standard Acute Contract should not be used as a mechanism to ... cap activity' (Department of Health 2011a). At the time of writing, details on how this restriction might be implemented are still pending. The Department has also said: 'All patients with a clinical need for a follow-up appointment in hospital should have one. We have not set targets to reduce the number of follow-up appointments and have no plans to do this' (Pulse 2011), but has hitherto stopped short of directly prohibiting commissioners from doing so.

If commissioners are to be restricted in their ability to adjust the overall volume and pace of treatments being offered by hospitals, they may concentrate ever more on two other rationing tools. These are limiting who can be referred for treatment in the first place through the introduction of referral thresholds, and developing 'exclusion lists' of treatments for which routine NHS funding will not be made available.

## Rationing by selection: eligibility thresholds

It is general practitioners who drive much of the demand for hospital services; it is not surprising that the Secretary of State is placing responsibility for budgets in their hands. But GPs vary widely in their clinical decisions and many of the factors that determine whether a patient is referred to hospital fall outside of narrowly conceived clinical concerns. Such factors include the facilities and

expertise contained in a practice, and the GP's relationship with the patient and with the relevant hospital specialist. These are in addition to a series of cognitive and psychological influences, including the practitioner's willingness to tolerate risk, their fear of accusation of malpractice, their sensitivity towards patient demands, their attitudes towards hospital admission, and their beliefs about the benefits of referral to particular services (Foot *et al* 2010). The result is the renewed emphasis by commissioners on attempting to control GP referrals by introducing referral guidelines and thresholds.

Such controls range from PCTs prescribing the criteria that GPs ought to consider when making a decision to imposing compulsory, quantified scores to test eligibility for referral. For example, the pre-operative Oxford Hip and Knee Scores are used as a tool for 'grading' the degree of severity of someone's joint problem, incorporating the results of diagnostic imaging with information on stability and mobility of the joint, and the nature and extent of the pain they are suffering; the resulting score determines whether a patient is eligible for referral. A fifth of the 692 orthopaedic surgeons responding to a BBC survey in spring 2011 reported that referral thresholds had been raised by their local PCT (Jeffries 2011). The policies of half a dozen PCTs serve to illustrate the geographical variation in thresholds for tonsillectomies (a long-contested procedure) and bariatric surgery (*see* box opposite).

Such thresholds effectively limit clinical freedom, and so are controversial among professionals as well as patients. The definition of an appropriate referral can be contentious. For example, a study of musculoskeletal referrals in the early 1990s found that 43 per cent of referrals rated as 'probably' or 'definitely' unnecessary by a GP were rated as 'definitely appropriate' by a consultant (Roland *et al* 1991). And while some commissioners have adopted the types of scoring systems described above, others have disputed their reliability or appropriateness (*see*, for example, Solutions for Public Health 2010). The resistance clinicians may have to such interventions can be exacerbated or relieved depending on the type of enforcement tool used by the commissioner.

The methods for enforcing thresholds come in more or less authoritarian forms (Imison and Naylor 2010). At the softer end, commissioners produce guidelines that should inform practitioners' referral decisions, and in some cases involve clinicians in their development to ensure that the guidelines are clinically informed, and in an effort to win hearts and minds. At the sharp end, just over a quarter of PCTs have referral management centres (Pulse 2011), which are used to review and potentially redirect or reject some or all referrals made by GPs. Somewhere between the two lie peer review processes, in which the comparative referral rates of individual GPs and/or their practices circulated within an organisation are shared with other local practices as well as the commissioner.

## Thresholds for referral: tonsillectomy and bariatric surgery

*Illustrative examples drawn from the policies of seven commissioners from across England in November 2011*

### **Tonsillectomy (removing tonsils)**

Commissioners used a number of different criteria for limiting access to tonsillectomy. Common among them is a requirement about the number of episodes of the illness the patient must have suffered in recent years in order to be eligible: these range from a requirement for seven severe cases in the previous year to five such cases, or a fewer number of cases but experienced consistently over a series of years – three or more cases in each of the previous three years, for example. Some commissioners require a period of six months ‘watchful waiting’ before the referral; others do not. For some, severe halitosis that has been demonstrated to be tonsil-related is grounds for a referral; others make no mention of this. One commissioner’s policy required that the patient must have previously suffered from quinsy in order to be eligible for tonsillectomy.

### **Bariatric surgery (weight-loss surgery)**

Here the commissioners usually established a minimum body mass index (BMI) to determine eligibility for surgery, which is higher or lower depending on whether the patient suffers from co-morbidities whose treatment could be improved by reducing their weight. For example, a patient with a BMI of 50 or over, or a patient with a lower BMI (45, for example) who was also suffering from diabetes or hypertension could be eligible, so long as they had participated in a programme combining a range of non-surgical treatments and support. But there are differences between areas. For example, to qualify for surgery in one area, the requirements include a BMI of more than 50 while in another the requirement is for a BMI of 35. Some name the co-morbidities necessary for eligibility; others do not. Some bar patients who have not sustained a weight loss of 5 per cent or 10 per cent, over three or six months prior to the referral; others do not. Commonly, patients are expected to be 18 or over, but in one area, patients had to be at least 20. For some the maximum age is 60, for another it is 65, and others have no stated maximum age limit.

What is made of this information, and how it might go on to change future behaviour patterns, is the subject of local negotiation. Here data indicators are used as a tin-opener rather than a dial (Carter *et al* 1992), ie, they present an opportunity to examine what is happening rather than offering a precise reading.

Of course GPs can only control decisions about whether to refer a patient to hospital. Once that referral has been made, decisions about whether to treat patients are at the discretion of hospital clinicians. We explore attempts by hospitals themselves to influence these decisions in the section on ‘Decoding provider rationing’.

## Rationing by deflection: risks to NHS continuing care?

There is a decades-long history of tension between the NHS and local authorities over who should fund the care costs of individuals who have ongoing health care needs outside of hospital. As with other aspects of NHS care, primary legislation has not prescribed the NHS’s responsibilities in precise terms, stating only that the Secretary of State must provide ‘services or facilities for the prevention of illness and the after-care of persons who have suffered from illness as he considers appropriate as part of the health service’. However, following various legal directions by the Secretary of State, if somebody’s ‘primary need is a health need’, they may be eligible for NHS continuing care, which means their PCT must fully fund their health care, personal care and (if they are resident in a care home) board and lodging costs as long as they are required.

If the individual does not meet the criteria for NHS continuing care, then responsibility for their care will fall (at least in part) on the local authority. This means the local authority may foot part of the bill, and since social care is not fully funded by local authorities (funding is contingent on local decisions about which ‘needs’ bands will be funded in that area, and means testing), the individual patient may themselves have to pay for much of their care costs.

Through legal challenges and investigations by the NHS ombudsman in the mid-1990s to the mid-2000s, local NHS commissioners were repeatedly found to have used overly strict eligibility criteria in denying funding to individuals, effectively deflecting the costs of this care onto local authorities and the individuals themselves (for a summary of cases *see* Alzheimer’s Society 2011 and Dow 2009). In 2007 the Department of Health published for the first time national guidance on establishing eligibility for the funding (which had previously been left to the discretion of strategic health authorities and their predecessors), but this remains very broad in its phrasing, with plenty of space for different local interpretations.

Eligibility assessments for continuing care funding are carried out by a ‘multidisciplinary team’ commissioned by the PCT, which may or may not include a member of PCT staff. Many PCTs do, however, have their own ‘panels’ that are intended to review the quality of the assessment decisions. As with other PCT decisions on the funding of care, patients can appeal against the decision made (initially to the PCT, then to the strategic health authority (SHA), which can refer the decision to an ‘independent review panel’, and ultimately to the NHS

ombudsman). A survey of PCT clusters in 2011 found that the number of appeals made against decisions increased by 9 per cent in 2009/10 to 2010/11, as did the proportion of those appeals that were successful (40 per cent compared to 33 per cent the previous year) (Samuel 2011b).

The overall numbers of patients receiving funding for continuing care increased each year between 2009 and autumn 2011 (both absolutely and as a proportion of the population) against what we might expect in a period of financial austerity. However, there are considerable variations between PCTs (in quarter two of 2011, for every 50,000 individuals in Berkshire West, 17 received continuing care funding, compared to 158 per 50,000 in Barking and Dagenham (Department of Health 2011b)), and the overall rise in numbers masks the fact that in more than a quarter of PCTs, the numbers receiving funding fell by 10 per cent in the final quarters of the financial years of 2010 and 2011 (Samuel 2011a). It is difficult to draw any firm conclusions from these figures without accompanying data on any changes in the proportion of the population with relevant needs at that time to put them into context.

However, it may be telling that the Department's deputy director for social care felt the need to write to the SHA and PCT chief executives in summer 2011 to 'remind colleagues of the importance of the national processes set out in the National Framework for determining eligibility for NHS Continuing Healthcare', which are 'particularly important at a time where local organisations may have local Quality, Innovation, Productivity and Prevention (QIPP) initiatives, which may be looking to make efficiencies in this area' (Ayling 2011). The national director for social care at the Association for Directors of Social Services reported in summer 2011 that: 'In many areas, continuing care is working very well, in others it's not. PCTs have used it as an obvious area to cap costs', and that some PCTs were delaying assessments for 'considerable periods of time, which means the individual is paying for the costs or their local authority is' (Samuel 2011a).

## Exclusion lists

The starkest and most visible form of rationing is the 'exclusion list'. The Secretary of State is required by law to promote a 'comprehensive' health service, but at no point has that phrase been defined in terms of a list of treatments that should be provided by the NHS. NICE guidance and national policies impose certain requirements on commissioners, but for the most part they must decide what treatments to make available in the context of local need and their finite budget. None publishes a full list of what it does fund for local NHS patients, but many now list the treatments they will not fund. More than a third of PCTs surveyed in 2011 had expanded the number of treatments for which they were withholding funding (Moberly 2011).

Such exclusions tend to be justified on the grounds that the treatments represent poor clinical value (they don't offer enough bang for their buck, or simply not much bang at all), or that they are not clinically necessary because they are principally cosmetic, for example. Surgery for lower back pain, hysterectomies for heavy menstrual bleeding, and grommets are common examples of the first group. Examples of the second include cosmetic surgery such as breast enlargement or reduction, tattoo removal or the removal of varicose veins.

The content of such lists varies between areas. Looking at 'low-priority treatment' lists overall (which can include both outright exclusions and treatments subject to eligibility thresholds), the Audit Commission estimated that PCTs have collectively identified approximately 250 different procedures for which funding is in some way restricted (Audit Commission 2011). Just keeping track of which commissioner is funding what can be a real challenge for clinicians and managers based in provider organisations (*see*, for example, the box below).

Many PCTs have in practice formed networks at local and regional levels to pool effort in this area and to present a more consistent set of policies to their local populations. A survey of PCTs' arrangements for considering funding for new medicines conducted in 2008 found that half of PCTs shared a committee with at least one other PCT, and a fifth with three or more other PCTs (Adelphi

### Variations in funding of IVF

A survey of commissioners in England, Wales and Scotland published by the All-Party Parliamentary Group on Fertility in June 2011 found considerable variation in local funding policies on IVF, despite NICE guidance recommending that in England three cycles should be funded.

- Five organisations stated that they did not offer any funding for IVF.
- Of the PCTs (and their equivalents) offering IVF to patients, just over a quarter (27 per cent) offered three cycles of treatment, with 39 per cent offering one cycle and 27 per cent offering two.
- There were also variations in the commissioners' criteria for eligibility for the treatment: for most the upper age limit was 39, but for some it was lower – for example, 35; some required that neither parent had a living child (in contrast to the NICE requirement that only the woman does not have a child); some specified that the woman or both parents must be non-smokers; and some applied upper BMI limits on both of the parents.

Source: All-Party Parliamentary Group on Infertility 2011

Research UK 2009). PCTs also informally share information through contacts or established networks. The 2008 survey found that almost half of committees shared information with neighbouring PCTs about their decisions ‘extremely frequently’, though there was considerable regional variation. The Public Health Commissioning Network includes on its website a list of published ‘priorities statements’ of various PCTs as part of its ‘Essential Resources Series’ (Public Health Commissioning Network 2010). Some PCTs and PCT networks employ independent organisations to support their decision-making in this area while others rely on internal staff. For example, the Cumbria and Lancashire Commissioning Business Service is funded by, and accountable to, six PCTs in that area and the South Central Priorities Committee is supported by ‘Solutions for Public Health’, a not-for-profit public health organisation.

The advent of clinical commissioning groups could lead to a fragmentation of existing networks of decision-making. At the time of writing there looked set to be around 270 clinical commissioning groups (compared to 152 PCTs), which may themselves have no connection with existing networks, and could dramatically increase the number of different policies for treatment funding within the NHS and accompanying accusations of a ‘postcode lottery’. However, in practice the development of clinical commissioning groups and their policies is taking place under the supervision of PCT clusters; some clinical commissioning group members will no doubt have previously sat on professional executive committees in PCTs; and some new lay members will have played non-executive director roles in predecessor organisations. This means there is the potential for clinical commissioning groups to draw on existing policies and networks.

Whoever makes such decisions on treatment funding, they will surely continue to attract controversy. And yet despite their headline-grabbing nature, the scale of savings to be secured by such rulings is actually relatively modest. The Audit Commission estimated that if all commissioners performed as well as the top 25 per cent in terms of limiting access to the list of ‘low-priority treatments’ developed by Croydon PCT and used by many others (*see box overleaf*) (Audit Commission 2011), there could be a national saving of £441 million a year. The report found potentially significant gains for some individual PCTs (topping £12.5 million in one case), but in terms of overall savings to the health service, such an approach represents a drop in the ocean of the estimated £15–20 billion of efficiency savings the NHS needs to secure over next few years.

Because outright refusals to fund treatment are the most direct and visible form of rationing, they are also the most open to challenge. And indeed there are systems in place to deal with such challenges. In the next section we explore how such appeals by individuals are dealt with.

## The Croydon list of low-priority treatments

### **Relatively ineffective procedures**

Back pain: injections and fusion  
Dilation and curettage for women under 40  
Grommets (surgery for glue ear)  
Jaw replacement  
Knee wash-outs  
Spinal cord stimulation  
Tonsillectomy  
Trigger finger

### **Effective procedures where cost-effective alternatives should be tried first**

Anal procedures  
Bilateral hip surgery  
Carpal tunnel surgery  
Elective cardiac ablation  
Hysterectomy for heavy menstrual bleeding

### **Effective interventions with a close benefit or risk balance in mild cases**

Cataract surgery  
Cochlear implants  
Dupuytren's Contracture (tightening of tendons)  
Non-surgical female genital prolapse/stress incontinence  
Surgical female genital prolapse/stress incontinence  
Hip and knee revisions  
Knee joint surgery  
Other joint prosthetics or replacements  
Primary hip replacement  
Wisdom teeth extraction  
Potentially cosmetic interventions  
Aesthetic surgery – breast; ear, nose and throat; plastics; ophthalmology  
Incisional and ventral hernias  
Inguinal, umbilical and femoral hernias  
Minor skin surgery for non-cancerous lesions  
Orthodontics  
Other hernia procedures  
Varicose veins

## Exceptional funding requests

Following a key principle of administrative law, PCTs are legally required to allow for exceptions to general policies on treatment funding (R v NW Lancashire HA [1999]). This means PCTs must have a system that allows patients to argue for a treatment that the commissioner has explicitly decided not to fund for particular indications (an exceptional funding request), or which is not normally provided and for which there is no policy about its funding (an individual funding request).

The number of such requests received by commissioners each year can be considerable: a Department of Health survey of PCTs conducted in 2008 estimated from its responses that some 26,000 applications for exceptional funding had been received by PCTs in the previous year, on average 177 requests per PCT, though there was significant variation between areas (from one PCT receiving just one request to another receiving 1,017 applications) (Richards 2008). On average, nearly two-thirds of requests for cancer treatments (64 per cent) and three-quarters of requests for non-cancer treatments (74 per cent) were approved, but agreements to fund such requests also varied widely, from one PCT not granting any approvals to six PCTs approving funding for 100 per cent of their requests (Richards 2008). A separate survey focusing on the funding of treatments for rarer cancers found that in the same year two neighbouring Essex PCTs had approved 0 per cent and 96 per cent of requests respectively (Rarer Cancers Forum 2008).

One reason for such variation may be the different ways in which PCTs interpret the term 'exceptional'. Research published in 2008 found that, although most PCTs (85 per cent) had a written protocol to govern the processing of exceptional funding requests (Rarer Cancers Forum 2008), almost a third (30 per cent) did not have a definition of 'exceptionality' (Adelphi Research UK 2009). But most commissioners have developed such a definition, and although there is considerable variation in the content of the definitions (Rarer Cancers Forum 2008), many have adopted a formulation developed by a public health consultant and publicised by the NHS Confederation (NHS Confederation 2008). This states: 'In making a case for special consideration, it needs to be demonstrated that: the patient is significantly different to the general population of patients with the condition in question; and the patient is likely to gain significantly more benefit from the intervention than might be normally expected for patients with that condition. The fact that a treatment is likely to be efficacious for a patient is not, in itself, a basis for exceptionality.'

---

# The legal framework for policy decisions

---

In all of this, the courts have played an important role in shaping the context within which funding decisions are made. They have pronounced on both the criteria for decisions about funding, and the way in which such decisions ought to be made. A series of judicial review cases has, in very general and sometimes ambiguous terms, ruled on appropriate criteria for judging a case as exceptional (*see* box below).

## Defining exceptionality: principles emerging from legal cases

- An ordinary reading of the term ‘exceptional’ should be applied (eg, a dictionary definition).
- Features of exceptionality should be reviewed ‘in the round’, rather than individually.
- The index case should be compared against the cohort of people eligible for treatment, but he cannot be deemed unexceptional because he is representative of a group of patients. He does not have to meet a standard of uniqueness.
- In the absence of direct clinical implications, social factors do not have to be considered in the assessment of exceptional circumstances.
- Demonstrating an overriding clinical need for treatment presents a strong case for being considered exceptional.
- Demonstrating features which suggest the index case is more likely to benefit from treatment than others can, but does not necessarily, make the index case exceptional.
- The patient’s prognosis need not be a consideration, but survival for a short period of time can make one exceptional, and the example provided is where care arrangements need to be made for a young child.

Reproduced from Ford (2012).

Another source of recourse for patients wishing to challenge the substance of commissioners' decisions is the European Convention on Human Rights, which is enshrined in the UK Human Rights Act. The most relevant clause is Article 2, which states: 'Everyone's right to life shall be protected by law'. The European Court of Human Rights has ruled that this right 'enjoins the state not only to refrain from the intentional and unlawful taking of life, but also to take appropriate steps to safeguard the lives of those within its jurisdiction', but that this obligation 'must be interpreted in a way which does not impose an impossible or disproportionate burden on the authorities' (*Osman v United Kingdom* (1998) 29 EHRR 245, paragraph 116 cited in Foster 2007). In practice, although resource allocation cases have been brought to court citing the Act, the courts have been very reluctant to make rulings on these grounds. As one commentator concludes: 'in general a patient who demands certain treatment will find little assistance in the courts' (Wicks 2009; *see also* Newdick 2005 and Foster 2007). For the time being at least, the legal oversight of commissioners' decisions remains focused on the processes of decision-making, rather than on the outcome of such decisions.

### Legal requirements on commissioners' decision-making processes

The NHS Constitution gives all patients the right to have local decisions about funding medicines and treatments made on a rational basis. Since 2009 PCTs have been required by law to set out on their website their processes for adopting general policies on what they will fund (though not the policies themselves) and, if requested, to provide a written statement setting out their reasons for any general policy on not funding an intervention (Department of Health 2009). The Department of Health has also commissioned a series of nine guiding principles that commissioners are expected (though not legally bound) to follow when making local funding decisions (National Prescribing Centre and Department of Health 2009). In summary, these are to:

- establish decision-making groups that are appropriately skilled
- establish robust decision-making procedures
- define criteria for decision-making, with decision-making based on the best available evidence, and taking into account ethical frameworks and legal requirements
- document how decisions were made and the rationale for the outcome
- make decisions in a reasonable timeframe, but without compromising the process

- establish an appeals process for individual funding requests independent of the original process
- ‘take reasonable steps to engage with stakeholders including the wider NHS, patients and the public to help increase understanding of local priority-setting about medicines’
- communicate clearly the process, decisions and rationale for the decisions
- establish an assurance process to monitor the decision-making arrangements.

These requirements contain strong echoes of the Accountability for Reasonableness Framework set out in the section on ‘Competing criteria and values’.

As public bodies, commissioners’ decision-making processes are also subject to challenge by patients or others through judicial review. The courts cannot order the outcome of the decision to be changed; but they can order the commissioner to start from scratch and take the decision again, taking account of the courts’ guidance. Traditionally there are three grounds on which a public authority’s decision can be challenged through judicial review: if it is found to be illegal (eg, the body is acting in a way that is inconsistent with its statutory powers), to be irrational, or to have been subject to some kind of procedural impropriety (Council of Civil Service Unions v Minister for the Civil Service [1985] AC 374). This last requirement is echoed in the guiding principles published by the Department of Health, and essentially requires that commissioners ‘develop transparent priorities frameworks which treat patients equally, fairly, and consistently, together with a mechanism for considering “exceptional” cases’ (Newdick 2005).

A decision is considered irrational by the courts if it is ‘so unreasonable that no reasonable authority could ever have come to it’ (Associated Provincial Picturehouses Limited v Wednesbury Corporation [1947]). Most recently, courts have interpreted this to mean decisions with ‘flawed logic’, and courts are increasingly seeking to identify adequate *reasons* for the decision being made as it was (Newdick 2005). For example, one ruling in this area established that when setting priorities, it is ‘vital’ that a health authority should ‘assess the nature and seriousness of each type of illness’ and ‘determine the effectiveness of the various forms of treatment for it’ (Auld LJ [1999] Lloyds Rep Med, 399, 408 in Newdick 2005).

In terms of the legality of commissioners’ decisions, the statutory powers and duties of commissioners are established in the legislation governing the NHS. Under the NHS Act 2006, the Secretary of State has a duty to promote

a comprehensive health service. In recent years this was delegated to PCTs (Newdick 2005; Foster 2007). The courts have decided that this duty, which has been retained in the Health and Social Care Act 2012, is ‘limited to providing the services identified to the extent that he considers that they are *necessary* to meet *all reasonable requirements*’, given the available resources (*R v N and E Devon Health Authority, ex p Coughlan* [1999], *Lloyds Rep Med* 306, paras 23–25, in Newdick 2005).

---

# Decoding provider rationing

---

When a purchaser – PCTs in the past, clinical commissioning groups in the future – raises the threshold for surgery or limits the use of expensive drugs, there is a direct relationship between the decision and what happens to patients. The transparency condition is met; PCTs work under a set of rules about due process and we know when explicit rationing is taking place. In the case of providers the picture is much more opaque and becomes steadily more so as we move from the level of institutions to that of the individual clinicians delivering care to patients. The inverse knowledge law applies: we have the barest knowledge in precisely those areas – hospitals – where some of the most critical decisions about the allocation of resources to patients are made.

There is a further important difference between purchaser and provider strategies. While commissioners try to limit the volume of services and the prices they pay for them, providers try to limit the intensity of care by cutting the costs per patient through reductions in the amount and price of inputs such as equipment and staff time.

There is no shortage of information about what is happening to hospital budgets. The headlines proclaim the fiscal plight of trusts and the pressure on them to make multi-million savings. Similarly, there is no shortage of warnings about the consequential impact on staffing. So, for example, a Royal College of Nursing survey reported that almost 75,000 nurses expected to lose their jobs, have their hours cut or see their roles downgraded in 2012 (Royal College of Nursing 2011). Dire reports of falling standards will inevitably multiply in coming months. However, occupational groups – doctors, nurses and others – have historically had a self-interest in crying havoc (Klein 1983): protecting jobs and protecting patients can all too easily become confused. The difficulty therefore lies in interpreting what budget and staffing cuts mean for patients. The impact on patient care cannot be read off the figures of spending or staff cuts.

The broad strategies adopted by trusts to meet their retrenchment targets are clear enough (Appleby 2011). They include reducing capacity by closing down wards and services; and improving efficiency by, for example, reducing lengths of stay and cutting down on back-office staff. Different strategies are likely to have different impacts on services to patients, all the more so as the context, the scale of the fiscal challenge, and implementation all vary from trust to trust. This became very clear from a number of informal interviews with medical and lay managers,

as well as clinicians, designed to elicit illustrative insights (no more) into what is happening at the coalface of the NHS.

Some trusts are pursuing a strategy of cutting out chunks of activity, as part of a reconfiguration of services between hospitals. This does not necessarily mean any diminution of total activity in the system as a whole, or of quality; indeed, such changes are often justified in terms of improving quality. And the staff redundancies that follow the closing of wards or services do not necessarily have implications for patient care. The question of rationing arises only if consequent problems of access make it difficult for some patients to use the service (rationing by deterrence). However, this strategy may not be an option for smaller trusts, where there is strong interdependence between different units. Similarly, the redesign of disease pathways may actually improve patient care.

In the case of ‘efficiency savings’, a semantic chameleon phrase, much depends on the nature of specific schemes. Two examples make the point. In one case a new system, designed to save £10 million a year, is being introduced to allow consultants to email their reports directly to GPs, so permitting secretarial staff to be cut. In another case, however, secretarial staff have been cut without any accompanying changes in technology, thus diverting clinicians (in the view of some doctors, at any rate) from spending time with patients. In the first case, cuts may actually lead to an improvement in services to patients; in the latter case, they may be a threat to quality.

Similarly, there is a crucial difference between cutting or diluting staff without changing the working environment and adopting the same strategy while changing that environment. If wards are closed, then fewer nursing staff should not mean a lower quality of care: if there are fewer nurses but the number of patients remains constant, then clearly there is a threat to the quality of care. Hence, the problem of interpreting the headline figures and drawing conclusions about whether they do or do not imply rationing by dilution.

Some changes may affect patients directly. Examples of such economy measures include reducing the frequency of outpatient appointments and consultant-to-consultant referrals, no longer offering routine mammograms to women on hormone replacement therapy and reducing transport for some patients. But rationing by denial appears to be very much the embryonic exception in the case of hospital services, involving for the most part only marginal and petty-cash activities. At first sight ‘good housekeeping’ measures, like denying consultants freedom to buy whatever surgical instruments take their fancy without regard to cost, do not appear to have any direct implications for patients, unless the result is to buy poor-quality equipment. But seemingly minor changes at the edges may

have a cumulative effect. If secretarial support for doctors is eliminated, if the hours allocated to consultants for supporting professional activities (such as audit and designing protocols) are reduced, if the staff producing patient information leaflets are cut, there may be a potentially demoralising effect. ‘The system runs on good will’, as one doctor remarked, ‘and that good will is now under pressure.’

These examples illustrate a key difference between the purchasers and hospital providers of health care. In the case of the purchasers, the trick is to limit patient demands in order to constrain spending. In the case of providers, the trick is to limit demands once the patient is in hospital. In other words, it is largely clinical decisions that drive spending once patients are in hospital. The following quotation illustrates the point:

*The real issue is not whether to perform the appendectomy; it is whether to fund countless marginal interventions that are potentially part of the procedure – marginal blood tests and repeat tests; precautionary, preventive antibiotic therapy before surgery; the number of nurses in the operating room; and the backup support on call or in hospital. Even more decisions about marginal elements will arise during the recovery phase – exactly how many days of hospital stay are permitted, how often the physician should make rounds, how many follow-up tests there should be, and so on. Many of these are predicted to offer more benefit than harm, but with margins so small that one could argue that resources ought to be used elsewhere.*

(Veatch 1992)

What goes for appendectomy also goes, of course, for other procedures. And what goes for doctors also applies, in spades, to nurses. In short, clinical decisions about how to use resources are an integral, routine part of clinical practice. Much effort is therefore being invested by management in influencing the behaviour of clinicians. Hospital drug formularies have of course long limited clinicians’ discretion. Benchmarking is used to compare the activity of consultants in, for example, their use of scans or expensive drugs. Consultants, in turn, are expected to control junior doctors, who are apt to order too many pathology tests or MRI scans. However, the emphasis appears to be on persuasion through comparison rather than prescription, far less outright proscription: ‘It’s erecting hurdles, not denying services,’ in the words of one clinical director; or, as a manager put it: ‘We want the clinicians to make the rationing decisions.’

The investment in changing clinical behaviour – and the introduction of clinical budgets that require doctors to exercise fiscal discipline (Foot *et al* 2012) – is recognition that it is medical decisions that largely determine who gets what in hospital settings. So, to take the example of one trust, it is clinical directors who make decisions about which drugs are affordable. Such decisions involve

not only considerations of evidence about effectiveness and cost but also judgements about what will work in a particular hospital environment (Jenkins and Barber 2004).

## The secret garden of clinical rationing

Managerial strategies are reasonably visible. But we know remarkably little about what has been called ‘bedside rationing’ by clinicians. Indeed, the most prevalent form of bedside rationing may also be unconscious. An early and still rare attempt to study hospital rationing (Aaron and Schwartz 1984) concluded that British physicians internalised resource constraints in, for example, their parsimonious use of cancer drugs and diagnostic tests when compared to their American counterparts. Lack of resources was translated into good-practice norms, the authors argued. The conclusion was quite persuasive, even though the method of reaching it was questionable, resting on the assumption that American medicine provided an appropriate model of good practice. And it was reached before NICE and proliferating clinical guidelines and protocols limited the scope for idiosyncratic decisions, and good-practice norms became more evidence based.

Again, the NHS in England is not unique in the practice of bedside rationing. It is universal. In a survey of physicians in four European countries – Norway, Switzerland, Italy and the United Kingdom – 56.3 per cent reported that they did indeed ration interventions (Hurst *et al* 2006). The most frequently rationed interventions were MRI and screening tests. The most frequently mentioned criteria were small expected benefits and low chances of success, but a majority of respondents also reported that they were more likely to refrain from using an intervention if the patient was over 85. Surprisingly, the survey showed no correlation between the level of health care spending in the four countries and physician responses: doctors in the highest-spending country, Switzerland, reported the most rationing. However, given a low (43 per cent) response rate from the doctors sampled, and the different incentives to practitioners in the countries concerned, perhaps not too much should be made of this.

A review of the literature identified 15 survey-based studies in the same group of countries plus the United States, Canada and the Netherlands (Stretch *et al* 2009). Overall, 60.7 per cent of physicians in the pooled data, much the same proportion as in the four-country study, reported a willingness to ration. But the review advised caution in interpreting the results. Much depended, unsurprisingly, on the specific wording of the questions. Different wordings produced rather different results: for example, doctors appeared to be more comfortable with accepting that cost considerations had to come into play than with the explicit notion of rationing or withholding beneficial treatment.

Do qualitative studies provide more insights into what actually happens at the bedside? Turning to another literature review (Stretch *et al* 2008), perhaps the most striking fact is that a trawl through the international literature produced only nine studies with sufficient methodological rigour to be included. All of the four UK studies involved only GPs; indeed, only three of the nine focused on hospital clinicians. The review confirmed that ‘physicians see rationing as [a] matter of fact that is widely prevalent in everyday medical practice’. Interestingly, among the very varied criteria cited by physicians were two that are conventionally regarded as ethically questionable: the patient’s ability to exercise pressure and the patient’s contribution to society.

The last finding gives extra edge to perhaps the most interesting conclusion drawn by the review of qualitative studies. This is that most of the reported rationing strategies, such as deferral and deflection, were implicit. This raises an issue that has already been touched on: the appropriate balance between explicit and implicit decision-making. Can implicit rationing be accepted if the result is to cloak decisions that would be regarded as unacceptable if made public? In the concluding discussion that follows, we explore this and other issues further.

---

# Concluding discussion

---

In September 2011 Professor Sir Bruce Keogh, Medical Director of the NHS in England, wrote a letter reiterating Department of Health policies on limiting access to services (Keogh 2011). Addressed to the medical directors of strategic health authorities (SHAs), it was prompted by ‘...press and parliamentary interest in PCTs looking to address their current financial challenges through restricting patients’ ability to access or be referred to some services’. The guiding principle, the letter stressed, was that commissioners must not introduce outright blanket bans for interventions or treatments, but must be sensitive to individual circumstances. Elaborating the point, the letter argued:

*Decisions on the clinical effectiveness of treatments and interventions must take account of a patient’s individual circumstances – different interventions will provide different levels of benefit depending on the individual. Any decision to restrict access to treatment or intervention must be justified in relation to a patient’s individual circumstances, and any individual must be able to challenge such decisions through an exceptional cases review process ... there will be patients for whom their individual clinical circumstances mean an intervention is likely to be of higher value in their case in comparison to other individuals with similar conditions. This means that decisions should not be made solely on the basis of cost, and any refusal to offer the intervention in question must be fair and consistent... In line with this, individuals should always be entitled to argue that their treatment is likely to be of such exceptional benefit that it deserves to be funded by the NHS and to have access to a review process.*

The letter has been quoted at some length because, although the word ‘rationing’ is never used, it illustrates key themes discussed in previous sections. By underlining the heterogeneity of patients, and the variations in their capacity to benefit from interventions, the letter also underlines the limitations of a health-maximising, cost-utility approach. But there remain unresolved questions. If cost is not to be the sole consideration, what other criteria are to be used? How ‘exceptional’ does a benefit have to be to qualify for NHS funding, and in what currency is the benefit or exceptionality to be measured? Should it be increased life expectancy or the quality of that life? And should any such gains be valued more highly in early years or in the last months of life? Even if the technical debate about the strengths and weaknesses of Quality Adjusted Life Years (QALYs) is ignored, even if no money figure is put on the gains in the extent or quality of life, the underlying issues remain. And surely cost, even if it is not the sole determinant, cannot be ignored: can the NHS offer an open cheque?

Given these unresolved (perhaps unresolvable) issues, it is not surprising that the NHS medical director's letter stresses another theme emerging from our analysis: the importance of due process as defined by the courts, with echoes of the Accountability for Reasonableness Framework. However, while the courts can ensure consistency of process – such as ensuring that commissioners have examined all relevant facts – they do not address substantive issues, as noted previously. So, while recognising the importance of process and the requirements of accountability for reasonableness, a fundamental question remains. Should it be the responsibility of individual commissioners to articulate 'a clear and consistent position describing the circumstances in which an individual would receive treatment' or 'a fair and transparent basis for refusing treatment'?

At present the clarity is lacking. Consider the dramatic variations between PCTs in the rate of approval for exceptional funding requests, previously noted (*see* p 25). There may be many reasons for such variations: the criteria used, the nature of the funding requests themselves, the quality of the available evidence, and the financial situation of the PCTs concerned. We simply do not know: this is yet another example of how lack of information inhibits our understanding of rationing in practice. But it does suggest that there is a case for more precise, national definitions of the criteria to be used when determining exceptional funding requests while allowing purchasers to interpret those criteria in the light of local affordability or other factors, always provided that these can meet the transparency condition and be publically justifiable. A starting point could be the principles emerging from legal cases, as set out in the box entitled 'Defining exceptionality: principles emerging from clinical cases' (*see* p 26), and the work of the NHS Confederation (Austin 2008) in devising an appropriate framework.

### Is there a case for postcode rationing?

We have seen that commissioning bodies (PCTs hitherto) do indeed vary in the decisions they make. They have different policies about the introduction of thresholds for operations. They vary in the conditions they impose for IVF treatment, tonsillectomies and bariatric surgery. They differ in their lists of excluded interventions that are deemed to yield only marginal benefits. There are signs of convergence, reflecting increasing collaboration between PCTs drawing on the same body of evidence and NICE guidelines. But postcode rationing – meaning that patients with the same condition but with different addresses may receive different treatment – remains a live issue. Indeed, it could become even livelier with the replacement of PCTs by clinical commissioning groups. This will increase the number of commissioners and risks cutting links with existing networks, so compounding existing diversity in decision-making.

Not surprisingly, therefore, there is pressure to limit the discretion of commissioners. In the summer of 2011 the British Medical Association's annual representative meeting voted for a motion calling on the government to produce 'a consistent set of national standards' for low-priority treatments. This followed strong criticism from doctors of the variation between PCTs in the definition of, and policies on, low-priority procedures. The sensitivity of the medical profession on this point is unsurprising. The introduction of clinical commissioning groups creates a 'double agent' dilemma (Shortell *et al* 1998) for general practitioners. Traditionally, they have seen themselves as the patient's agent. Now, however, they will also have to see themselves as the agents of the collective of practitioners responsible for their budget. Can they maintain the trust of their patients if there is the suspicion that clinical decisions are driven by budgetary considerations?

The demand for limiting the discretion of commissioners has strong public resonance, since postcode rationing seems at odds with the whole concept of a national health care system and Nye Bevan's 1948 ambition to 'generalise the best' – in other words, to ensure that people's treatment would not depend on where they lived. However, 60-odd years later, people's treatment does very often depend on where they live. Concentrating exclusively on decision-making by commissioners and the resulting postcode rationing risks overlooking one crucial fact: the variations in what local 'health economies' deliver – the combined effect of decisions by commissioners and the practices of providers.

Geographical variations in the treatments offered by NHS providers are gross, well-documented and have a long history (Ham 1988). A few may be directly influenced by PCT priorities: thus there is a 38-fold variation in the rate of bariatric surgery between different PCT populations, falling to a 12-fold variation when the top five and the lowest five PCTs are excluded (Right Care 2010). Others cannot be attributed simply to explicit PCT decisions about priorities. The geographical variation in spending on hip replacements between different populations, after allowing for need factors, is 14-fold, falling to four-fold when excluding outliers. Variations are evident not just in elective surgery where patient preferences may play a role or where the private sector provides an alternative. Nor are they a particular feature of contested, low-effectiveness interventions like tonsillectomy (Appleby *et al* 2011). They are the norm.

The NHS is not unique in this respect. Variations are a much-studied international phenomenon. In the case of the United States, high rates of health care utilisation are associated with over-use. If beds and scanners are available, they will be used: supply creates its own demands (Wennberg 2010). In the case of the NHS, over-use may be less of a problem. But neither can it be automatically assumed that low rates of utilisation are necessarily evidence of

inadequacy – and therefore a form of implicit rationing. Such judgements depend on a clinical consensus about the appropriate criteria for decisions to treat, and how – a consensus that in many cases does not appear to exist: doctors are as heterogeneous in their judgements as patients are heterogeneous in their needs.

There is little understanding of the variations in the NHS and what causes them. Some may indeed be attributable to commissioning decisions, as in the case of bariatric surgery cited above. But given that variations were the norm in the NHS long before commissioning was ever dreamed of, they would seem to point to path dependency as the cause: past differences in the distribution of resources and local clinical cultures continue to shape activities, which may have no apparent rationale in the present. Some argue that in a world where the NHS is trying to respond to demands as well as needs, preference-sensitive variations should be distinguished (and treated differently) from supply-sensitive variations, since the former reflect consumer rather than professional decisions. However, given that consumer preferences may well have been shaped over time by the available supply, as mediated by the local clinical practices, it is difficult to sustain the distinction.

From a health economy perspective, therefore, it is both rational and predictable that commissioners have different priorities and different rationing strategies. Consider the example of the 14-fold variation in hip replacements. For a PCT at the top end of the distribution, it would seem rational to introduce a threshold designed to limit the number of operations, calculating that the resources could be used better elsewhere. Conversely, it would seem perverse for a PCT at the other end of the distribution to introduce such a form of rationing by selection. If anything, more such judgements could be made in future, informed by the use of patient-related outcome measures (PROMS) that allow the health improvements following interventions to be measured (Devlin and Appleby 2010). There is a paradox here. Given the existing geographical variations in the availability of treatments, it would be extraordinary if commissioners did not vary their priorities.

So postcode rationing, in the widest sense, cannot be condemned out of hand. Again, the distinction between priority-setting and rationing is relevant. The argument here is that variations in priority-setting are not only defensible but also desirable to the extent that they represent attempts to rebalance local health economies. When priorities about services, patient groups or specific interventions are translated into decisions about individual patients – rationing in the strict sense – a different set of considerations come into play. Here we face a dilemma, with a clash between two desirable aims of policy pulling in opposite directions. On the one hand, there would be widespread support for the principle that the same criteria should be applied to all similar cases regardless

of where people live. So, for example, the conditions of eligibility for IVF or bariatric surgery should be universal. On the other hand, there would also be widespread support for the principle that decisions about resources should take local circumstances into account. So how can the highly desirable aim of treating everyone alike, irrespective of where they live, be reconciled with the fact that both needs and resources for particular services vary and that the conditions of eligibility (and other rationing decisions) may have to take this into account?

Complicating the issue further, there is another dimension to the evidence about variations, so far not analysed and generally ignored in discussions of priority-setting and rationing. These are the variations in the *quality* of the services provided. Although Bevan's already quoted phrase about generalising the best is in some senses an oxymoron, it is a reminder that the goals of the NHS cannot be expressed in quantitative terms alone. So variations in quality matter. And they are considerable. So, for example, NICE guidance is that people with a suspected high-risk transient ischaemic attack (TIA) should receive specialist treatment within 24 hours. In 2009–10, the variation between PCTs was 50-fold, and 10-fold after excluding the outliers (Right Care 2011). This may be an extreme case, and there are some problems of data interpretation. But it is certainly not atypical in demonstrating qualitative variations.

Most worrying perhaps, if less amenable to precise quantification, is the evidence about variations in the quality of care provided for one particular group in the population: older people. Most of the debate about priority-setting is about specific interventions, access to treatment, the availability of different drugs and so on. Much less attention is paid to the quality of the environment in which care is delivered. Yet a succession of reports has drawn attention to the sometimes appalling conditions in the care offered to older people. Two recent examples will suffice, though the point could be illustrated by a stream of reports over the decades. Reporting on a sample of cases, the Health Service Commissioner for England (2011) instanced 'the dismissive attitude of staff, a disregard for process and procedure and an apparent indifference of NHS staff to deplorable standards of care'. An inspection by the Care Quality Commission (2011) picked up cases of care that were putting people at unacceptable risk of harm in one out of five hospitals.

The quality of care cannot be automatically read off the quantity of resources. As the introduction to the CQC's inquiry comments: 'Having plenty of staff does not guarantee good care (we saw unacceptable care on well-staffed wards, and excellent care on understaffed ones) but not having enough is a sure path to poor care.' In short, the causes of poor-quality care need to be teased out carefully in individual cases to identify the relative contribution of inadequate resources and

inadequate management. If the explanation is found to lie in the inadequacy of resources, then poor-quality care is clearly an instance of rationing by dilution.

There remains another important issue, so far not discussed. As we have seen, there is no agreed formula for determining priorities. Cost-utility analysis, QALYs and techniques like programme budgeting are all useful tools in what is bound to be an on-going debate that will continue to generate controversy. Even if there is uniformity in the criteria used, there will still be disagreements about the weight to be given to them and how conflicts between them are to be resolved – hence the importance of process; but hence, too, the importance of who is involved in the debate. The next section addresses this issue.

### Accountability for reasonableness: but to whom?

There is, then, a strong case to be made in defence of local variations in priority-setting and the consequential rationing of access to treatments, provided that they reflect the variations in the local health economies. Much depends, however, on the quality of decision-making; here the courts have an important fall-back role to play as guardians of process to ensure that decisions are not arbitrary or inadequately justified, as we have seen. Even more, perhaps, depends on how the decision-makers are perceived by the public. If postcode rationing has had such a bad press and become a pejorative phrase, it is in part because PCTs were perceived to lack legitimacy. Much has been made of a ‘democratic deficit’.

The coalition government’s initial programme of action (HM Government 2010) included a proposal for directly elected individuals on PCT boards. With PCTs replaced by clinical commissioning groups, two months later, this notion was dropped. Clinical commissioning groups will have two lay members: one with a lead role in championing patient and public involvement, the other with a lead role in overseeing governance arrangements. Neither will be elected. The main instrument for strengthening ‘the local democratic legitimacy of the NHS’ (Secretary of State for Health 2010) is to be the involvement of local authorities. Health and wellbeing boards will bring together local authorities and NHS commissioners. The boards will be involved in reviewing commissioning plans but will not have a right of veto.

It remains to be seen how this system will work and whether it will indeed put a democratic gloss on decisions about priorities and rationing. It seems doubtful. More to the point may be the emphasis on patient and public involvement, a theme much invoked by governments over the decades and now played fortissimo. But just how should the public be involved in decisions about priority-setting and rationing?

The evidence suggests that the public is ambivalent about direct involvement in rationing decisions (Lomas 1997; Russell *et al* 2011). A recent study shows that while there is support for the right to be involved, most members of the public think that rationing and prioritising decisions should be taken by doctors (Chisholm *et al* 2009). In this respect, if no other, the Secretary of State's decision to put commissioning in the hands of general practitioners seems to be in accord with public opinion. Whether public enthusiasm for leaving it to the doctors will be sustained if priority-setting is farmed out to private organisations is another matter.

The rhetoric of public involvement, as a legitimising device for rationing decisions, needs to be used with care. There may well be scope for presenting evidence to 'citizens' juries', whose members can cross-examine the experts and discuss the issues in depth. But the 1990s wave of enthusiasm for deliberative democracy (McIver 1998) seems to have spent itself. The case of NICE's Citizens' Council may explain why. Its proceedings are impressive, a record of thoughtful discussion. But they are impressive because of the time and resources invested in the process: replication by a multitude of commissioning groups would be beyond their administrative and budgetary capacity.

If direct involvement by the public is problematic, there is the alternative of proxy involvement by organisations acting for the public. The history of bodies designed to articulate the interests of patients and public is not a happy one, with an ever-accelerating rate of institutional euthanasia. Over the past decade, Community Health Councils have been abolished in favour of Patient and Public Involvement Forums, only for these to give way to Local Involvement Networks. Now the Networks are to metamorphose into local HealthWatch, with a requirement that membership should be representative of different users. HealthWatch England, based in the Care Quality Commission (CQC), is to provide central support.

If past history suggests a short shelf life and at best variable impact, one aspect of the new institutional arrangements gives cause for cautious optimism. This is the link of HealthWatch with the CQC. Potentially this could give local HealthWatch agencies what their predecessors have conspicuously lacked: access to the information and analytical capacity needed to put the performance of their health economy into a comparative perspective, so providing context for one-off decisions about priorities and rationing. Not only that: drawing on the CQC's battery of indicators, as well as being informed by the experience of patients, they should be better able to identify rationing by dilution. Similarly, local authority scrutiny committees – whose power to question the performance of NHS commissioners and providers has survived – will in future have readier access to independent analysis and support, as a result of the public health function moving from the NHS to local government.

The importance of the potential for an increased capacity to analyse what is happening lies in the fact, stressed throughout this paper, that priority-setting and rationing are a continuing, ever-present feature of the NHS, as of all health care systems. It is dramatic, one-off, often tragic, cases that may capture the headlines (Ham and Pickard 1998) and thus structure debate. Similarly, public attention tends to be caught by commissioner decisions to restrict access to drugs or interventions. But the reality is, as we have argued, complex, messy and opaque.

## Issues for future debate

This paper has been an exercise in exploration, not prescription. We therefore conclude by drawing out some implications from the issues identified:

- Different commissioners will continue to make different decisions about priorities, and to be criticised accordingly. But given the variations in local circumstances and services, some differences are inevitable. The critical test here is how the decisions have been taken and whether they can be publicly justified in the light of evidence.
- Given the local variations in the lists of treatments that are not funded by commissioners, there is a temptation to argue for a national menu. The experience so far argues against such a policy. The local exclusions from the menu tend to be marginal interventions so that the savings are marginal also. And the bureaucratic costs of running a system for dealing with exceptional funding requests are considerable. There is little reason to expect a national system to be so very different. Moreover, commissioners would almost certainly have to be selective in their use of the national menu to reflect local circumstances, so uniformity would still remain an illusory policy goal.
- Although the case for a national menu is weak, there is a strong argument for devising national criteria to be used in local decision-making when dealing with exceptional funding applications. However, too much should not be expected. Uniformity is neither achievable nor perhaps even desirable. Given the heterogeneity (and fallibility) of medical judgements, and the ambiguity of much evidence for medical interventions, seeming anomalies are to be expected.
- There is a tension at the heart of Mr Lansley's new-model NHS: the double-agent dilemma. On the one hand, the introduction of clinical commissioning groups accurately reflects public confidence in the medical profession, as distinct from managers or politicians, to make rationing decisions. On the other hand, in the new model, the public's trust may be tested by the suspicion that decisions are being taken for financial rather than clinical

reasons. It remains to be seen whether public trust will make the rationing decisions of clinical commissioning groups acceptable, or whether those decisions will undermine that trust.

- Given hard times ahead for the NHS in England, rationing by dilution is likely to be an evermore important phenomenon, not limited to care for older people or other vulnerable groups. It is the most insidious, because it is the least visible, form of rationing. The challenge to HealthWatch and local scrutiny committees, therefore, is to give it visibility and thus to compel justification. Some forms of dilution may indeed be acceptable – but the case has to be argued.
- The case of older people illustrates the importance of analysing the effects of priority-setting and rationing on the quality as well as the quantity or accessibility of care. But it also raises a question about consistency in decision-making. On the one hand, a high value has been attached to improving both the quantity and the quality of the last few months of life for cancer patients. On the other hand, this does not appear to translate into attaching an equally high value to the quality of life of older patients in hospital (many of whom also will not have long to live). Is there some systemic bias in existing approaches to deciding on priorities that favours interventions as distinct from focusing on the quality of care? And if so, how might this be overcome?
- Most of the day-to-day rationing decisions about who should get what in the NHS are taken by clinicians – doctors and nurses – as part of their routine daily activities. That is as it should be. But we know very little about the secret garden of clinical rationing. While it would be unreasonable to expect clinicians to account individually for such decisions, we need a better understanding of the collective impact of those decisions on the care patients get as a result. Filling in this and other black holes in our existing knowledge could lead to an improved understanding of the impact of decisions about resources on patients, and thus to a better-informed debate about priority-setting and rationing.

---

# References

---

- Aaron HJ, Schwartz WB (1984). *The Painful Prescription*. Washington, DC: The Brookings Institution.
- Adelphi Research UK (2009). 'A Comprehensive Survey of PCTs to Evaluate Local Decision-making Processes for Funding New Medicines.' National Prescribing Centre website. Available at: [www.npc.nhs.uk/local\\_decision\\_making/resources/npc\\_survey\\_report](http://www.npc.nhs.uk/local_decision_making/resources/npc_survey_report).(accessed on 28 March 2012).
- Airoldi M, Bevan G (2010). *Improvement in Practice: Commissioning with the community*. London: The Health Foundation.
- All-Party Parliamentary Group on Infertility (2011). 'Holding back the British IVF revolution? A report into NHS IVF provision in the UK today'. Available at: [www.garethjohnsonmp.co.uk/uimages/File/appg\\_IVF\\_report.pdf](http://www.garethjohnsonmp.co.uk/uimages/File/appg_IVF_report.pdf) (accessed on 14 December 2011).
- Alzheimer's Society (2011). 'When Does the NHS Pay for Care? Guidance on eligibility for NHS continuing healthcare funding in England and how to appeal if it is not awarded. Version 4'. Alzheimer's Society website. Available at: [http://alzheimers.org.uk/site/scripts/download\\_info.php?fileID=75](http://alzheimers.org.uk/site/scripts/download_info.php?fileID=75) (accessed on 12 January 2012).
- Appleby J (2011). *How is the NHS Performing? Quarterly monitoring report*. London: The King's Fund.
- Appleby J, Raleigh V, Frosini F, Bevan G, Gao H, Lyscom T (2011). *Variations in Health Care: The good, the bad and the inexplicable*. London: The King's Fund.
- Appleby J, Devlin N, Parkin D (2007). 'NICE's cost effectiveness threshold: how high should it be?'. *British Medical Journal*, vol 335, pp 358–9.
- Associated Provincial Picturehouses Limited v Wednesbury Corporation [1947] EWCA Civ 1. Available at: [www.bailii.org/ew/cases/EWCA/Civ/1947/1.html](http://www.bailii.org/ew/cases/EWCA/Civ/1947/1.html) (accessed on 14 December 2011).
- Audit Commission (2011). *Reducing Spending on Low Value Clinical Treatments*. Health Briefing April 2011. Audit Commission website. Available at: [www.audit-commission.gov.uk/sitecollectiondocuments/downloads/20110414reducingexpenditure.pdf](http://www.audit-commission.gov.uk/sitecollectiondocuments/downloads/20110414reducingexpenditure.pdf) (accessed on 14 December 2011).
- Austin, D (2008). *Priority-setting: Managing individual funding requests*. London: The NHS Confederation.

- Ayling G (2011). 'Dear Colleague Letter. NHS Continuing Healthcare (NHS CHC) National Framework 2007 – revised 2009'. Department of Health website. Available at: [www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/documents/digitalasset/dh\\_128295.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_128295.pdf) (accessed on 12 January 2012).
- Bevan G (2009). 'The search for a proportionate care law by formula funding'. *Financial Accountability and Management*, vol 25, no 4, pp 391–410.
- Blustein J, Marmor TR (1992). 'Introduction to rationing'. *University of Pennsylvania Law Review*, vol 140, no 5, pp 1539–42.
- Brambleby P, Jackson A, Gray JM (2008). *Programme-based Decision-making for Better Value Healthcare*. Oxford: NHS National Knowledge Service.
- Calabresi G, Bobbitt P (1978). *Tragic Choices*. New York: Norton & Co.
- Care Quality Commission (2011). *Dignity and Nutrition Inspection Programme: National overview*. London: CQC.
- Carter N, Klein R, Day P (1992). *How Organizations Measure Success: The use of performance indicators in government*. London: Routledge.
- Chisholm A, Briggs K, Askham J (2009). *Not NICE: Can PCTs engage patients and the public in commissioning new health technologies?*. Oxford: Picker Institute.
- Co-operation and Competition Panel (2011). 'Review of the Operation of "Any Willing Provider" for the Provision of Routine Elective Care. Final Report'. Co-operation and Competition Panel website. Available at: [www.ccp-panel.org.uk/content/cases/Operation\\_of\\_any\\_willing\\_provider\\_for\\_the\\_provision\\_of\\_routine\\_elective\\_care\\_under\\_free\\_choice/280711\\_AWP\\_Review\\_Final\\_Report.pdf](http://www.ccp-panel.org.uk/content/cases/Operation_of_any_willing_provider_for_the_provision_of_routine_elective_care_under_free_choice/280711_AWP_Review_Final_Report.pdf) (accessed on 12 December 2011).
- Coulter A, Ham C (eds) (2000). *The Global Challenge of Health Care Rationing*. Buckingham: Open University Press.
- Council of Civil Service Unions v Minister for the Civil Service* [1985] AC 374.
- Daniels N, Sabin J (2002). *Setting Limits Fairly: Can we learn to share medical resources?*. New York: OUP.
- Department of Health (2011a). *Patient Choice. The Department of Health's response to the Co-operation and Competition Panel's report on patient choice* [online]. Available at: [www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/documents/digitalasset/dh\\_131089.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131089.pdf) (accessed on 12 December 2011).
- Department of Health (2011b). 'Table Showing People Receiving Continuing NHS Care, by PCT, 2009–2011'. Department of Health website. Available at: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_103162](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_103162) (accessed on 12 January 2012).

- Department of Health (2009). 'Directions to Primary Care Trusts and NHS Trusts Concerning Decisions about Drugs and Other Treatments'. Department of Health website. Available at: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsLegislation/DH\\_096067](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsLegislation/DH_096067) (accessed on 14 December 2011).
- Department of Health (2007). *The National Framework for NHS Continuing Healthcare and NHS-funded Nursing Care* [online]. Available at: [http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_076288](http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_076288) (accessed on 13 April 2012).
- Devlin N, Appleby J (2010). *Getting the Most out of PROMs*. London: The King's Fund.
- Devlin N, Parkin D (2004). 'Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis.' *Health Economics*, vol 13, no 5, pp 437–52.
- Dolan P, Lee H, King D, Metcalfe R (2009). 'Valuing health directly'. *British Medical Journal*, vol 339, p 2577.
- Donaldson C (2008). 'Moving forward on rationing: an economic view'. *British Medical Journal*, vol 337, p 1872.
- Dow J (2009). 'Continuing care: The story so far'. *Journal of Integrated Care*, vol 17, no 4, pp 12–15.
- Foot C, Sonola L, Maybin J, Naylor C (2012). *Service Line Management: Can it improve quality and efficiency*. London: The King's Fund. Available at: [www.kingsfund.org.uk/publications/slm\\_paper.html](http://www.kingsfund.org.uk/publications/slm_paper.html) (accessed on 7 February 2012).
- Foot C, Naylor C, Imison C (2010). *The Quality of GP Diagnosis and Referral*. London: The King's Fund. Available at: [www.kingsfund.org.uk/current\\_projects/gp\\_inquiry/dimensions\\_of\\_care/diagnosis\\_and.html](http://www.kingsfund.org.uk/current_projects/gp_inquiry/dimensions_of_care/diagnosis_and.html) (accessed on 12 December 2011).
- Ford A (forthcoming) 'The concept of exceptionality – a legal farce?'. *Medical Law Review*.
- Foster C (2007). 'Simple rationality? The law of healthcare resource allocation in England'. *Journal of Medical Ethics*, vol 33, pp 404–7.
- Gillon R (1994). 'Medical ethics: four principles plus attention to scope'. *British Medical Journal*, vol 309, p 184.
- Ham C (ed) (1988). *Health Care Variations: Assessing the evidence*. London: The King's Fund.
- Ham C, Pickard S (1998). *Tragic Choices in Health Care*. London: The King's Fund.

- Health Service Commissioner for England (2011). *Care and Compassion? Report of the Health Service Ombudsman on ten investigations into NHS care of older people*. HC778. London: The Stationery Office.
- HM Government (2010). *The Coalition: Our programme for government*. London: HM Government.
- House of Commons Health Committee (2007). *National Institute for Health and Clinical Excellence*. First Report, Session 2007–08. HC27. London: The Stationery Office.
- Hurst SA, Slowther A-M, Forde R, Pegoraro R, Retter-Thell S, Perrier A, Garrett-Mayer E, Danis M (2006). 'Prevalence and determinants of physician bedside rationing'. *Journal of General Internal Medicine*, vol 21, pp 1138–43.
- Imison C, Naylor C (2010). *Referral Management: Lessons for success*. London: The King's Fund. Available at: [www.kingsfund.org.uk/publications/referral\\_management.html](http://www.kingsfund.org.uk/publications/referral_management.html) (accessed on 12 December 2011).
- Jeffries B (2011). 'Surgeons' alarm over waiting'. BBC News website. 5 April. Available at: [www.bbc.co.uk/news/health-12964360](http://www.bbc.co.uk/news/health-12964360) (accessed on 12 December 2011).
- Jenkins KN, Barber N (2004). 'What constitutes evidence in hospital drug decision making?' *Social Science & Medicine*, vol 58, pp 1757–66.
- Klein R (1983). 'The NHS and the theatre of inadequacy'. *Universities Quarterly*, vol 37, no 3, pp 201–15.
- Klein R, Day P, Redmayne S (1996). *Managing Scarcity: Priority setting and rationing in the National Health Service*. Buckingham: Open University Press.
- Keogh B (2011). 'Dear SHA Medical Director, Re-Access to services'. Email to SHA medical directors, 21 September.
- Lomas J (1997). 'Reluctant rationers: public input to health care priorities'. *Journal of Health Services Research & Policy*, vol 2, no 2, pp 103–11.
- Maynard A, Bloor K (2009). 'NICE wobbles'. *Journal of the Royal Society of Medicine*, vol 102, pp 212–13.
- McIver S (1998). *Healthy Debate? An independent evaluation of citizens' juries in health settings*. London: The King's Fund.
- Mechanic D (1997). 'Muddling through elegantly: finding the proper balance in rationing'. *Health Affairs*, vol 16, no 5, pp 83–92.
- Moberly T (2011). 'Third of PCTs increase treatment restrictions'. GP website. Available at: [www.gponline.com/News/article/1081588/Online-exclusive-Third-PCTs-increase-treatment-restrictions/](http://www.gponline.com/News/article/1081588/Online-exclusive-Third-PCTs-increase-treatment-restrictions/) (accessed on 12 December 2012).

National Prescribing Centre and Department of Health (2009). 'Defining guiding principles for processes supporting local decision making about medicines' [online]. Department of Health website. Available at: [www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_093413](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_093413) (accessed on 14 December 2011).

Newdick C (2005). *Who Should We Treat? Rights, rationing and resources in the NHS*. Oxford: OUP.

NHS Confederation (2008). *Priority Setting: Managing individual funding requests*. London: NHS Confederation. Available at: [www.nhsconfed.org/Publications/Documents/Priority%20setting%20managing%20individual%20funding%20requests.pdf](http://www.nhsconfed.org/Publications/Documents/Priority%20setting%20managing%20individual%20funding%20requests.pdf) (accessed on 14 December 2011).

NICE (2009). *Social Value Judgments: Principles for the development of NICE guidance*. 2nd ed. London: NICE.

Oliver A, Sorensen C (2005). 'The limits and challenges to the economic evaluation of health technologies' in Costa-Font J, Courbage C, McGuire D (eds), *Economics of Innovation and New Health Care Technologies*. Oxford: OUP.

Public Health Commissioning Network (2010). 'List of Priorities Statements'. Public Health Commissioning Network website. Available at: [www.phcn.nhs.uk/PCT%20priorities%20statements.pdf](http://www.phcn.nhs.uk/PCT%20priorities%20statements.pdf) (accessed on 14 December 2011).

Pulse (2011). 'Hospitals axe 1.2m follow-up appointments in a year'. Pulse e, 16 November. Available at: [www.pulsetoday.co.uk/newsarticle-content/-/article\\_display\\_list/13052928/hospitals-axe-1-2m-follow-up-appointments-in-a-year](http://www.pulsetoday.co.uk/newsarticle-content/-/article_display_list/13052928/hospitals-axe-1-2m-follow-up-appointments-in-a-year) (accessed on 12 December 2011).

R v NW Lancashire Health Authority, ex parte A, D and G [1999]. Lloyd's Rep Med 399.

Raftery J (2009). 'NICE and the challenge of cancer drugs'. *British Medical Journal*, vol 329, pp 224–7.

Rarer Cancers Forum (2008). *Taking Exception. An audit of the policies and processes used by PCTs to determine exceptional funding requests* [online]. Rarer Cancers Forum website. Available at: [www.rarerancers.org.uk/images/stories/cdf/p8and9/taking%20exception.pdf](http://www.rarerancers.org.uk/images/stories/cdf/p8and9/taking%20exception.pdf) (accessed on 14 December 2011).

Rawlins ND, Culyer AJ (2004). 'National Institute for Clinical Excellence and its value judgments'. *British Medical Journal*, vol 329, pp 224–7.

Richards M (2008). *Improving Access to Medicines for NHS Patients: A report for the Secretary of State for Health by Professor Mike Richards* [online]. London: Department of Health. Available at: [www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/@dh/@en/documents/digitalasset/dh\\_089952.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_089952.pdf) (accessed on 14 December 2011).

- Right Care (2011) *The NHS Atlas of Variation in Healthcare*. 2nd ed. London: Right Care.
- Right Care (2010). *The NHS Atlas of Variation in Healthcare*. London: Right Care.
- Robinson S, Dickinson H, Williams I, Freeman T, Rumbold B, Spence K (2011). *Priority Setting: An exploratory study of English PCTs*. London: The Nuffield Trust.
- Roland MO, Porter RW, Matthews JG, Redden JF, Simonds GW, Bewley B (1991). 'Improving care: a study of orthopaedic outpatient referrals'. *British Medical Journal*, vol 302, no 6785, pp 1124–8.
- Royal College of Nursing (2011). *Views From the Frontline: RCN Employment survey 2011*. London: RCN.
- Russell J, Greenhalgh T, Burnett A, Montgomery J (2011). 'Public and patient involvement in individual healthcare rationing'. *British Medical Journal*, vol 342, pp 1399–401.
- Samuel M (2011a). 'Councils face mounting pressures as PCTs cut continuing care'. *Community Care*. Available at: [www.communitycare.co.uk/Articles/10/08/2011/117280/councils-face-mounting-pressures-as-pcts-cut-continuing-care.htm](http://www.communitycare.co.uk/Articles/10/08/2011/117280/councils-face-mounting-pressures-as-pcts-cut-continuing-care.htm) (accessed on 12 January 2012).
- Samuel M (2011b). 'PCTs facing more appeals against continuing care decisions'. *Community Care*. Available at: [www.communitycare.co.uk/Articles/15/10/2011/117663/pcts-facing-more-appeals-against-continuing-care-decisions.htm](http://www.communitycare.co.uk/Articles/15/10/2011/117663/pcts-facing-more-appeals-against-continuing-care-decisions.htm) (accessed on 12 January 2012).
- Secretary of State for Health (2011). *Government Response to the NHS Future Forum Report*. Cm 8113. London: The Stationery Office.
- Secretary of State for Health (2010). *Equity and Excellence: Liberating the NHS*. Cm7881. London: The Stationery Office.
- Shortell SM, Waters TM, Clarke KW, Budetti PP (1998). 'Physicians as double agents: Maintaining trust in an era of multiple accountabilities'. *Journal of the American Medical Association*, vol 280, pp 1102–8.
- Solutions for Public Health (2010). 'Milton Keynes, Oxfordshire, Berkshire East, Berkshire West & Buckinghamshire (MOBBB) Priorities Committee. Minutes of meeting held on Wednesday 15 December 2010'. Oxfordshire PCT website. Available at: [www.oxfordshirepct.nhs.uk/professional-resources/priority-setting/minutes/documents/MOBBBminutesDecember2010.pdf](http://www.oxfordshirepct.nhs.uk/professional-resources/priority-setting/minutes/documents/MOBBBminutesDecember2010.pdf) (accessed on 12 December 2011).
- Stretch D, Persad G, Marckmann G, Danis M (2009). 'Are physicians willing to ration health care? Conflicting findings in a systematic review of survey research'. *Health Policy*, vol 90, pp 113–24.

- Stretch D, Synofzik M, Markmann G (2008). 'How physicians allocate scarce resources at the bedside: a systematic review of qualitative studies.' *Journal of Medicine and Philosophy*, vol 33, pp 80–99.
- Sullivan R, Peppercorn J, Sikora K *et al* (2011). 'Delivering affordable cancer care in high-income countries.' *The Lancet*, vol 12, pp 933–79.
- Syrrett K (2003). 'A technocratic fix to the "legitimacy problem"? The Blair government and health care rationing in the United Kingdom.' *Journal of Health Politics, Policy and Law*, vol 28, no 4, pp 715–46.
- Veatch RM (1992). 'The Oregon Experiment: Needless and real worries' in Strosberg MA *et al* (eds), *The Oregon Plan and Beyond*. Washington, DC: The Brookings Institution.
- Watts v Bedfordshire PCT, Case C-372/04. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62004J0372:EN:HTML> (accessed on 2 May 2012).
- Wennberg J (2010). *Tracking Medicine*. Oxford: OUP.
- Wicks E (2009). 'Right to receive treatment in accordance with the Human Rights Act 1998.' *British Journal of Nursing*, vol 18, pp 1192–3.
- Williams A (1997). 'Intergenerational equity: an exploration of the "fair innings" argument.' *Health Economics*, vol 6, no 2, pp 117–32.
- Williams A (1992). 'Cost effectiveness analysis: is it ethical?'. *Journal of Medical Ethics*, no 18, pp 7–11.
- Wilson G (2011). 'Health Secretary bans hospital minimum waiting times'. *Sun*, 14 November. Available at: [www.thesun.co.uk/sol/homepage/news/politics/3933556/Health-Secretary-bans-hospital-minimum-waiting-times.html](http://www.thesun.co.uk/sol/homepage/news/politics/3933556/Health-Secretary-bans-hospital-minimum-waiting-times.html) (accessed on 12 December 2011).
- Yorkshire and Humber Public Health Observatory (2010). 'Prioritisation Frameworks and Tools' Yorkshire and Humber Public Health Observatory website. Available at: [www.yhpho.org.uk/resource/item.aspx?RID=69847](http://www.yhpho.org.uk/resource/item.aspx?RID=69847) (accessed on 13 April 2012).