

Organisations such as the National Institute of Clinical Excellence seek to assess the value for money of new health care technologies. Assessment commonly requires the use of thresholds or benchmark levels of cost effectiveness. Key issues that consequently need to be resolved include: the basis on which thresholds should be determined, how explicitly these should be stated and whether UK health care thresholds should be comparable to those elsewhere in the public sector, or in other countries.

Based on a seminar jointly hosted by the Office of Health Economics and the King's Fund in March 2002, this book follows the discussion between those working for NICE, the King's Fund, the Office of Health Economics, the Department of Health and eminent academic economists.

The use of thresholds

The book begins by examining the role of cost-effectiveness thresholds and issues of openness and transparency of their use. NICE is asked, and answers, whether it has a threshold of around £30,000 per QALY given the decisions it has made since its inception.

How are thresholds used elsewhere?

It is argued the common health care valuation tool, QALY, may be outdated and complicating the process of transferability and comparability with the UK public sectors. Equally, comparability to NICE's equivalents overseas is limited given the differing objectives and values of these decision making processes.

Ethical issues regarding a cost-effectiveness threshold

Ethical issues are associated with use of thresholds, as, whether implicit or explicit they invoke rationing decisions. To be acceptable public support will be needed and a communal rather than individual value placed on health care.

Public involvement

How are the public to be involved in health care decision making? The objectives and the difficulties of including public opinion are examined. The NICE proposal of using a Citizens Council (subsequently established) is discussed.

Next steps

Finally, how can bodies like NICE become more sophisticated in their decision making processes. This requires more research into creating a common currency for health care costs and outcomes across the whole of the NHS, valuing QALYs through a greater understanding of the public's 'willingness to pay', incorporating other factors into the decision-making process and looking at the process of implementing decisions at a local level.

COST-EFFECTIVENESS THRESHOLDS

Economic and ethical issues

Editors

Adrian Towse

Clive Pritchard

Nancy Devlin

 King's Fund



COST-EFFECTIVENESS THRESHOLDS: ECONOMIC AND ETHICAL ISSUES

Edited by Adrian Towse, Clive Pritchard and Nancy Devlin

King's Fund

King's Fund
11-13 Cavendish Square
London W1G 0AW
www.kingsfund.org.uk



Office of Health Economics
12 Whitehall
London SW1A 2DY
www.ohe.org

© King's Fund and Office of Health Economics, November 2002

Any part of this publication may be reproduced without permission for educational and non-profit purposes only if the source is acknowledged. All other rights reserved, including the right of reproduction in whole or in part in any form.

Price £10.00

ISBN 1 899040 57 9

Printed by BSC Print Ltd, London.

Office of Health Economics

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

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

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and income from commissioned research and sales of its publications.

King's Fund

The King's Fund is an independent charitable foundation working for better health, especially in London. We carry out research, policy analysis and development activities, working on our own, in partnerships, and through grants. We are a major resource to people working in health, offering leadership and education courses; seminars and workshops; publications; information and library services; a specialist bookshop; and conference and meeting facilities.

Acknowledgements

These papers are based on the proceedings of a workshop jointly organised by the OHE and the King's Fund and hosted by the King's Fund on 1 March 2002.

We are grateful to Tony Culyer, Peter Littlejohns, Alan Williams and Anna Coote for each chairing one of the four sessions.

CONTENTS

Contributors	5
Glossary	6
Executive Summary	7
Introduction	9
Tony Culyer	

Section 1: The use of thresholds

1	An introduction to the use of cost-effectiveness thresholds in decision making: what are the issues? Nancy Devlin	16
2	Does NICE have a threshold? An external view Adrian Towse and Clive Pritchard	25
3	Does NICE have a threshold? A response Peter Littlejohns	31
4	Discussion: The use of thresholds	38

Section 2: How are thresholds used elsewhere?

5	Valuing life years and QALYs: 'transferability' and 'convertibility' of values across the UK public sector Graham Loomes	46
6	Overseas approaches to decision making Clive Pritchard	56
7	Discussion: How are thresholds used elsewhere?	69

Section 3: Ethical issues regarding a cost-effectiveness threshold

8	Ethical issues associated with rationing and the use of thresholds David Cook	76
---	--	----

CONTENTS

9	A commentary on ethical issues Alan Williams	85
10	Discussion: Ethical issues regarding the cost-effectiveness threshold	91
Section 4: Public involvement		
11	Public involvement in health care decision making Robin Clarke	96
12	The Citizens Council Peter Littlejohns	104
13	Discussion: Public involvement	107
Section 5: Next steps		
14	What research is required? Nancy Devlin and Adrian Towse	114
Section 6: Appendices		
	Appendix 1 Supporting analysis for Table 1	122
	Appendix 2 What should NICE's threshold be? A survey of the opinions of workshop delegates	135
	Appendix 3 List of attendees	137

CONTRIBUTORS

Robin Clarke

Senior Research Fellow, Public Involvement, Institute for Public Policy Research (IPPR)

David Cook

Chaplain, Green College, University of Oxford

Tony Culyer

Professor of Economics, University of York, and Chairman of the Office of Health Economics Policy Board

Nancy Devlin

Professor of Health Economics, City University
Formerly Research Fellow, Kings Fund

Peter Littlejohns

Clinical Director, National Institute for Clinical Excellence

Graham Loomes

Professor of Economics, School of Economics and Social Studies,
Centre for the Economic and Behavioural Analysis of Risk and
Decision (CEBARD), University of East Anglia

Clive Pritchard

Health Economist, Office of Health Economics

Adrian Towse

Director of the Office of Health Economics

Alan Williams

Professor of Economics, University of York

GLOSSARY

ICER	Incremental cost-effectiveness ratio
IPPR	Institute for Public Policy Research
LYG	Life year gained
MS	Multiple sclerosis
NICE	National Institute for Clinical Excellence
PBAC	Pharmaceutical Benefits Advisory Committee (Australia)
PHARMAC	Pharmaceutical Management Agency (New Zealand)
QALY	Quality-adjusted life year
VOLY	Value of a life year
VORLE	Value of remaining life expectancy
VPF	Value of preventing a fatality
WTP	Willingness to pay

EXECUTIVE SUMMARY

- Decisions about what is ‘acceptable value for money’ in health care requires some benchmark, or ‘threshold’, level of cost effectiveness. Key issues to be resolved by decision makers, such as the UK’s National Institute of Clinical Excellence (NICE), include: On what basis should an appropriate threshold be determined? Should the threshold be stated explicitly? Should the threshold applied for health care be consistent with other public sector decisions? How (and what) factors other than cost effectiveness should be taken into account in decision making?
- NICE has made it clear that there is no single cost effectiveness benchmark above which health technologies are automatically rejected and below which technologies are accepted. However, decisions made by NICE to date, as well as its published statements, imply that a health technology is more likely to be viewed favourably, subject to other relevant considerations, if it costs less than £30,000 for each (equivalent) year of perfect health it produces. Questions remain as to whether this is the right threshold and as to what grounds there might be for exceptions to such a threshold.
- One approach to selecting a threshold is to base it on the preferences of the general public and the value they attach to health improvements. This approach has been widely used in other sectors – such as road safety – for many years. Simply transferring and converting these values to the health sector is problematic, but the techniques used to obtain the values can be applied to health care.
- Judging what is good value for money is an issue for all health care systems that use cost-effectiveness evidence in their decisions. However, differences between the health systems in each country, the way such decisions are managed, and in the willingness and ability of a country’s citizens to pay for health improvements, means cross-country comparisons of thresholds are of limited relevance to the UK.
- The choice of a threshold invokes ethical, as well as efficiency, issues. An explicit threshold is more likely to result in rationing

EXECUTIVE SUMMARY

decisions that are transparent and consistent, but to be accepted as valid it will require careful justification and command public support. Basing a threshold on the 'average' values of the general public raises the challenges of how (and whether) diverse local and individual views and preferences can be taken into account.

- A key challenge, therefore, is to find appropriate means by which the public's views and values on value for money can be sought. Options include surveys to obtain the general public's valuation of health improvements, and the use of a citizens council (subsequently established by NICE on 8 November 2002).
- Further research is also required to clarify what factors other than cost effectiveness are relevant to decisions. For example, what does 'fairness' mean in this context and how much 'health gain' are we prepared to give up to achieve it?
- The way in which national decisions about value for money are implemented locally, and how this affects the mix of health services provided in practice, needs to be explored.

Introduction: Ought NICE to have a cost-effectiveness threshold?

TONY CULYER

The papers in this book are based on the proceedings of a workshop jointly organised by the OHE and the King's Fund and hosted by the King's Fund on 1 March 2002.

The authors of the papers that make up this book were asked to address a specific question: 'Ought NICE to have a threshold incremental cost-effectiveness ratio (ICER)?' – with an implicit second question 'Should the threshold be explicit?'

Since its inception, NICE (the National Institute for Clinical Excellence) has made a very firm and public commitment to maximise the openness and transparency of its criteria and the processes through which they are developed and applied. It therefore seems uncontroversial that any such threshold ought to be explicit.

In Chapter 1, Nancy Devlin argues that economists are in favour overwhelmingly of explicitness and openness. That does not mean, however, that there are not pros and cons, benefits and even costs associated with transparency, as anyone who ever tried to run a transparent organisation will know only too well. Nor are the costs trivial. There are a whole host of issues that would arise for NICE were it to adopt a threshold formally. These issues relate to its credibility and very survival.

For example, suppose that NICE were to be very explicit on a cost per quality-adjusted life year (QALY) threshold of say £32,500 and then, in the very next decision that it made, it were to approve a technology with an ICER of £40,000. One can imagine the headlines: 'Inconsistencies in NICE'; 'Politicians get at NICE'; 'NICE is corrupt', and so on. In fact, the higher figure may be a perfectly consistent application of a rather more sophisticated decision rule than an automatic cut-off value, or it may result from the application of multiple-decision guidelines. The issues about managing an environment in which decision rules are transparent but complex, and indeed whether this environment is manageable at all, are important. There is an issue about how ahead of the field NICE can afford to be in terms of sophistication.

INTRODUCTION

Whatever the rightness or wrongness of NICE's current commitment to openness, the focus in the OHE/King's Fund workshop was on the implicitness of any threshold – if there was one – in the past. Inferring what the threshold was, or the range in which it may have sat, occupied much of the discussion and is the main topic of the opening chapters.

The discussion at the workshop rapidly fell into three approaches to thresholds:

- inferring thresholds from past decisions;
- setting thresholds in order to exhaust a budget optimally (given an arbitrary budget, this may not generate an overall optimum);
- setting thresholds optimally in order to determine an optimal budget.

Approach 1

The first approach (attempted by Adrian Towse and Clive Pritchard in Chapter 2) necessarily assumes that the past decision makers have been implicitly or explicitly operating according to some decision rule. If one knows what the rule was and can confidently believe that it was consistently applied and that there were no other rules applied with which it might conflict, then a threshold may be inferred. Towse and Pritchard find a fairly narrow range in which it appears to have sat.

An interesting topic for further enquiry would be to ask what one might legitimately infer in the presence of more than one allocation rule. For example, suppose that to the objective of maximising health gain one added maximising some version of an equitable distribution of that gain. What issues arise in correctly inferring the efficiency threshold in such circumstances? Unlike the second and third approaches, this first one is not (at least not directly) normative in nature. It seeks to discover what is and the fact that the 'what is' is a value does not itself make the analysis normative in character.

Approach 2

The second approach is probably what members of NICE and its

Appraisal Committee would recognise as the role for ICERs – if there were one at all. Here, as Nancy Devlin outlines, the objective is to identify the most cost-effective interventions (i.e. those with the lowest £ per QALY) that can be undertaken given the budget available to the NHS. The challenge for NICE would therefore be whether the technologies it is asked to appraise have a better or worse £ per QALY than interventions currently undertaken by the NHS. However, as Devlin notes, this requires information on what is happening elsewhere in the NHS which is simply not available.

Approach 3

The third approach is far more ambitious than the others. It takes us into the unfortunately named territory of ‘affordability’. In particular, it seems to challenge the current positions of NICE and of politicians of all parties – namely, that the overall budget for the NHS (or any part of it) is not for NICE to set, but for properly accountable politicians to determine.

It’s difficult to argue against such a view, not only on grounds of democratic accountability, but also on grounds of lack of information. NICE simply does not have (and nor is it mandated to acquire) the kind of information about outputs in non-health sectors that it would need to form necessary judgements about the marginal costs and benefits of health spending versus spending in other areas of public services. So, if NICE is to have thresholds at all, they are to be seen in a rather partial equilibrium context. However, that need not be a bad thing. After all, whether the NHS budget is too small or too large, it seems a good thing to spend it in such a way as to maximise the objectives of the NHS, i.e. to maximise health gain.

One topic in our discussion of what a threshold ICER ought to be was neglected: that of second-best considerations. This refers to a situation in which one has what seems like an excellent rule that would appear to serve the social good, but which in practice could actually go against the social good because of the imperfect nature of the world in which it operates.

The classic example of this in health economics is the so-called excess burden of health care insurance (despite the obvious benefits of

INTRODUCTION

insurance). The situation arises in the form of moral hazard, where the additional demand generated when demand elasticities are high means that the value of treatments exceeds their marginal cost. This may not constitute the inefficiency it appears to be, and therefore will not need remedial action, if there is a systematic tendency for health care to be under-demanded (for example, for reasons of externality).

NICE has already met charges of this sort. For example, health authorities frequently say that because the Government prioritises NICE recommendations for the NHS and requires them to be funded when requested by clinicians, other local priorities (which might have a higher marginal cost-effectiveness payoff than the one being recommended) are excluded. There is therefore a question of the context in which these things are introduced. It may be that the ideal rule turns out not to be the best rule in practice after all.

In Chapter 5, Graham Loomes sets up a fundamental challenge in questioning the use of QALYs as the denominator of incremental cost-effectiveness ratios (ICERs), partly because he believes they represent an outdated technology of benefit measurement, and partly because the obvious alternative to QALY-based ICERs enables precisely the kind of inter-sectoral comparison that may enhance our ability to attain a higher level of overall efficiency – or at least overall consistency. Nor does it require us to specify what a QALY is – with all its weights, inclusions and exclusions. Moreover, it can have the happy (for some) characteristic of being based on the population's preferences. It remains to be seen whether his belief is correct that it is more socially acceptable today to elicit monetary willingness to pay for additional health than the relative value of QALYs.

In Chapter 6, Clive Pritchard suggests that inferring thresholds is, if anything, more hazardous in other countries than it is for England and Wales. But there is a kind of fascination to be had in discovering what appear to be fairly large differences and trying to account for what might determine them. A potent source is likely to be the fact that most decision makers have more than one decision rule. Efficiency is one such rule. Other charges which NICE has from the Secretary of State – to do with equity and the impact on the NHS – doubtless exist elsewhere too. We talk about a (singular) threshold but

INTRODUCTION

the common reality is that there are many thresholds, or criteria that may be interpreted as giving rise to thresholds.

The main issue that prompts discussion in Chapter 10 is whether there is an intrinsic conflict between the individual-based balancing of cost and benefit undertaken by:

- a physician acting as the (ideal) agent of a patient (with whatever values they hold) and
- the collective balancing of cost and benefit (defined as health gained or health forgone), together with an equity criterion.

This is discussed with reference to Chapters 8 and 9, by David Cook and Alan Williams respectively. In the discussions, some believe there is a conflict; others think that the collective decision can be seen as context setting for the exercise of constrained individual decisions (within the NHS). The issue is not resolved and merits further thought.

Chapters 11 and 12, by Robin Clarke and Peter Littlejohns respectively, relate to public involvement. The ensuing discussion in Chapter 13 centres on two themes. The first is whether methods of eliciting the public's preferences necessitate an inverse relationship between quantification and depth. The second is whether the proposed Citizens Council of NICE would be able to determine a threshold for use by NICE's Appraisal Committee.

It seems extremely unlikely that the Citizens Council would be given a brief to determine threshold ICERs. This is not because its members will not be representative of the community at large. Nor is it because they will be inherently incapable of doing the sophisticated thinking that all ICER manipulators have to work through. In fact, it is for a far more basic and compelling reason: that they will not be accountable. They will also have the other disadvantage that stops NICE from legitimately deciding affordability questions: an inability to assess the opportunity cost in the rest of the public and private sectors.

In Chapter 14, Nancy Devlin and Adrian Towse outline some of the research questions that are raised in the course of the discussion chapters. The agenda is fascinating, not least because so much of it is

INTRODUCTION

at the intersection of economics, politics and philosophy, not to mention medicine, epidemiology and biostats. The papers presented here are an appetite whetter and I look forward to future conferences and publications at which the fruits of further thought are presented.

Section 1

THE USE OF THRESHOLDS

Chapter 1

An introduction to the use of cost-effectiveness thresholds in decision making: what are the issues?

NANCY DEVLIN

This chapter looks at:

- economic evaluation methods;
- what is meant by the term threshold;
- whether we should be explicit about the value of a QALY;
- whether thresholds should differ according to the cost-effectiveness methods used;
- whether health threshold decisions should be consistent with other public sector decisions;
- how other objectives should be recognised in decision making.

Introduction

Notwithstanding the vastly increased funds devoted to it,¹ rationing in the NHS is still, and always will be, unavoidable.² A decision to devote resources to any one treatment or health care service inevitably means those same resources are no longer available to generate health outcomes in other ways. Economic evaluation (weighing up the benefits and costs of each alternative) provides one means of informing difficult decisions about which services and treatments should get priority.

Economic evaluation methods

The most widely used method of economic evaluation is cost-effectiveness analysis. Under this approach, each option is described in terms of the resources required to obtain one unit of improved outcome – commonly ‘cost per life year gained’ or ‘cost per quality-adjusted life year (QALY) gained’.³ The latter is preferred by analysts

and decision makers because it enables the comparison of health services with dissimilar types of outcomes. Attempts to systematically incorporate such information into health care priority-setting (with varying degrees of success) include two organisations in New Zealand: the Health Funding Authority⁴ and PHARMAC (Pharmaceutical Management Agency)⁵; Oregon⁶ and, in the UK, NICE.⁷

There are many ethical issues surrounding the use of economic information in such decisions, and indeed disagreement over whether rationing should involve explicit priority-setting at all, or should instead remain implicit.⁸ Explicit approaches to rationing have the merit of improving the accountability of decision makers, improving the transparency of decisions to the public and patients, and facilitating decisions that are consistent, impartial and capable of achieving the stated objectives of the health system.⁹ It is these characteristics that have led many countries to introduce (to varying degrees) explicit ways of deciding 'who gets what' in their health systems.

Interpreting economic evidence

If economic evaluation is to play a role in deciding these priorities, a key consideration is the interpretation of economic evidence. Cost-effectiveness results (the addition to cost per life year, or quality-adjusted life year, gained) have no absolute meaning: whether one option is desirable on economic grounds depends on its comparison with something else.

If more than one option exists for treating a particular disease, cost-effectiveness analysis can rank the alternatives and suggest which option yields the greatest improvement in health per pound spent. But priority setting involves a wider set of choices and comparisons. In this case, the decision as to whether any given treatment represents good value for money relies on its comparison with some benchmark or threshold.

The threshold

To explain what the threshold is, imagine a world in which policy makers have access to complete information on the cost and health

outcomes of every possible treatment, and where all that matters in their choices is obtaining the greatest improvement in health possible given their total budget. All health services could be listed in descending order from those that have the lowest £ per QALY gained at the top (the most desirable) to those with the highest £ per QALY gained at the bottom (the least desirable).

Decision makers would work their way down the list, and continue to approve each service until the point where their total budget is exhausted. At that point, the cost per QALY of the very last service that was funded would reveal society's willingness to pay to gain an extra QALY. If society valued QALYs more than that, it would increase the health budget (and a further service, with a higher cost per QALY gained, would be approved); if it valued QALYs less than that, the opposite would occur.

Barriers to accessing information

In practice, the world is much more complex. First, only a fraction of health services are subject to rigorous clinical evaluation – let alone economic evaluation – so decision makers have little information. Second, value for money (in the sense of maximising QALYs) is not all that matters in such decisions: judgements may be influenced to some, or a greater extent, by other aims, such as reducing the inequalities in health, rather than maximising the total gains in health.

For both the reasons stated above, the questions of what the threshold is, and what it should be, cannot be revealed in this manner, and some alternative means of establishing the value that society places on a QALY is required.

Estimating costs and values

In summary, the interpretation of cost-effectiveness results depends critically on the threshold with which they are compared, but that threshold can only be estimated. Therefore, any health system seeking to incorporate cost-effectiveness evidence into its decisions faces some important issues.

There are four main questions that need to be answered:

1. should we be explicit about the value of a QALY?
2. if cost-effectiveness methods vary, should a different threshold apply?
3. should the threshold for health decisions be consistent with decisions made elsewhere in the public sector?
4. maximising health is not all that matters – but how should other objectives be recognised in decision making?

1. Should we be explicit about the value of a QALY?

Economic evaluation can help inform priority setting – but the basis upon which judgements are made about good value for money must be clarified. Otherwise, these decisions will not be transparent, risk being inconsistent, and may not be particularly explicit.

In Chapter 2, Adrian Towse and Clive Pritchard consider the threshold implicit in the guidance that NICE has issued to date. Comments made by the chairman of NICE at its last annual general meeting suggest a threshold (although the term itself was not used) of £30,000,⁹ although NICE maintains ‘The Institute does not have such a threshold.’¹⁰ If transparency and explicitness are desired procedural characteristics, it will be important for NICE to be explicit about the way it uses cost-effectiveness evidence, to provide a defensible basis for its threshold, and to seek a means by which the opinions of the public might be sought on its appropriateness.

Methods used in other areas of the UK public sector to estimate the value of health outcomes may provide a way forward. In Chapter 5, Graham Loomes considers whether approaches used to determine the value of a statistical life, and the values themselves, have relevance to the valuation of health outcomes from health care.

International thresholds

In Chapter 6, Clive Pritchard discusses the way thresholds are used in priority setting overseas. The wide range of thresholds reflects

different methods of estimation, different systems for funding health care, and different levels of economic prosperity. Cutler and McLellan cite a threshold value of a year of life gained free from disease as US\$100,000.¹¹ At the other end of the spectrum, New Zealand's PHARMAC uses a threshold of NZ\$20,000.¹² Converted to GBP, using purchasing power parities, the value of a QALY in each case ranges from approximately £9,000 to £65,000. The value of a QALY will be determined by the willingness and ability of UK taxpayers to pay for improvements in health.

What remains unresolved, from overseas experience, is how best to engage the UK public in establishing this value. Robin Clarke and Peter Littlejohns consider public involvement in Chapters 11 and 12 respectively.

2. If cost-effectiveness methods vary, should a different threshold apply?

The lack of information on costs and health outcomes poses other problems for being explicit about the threshold. For many health services, quality of life information of the kind required in the calculation of QALYs is simply not available from existing research.¹³

Of the first 22 guidance reports, only half cited cost per QALY evidence on cost effectiveness.¹⁴ Of the eight issued subsequently, only one presented evidence on cost per QALY gained. Instead, four reported cost per life year gained, one reported costs per year of remission, and three reported inadequate data to perform cost-effectiveness of any kind.¹⁵ This demonstrates the difficulty faced by NICE in undertaking analyses to provide a consistent basis for decision making and suggests different thresholds need to operate where health outcomes are represented in different metrics.

Differences in method

Even where quality of life information is available, it is possible that different approaches to its measurement (i.e. describing the quality of life in each health state) and to valuing each state (i.e. the 'weights' used to quality adjust each year of life) compromise the

comparability of the QALY estimates. Differences in method are known to reduce the comparability of results between studies. They also reduce the comparability of results against a threshold.

Wider costs and perspectives

Similarly, some studies might estimate the costs in a cost per QALY study from the point of view of the NHS. Yet other studies might take wider costs and perspectives (for example, that of patients or other areas of government activity) into account.

If what is measured in both the denominator and the numerator of a cost per QALY gained calculation differs from study to study, this makes the interpolation of implicit thresholds difficult and suggests that different explicit thresholds will be appropriate in each case.

Establishing a single threshold value may be problematic for another reason. O'Brien et al provide evidence that consumers' willingness to accept monetary compensation for relinquishing the benefits of an existing health service (the 'selling price') is greater than their stated willingness to pay to obtain those same benefits by funding a 'new' service (the 'buying price').¹⁶ The possibility that the value of a QALY differs between investments and disinvestments in health care poses a further challenge for the choice of an appropriate cost-effectiveness threshold.

3. Should the threshold for health decisions be consistent with decisions made elsewhere in the public sector?

A problem with using cost per QALY gained calculations is that this form of economic evaluation is popular in the health sector, but rarely used outside it.

If different methods and benchmarks are used to judge what is good value in the health sector in comparison with another sector sharing similar objectives (land transport safety, for example), this may mean that there is persistent allocative inefficiency between the two domains of governmental activity. It therefore might be possible to

increase gains in length or quality of life by shifting resources between these budgets.

4. Maximising health isn't all that matters – but now should other objectives be recognised in decision making?

There are at least two objectives of the NHS:

- to improve health;
- to reduce inequalities in health.

The pursuit of multiple objectives complicates the pursuit of explicit rationing: it suggests there are trade-offs, and that judgements concerning these trade-offs need to have a consistent and well-argued basis. There are two possible approaches:

Weight the QALYs

Weight the QALYs gained by particular subgroups. Such weights would reflect a view that improvements in health experienced by some groups in society should 'count for more' than improvements in the health of others. Using this approach, equity considerations are incorporated into economic evaluations, and comparisons can be made against an agreed threshold. However, it is not clear what the appropriate subgroups are, nor what the correct weights should be.

Separate equity and efficiency

Keep efficiency considerations (the amount of health gain per £ spent) separate from considerations of who gains the QALYs (equity). However, specify the extent to which we are prepared to accept trade-offs against the threshold (i.e. accept options with a cost per QALY above the threshold) on the grounds of equity.

Equity-related concerns regarding the value of a QALY are discussed in detail by David Cook and Alan Williams in Chapters 8 and 9 respectively.

Conclusion

My own answers to the four questions I have posed above are: (1) yes (2) no (3) yes (ideally) and (4) as explicitly as possible. Each of these issues – and still others pertaining to the use of thresholds in health care rationing – is addressed in greater depth in the following chapters, not always with the same conclusions.

These issues are relevant to any health system using economic evaluation, but particularly so to the NHS. The establishment of NICE arguably represents the most ambitious and rigorous attempt internationally to adopt explicit priority setting. Resolving the issues surrounding the use of cost-effectiveness evidence will be important in building on the considerable gains already made by NICE in moving towards a transparent, defensible decision-making process.

REFERENCES

- 1 Appleby J & Coote A (eds). NHS Funding. Chapter 2 in *Five-Year Health Check: A review of Government health policy 1997-2002*. London: King's Fund, 2002.
- 2 Maynard A & Bloor K. *Our Certain Fate: Rationing in health care*. London: Office of Health Economics, 1999.
- 3 Gold MR, Siegel J, Russell LB & Weinstein MC (eds). *Cost Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 4 Health Funding Authority. *How Shall we Prioritise our Health and Disability Support Services?* A discussion paper. Wellington: HFA, 1998.
- 5 www.PHARMAC.gov.nz
- 6 www.ohppr.state.or.us/hsc/prioritized_hsc.htm
- 7 www.nice.org.uk
- 8 See Doyal L. That rationing in the NHS should be explicit: the case for. *BMJ*, 1997; 314: 1114-8 and Coast J. That rationing in the NHS should be explicit: the case against. *BMJ*, 1997; 314: 1118-22.

- 9 McDonald G. Opening up the NICE debate. *BMJ*, 2001; 3 October. www.bmj.com/cgi/eletters/323/7325/743.
- 10 NICE. Evidence to the Health Select Committee. London; NICE, 2002.
- 11 Cutler DM & McLellan M. Is technological change in medicine worth it? *Health Affairs*, 2001; 20(5): 11-29.
- 12 Personal communication.
- 13 Devlin N, Appleby J & Parkin D. *Patients' Views on Explicit Priority Setting: What are the implications for NICE?* London: King's Fund, 2002.
- 14 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ*, 2001; 323: 1300-1303.
- 15 Appleby J & Coote A (eds) *Rationing in the NHS. Chapter 4 in Five-Year Health Check: A review of Government health policy 1997-2002*. London: King's Fund, 2002.
- 16 O'Brien BJ, Gersten K, Willan AR & Faulkner LA. Is there a kink in consumers' threshold value for cost effectiveness in health care? *Health Economics Letters*, 2002; 6 (1).

Chapter 2

Does NICE have a threshold? An external view

ADRIAN TOWSE and CLIVE PRITCHARD

This chapter looks at:

- NICE's criteria for making decisions;
- reasons for thinking NICE has a £30,000 per QALY threshold;
- NICE's publicly stated view on thresholds;
- patterns occurring in NICE appraisals;
- issues relating to the role of thresholds.

What are NICE's criteria for making decisions?

The criteria in the original announcement setting up NICE are as follows:

‘In reaching its judgement the Institute will have regard to the factors listed in the Secretary of State and National Assembly for Wales’ directions, namely:

- the Secretary of State's and the National Assembly of Wales' broad clinical priorities (as set out for instance in National Priorities Guidance and in National Service Frameworks, or any specific guidance on individual referrals);
- the degree of clinical need of the patients with the condition under consideration;
- the broad balance of benefits and costs;
- any guidance from the Secretary of State and the National Assembly for Wales on the resources likely to be available and on such matters as they may think fit.’¹

Our understanding is that NICE has not been given ‘any guidance from the Secretary of State... on the resources likely to be available...’ The ‘broad balance of benefits and costs’ is in large part what we have been talking about so far.

That leaves two other factors, in addition to economics, that NICE has to take into account when looking at whether a technology should be made available. The first is the broad clinical priorities set out in the National Service Frameworks and any other announcements. The second is ‘the degree of clinical need of the patients with the condition under consideration’. These two priorities provide very clear directions as to what NICE should take into account in addition to the issue of the costs and benefits associated with the treatment.

Why do we think that there is a £30,000 per QALY threshold?

There are several reasons why we think NICE is operating some form of threshold of £30,000 per QALY. Two of them are as follows:

Firstly, in the Orlistat appraisal determination, there is a very explicit reference to a range of £20,000-30,000 per QALY:

‘to attain a sufficient level of cost-effectiveness, in the region of a cost per QALY gained of between £20,000 and £30,000, people treated with Orlistat have to lose about 5% of body mass for each three months that they are maintained on treatment, or achieve a cumulative loss of at least 10% of body weight from the start of treatment over the first six months.’²

The Appraisal Committee concluded that the technology should only be available to those who meet this weight reduction criteria. The clear implication is that this £20,000-30,000 per QALY range is where it was trying to target the intervention.

Secondly, in the Department of Health’s recent circular on the multiple sclerosis risk-sharing arrangement, by way of background and putting the deal in context, it makes it very clear that, while

‘... NICE has not adopted a standard ‘threshold’ for its judgements... retrospective analysis of appraisal determinations in its first year of operation, as summarised by Sir Michael Rawlins at NICE’s annual public meeting, suggests that positive recommendations were in general associated with a cost per QALY of £30,000 or less; higher cost per QALY

figures were accepted only if there were special factors... not covered by the formal modelling.’³

What is the publicly stated NICE view on thresholds?

In a recent press statement issued after Michael Rawlins and Andrew Dillon had given evidence to the House of Commons’ Health Select Committee, NICE said that Michael Rawlins told the Committee that:

‘This was also a great opportunity to again put on public record that the Institute does not have a cost threshold beyond which a technology would be automatically rejected.’⁴

The phrase ‘a cost threshold beyond which a technology would be automatically rejected’ is quite consistent with a threshold that does not lead to automatic acceptance or rejection, but which informs decision making.

If we analyse the appraisals that have taken place, what kinds of patterns can we see?

In trying to make sense of what has happened to the end of May 2002 we have been through the individual appraisal determinations. In some cases, there is no reference to quality of life, or anything that one can attempt to turn into an estimate of a quality-adjusted life year. In others, there is either a life year or a cost per episode avoided, which we have converted into a cost per QALY (on eminently reasonable assumptions, which are set out in Appendix 1 on page 122).

We have divided the determinations into those where there is:

- a cost per QALY range given and the technology has been accepted;
- a range within which the technology is restricted to some patient groups within that group, i.e. it is not possible to separate out the cost per QALY for the patient group that it is allowed for from the cost per QALY for the patient group for which it is not;
- a cost per QALY given and the technology has been rejected.

Table 1 Results of appraisals to end May 2002

Cost per QALY	Accepted	Restricted	Rejected
< £20K	15	3	2
£20-30K	4	5	1
> £30K	3	4	4

The results are set out in Table 1, above. If we start from the hypothesis that there is no relationship between the cost per QALY, on the vertical axis, and whether it is likely to be accepted, restricted or rejected, on the horizontal axis, and do a simple statistical test, then we reject that hypothesis. The statistical test therefore tells us that there is a positive relationship between the cost per QALY and the likelihood of the technology being accepted, restricted or rejected. The test results are significant at the 5 per cent level.

Exceptions

However, there are exceptions. It is not simply that if a technology costs less than £20,000 per QALY it is accepted, if it is between £20,000 and £30,000 it is restricted and if it is above £30,000 it is rejected. To give an example, the Riluzole appraisal showed that the cost of the treatment was over the £30,000 per QALY, yet it was accepted by the NICE Appraisal Committee for use by the NHS. The appraisal says:

‘The Committee took account of the severity and relatively short life span of people with ALS and in particular... the values which patients place on the extension of tracheostomy-free survival time.’⁵

Special factors

Special factors are taken into account, and this is an example of one such factor. The cost per QALY was similar to that for use of zanamivir by normal (i.e. not high risk) patients with influenza. Yet the use of zanamivir for normal patients with influenza was not recommended.

The importance of special factors is again referred to explicitly in the health circular issued by the Government on the multiple sclerosis (MS) scheme. In the preamble setting out the deal, the Department of Health refers to ‘the ‘special factors’ which might be considered to be relevant’. Effectively the Department of Health is saying that there were special factors that NICE indicated were relevant to an assessment of the willingness of the NHS to pay for a QALY for an MS patient and, by implication, had NICE had the deal on the table that it now reached with the companies, this was the sort of cost per QALY threshold it would be prepared to accept.

The Department is saying that the value of these two unquantified special factors that the NICE appraisal referred to: first, the severity of relapses, and second, the possible cost offsets of personal social services as a consequence of treating people – is an extra £6000 per QALY. It is therefore appropriate to go above the £30,000 limit to £36,000.

So what? Where does that get us to in terms of the role of thresholds?

The evidence clearly suggests that, to date, both £20,000 and £30,000 are regarded as significant figures by the NICE Appraisal Committee. It is also very clear from the terms of reference given by the Secretary of State that costs and benefits are meant to be taken into account by the Appraisal Committee. What is also important is that cost effectiveness is not the only criteria that is being used. There are exceptions to the cost-effectiveness threshold. The Secretary of State’s terms of reference require other factors to be taken into account, and hence these exceptions are to be expected.

This raises a number of issues, for example:

- is there a lack of transparency;
- how are we accounting for the special factors, which by definition will vary from technology appraisal to technology appraisal;
- to what extent should these special factors be integrated into the cost per QALY;

- should we be capable of valuing these, either as cost offsets or in some way getting the measure of cost or of health benefit to reflect all of these extra dimensions or special factors so they can be incorporated in some way into the cost per QALY measure?

NHS limitations

Part of the difficulty in discussing explicit rationing is that NICE is currently in the front line within the NHS in terms of getting across the message that, even with extra resources, the NHS cannot do everything. The more NICE comes out of the closet in terms of acknowledging its explicit rationing role, the greater the risk that the focus then moves to NICE as an institution rather than on how choices should be made about what the NHS will provide.

Finally, there is an issue of legitimacy – notwithstanding the rationing role NICE has been given. Currently, members of the appraisal committees have often been selected because of their technical competence, rather than because they are particularly good judges of the preferences of the nation and of what the typical NHS patient or citizen would regard as reasonable or unreasonable for the NHS to pay for. There is an issue about the extent to which it is reasonable for the appraisal committees to weigh up the values to be placed on any special factors, or to decide what is the threshold.

REFERENCES

- 1 NICE. *Guide to the Technology Appraisal Process*. London: NICE, 2001.
- 2 NICE. *Guidance on the Use of Orlistat for the Treatment of Obesity in Adults*. Technology Appraisal Guidance No 22. Para 4.2.3. London: NICE, 2001.
- 3 Department of Health. *Cost Effective Provision of Disease Modifying Therapies for People with Multiple Sclerosis*. London: Department of Health, 2002.
- 4 NICE. *NICE Present Evidence to Health Select Committee*. London: NICE, 2002.
- 5 NICE. *Guidance on the Use of Riluzole (Rilutek) for the Treatment of Motor Neuron Disease*. Technology Appraisal Guidance. No 20. London: NICE, 2001.

Chapter 3

Does NICE have a threshold? A response

PETER LITTLEJOHNS

This chapter looks at:

- defining ‘threshold’;
- whether a threshold exists;
- where the £30,000 figure comes from;
- modifying factors that influence NICE decisions;
- whether health care is really a ‘bottomless pit’ that requires rationing.

I should first point out that nothing should be inferred from the fact that the title of this chapter includes the term ‘threshold’. It does not mean that there is a threshold; indeed, the Institute has never had an explicit threshold. I will focus on responding to Adrian Towse and Clive Pritchard on the issue of whether NICE has used thresholds in the past. But before we get into these discussions, it is very important to be explicit about the terminology: slack terminology leads to slack science.

What is a threshold?

The definition of a threshold, as given by the *Oxford Shorter English Dictionary*, is:

- i) the border or limit of a region; ii) a lower limit of some state; iii) the magnitude that must be exceeded for a condition to occur.

Does a threshold exist?

In the context of NICE, Sir Michael Rawlins, Chairman of NICE, has stated that the Institute would ‘need to be very clear in its reasons for supporting technologies with cost-effectiveness ratios higher than £30,000 per QALY’.

This statement was made at the 2001 Annual General Meeting, based on his review of NICE's appraisals over the previous year. He arrived at this conclusion using the same publicly available data that James Raftery subsequently used for his analysis published in the *British Medical Journal*.¹ Given that, as Chairman of NICE, Michael Rawlins is a member of the Appeals Committee but is not a member of the Appraisals Committee, he was speaking as an external observer of the appraisals process.

As Towse and Pritchard note, in evidence to the House of Commons Select Committee on NICE in February 2001, Michael Rawlins has again stated that a threshold does not exist. The full quote is as follows:

‘This was also a great opportunity to again put on public record that the Institute does not have a cost threshold beyond which a technology would be automatically rejected. Despite our clear statements on this subject, the issue of a threshold seems to have become an urban myth, and that is what it is – a myth.

This myth has resulted from data taken out of context and inaccurate comments and reporting by individuals who weren't present when I presented a review of the completed appraisals at the Institute's Annual Public Meeting last year.'

To date, this is the formal position of the Institute on this issue.

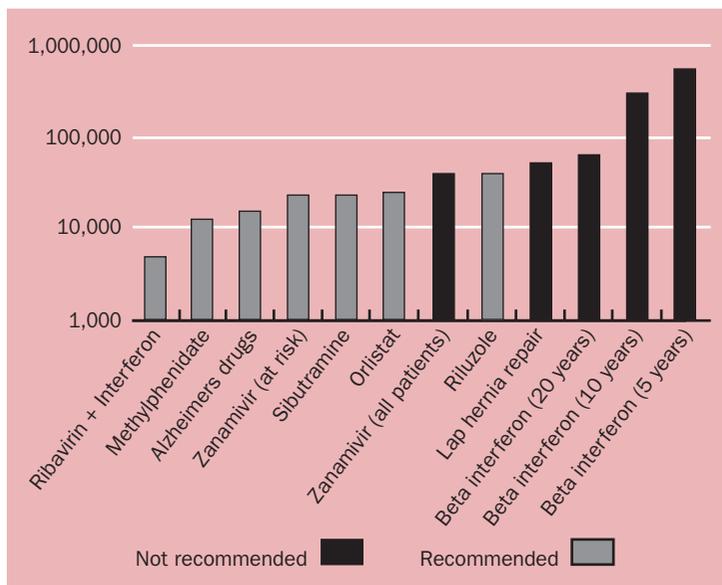
Where does the figure of £30,000 come from?

The basis for the interest in the £30,000 per QALY threshold is the result of the sort of simple analysis using QALYs, shown in Figure 1.

If there were a simple threshold, you would expect to see findings that can be described by Figure 2. You would suddenly hit the £30,000 level, or whatever the figure is supposed to be, shoot up, and so you would either get rejection or acceptance.

You seldom see straight lines in nature; curves are more often the norm. In seeking to apply this concept to the threshold issue, it is

Figure 1 NICE decisions by cost (£) per QALY ratio



appropriate to consider the results of the Committee's deliberations on the cost effectiveness of interventions as having a distribution around a central position. Figure 3 presents a range of results allowing one to infer where the average (mode, mean, midline or median – whatever one wants to consider the appropriate description of average) would occur. Others have already raised the point that there will be modifying factors to any final decision on whether a technology is considered cost-effective. We now need to concentrate on teasing out what those factors are.

Modifying factors that influence decisions

As an outsider of the actual decision-making process of the Appraisal Committee, but by virtue of my executive position at NICE being an observer of the Appraisal Committee making decisions, there seem to

Figure 2 **The threshold approach to accepting/rejecting appraised technologies**

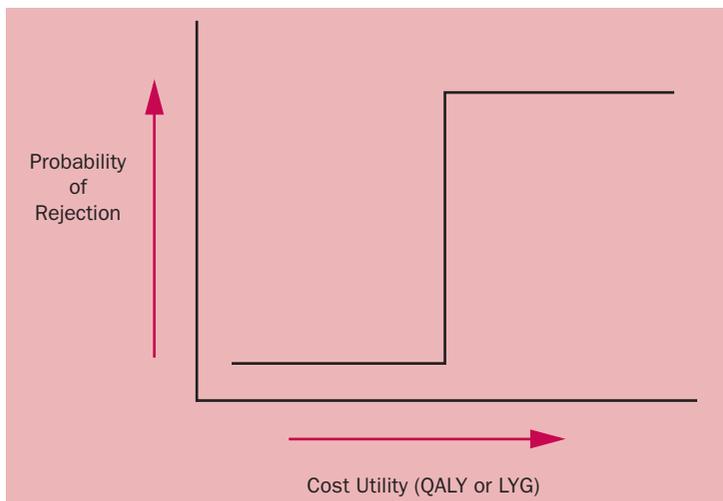
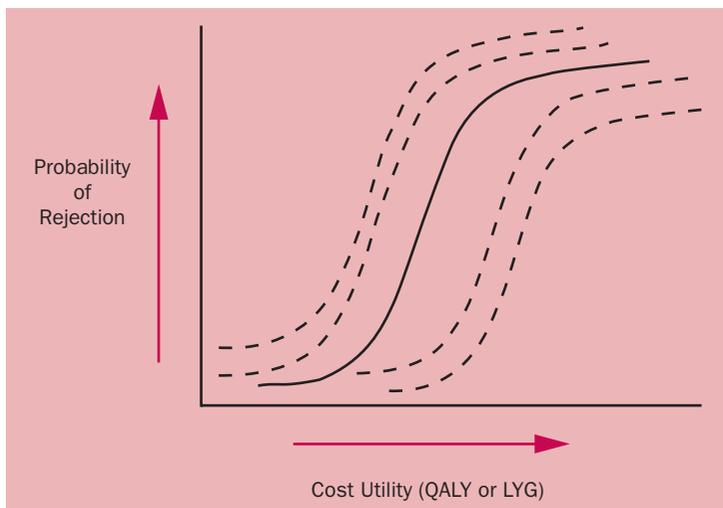


Figure 3 **Probability of rejection for cost ineffectiveness**



me to be two aspects that need to be addressed, one intrinsic and the other extrinsic to the cost-effectiveness analysis.

Intrinsic to the analysis is the presence and robustness of the data. The technology appraisal process depends on good data, but there are occasions when the exact data that is required is not available. In addition, there is much data of dubious quality. This means that assumptions have to be applied to the models and the results may differ, depending on the views (and hence numerical value) given to them by different stakeholders. In addition, because of the number of variables that can be involved, subsequent sensitivity analysis can lead to considerable variation in the final estimate. I have seen confidence intervals ranging from zero to infinity.

Value judgements

Extrinsic to the analysis, are the value judgements that often have to be applied. QALYs are intended to capture the relevant perceptions of the importance of differing health states across different diseases, but we know that not all aspects are covered. Factors that are now taken into account intuitively will have to be quantified if we are to move to an explicit presentation of the results. For example, issues that have been discussed in relation to the appropriate application of cost-effectiveness methodology include the degree of clinical need, the uniqueness of some interventions, and whether differing weight should be given to the start and end of life.

Furthermore, there are issues of equity, fairness and choice that fall outside current analyses of cost-effectiveness. The Institute wishes and needs to be explicit about the assumptions underlying the final decisions. This is what the public, professionals and scientists expect. We have a Research and Development (R&D) Sub-committee chaired by Tony Culyer and, in conjunction with the NHS R&D Methodological Group, we are identifying a research agenda to start to explore how our committees can address this quality issue and how those final decisions are made.

Although it is imperative that the Institute bases its decisions on research, we are also aware that research is only the beginning. If individual stakeholder groups hold very strong views, it is difficult for

even the best of research to shift opinion. But being successful in modifying extreme views is very important if the Institute is to succeed. The topics we are discussing are not only challenging in scientific terms, but are very emotive for patients and the public.

Health care: the bottomless pit?

It is important to be very precise about terminology. We need to speak in simple terms so that non-economists can understand the debate, but should not be simplistic, or the aims and objectives of the process may be destroyed.

The papers presented in this book discuss rationing as an issue. There have been debates on numerous conference floors about whether or not NICE is actually a rationing organisation. I will continue that atmosphere of controversy, from a non-health economist's perspective, and quote some professional colleagues of mine – epidemiologists from Bristol, the so-called 'Bristol optimists' – who question whether 'rationing' (as health economists describe it and the public currently understand it), is actually a legitimate stance.

In a *British Medical Journal* article,² Frankel *et al* challenge the 'bottomless pit' analogy of health care. They conclude that:

- the rationing debate has been conducted almost exclusively through assertion and political analysis;
- conventional assumptions of an imbalance between demand and supply are not met by evidence;
- pessimism about adverse future trends in demand arising from an ageing population, the costs of innovation, and rising public expectations are similarly unsupported by good evidence;
- many perceived deficiencies in health care are attributable to issues other than overwhelming demand, such as the unwillingness of the public to accept limits of effectiveness and the self interest of professionals;
- the proposition that the limits to demand lie within the capacity of a properly resourced NHS should be tested.

We are talking here about being much more precise about what cost-effectiveness is, and what society can expect from certain interventions. That is the perspective from which the Institute comes. It works from the point of view of patients, the public and the NHS, and we have to address the broad issues as well as the scientifically important underlying assessments undertaken by health economists.

REFERENCES

- 1 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ*, 2001; 323: 1300-1303.
- 2 Frankel S, Ebrahim S & Davey Smith G. The limits of demand for health care. *BMJ*, 2000; 321: 40-45.

Chapter 4

Discussion: The use of thresholds

The discussion touched on four main themes:

- what the threshold should be;
- explicitness of the threshold;
- the use of cost-effectiveness ratios in decision making;
- comparability with other parts of the public sector.

What should the threshold be?

John Henderson: It would be of interest to know, if all NHS interventions were ranked according to their cost-effectiveness in the kind of league table discussed by Nancy Devlin, how many items on the list the NHS budget would be able to purchase. Would the budget be exhausted at a cut-off of £30,000 per QALY or a different figure? That work still needs to be done and, as well as indicating how far the budget would go, would raise a number of research questions. Alan Williams is busy working on a vignette for the NHS R&D methodology programme. It is possible that it might be one of the things that came out of that.

Alan Williams: We have been talking mostly about inferring thresholds from what people actually do. For many purposes, however, you need to inject a threshold in order to guide what people should do. We have the thing the wrong way round, it seems to me. The real crunch issue is how could we set about finding out what the public think would be an appropriate level for the threshold. This is the fundamental issue, and at the moment, I do not see how it could be addressed.

Ron Akehurst: I am a member of the Appraisal Committee of NICE. I have some problems with the idea of a threshold *per se*. I have more of a 'smudge' in mind – which might correspond to £20,000-£30,000 or some other figure. The non-linearity discussed by Peter Littlejohns would remain even in the absence of factors other than cost-effectiveness that influence NICE's decision, purely because there is uncertainty about what the data are indicating.

Explicitness about the threshold

Ron Akehurst: I see the smudge as reflecting increasing discomfort as the cost goes higher and higher. I would be in favour of making some sort of explicit statement about that. It then gives the outside world some guidance as to what is expected. It also gives the outside world something to start to research around.

Going back to the discussion we have just had about league tables, I would be very, very unhappy to take the league tables we have from the existing databases. The quality and appropriateness of the studies that make those up are so enormously variable, and I do not think that you can do much with them. But if we started to be explicit about the sorts of ranges we were talking about, we could start to direct research more closely to seeing how different technologies compare with that.

Chris Heginbotham: I very much agree. It is impossible to draw a firm, single threshold and there is bound to be some variance. I want to pick up on the point that there are other factors you have to take into account. If you are to be explicit about a threshold or some sort of smudge, you will also have to be explicit about all the other factors that are affecting that threshold.

For example, the appraisal on implantable cardioverter defibrillators (ICDs) demonstrated that NICE was approving those for use, but also said that the indicative level should be 50 per 1 million patients. That is setting another threshold. It is putting another cap on the budget, which clearly then influences the threshold that was used previously for deciding on whether they should be approved. That would be true for any other budgetary requirements that were placed on the use of a technology, whatever it is. I think we have to be explicit about all those other factors which will also have to be taken into account.

Jack Dowie: This conversation is going down a route I do not like, in the sense that it is diverting attention from the alternative to an explicit and formal way of making these decisions – with the implication that some other ‘taking into account’, ‘considering’, and

all of those words, would be better than doing this explicitly. It is going down the route of what I call acronymically 'partial and non-comparative evaluation'. If we go down that route, we will spend all our time not discussing the real issue: namely, what the incremental benefit is from a rational fudging tactic of the sort that NICE is going along with.

Rational fudging is only possible because of the political context in which NICE is located. It is a massive improvement on the past, but it is not anything like the rational approach that most of us would like. It is rational in one sense, however, to fudge at the moment.

Keith Tolley: I agree with Ron's point that the smudge factor should be there. What strikes me, however, is that you have to be explicit about a threshold, whether it is smudged or fixed, because policymakers will set up a threshold for you if you do not do something about it. If you take the beta-interferon example, there is a threshold there of £36,000. If we do not actually set up a smudge threshold and be explicit, a threshold will come about anyway. It is inferred already. In the Department of Health negotiations with the companies we have an inferred threshold. A fixed threshold will come about because of those sorts of pressures, unless we are explicit about a smudged one. Either way, it has to be explicit.

Alan Williams: Keith is right. Nature abhors a vacuum. These people have to make decisions and they will make the decisions. Out of their decisions, wittingly or unwittingly, there will come to be a pattern of activity. They will be trying to be consistent and fair between one and the next, so it will happen. It is the question somebody posed earlier: what is the legitimate authority of an appraisal committee to decide what the British public wants?

Tony Culyer: That is a rhetorical question. We all know what we think the answer to that is. I agree: I think the answer is that it has none. It has no legitimate authority to decide that sort of question on behalf of the British public. It seems to me that the question is equivalent to the question of affordability, as it is commonly put. NICE is very clear that it is not competent to make that judgement. But take us to the next stage. Where does that get us?

Alan Williams: If you take that line, then the NICE threshold should be set by the Cabinet.

Richard Cookson: I want to pick up on an earlier point, that if you are explicit about a threshold, you therefore have to be explicit about everything else. Is there not an issue of diminishing marginal utility of explicitness going on here? Perhaps there is an optimal level of explicitness – because it costs money to invest in all these techniques.

The use of cost-effectiveness ratios in decision making

Adam Oliver: I question the usefulness of incremental cost-effectiveness ratios. They are only applicable when a new intervention is cost increasing. If it is not cost increasing and it is more beneficial, then it is dominant and you just apply it. So it is really applicable when you are talking about cost-increasing interventions. In an incremental cost-effectiveness ratio, however, you never assess the opportunity costs of that extra cost. So the more cost-increasing a new intervention is, the greater the unassessed opportunity costs. The more cost-increasing it is, the more biased the ratio is, and therefore the ratios are biased towards cost-increasing interventions.

If you recommend a new intervention to a health authority that is cost increasing, the health authority has to give things up to pay for that new health intervention. But what they give up is not assessed within the ratio, so there are unassessed opportunity costs.

John Henderson: I think it brings us back to the issue of what the NHS can afford to do. It perhaps poses the question: is NICE there to help the NHS spend its money so as to maximise health, given the amount of resources that are there for the NHS? Or is NICE also implicitly saying how much the NHS budget ought to be, by adopting a particular threshold based on, say, willingness to pay?

Martin Buxton: I am also a member of the NICE Appraisal Committee. As someone who has struggled to work out what we were doing on the Committee and what would guide us, it seems to me that one legitimate starting point is the view that what we are trying to do is help the NHS make the best use of a fixed budget.

If you start from that point at a particular time, then you logically end up in Nancy Devlin's position – that you want to know the cost-effectiveness of everything the NHS might do, and then you find the point at which you have spent the budget. We know that there are difficulties but, conceptually, one could go down that route. I think that is what people are beginning to suggest we should do.

As someone who has tried to provide empirical cost-effectiveness analyses, I do not think that in a relatively short time period – even with the help of the methodology programme – you will find that the cost-effectiveness of everything that the NHS provides can be estimated with reasonable accuracy. I do not think that approach will ever give us the answer, even if conceptually we think it might be the right way to do it.

I do agree with Adam Oliver that the most appropriate thing, in the circumstances, is to ask 'What is the opportunity cost?' – in terms of 'What can we throw out of the system, or what – at the margin – will we not do if we adopt this?' That is where I think the empirical effort is going.

In a practical sense, it would be a lot easier to do that if NICE also looked at things that it felt should cease to be done in the NHS. We would then be saying that we knew there was a chunk of money that could be released that was currently achieving poor cost-effectiveness.

Comparability with other parts of the public sector

John Henderson: Consistency across the public sector is another important issue. There are a lot of research issues that need to be addressed. If other government departments whose policies have health effects were to rank their projects on our league table of cost per QALY, where would they be?

Nick York: Within the Department of Health, the programme decisions that we make on capital expenditure, within the Treasury's guidance on appraisal of capital expenditure, make implicit assumptions about the cost-effectiveness and the cost and benefits of different investment decisions. I have not seen an analysis that

DISCUSSION: THE USE OF THRESHOLDS

reconciles any implicit threshold being used by NICE with other spending on health – whether in road traffic accidents, education or housing – in terms of value for money.

Section 2

HOW ARE THRESHOLDS USED ELSEWHERE?

Chapter 5

Valuing life years and QALYs: transferability and convertibility of values across the UK public sector

GRAHAM LOOMES

This chapter looks at:

- whether the value of life used in transport investment decisions is transferable to other sectors;
- whether ‘willingness to pay’-based monetary values can be converted into a value of life year or QALY;
- difficulties in reconciling the results of the different approaches;
- the future research needed to investigate people’s preferences and values.

Introduction

Although cost-effectiveness analysis is routinely used (for example, by NICE) to inform resource allocation in the NHS, in many other parts of the UK’s public sector, the preferred method of economic evaluation is cost-benefit analysis. Ideally, this would use stated preference-based values – principally, willingness to pay (WTP) values – for the benefits from health, safety or environmental interventions. This is the approach that the Treasury has endorsed, and for which the (then) Department of Environment, Transport and Regions (DETR) commissioned a guide to best practice.¹

Notwithstanding the reservations that one should have about the practicalities of eliciting monetary values for these benefits, there is now a considerable body of research on the value of preventing a fatality (VPF) and, more recently, the value of preventing injuries of varying degrees of severity.

Are Highway Authority values transferable to other sectors?

At present, the values that are used by the highway authorities in

appraising road traffic interventions are approximately £1 million for every expected fatality prevented and about 10 per cent of that – roughly £100,000 – for preventing serious injuries. The latter is a very broad category ranging from, at one extreme, paraplegia to, at the other extreme, outcomes that might involve lacerations and/or fractures but not necessarily a protracted period in hospital or permanent disability.

Such values are regarded, in some areas of the public sector, at least, as providing a ‘peg’ on to which one might hang other values for other purposes. This paper raises two main questions. First, are these values directly transferable to other sectors? Second, might they be converted into a value of a life year (VOLY), or indeed into a value for a QALY?

Transferability

On the question of transferability, the first issue to consider is whether one would wish to take account of physical differences in the population affected by different health and safety hazards. The most obvious thing one might look at is differences in the age distributions of the beneficiaries of various interventions and their existing health state. The relevant factors might be collectively referred to as ‘the profile of harm’, which is mitigated: this includes not just the degree of prematurity of the death, but also the typical pattern of states of ill health prior to the death.

Characteristics of the hazard

A second issue concerns what might be called ‘the characteristics of the hazard’. Psychometric research² has investigated why it is that people perceive certain risks that are statistically quite small to be very important, while others that are much larger are perceived to be somehow less attention-grabbing and therefore less likely to be given priority by the political process.

A whole host of potentially influential factors have been identified, for example:

- the extent to which people feel that the risks to which they are exposed have been incurred voluntarily or have been forced upon them;
- the degree of control people feel they have;
- how familiar they are with the hazard and how much or little is known about the true nature and magnitude of the risk;
- how insidious it is;
- how far the risk is one's own responsibility, as opposed to something for which others might be blamed.

It is at least possible that if people feel strongly about some characteristics that may be present to very different degrees in different sectors, they may want public policy to use different values for preventing fatalities or injuries in those different sectors.

Public attitudes to risk

A specific question about the extent to which values for safety can be transferred directly to health (or indeed, how far the same values can be used in different parts of the health sector) relates to whether people treat man-made risks – risks that are the product of either our own or other people's activities – as somehow different from the way that nature deals out the cards. Although it would be far too simplistic to characterise most transport deaths and injuries as largely the result of our own or others' stupid, negligent, reckless or criminal behaviour, while attributing the majority of health problems to pure chance or plain bad luck, many people may perceive a sufficient difference along these lines to constitute an argument for different values to be used.

Statistics vs. individuals

There may also be a difference between people's attitudes to measures designed to reduce the numbers of 'statistical' deaths and injuries, and those aimed at helping identifiable individuals. Many safety or preventive measures have the characteristic that it is not known in

advance – and often, not even after the event – which individuals will be the beneficiaries of those measures: the most that can normally be hoped for is that statistical analysis of the data on deaths and injuries will indicate some significant likelihood that the measures have had a beneficial effect. By contrast, the identifiability of candidates for health care treatments may lend weight to the case for resources to be allocated to such treatments.

Do values differ depending on the hazard?

Such considerations may be used as arguments for not simply taking an established transport safety-based value and applying it across the board. However, the evidence on this matter is still relatively limited. A study commissioned by a consortium of government agencies and departments led by the Health and Safety Executive³ found no very great difference in people's attitudes to the value of preventing death by fire (either in the home or in public places) or death on the railways, relative to the value for preventing death on the roads.

This is perhaps not surprising as it could be argued from a psychometric point of view that the hazards examined in that study were not very different: the age distribution of the people concerned is much the same, and many of the other characteristics are quite similar, with all the hazards being quite familiar, and a high degree of voluntariness (i.e. the degree of choice as to whether to risk exposure to the hazard). The study findings are also contrary to the values implicit in current policy proposals such as the introduction of Automatic Train Protection, which entails a value of something in the region of £10-15 million for each railway passenger death prevented. When asked directly, people did not want to distinguish so greatly: they thought that much the same value should be attached to preventing fatalities in all of those contexts. Certainly, a 15:1 differential was wildly out of line with the sentiments expressed by the great majority of participants in that study.

In the light of those findings, the Health and Safety Executive has commissioned a further study to identify whether there are any dreaded risks – risks whose characteristics are so aversive that people would want to have a substantial premium attached to the value of avoiding them.

Air pollution

At present, it is too early in the life of that project to be able to report any results, but meanwhile my colleagues and I are currently involved in a project for the Department of Environment, Food and Rural Affairs looking at valuing the health benefits of reducing air pollution. One of the issues to be addressed is that the majority of deaths attributable to air pollution – and in particular, to days of abnormally high levels of air pollution – involve elderly people who are already in very poor states of health.

Indeed, it has been suggested that such people may, on average, have their lives shortened by less than a week as the result of a ‘bad air day’. If this is the case, we might not want to use the same £1 million value for a death prevented from air pollution as for a death prevented on the roads, where the average loss is in the region of 40 years of life, most of which is likely to have been spent in a good state of health. One might want some kind of modification of the value to take account of the much shorter time lost and the lower quality of that time.

Can WTP-based monetary values be converted into a VOLY or QALY?

Such a consideration leads to the second question this paper seeks to raise: namely, whether the WTP-based monetary values for preventing a fatality on the roads are capable of being converted into a value of a life year (VOLY), or indeed into a value of a quality-adjusted life year (QALY).

As mentioned above, given the age distribution of road accident victims, preventing a road fatality gives, on average, an extra 40 years of life. If we were to perform the kind of exercise that economists are keen on – that is, to calculate how much a net present value of £1 million pounds would be per year, spread over 40 years and appropriately discounted for people’s time preference, we would find (by a seeming coincidence that might bring great cheer to NICE) that this produces a figure in the region of £30,000 per year.

So a VOLY, if that is how it is to be computed, would appear to be about £30,000. If we further suppose that most of those 40 years are of reasonably good quality, then it would give something not very much more than £30,000 per QALY.

VORLEs

If this concept of a VOLY were valid, it would allow us in principle to compute the value of remaining life expectancy (VORLE) for any age distribution of potential victims. For example, if we were simply to apply this same figure for a life year to an intervention in the case of somebody who has a life expectancy of some 15 years in normal health, it would give a value of preventing a fatality for such an individual of just over £400,000, and so on.

Moreover, if this value of a life year in normal health is taken to be the value of a QALY, it suggests that we should then make some appropriate adjustment in order to allow for the fact that people may be in less than full health. If we were to apply that type of analysis to the air pollution case referred to earlier, and if we were to take the (possibly rather extreme) estimate that, on average, bad air days bring forward by less than a week the deaths of people who are in an already compromised state of health, then it would mean that the value to be applied to the benefit of reducing air pollution to the extent of avoiding one fatality from a bad air day would be no more than a couple of hundred pounds.

It has been suggested that this approach offers the prospect of being able to convert into money form, almost at a stroke, the entire output from the enormous industry of calculating health state indices and QALYs, which would allow us to apply cost benefit analysis where previously only the antediluvian cost-effectiveness analysis has been used by unreconstructed health economists!

Difficulties with the VOLY/VORLE analysis

Such an approach does, however, raise a number of difficulties. In particular, this kind of computation is at odds with the bulk of the evidence that has been collected from the WTP-based attempts to estimate VPFs and to explore the relationship between VPFs and age.

Jones-Lee⁴ reported one estimate of that relationship on the basis of people's willingness to pay to reduce their risks of being killed in road injuries. Work from other sources – some of it summarised in Pearce⁵ – seems to support the general conclusion reached by Jones-Lee: namely, that after controlling for other factors (particularly for income, as the most significant other factor that explains people's WTP responses), the relationship between VPF and age is best characterised by an inverted U-shaped function.

That is to say, all other things being controlled for, VPF tends to increase somewhat between the ages of 20 and around 45 to 50, then levels out or perhaps falls slightly, but does not seriously decline until people are into their 70s, at which point the value may start to decline rather faster, but not so fast as to tend towards zero, even at the age of 100.

This, of course, is not at all what the VOLY/VORLE analysis would entail – namely, a steady reduction in VPF as people get older and therefore have fewer remaining years to expect. By contrast, if the WTP-based evidence is to be believed, it would suggest a rather different way of modelling people's values.

Love of life

One way of explaining why the VPF does not peter out altogether as people approach the end of their life is that there may be a kind of love-of-life element, reflecting a desire by most of us, especially those of us who do not expect to go on to a better place for the rest of eternity, to hang on to the less-than-totally-ideal place that we are in at the moment. Indeed, as we get older, we may discover to our surprise that things we did not think would be worthwhile are sources of pleasure (or at least, consolation!), and we might be willing to stump up quite a lot until rather late in our years, or until we are in such a miserable state that even oblivion seems like a better alternative.

Combining a love-of-life lump sum element with a declining VORLE component may thus be a better way to model the VPF-age relationship from middle age onwards. Of course, in that simple form it does not

accommodate the initial rise from 20 to 45. But if we also allow for people's love-of-life element to grow during those years – perhaps reflecting many people's sense of the importance of staying alive during child-rearing, perhaps simply capturing a maturation process – it may be possible to account for the inverted-U relationship in terms that are psychologically plausible and readily recognisable by most people.

Can there be a single value for a QALY?

However, if that is the appropriate model, then one cannot reasonably expect ever to get a value of a QALY that fits all sizes and all purposes. One illustration of the possible limitation or inappropriateness of a single-value-of-a-QALY approach may be illustrated by an interpersonal dilemma.

My colleague Robert Sugden put a scenario to me along the following lines: 'Suppose a fire engine is called out and arrives at the scene to find that there are two different buildings on fire. In one building there is a 20-year-old man and in the other building there are five 80-year-old pensioners.

Would the fire officer be justified in saying: 'The 20-year-old has more years of good quality life left than the other five put together. Given that we have only got one hose, (that is the nature of the economic decision – not enough resources to fulfil all the demands!) we should let the five die in order to save the one?'

It is hard to discount the fact that these are five human beings, each with their own 'love-of-life lumps' that they are nurturing in the smoke in the other building. It would be a dreadful dilemma, and one in which it is far from clear that a single-value-of-a-QALY approach would adequately capture either the values of the individuals concerned or those of the population at large.

Future research

So if there is to be a programme of research undertaken in this area, a searching analysis of the theoretical basis of the standard VOLY or QALY computation versus alternative models of human preferences

and decision making is a topic that merits some serious investment and investigation. We need to think critically and imaginatively about ways in which the value of a QALY, if such a thing does – or can – exist, might be estimated. Much can be learned from the WTP methods used to value benefits in other sectors – not least, the difficulties and challenges of undertaking such work.

Conclusion

We may have many reservations about the efficacy of attempts to elicit money values for health and safety benefits – as indeed those who have been involved will have about the attempts to get at health state indices through stated preference methods. However, we need to explore the possibilities and the inevitable limitations rather than rely on some mixture of faith and assumption, either to transfer values from one context to another or to generate VOLY/QALY figures that may have no sound support in the actual structure of people's values and preferences.

Even less acceptable, of course, is the idea that a value of a QALY should be inferred from past policy decisions arrived at by committees using value judgements that are far from transparent to the rest of us, and that may not even have been clearly articulated by those involved. Whatever the practical difficulties and shortcomings, we should at least try to investigate the preferences and values of the population as part of determining any sort of threshold or figure that may be used in health sector decisions, and thereby try to locate such decisions within the same broader framework applied to other parts of the UK public sector.

REFERENCES

- 1 Bateman I et al. *Economic Valuation with Stated Preference Techniques: A manual*. Cheltenham: Edward Elgar, 2002.
- 2 Slovic P, Fischhoff B & Lichtenstein S. Facts and fears: understanding perceived risk in Schwing R & Albers W (eds). *Societal Risk Assessment: How safe is safe enough?*, New York: Plenum Press, 1980; 181-216.

- 3 Chilton S *et al.* Public perceptions of risk and preference-based values of safety, *Journal of Risk and Uncertainty*, 2002; 25(3): 211-32.
- 4 Jones-Lee MW. *The Economics of Safety and Physical Risk*. Oxford: Basil Blackwell, 1989.
- 5 Pearce D. *Valuing Lives, Valuing Health: Towards consistent transfer estimates in the European Union and accession states* (unpublished), 2000.

Chapter 6

Overseas approaches to decision making

CLIVE PRITCHARD

This chapter looks at:

- how economic evaluation is used overseas;
- whether their reimbursement decisions are simply yes/no;
- why economic evaluation is used;
- what impact it has had;
- whether a cost-effectiveness threshold is used;
- comparisons of NICE with authorities overseas.

Introduction

From an international perspective, NICE is by no means unique in using an economic criterion for making high level decisions about the allocation of resources in health care. The purpose of this chapter is to outline how economic evidence is being used in other countries, making some comparisons with the work of NICE, discussing the issue of a cost-effectiveness threshold and drawing some lessons from the experience gained to date.

How is economic evaluation used?

Australia was the first country formally to incorporate economic analysis into decisions about the use of health care resources. Since 1993, companies seeking public reimbursement of a new pharmaceutical have been under a mandatory obligation to submit economic data to the Pharmaceutical Benefits Advisory Committee (PBAC). A number of other governments have now introduced equivalent requirements. New Zealand established the Pharmaceutical Management Agency (PHARMAC) in 1993, with a remit to consider cost-effectiveness, among other criteria, when making recommendations for the purchase of pharmaceuticals.

Since 1995, the province of Ontario has required pharmaceutical companies to present economic evidence on their products (or justify its absence from their submission), a policy adopted by British Columbia in 1996. Within Europe, Finland has required all new drugs to be subject to economic assessment since 1998, as has Norway from 2002. In the USA, as of 1997 and 1999 respectively, new drug listings on the formularies of Foundation Health Corporation and Blue Cross/Blue Shield have been conditional upon the provision of economic evidence.

In other countries, companies can be asked to submit economic evidence when a reimbursement decision is to be made but this power is used on a discretionary basis. This is the case in Portugal, Denmark and France. In each of these countries, requests for economic submissions to support reimbursement decisions are made relatively sparingly. The system that comes into force in the Netherlands in 2003 applies only to a defined group of new drugs: namely, those that cannot be included in an existing therapeutic cluster under the reference-pricing scheme.

The primary concern in Italy appears not to be reimbursement. On the contrary, cost-effectiveness is now called for as an element of the data required by the price negotiation procedures. Favourable cost-efficacy is expected to be demonstrated in price negotiations for drugs targeted at diseases with no existing treatment, when current therapies are inadequate and when a cost-benefit advantage is claimed over other drugs for the same indication.¹ Although the use of economics in policy making outside the UK is overwhelmingly focused on pharmaceuticals, it should be noted that the Australian Medical Services Advisory Committee has begun to use cost-effectiveness in making recommendations about the listing for public subsidy of new devices and procedures.

Are reimbursement decisions simply yes/no?

Targeting treatment

PHARMAC is one of a number of authorities that have applied economic evaluation not only to the decision to list a drug or not, but

also to the task of restricting access to the subsidy for a drug to those patients for whom it is most cost-effective.

In British Columbia, different categories of reimbursement are possible. As well as full benefit or non-benefit status, drugs can be assigned to restricted benefit status, which means that they can only be prescribed by specialists and/or are limited to particular patient groups. Similarly, the PBAC can place a drug on 'restricted benefit' or 'on authority'. The former limits the patients to whom the drug can be prescribed, while the latter requires the prescriber to provide an assurance that the drug is being used within the (restricted) approved indications.

Influencing prices

Secondly, requirements for economic analysis have been used as a tool to exert downward pressure on prices. Where restrictions are placed on the use of a drug (for example, prescription only by a specialist, or to patients who meet specified criteria), PHARMAC is willing to countenance the offer of a price reduction in exchange for a removal of those restrictions. In Australia, a price reduction may be necessary to reduce a drug's cost-effectiveness ratio from what the PBAC considers an unacceptably high level to a more reasonable sum. Consequently, the official advice for some drugs will be 'recommended at a lower price'.

Price-volume agreements

Among other measures given impetus by the implementation of a 'fourth hurdle', Birkett *et al*² report on the use of price-volume agreements in Australia to allow for the possibility that drugs will be prescribed for patients outside those in whom it is considered acceptably cost-effective. When use is higher than estimated, a price reduction comes into effect. The authors characterise this as a form of risk sharing, a term also applied by Braae *et al*³ to the expenditure caps agreed between companies and the New Zealand government. These simply dictate that the company refund any amount above the agreed expenditure limit, or 'cap'.

Why is economic evaluation used?

Requirements for economic data to be used in the decision making process clearly indicate a desire amongst policy makers to achieve greater value for money in health care and, in particular, from the drugs budget. In addition, it is clear that, in some countries, they have been driven at least partly by cost containment pressures. Rising drug costs have been seen as the main source of pressure driving up overall health care expenditure.

Controlling pharmaceutical expenditure

In Ontario, the authorities sought to impose (but later dropped) a condition that if a new drug was to be added to the provincial formulary, the company must identify offsetting cost savings from within the drugs budget in order that the new drug would have a neutral impact on the budget.

In British Columbia, although Anis *et al*⁴ emphasise that the aim of the Pharmacoeconomic Initiative was to maximise health benefits, it is clear that the policy was influenced by other considerations. They note that policy makers had become concerned with controlling pharmaceutical expenditure because of its 'disproportionate' rate of increase relative to health care spending generally. In this context, the requirement for economic evaluation is regarded as one among a range of cost containment tools.

Similarly, PHARMAC was established against the background of a growing drugs bill. It was allocated the task of 'finding new and more effective ways to manage expenditure growth'. It appears that new drugs can only be added to the Pharmaceutical Schedule if unanticipated savings are made elsewhere, perhaps by price reductions for other drugs or savings in other parts of the health care system. As the 2000 Annual Review states, 'expenditure reductions ... are often PHARMAC's only source of discretionary spending for new pharmaceutical developments or widened access to existing ones'.⁵

What impact has economic evaluation had?

George *et al*⁶ reviewed all 355 submissions to the Australian PBAC over the period January 1991 to June 1996 (voluntary submissions were possible before 1993). Of the 355 submissions, 73 were re-submissions, indicating the previous failure of a submission and therefore, at the minimum, a delay in the drug being listed on the Pharmaceutical Benefits Schedule (PBS). Of the 26 submissions for which cost per life year gained data were available, eight were rejected and, in three cases, the drug was recommended at the lower than nominated price, indicating that some prices may have been set at a lower level than would have applied without the PBAC evaluation.

From this small group of submissions (8 per cent of the total), the authors concluded that the decisions made have been broadly consistent with the use of economic efficiency as a decision making criterion.

Government controls

It is apparent, however, that the PBAC's recommendation is not the only factor in determining a drug's listing. For example, the decision not to recommend Herceptin was overruled by the Health Minister.⁷ In the case of Viagra, the opposite has occurred – that is, the political intervention has been to deny coverage, on the grounds of cost, when the PBAC has made a recommendation in favour of reimbursement on grounds of cost-effectiveness.⁸ The government possibly expected the scheme to exert greater control than has actually been the case on total drug costs, which rose nearly 20 per cent in the year to the end of June 2001 compared with the previous year.⁹

In New Zealand, Braae *et al*³ note that, following growth of nearly 20 per cent each year in government expenditure on pharmaceuticals during the 1980s, the government introduced a number of measures to bear down on drug costs but the underlying growth rate remained close to 10 per cent. Between 1993 (when PHARMAC was established) and 1998, growth of government expenditure on drugs averaged 5 per cent, and expenditure fell by around the same percentage in the year to the end of June 1999. In the year to June 2001, expenditure grew

by about 2 per cent, compared with the 9 per cent PHARMAC estimates spending would have grown in its absence¹⁰.

Given that the organisation has continued to add new drugs to the formulary and to broaden access to drugs already listed, it is likely that New Zealanders are deriving better value for the public dollars spent on drugs than would be the case in the absence of the system. It is less clear whether, by managing drugs separately from other components of health care, overall expenditure on drugs is optimal, given total health spending. A further unanswered question is the extent to which equity in access to medicines may have been sacrificed by transferring the burden of paying for drugs away from the public budget to individual patients.

The impact of PHARMAC cannot, however, be attributed solely to the policy of restricting reimbursement to new products that satisfy a cost-effectiveness criterion. Although more new chemical entities were denied listing in the year to end June 2001 (32) than were accepted (20),¹⁰ other measures have had an impact on government drug expenditure. It is estimated that most of the savings up to the end of June 2001 were due to the system of tendering: a scheme intended to increase price competition in generic markets, whereby 'sole supply' or 'preferred brand' agreements are reached.

In the case of British Columbia, Anis and Gagnon¹¹ provide details of 95 submissions made to the provincial drug plan between January 1996 and April 1999. Of the 88 submissions reviewed by the Pharmacoeconomic Scientific Committee (seven included no economic analysis), the rejection rate was around 74 per cent. In only nine submissions was a recommendation made for full benefits status and, in a further 14, restricted-benefits status was recommended.

Is there a cost-effectiveness threshold?

Australia

The logic of using economic evaluation to enhance efficiency in the provision of health care is that funding will cease to be provided for treatments with an excessively high cost-effectiveness ratio. What

constitutes 'excessive' is normally conceptualised in terms of a threshold or ceiling cost per QALY ratio. One approach to investigating decision makers' threshold cost-effectiveness ratio is to derive an implied value based on decisions already made. The only published attempt outside the UK to do this relates to decisions made by the PBAC.

George *et al*⁶ found that the PBAC's recommendations were consistent with an economic efficiency criterion of cost per life year gained. (The number of studies with cost per QALY ratios was too small to draw any conclusions). A comparison of those drugs recommended for listing and those not recommended (including 'recommended at lower price') showed a significant difference in cost per life year gained between the two groups. The maximum cost at which a drug was recommended with its nominated price was A\$75,286 per life year gained while only one drug did not receive such a recommendation below an incremental cost of A\$39,821 per life year gained. The threshold cost per life year gained appeared to fall somewhere between A\$42,000 (the lowest ratio at which a 'reject' recommendation was made) and A\$76,000.

The authors point out that the PBAC's decisions could not have been based solely on economic efficiency because there were a number of apparent inconsistencies for drugs whose cost-effectiveness fell between A\$42,000 and A\$76,000. Other caveats about these results are that, since they take no account of quality of life, the implied cost per QALY threshold may lie above the upper end of the range of A\$76,000 per life year gained and, secondly, that the sample constituted less than 10 per cent of submissions.

New Zealand

Although cost-utility analysis is 'almost always being used to assess new medicines' by PHARMAC,³ there is no published review of decisions made according to cost per QALY ratio. However, some unpublished evidence suggests that an implied threshold may be more difficult to identify, not because of variations in the considerations applied to individual drugs, but because the ability to add new drugs to the formulary depends on the budgetary situation for that particular year. With the addition of new drugs being

Table 2 Average cost per QALY ratios of PHARMAC's decisions

Year	Cost per QALY ratio	Comments
2000/2001	NZ\$20,700	Savings in this year allowed wider access to long-acting beta ₂ -agonists for asthma, costing NZ\$40,000 per QALY
1999/2000	-NZ\$1,800	Estimated on the basis of the potential savings of widening access for olanzapine to patients unresponsive to or intolerant of risperidone. The saving of NZ\$5748 per QALY gained was offset by the listing of beta-interferon, not considered good value for money at more than NZ\$80,000 per QALY, but included on the Pharmaceutical Schedule after a political intervention ⁵
1998/1999	NZ\$16,100	This was due in part to the listing of risperidone, clozapine and olanzapine for schizophrenia at NZ\$43,000 per QALY

Data supplied by Scott Metcalfe and Matthew Brougham (PHARMAC, NZ). Cost per QALY ratio shows indicative estimates only.

dependent on savings made elsewhere, the larger the savings made in any one year, the more drugs can be added, and the higher the implicit cost per QALY threshold is likely to be.

The cost per QALY ratios implied by some PHARMAC decisions are presented in Table 2, above. They are broadly consistent with a ceiling of NZ\$20,000 (around £9000 at 2000 Purchasing Power Parity values) informally regarded as indicating what constitutes a 'good buy'. Investments cover both the listing of new chemical entities and the expansion of access to chemicals already listed. For example, there were 19 decisions (not necessarily the number of chemical entities) to de-restrict or expand access in the year to June 2001.

Comparison with NICE

In terms of its underlying policy objectives, NICE is concerned with

value for money across a broad spectrum of health care interventions, as opposed to drugs alone, and does not attempt to evaluate all new technologies. The other major policy goal of NICE, which is specific to the Institute, is to reduce inequalities in access to treatment in different regions, or 'postcode prescribing', as it has been dubbed in the context of pharmaceuticals.

NICE has not been concerned with controlling overall expenditure on pharmaceuticals in the way that policy in New Zealand or British Columbia has been driven. According to NICE's estimates, its recommendations on drugs to the end of May 2002 could give rise to increases in expenditure of over £300 million a year, a figure that can be expected to increase over time as more drugs are appraised. The impact of NICE on total expenditure in practice remains to be seen, but the mandatory status of its recommendations from 2002 should ensure that the requisite funding is at least made available.

The PBAC

The PBAC is similar to NICE in not focusing on the overall drugs budget, although the Viagra decision shows that government policy makers are concerned with this. Indeed, it has been argued that drug costs in Australia may now be growing at an 'unsustainable rate'² despite the cost-effectiveness hurdle, which raises questions as to whether the hurdle is working efficiently or whether concern is focused on the absolute growth rate of one component of health spending (pharmaceuticals) rather than the overall cost-effectiveness of health care expenditure.

To date, decision-making bodies outside the UK have appeared to be more willing than NICE to reject submitted products for reimbursement. The multiple sclerosis drugs were the first example among NICE's full technology appraisals of pharmaceuticals to have received a complete rejection by NICE (although these products are now provided on a risk-sharing basis). As in other jurisdictions, NICE has frequently made recommendations to restrict the use of a drug to a more limited group of patients than that for which it is licensed.

Whether the policy objective of achieving greater value for money from pharmaceuticals has been satisfied in other countries is not entirely clear. At purchasing power parities, the PBAC's threshold appears to lie between £21,000 and £38,000 per life year gained. This compares with the implied but disputed threshold for NICE's decisions of around £30,000 per QALY gained – again, based on incomplete cost per QALY coverage for the technologies appraised by NICE. (Figures for cost per QALY or life year gained were cited in 27 of the first 41 appraisals.)

Few technologies have exceeded this figure but, assuming it is maintained, NICE may perhaps be considered less generous than the PBAC if (as is possible) the latter has a cost per QALY above the top end of the range for cost per life year gained of £38,000.

Over time, therefore, NICE's rejection rate may be expected to increase. We should, however, be wary of these comparisons, particularly given problems of generalising from the small sample used in the analysis of PBAC recommendations and the fact that other considerations come into play when both of these bodies make their decisions. For example, George *et al*⁶ note that some decisions to list a drug on the Australian Pharmaceutical Benefits Schedule may have been influenced by the 'rule of rescue' whereby the treatment in question was the only one available for a life-threatening condition among a small disadvantaged group. A similar consideration led NICE to recommend Riluzole for motor neurone disease, despite relatively poor cost-effectiveness.

Generally, when assessing the impact of bodies such as the PBAC and NICE, it should be borne in mind that wider political considerations will impact on decisions. In the UK, NICE's negative appraisal of drugs for multiple sclerosis led to the negotiation of a 'risk sharing' agreement with the Department of Health. NICE may or may not turn out to be a catalyst for other such schemes in the future. The use of economic evaluation by PHARMAC has coincided with the greater use of innovative agreements with the pharmaceutical industry.

PHARMAC

New Zealand has been most aggressive in this respect, using economic evaluation of new drugs as one of a package of measures

that also includes efforts to reduce the prices (and thus improve the cost-effectiveness) of established drugs. The contribution of PHARMAC's requirement for economic evaluation to its overall objective of controlling pharmaceutical expenditure is difficult to assess but it may well have given added impetus to some of the schemes it has devised in order to manage the drugs budget.

At first sight, the system administered by PHARMAC in New Zealand appears to come closest to the textbook approach to prioritisation by cost-effectiveness. In contrast to the situation described by George *et al* for Australia, where decision making takes place in the context of no fixed budget constraint, and 'an independent judgement about the marginal willingness to pay for health gains' is required, the textbook decision-making context is characterised by a fixed budget. Interventions are selected from a prioritised list, ranging from more cost-effective to less cost-effective, until the budget is exhausted.

This appears to correspond approximately to the environment faced by PHARMAC, which has a target for overall public expenditure on drugs, and evaluates all new claims for reimbursement on the basis of cost-utility analysis. It can then determine its strategy, encompassing the listing of new drugs, setting restrictions on their use, or de-restricting established drugs. From the cost-effectiveness of these options, combined with the budget constraint, emerges the threshold cost-effectiveness ratio.

Consistency and transparency

However, it is not clear how consistent or transparent this approach is, and as noted above, it may not be efficient to set a separate ceiling target for pharmaceutical expenditure if the overriding objective is to achieve the most efficient use of the overall health budget rather than one component of it which can often be substituted for other non-pharmaceutical treatments.

Conclusions

In a number of countries, it is mandatory for pharmaceutical companies to submit economic data when applying for the listing of

a new drug on the publicly reimbursed scheme. In other countries, economic criteria are used sparingly or are applied to a selected group of interventions. Economic evaluation has been used to influence prices (directly or indirectly through the implicit use of thresholds) and to target drugs at particular groups of patients for whom they are most cost-effective.

Whether policy makers rely on a strictly controlled budget or on an externally imposed threshold, an element of arbitrariness is inevitable. Any judgement about the appropriate cost-effectiveness threshold is in some ways an arbitrary one, made in the absence of knowledge about the cost-effectiveness of technologies that have not been evaluated and, more broadly, about the value of investments elsewhere in the economy.

If it is decided to place a tight limit on pharmaceutical expenditure, independently of health care overall, it may be difficult to take account of costs outside health care (or even outside the drugs budget) and the result may be a transfer of costs to the individual. It could mean that cost-effectiveness becomes subordinate to overall cost control and the acceptable cost-effectiveness ratio may vary from year to year for no apparent reason. Some drugs may receive approval in one year and not in another. Without information about the cost-effectiveness of other health care technologies, we cannot know how large the drugs budget ought to be.

Whichever approach is adopted, a rule of thumb is needed to guide policy in the absence of perfect information. The preferred approach will depend on the importance attached to different policy objectives. If decision makers are sceptical about the value of new drugs and if it is thought necessary to limit drug spending as a proportion of total health care spending, then the PHARMAC recipe may seem the appropriate one.

Alternatively, if decision makers are confident in their ability to select an appropriate value of health benefits and are less concerned about the overall cost impact on one component of health expenditure, then the approach adopted by PBAC or NICE may be preferred. In any event, the diffuse nature of decision making in health care means that

any centralised policy prescription will involve making trade-offs between a number of competing objectives.

REFERENCES

- 1 SCRIP. Italy strengthens price negotiation. *SCRIP* 2001; 2624: 6.
- 2 Birkett DJ, Mitchell AS & McManus P. A cost-effectiveness approach to drug subsidy and pricing in Australia. *Health Affairs*, 2001; 20(3): 104-114.
- 3 Braae R, McNee W & Moore D. Managing pharmaceutical expenditure while increasing access: the Pharmaceutical Management Agency (PHARMAC) experience. *PharmacoEconomics*, 1999;16(6): 649-60.
- 4 Anis AH, Rahman T & Schechter MT. Using pharmacoeconomic analysis to make drug insurance coverage decisions. *PharmacoEconomics*, 1998; 13(1pt2): 119-26.
- 5 PHARMAC. *Annual Review for the Year ended 30 June 2000*. Wellington: Pharmaceutical Management Agency Ltd, 2000.
- 6 George B, Harris A & Mitchell A. Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in Australia (1991 to 1996). *PharmacoEconomics*, 2001; 19(11): 1103-9.
- 7 SCRIP. Health minister overrules Australian PBAC on herceptin. *SCRIP*, 2001; 2687: 18.
- 8 SCRIP. Australia rejects Viagra subsidy. *SCRIP*, 2002; 2723: 20.
- 9 SCRIP. Australian pharma spend up 19.9%. *SCRIP*, 2002; 2711: 15.
- 10 PHARMAC. *Annual Review for the year ended 30 June 2001*. Wellington: Pharmaceutical Management Agency Ltd, 2001.
- 11 Anis AH, Gagnon Y. Using economic evaluations to make formulary coverage decisions. *PharmacoEconomics*, 2000; 18(1): 55-62.

Chapter 7

Discussion: How are thresholds used elsewhere?

The discussion centred on three topics:

- cost per QALY versus cost-benefit methodology;
- comparability between different parts of the UK public sector;
- the usefulness of international comparisons.

Cost per QALY versus cost-benefit methodology

Elizabeth Davies: It seems to me that there is a fundamental problem about the way in which the QALY methodology has been devised. It was refreshing to hear you coming back to real-life decisions that people might make about their health. If you look at the way QALYs were set up, there seem to have been a number of leaps – from disability and distress – to putting the time factor in. Then we ask people to make impossible trade-offs about situations they have no understanding of in order to arrive at a QALY measure.

A point was made about the QALY tables. If you look at some of the early estimates on which the whole idea was put forward, they are based on absolutely appalling data. For example, malignant glioma was one of the most expensive QALYs – £100,000 – and right at the top of a QALY league table. In subsequent studies, when health economists or clinicians looked at the actual cost and followed people over time rather than making estimates and models, they found that they had underestimated the costs of patients who did not receive the treatment and overestimated the cost of those who did.

I would make a plea for health economists to get into the real world asking people about the real sorts of decisions that they make, rather than the modelled, hypothetical world in which they seem to exist.

Keith Tolley: I was interested in Graham's last comments about cost-benefit analysis and putting a monetary valuation on things. I would like to clarify something in my own mind. Are you suggesting that

going down the QALY line in terms of thresholds is the wrong way for NICE, even if they are not explicit about it and that it should be a return to the fundamentals of cost-benefit analysis? You do not have to set thresholds then, because you are looking for a positive net benefit.

Graham Loomes: There is a very long answer to that, because it is a big issue. At the time when work on cost benefit analysis in health care using willingness-to-pay estimates to measure benefit was first underway, there was a different culture in this country. I recall having a PhD. student at the time, Carol Propper, who did work on what people would be willing to pay to reduce waiting times. Part of her empirical research was talking to people who had come to the University of York campus for conferences and summer schools.

The most common response was that people thought it was disgusting even to raise the question that people might pay money to jump the queue. But that was in the early days of Margaret Thatcher, and what was disgusting then seems not to be regarded as quite so disgusting a question to ask now!

In the face, therefore, of the difficulty of getting people to respond in the spirit in which we would want them to respond – that is, to find the value of the thing rather than to do market research for private insurance companies – the QALY approach was a viable alternative and the best alternative available at the time. That is my interpretation of how we got to use QALYs in healthcare.

Things have changed. We are now better at constructing questions that will get at monetary values. It should be possible to get monetary values relating to QALYs on the same basis as we do in other areas of the public sector at the moment, and that must be worth exploring.

Maybe I am wrong about this, but – and for the moment let us think of £30,000 as a number that comes from somewhere in the ether as the value of a QALY – if we use that as a cut-off point, are we not implicitly saying that it makes the benefit-cost ratio in this sector roughly comparable with the more explicit benefit-cost ratio in other sectors? If that is what is going on, then let us try to put the whole methodology on the same footing, if we can do it.

Comparability between sectors

Ron Akehurst: It seems to me that, in the end, you are saying the problem is exactly the same regardless of what approach is taken to valuing health outcomes. The issue is the things that we can do at the margin and what they displace – in other words, where the budgets bite, and what the marginal loss is.

The problem we have is that we make these decisions one at a time. We are never in a position where we have a big list and we can say ‘We will tick everything above this and reject everything below that.’ It seems to me that the question is a pragmatic one: what process can we adopt that enables us to make better decisions, given that we will always be in the kind of position that you describe?

If we look at some of the other aspects of what NICE is trying to do, it may at least start to address some of those issues. Part of the problem we have is that, at an individual health authority level, everyone is saying ‘We’ve more important things that we could do with this money, but we are being told by NICE that we have got to do this.’ Surely part of the drive to get much more standardisation, to lift the quality generally and to get much more equality across the country, is about trying to equalise those margins to try to take some of that argument away?

I do not see how we can avoid having some sort of consistency rule. That is likely to get us closest to the position of minimising the risk of displacing something that is more valuable than we are actually introducing. I do not mind whether we start from the position that you might come from, which says ‘let’s give it a value’, and then you plug that into the cost-benefit equation, or whether we work the other way round, and look at how much it will cost us per QALY. There is not much practical difference.

Graham Loomes: But there is a difference, is there not, if you are going across sectors? How do you do that cost per QALY estimate in transport? How do you do it in environment? How do you do it in other areas, where that same quality-adjusted life year technology does not exist and cannot obviously be applied? Whereas, if you could

put everything in monetary terms – if you could put things in terms of monetary benefits and monetary costs – you can at least have a benefit-cost ratio, which will be a great deal bigger than one, of course, but you can at least see how those ratios compare across sectors. It gives you that additional degree of comparability.

Adam Oliver: It seems to me that the QALY approach is only useful if you are looking at allocative efficiency within health care, or sectors that have some impact on health. If you were wanting to look at allocative efficiency across the whole of the public sector, or even across the whole of society, I believe you should be going for the cost-benefit approach – simply because you then have a comparable outcome measure across the whole of society.

My question to Graham is: do you think we should be going down that route, where we are comparing across the whole of society? If so, how would we develop NICE in that respect? Would we bring in people from education, housing, roads, and so on, in a huge NICE that looked at efficiency across the whole of the public sector?

Graham Loomes: First of all, if you believe that what is measured by the QALY in terms of benefits when compared with costs would give you allocative efficiency, and that the QALY is picking up everything you want to incorporate into your benefits of health care, then what you say would be all right within health. It is not obvious that is true. It might be that, if you try and pursue the valuation line, you find that those two things do not tally up at all, though there might be very good reasons why they do not. We do not know them yet because we have not looked at that issue.

Secondly, the idea of a kind of mega-NICE as an assembly of waifs and strays from across the public sector coming together is too horrendous to contemplate!

Nick York: I was struck by two things that Graham said. One was the interesting example of the fire engines. The other was the point about the possibility for quite large disparities in the opportunity costs of spending on different public sector interventions, such as road safety versus other alternatives.

From that, it seems to me you could conclude two things: first, that a simple decision rule (saying, for example, that if something costs more than £30,000 per QALY it should be rationed and if it does not, it should not be) will be unhelpful and second, that we need some transparent way of comparing across sectors.

I would therefore strongly favour greater explicitness about the value implied by decision-making approaches such as NICE uses. However, explicitness must not stop those people who have been asked by the elected politicians to make difficult decisions, with all the information in front of them, to do so.

You can have a threshold only as a sentinel indicator, which you use to identify major disparities in the opportunity costs of different choices in the public sector, without using a simplistic approach to making difficult decisions.

Graham Loomes: Without going over it all again, I would refer just to one thing. I would not exclude the possibility that using different values in different sectors could be appropriate. It could be in line with people's preferences. That is a possibility that we do not yet know enough about.

The idea would not necessarily be to impose uniformity across all sectors and all contexts by a single value. In an ideal world, there would be comparability in terms of benefit-cost ratios across different sectors, but allowing for the fact that the benefits may be valued differently in different sectors, for contextual reasons or for other perfectly good reasons to do with real preferences and values held by members of the population.

Overseas approaches to decision making

Richard Cookson: I am struck that in New Zealand the imputed cost per QALY threshold is about £9000 per QALY, which is less than a third of the imputed value here. I would therefore like to venture a hypothesis, which is that if you could impute values across the world, you would find that lower values tended to be associated with countries that have a smaller domestic drugs industry.

Jeremy Chancellor: I think that some of the concerns about thresholds, be they absolute or smudges, in the UK would be due to the notion that essentially there is a binary decision to be made: whether an intervention will be available or just not available at all.

There is quite a contrast here between the UK and certain other countries. In Australia, there is a much higher co-payment, certainly for drugs, than in the UK, possibly beyond the point that is efficient, because it could inhibit people using drugs. In France there is a situation of complementary insurance where the government will pay up to a certain amount for drugs, setting bands of reimbursement: zero, 35 per cent, 65 per cent, or 100 per cent. French people take out complementary insurance which pays for the balance.

It seems to me, therefore, that if we also had other means of financing interventions, the concern about thresholds would partly be defused and the value that they should take could be different.

Peter Littlejohns: Can I throw that question back to the audience? Are we saying that the differences within and across countries are so great that to try to draw conclusions from any comparisons is not possible? That is the sort of message I thought Clive was giving: that there were such different methodologies within systems, and that the reasons for them were so different, comparing them with the UK at this stage is not possible.

John Hutton: My question about this is: why would anyone believe in the first place that there is anything to learn from New Zealand about cost-effectiveness thresholds in the UK, apart from possibly the methods by which they are calculated? The whole business of international comparability is good for the travelling club, as Clive said, but it does not seem to add much to decision making.

Section 3

ETHICAL ISSUES REGARDING A COST-EFFECTIVENESS THRESHOLD

Chapter 8

Ethical issues associated with rationing and the use of thresholds

DAVID COOK

This chapter looks at:

- the morality of measuring outcomes and of rationing;
- alternative values to maximising total gain in health care;
- moral concerns: transparency, justification and consistency;
- reconciling public, political and professional concerns;
- public involvement in reaching an agreed framework for decision making.

Introduction

The WHO definition of health reveals the ideal aim for all health care systems as generating for its population ‘a state of complete physical mental and social well-being and not merely the absence of disease or infirmity.’¹ The NHS was designed to be free to all at the point of need, but ever-increasing demand and expectations, the success of medicine, new technologies and drugs create the necessity of resource allocation and questions of clinical and cost effectiveness.

The BMA has stressed that doctors have a responsibility both to individual patients and to use the most economic and effective treatments. In relation, then, to rationing, who gets what, for what purpose? Increasingly, rationing is politically encouraged, publicly acknowledged, scientifically and statistically grounded and professionally applied.

The morality of measuring outcomes

Selection itself raises problems about what is and should be measured, what counts as relevant, what are the criteria of measurement, who measures, what are the outcomes, and what counts as fulfilling them? Behind this is the inevitable question of

what is acceptable value for money? Should the aim be complete objectivity, or should it make a proper difference if the patient in question is your mother?

Problems with rationing and outcomes

There has been considerable debate about one significant attempt to bring statistical rationality into health care allocation by the use of quality-adjusted life years (QALYs). Behind the specific problems were some general concerns about all schemes of rationing by outcome. What is the balance between the individual and the community? Do we end up with allocation to each according to their ability to benefit rather than to each according to their need? Awareness of rationing may actually drive up demand. There is concern about any idea of 'one size fits all' health care rather than allowing different thresholds for different cases, a proper awareness of local priorities and the need to distinguish between local and national needs.

Others point to the hidden costs of rationing, which may lead to low morale, stress, loss of care and compassion, and reduced quality of relationships with patients. Rationing may undermine the very heart of medicine: the duty of care.

Alternative values to maximising total gain in health care

Rationing in relation to outcomes raises questions about whether or not and to what extent there may be other more important or alternative values to consider and apply. We could concentrate on reducing inequalities in health care, health care opportunities and the potential to fulfil health. We could stress equality and/or equity focusing on equal needs leading to equal access, a delivery system based on rights and legislation such as the Human Rights Act or the need to give good reasons for relevant differences in treatments.

Traditionally, the BMA has argued that no patient should be denied diagnosis or treatment because of age or other factors. Others go

further in arguing that a compassionate health care system ought to favour the disadvantaged, the elderly, the mentally and terminally ill. So, there may be other fundamental values that override efficiency and efficacy. These might make the basis of the delivery, and the standard of judgement, of any health care system such principles as the sanctity of life, avoiding personal bias, social value and contribution to society, randomness (such as first come, first served), or distinguishing between needs and wants.

Moral concerns

All of this brings us to a series of moral concerns about the theory and practice of rationing and the use of thresholds and outcomes, including the need for transparency, justification and consistency.

Transparency

There is no doubt that resource allocation and rationing do in fact take place at the micro and macro levels. The creation and practice of NICE is clearly one attempt to set some rational, objective and publicly expressed framework for rationing. This is surely a step in the right direction. For too long, rationing has been implicit rather than explicit. It has been at the vagaries of particular lobbies, powerful practitioners, enthusiasts or public pressure groups.

Western society will not tolerate secrecy or the appearance of secrecy. The delivery of health care must be open to public scrutiny. This is not just because tax payers foot the bill, but because all medical and economic decisions must be open to inspection. This is not simply one way of limiting cheating, bias and people taking advantage of and abusing the system. It is also a key way that the public are encouraged to face the reality of rationing and the need to allocate resources justly and fairly. This emphasis on transparency, openness and public scrutiny is one useful outcome of the demise of medical and economic paternalism.

Justification

In stressing the need for transparency, we are in fact arguing that

justification is vital. For people to accept rationing and the fact that it is necessary, good reasons must be given and accepted. The decisions of NICE, outcomes, the use of thresholds, rationing and bases of resource allocation decisions must be justifiable and justified. That implies some level of agreement about what is defensible and what is agreed. It also draws on some generally accepted and agreed grounds of what is factual, scientific, rational and objective.

Consistency

Consistency is crucial in the pursuit of science, medicine, economics and even political decision making. The alternatives – inconsistency and irrationality – make all understanding, reflection and evaluation impossible. For consistency in rationing decisions, principles and practice, there must be agreed, widely recognised and accepted valid criteria. There must also be solidly based, well-established results. The Cochrane initiatives and the development of evidence-based medicine have been a major step in the right direction.

All such decision-making principles and applications must be generalisable. By this I mean that in similar cases, other things being equal, the same process, principles, applications and results would follow and be accepted. If it does not work in every similar case, then how can it be claimed to work in this particular case? If it is not generalisable, then it is in danger of being invalid and morally flawed. It is certainly no basis for consistency and rational justification.

The public, political and professional spheres

Rationing touches on various levels of public, professional and political life and perceptions. Members of the general public are extremely concerned about the practice and reality of rationing. They are particularly incensed when their locality, family, or concerns seem to be disadvantaged either in comparison with other localities, families or concerns or with their expectations. This raises the proper question as to the accuracy and validity of such expectations and of such perceptions.

Members of the public form their perceptions of rationing from their own experiences of health care delivery, gossip, experiences of their

friends and families, and the media. Indeed, much of the battle for the hearts and minds of the public is fought in the media, especially in the world of the soap operas and the tabloid press.

Public perceptions and concerns intersect with professional perceptions, practices and concerns. There is widespread disillusionment with 'professionalism' in our society. The so-called 'professional foul' sums up the idea that the professional is someone who gets away with something wrong. Historically, a professional was someone who was part of a well-trained, qualified and supervised body of respectable practitioners. To be part of a profession was to have a stamp of approval and to be worthy of trust. The public accepted the professional as someone who provided a service and upheld high moral values in their practice.

The confidence of the general public in the medical – and particularly, the health economist – professionals needs to be safeguarded and recovered. Transparency, appropriate justification and better handling of the media may all have a part to play.

The waters are muddied by the inclusion of politics in health care delivery and rationing, and by the role it plays in this field. Politicians are deeply concerned about health care but are not above using health care horror stories for their own particular purposes. Any – and perhaps every – form of rationing is subject to political pressures and abuse. Somehow it may be possible to remove health care and discussions of resource allocation and rationing from the realm of the political infighting to that of a consensus about the necessity for health care, and its basic provision, on rational, publicly accepted and agreed grounds.

Until there is some move to disentangle the public, professional and political agenda so that areas of commonality – and those of genuine disagreement – can be identified and addressed, there is little hope that there will be a significant move to establish valid grounds for rationing. This means a conscious move towards some kind of agreement, and that will only happen if a model of partnership rather than of conflict is the norm.

Public involvement

There does seem to be a general level of public disquiet around the area of rationing and resource allocation. This is fuelled by the media, various claims and counter-claims of the effectiveness of drugs and treatments, and concerns about spiralling costs. One alternative to trying to present a more positive perception of NICE and its work in the media is to find some means of public involvement in decision making on resource questions.

Famously, this was attempted in Oregon, where the public was invited to rank various treatments and conditions in order of priority. Switzerland and Canada have tried to increase public involvement and participation with 'town hall' meetings, at which particular medical issues, such as xenotransplantation, have been described, discussed and voted on by local communities on a 'rolling roadshow' basis.

Alternatively, focus groups have been drawn together and presented with the necessity of decision making on such issues as insurance for the carriers of genetic-based diseases. All of these initiatives have seen genuine involvement among some of the public, but it is far from clear how effective such steps are in increasing the more general public awareness, helping counter misperceptions and misunderstandings and helping the public to take actual responsibility and to feel genuinely involved in the process of decision making about rationing and the allocation of resources.

The role of the media

It is clear that the press and other forms of broadcast media play a key role in public perception, confidence and approval or disquiet. The events of Alderhey, where the organs of dead children were removed for research without the consent of their parents, show how public confidence can be lost and that this effect can snowball to the extent that pathology – particularly paediatric pathology – has almost ground to a halt because of public disquiet.

It also reveals the knee-jerk reaction of politicians and Government. Of course one might say that the politicians are being sensitive to

public opinion and responsive in nature to public expressions of disquiet. There is no defence for improper, immoral abuse of medical power, but there needs to be a careful investigation into how we as a community view medicine, its delivery, practice and decisions about the use of medical resources.

One context in which we may be able to make a positive contribution is in the realm of the soap opera. Many social and medical issues are brought to the attention of the public by means of story lines in the soaps. The first mainline viewing time homosexual kiss, the reactions to HIV diagnosis and underage pregnancies are all examples in which health care issues have been the focus of public attention and helped create positive responses to such issues.

Given the public's insatiable desire for 'docu-dramas' based on hospitals, and the way that illness and life and death concerns are played out in the daily grind of the world of soap operas, it is not beyond the wit of health care economists and the medical profession to design story lines that present not just the dilemmas of rationing but the need for solid grounds and good decision making.

Some useful reminders

1. Recognising the differences

Regionalisation has been an attempt to recognise and respond to local differences. However, the idea that one system of health care allocation can be applied as a straight jacket rather than an overarching guide is dangerous. It imagines that there are no significant differences between rural and inner-city settings; hospitals and general practice; hospice care and acute medicine contexts; and that exactly the same standards can – and should be – applied in exactly the same way, regardless of significant differences.

While it is vital to have some objective and overall standard, this does not mean that it must – or should be – applied without careful reflection on what may be significant differences and subtleties. The key here is to distinguish what are relevant and important contextual differences and what are merely matters of preference or bias.

While it certainly adds to the complexity of rationing and resource allocation, to fail to take seriously crucial aspects of settings, practice, patient and professional differences will inevitably lead to failure and to local rejection or some indirect means of subverting the actual processes and standards of resource allocation. The lesson of waiting-list manipulation is instructive. Good threshold and outcome settings must be sensitive to local and specific differences.

2. Changing perceptions of clinical excellence

Twenty years ago, when pathologists took samples from dead children without consent, they were simply following a long-established practice. While we can now claim that this was a form of medical paternalism and a breach of the good practice of giving, taking and receiving consent, it is clear that most pathologists were not simply doing their own thing, regardless of the needs of others and medical niceties. Rather, there was an honest belief that to ask grieving parents for permission to remove organs or parts of tissue was intrusive, distressing and painful. Doctors felt that most parents would want to avoid any repetition of whatever had killed their child, and so would be pleased that their child would be a help to others.

The motives for this research in pathology and the taking of samples were often twofold: the need to discover the causes and developments of diseases, and the need to provide material to teach doctors how to recognise symptoms and diseases. Many pathologists are shocked that what was considered good practice 20 years ago is now vilified in the press; indeed, the resultant reaction is so severe that the very specialty is at risk, never mind the research into the causes and development of disease.

Good clinical practice is not a static state. It is constantly in the process of development and, hopefully, improvement. The rise of evidence-based medicine is not just some latest fad in medicine. It is a recognition that there are better ways of practising medicine, and thus good clinical practice changes.

In the world of rationing and resource allocation, it is vital to be sensitive to changes in clinical excellence. Indeed, the very creation of

NICE is in itself a move to improve and develop further clinical excellence. It is important not to make the mistake of solidifying some means of (and standards for) judging clinical excellence, outcomes and the use of thresholds to such an extent that there is a stultifying static, rather than a dynamic, sensitive, adaptive approach to resource allocation.

3. Taking pluralism seriously

We live in a pluralist society, with a variety of moralities and ethical values. We need to work towards a consensus in which we can agree what our key fundamental values are and how they should be applied. We need an agreed framework for decision making. That may include specifying certain core treatments, procedures and health provisions for all, and distinguishing that core provision from elective treatments and preferences that would be paid for differently.

Increasingly the perspective of the patient is the focus of health care delivery, and so that perspective must be paramount in rationing reflection and practice. There must be room for a proper subjectivity that treats people as subjects and not just as objects, statistics or units. Rationing must be relational and must incorporate an awareness of – and concern for – human relationships. This may be possible through critical reflection on what is in the best interests of a particular patient. We may well find that what is in the individual's best interests is almost exactly what is in the best interest of everyone in similar situations.

Conclusion

Moral reflection on rationing and resource allocation is not an optional extra: it is an essential part of ensuring not just that justice is done, but also that it is seen to be done. Public perception and the media will demand and expect no less.

REFERENCE

- 1 World Health Organization. *The First Ten Years of the World Health Organization*. Geneva, World Health Organization, 1958.

Chapter 9

A commentary on ethical issues

ALAN WILLIAMS

This chapter looks at:

- definitions of ‘need’;
- who should assess need;
- measuring effectiveness and valuing health;
- ‘postcode thresholds’;
- the need for trade-offs.

Introduction

Rationing and the use of thresholds have always been with us. Historically, they were encapsulated in ‘clinical priorities’ and ‘indications for treatment’. It has never been possible for doctors to do everything that might conceivably do somebody some good, so they have exercised their own judgement as to what was a reasonable interpretation of their duty of care to the various people they serve. Time and effort and other health care resources devoted to one patient could not be devoted to another, so they had to decide how to allocate these scarce resources so as to do the most good, as they saw it.

A key element in exercising that judgement – within the NHS at least – was responding to need rather than to willingness and ability to pay. But replacing willingness and ability to pay with the assessment of need does not abolish scarcity or the need for rationing and cut-off points.

Instead, what it does is to replace rationing by price with rationing by need assessment, and to replace a cut-off point (or threshold) determined by each individual’s willingness and ability to pay for their own treatment with one based on the whole community’s willingness and ability to pay for treatment for an unknown fellow citizen (who just might be themselves). It is this latter concept that is

the subject of our present discussions because that is what the cost-effectiveness threshold used by NICE represents.

Definitions of ‘need’

In these discussions, much turns on the meaning of ‘need’, which is a hotly disputed concept, too often used loosely and without clear definition, thus enhancing its polemical usefulness at the cost of obfuscating the key issues. The first of these key issues is whether we are referring to a need for health or to a need for health care.

A person could be said to have a need for health if their health is having an adverse effect on their capacity to flourish (i.e. to lead a fulfilling life). Whether this need for health could be met by offering them health care is, however, questionable. And whether offering them health care would be the most cost-effective way of meeting their need for health is even more questionable.

In the first case, the question to be answered would be: ‘How effective would the offered health care be?’, and in the second case the question to be answered would be: ‘Are these effects worth the sacrifices that others will consequently have to make?’. Thus, establishing a need for health care calls for rather more stringent tests to be applied than would be required for establishing a need for health. So it is important to be clear and explicit about what is meant by allusions to ‘need’ in discussions about health and health care. In the introduction to this paper, I used the word ‘need’ to refer to a need for health care, rather than the looser concept of a need for health.

Who should assess need?

The second key issue about the concept of need is who is to assess it. As a patient, I may assert that I need a certain drug but discover that the NHS will not provide it. If we were to investigate how my disappointment came about, we might find that my doctor does not believe that the drug I want would be effective in dealing with my health problem (thus answering the first of my earlier questions, about effectiveness, negatively). Alternatively, it may be that there are

some clinical guidelines that reserve this drug for later use after a more cost-effective treatment has been tried first. Or it may be that NICE has determined that the cost of this drug in relation to its benefits is so high that the sacrifices imposed on others by diverting part of the NHS budget to purchasing it for all the patients in my circumstances would simply be too high, and it is excluded.

Thus it is not my assessment of my need for health care that counts, but the assessment made by doctors, scientists and budget-holders (who may all be wrapped up in the same exceptional person, but more frequently are different people each playing a complementary role).

The separation of complementary roles

This separation of complementary roles is a third source of difficulty with need assessment because in a clinical encounter between a doctor and a patient the doctor is simultaneously playing several roles, which frequently conflict with each other. Respecting the autonomy of the patient might be held to mean doing what the patient wants, but, as in the preceding example, the fact that a patient wants a particular drug does not necessarily mean that the doctor will prescribe it.

As well as the doctor's duty of care to the patient, there is also the duty to maintain the integrity of the medical profession by behaving in a disciplined manner in accordance with the best available knowledge. In addition to all that, the doctor is part of a much bigger organisation or practice (of which he or she may also be the manager), which faces budget (and other resource) constraints and policies, and to which he or she is accountable.

In primary care (where patients typically have more scope for exercising initiative than is the case with hospital care), such policies are likely to concentrate on prescribing, referrals, appointment arrangements, home visiting, and so on, all of which may constrain the exercise of the 'duty of care'. And to cap it all, the doctor may also have family or other responsibilities that may also constrain his or her availability to fulfil all these other duties.

Balancing responsibilities

These personal, clinical and social responsibilities all have to be balanced, and always have had to be balanced. These problems were not created by the NHS, and they were certainly not created by NICE and its (alleged) threshold. What has happened recently is that they have been brought out into the open much more, as the medical profession has been exposed to much closer scrutiny, with concomitant demands for greater public (as opposed to intra-professional) accountability.

I do not believe that this challenge to the authority of doctors, as encapsulated in their changing role in the field of need assessment, is actually an ethical issue at all, but an additional psychological and emotional strain placed upon the medical profession by having to share this role with others. The ethical issues are the same as they always were.

Measuring effectiveness and valuing health

But the meaning, measurement and valuation of ‘need’ in priority-setting (in other words, rationing) and in the setting of a cut-off point, have a counterpart in the meaning, measurement and valuation of ‘health’ in priority-setting. Measuring effectiveness means measuring the impact on people’s health.

The measurement of health ranges from narrow biomedical parameters such as blood pressure or tumour size, which focus on the functioning of the body, through condition-specific measures, which focus on the particular aspects of a patient’s functional capacity and/or feelings that are of particular salience for that condition, to generic measures of health-related quality-of-life, and, ultimately (though quite impractically) to the WHO definition of health, which is more a measure of general wellbeing.

Since NICE has to make comparisons across many different conditions, and it needs to do so in some standard way so as to be even handed between rival claimants, it has no choice but to adopt some standard generic measure. And since it is in the business of

making recommendations from the perspective of the whole community, it needs to apply a set of values that represent the values of the whole community. Thus it is the values of the citizens, not the values of interested patient groups (or their champions among health professionals), that should dominate.

Having different rules for different people would only be ethical if these groups are defined by some characteristic that has moral relevance. Thus on grounds of distributive justice we may wish to have different cut-off thresholds for the old and the young, but not for light-skinned people and dark-skinned people.

‘Postcode thresholds’

An important issue that this poses is whether suffering from one particular disease rather than another is a morally relevant matter. Should we be providing much more elaborate terminal care for AIDS patients than for cancer patients (or vice versa), or should both be subject to the same criteria and thresholds? Is where one lives a morally relevant consideration? Should we take into account local needs and priorities, so that such terminal care is provided in some places but not in others, or would this be inequitable in a ‘National’ Health Service?

I believe it is inevitable that if budgets are decentralised (as they have to be) and if local bodies are setting priorities according to what they believe is best for their own communities (as they are being encouraged to do), then we have to accept ‘postcode rationing’ as an inevitable consequence. And behind postcode rationing lie ‘postcode thresholds’. Although NICE’s decisions are now mandatory, the way in which these are implemented locally, and what is foregone to fund these treatments, will vary. The tension between centralism and localism persists.

The need for trade-offs

Notions of fairness seem to be the dominating ethical concept in discussions about rationing and thresholds. Economists typically

contrast equity with efficiency. In the simplest possible terms, 'efficiency' means maximising something, and 'equity' means being concerned about how that something is distributed between people.

In health care, the standard assumption is that the NHS is trying simultaneously to maximise population health and reduce inequalities in the distribution of health, with health being measured as quality-adjusted life expectancy. In this consequentialist world, utilisation, access and resources are all seen as instrumental to these two primary objectives. And because neither of the primary objectives trumps the other, we will need to elicit what trade-offs people are willing to make when they come into conflict.

In philosophical discourse, such discussions usually end rather lamely with the observation that since all principles of justice are contestable, and none is absolute, some balance has to be struck between them. But economists pick up the story at this interesting point and seek to quantify what the general public thinks such trade-offs should be. This is essential if the need for transparency, the need for consistency and the need for justification are all to be met. It is the most constructive way to take pluralism seriously and to put it into practice.

Chapter 10

Discussion: Ethical issues regarding the cost-effectiveness threshold

Alan Williams: The big tensions raised by David Cook, but left unresolved, are: (1) that we need to adopt the patient perspective, and (2) that we need to set priorities for the whole of society. The question is: which is it to be?

Adam Oliver: John Butler's book *The Ethics of Health Care Rationing* refers to these as first-order and second-order rationing. The first order is population-level rationing (of the sort NICE is interested in), which involves deciding which treatments and which particular patient groups you choose to ration against or in favour of. The second order rationing involves allowing physicians to have a choice about when they are to operate or treat patients, and so on. Both types of rationing exist, and do not necessarily conflict with each other.

David Cook: It is possible to build into the value some allowance for individual patient preferences and wishes. This would make the values more complex, but more realistic. This approach would also be more comfortable to members of the medical profession who, in the end, will have to apply and deliver the system.

Jack Dowie: The two things are fundamentally in conflict. The best interests of the patient (singular) is the objective of private medical care. The best interests of the patients (plural) must be the objective of any publicly funded, resource-constrained system. You cannot respect both cost-effectiveness, including equity-weighting, and individual patient preferences in a National Health Service. They are completely logically-incompatible goals. Fudging about them is the root of the problems which all health care authorities, including NICE, are currently facing.

Ron Akehurst: One of the ways in which the conflict manifests itself in a NICE-type context is that we might do a cost-effectiveness analysis and say that there is a sub-group of patients, defined in some way (for example, by risk assessment) for which we are clear that this is likely to be cost-effective, and another part of the population for

which it is not. The problem is: if the clinician is using a set of selection criteria that do not fit with the analysis, in reality you end up achieving a cost-effectiveness that is much poorer. One of the ways in which you might get around this is to be more explicit about the rules, follow it up with guidelines and, eventually, follow it up with something like the Commission for Health Improvement.

David Cook: The role of the civil service is not to prescribe, but to offer various descriptions of scenarios. Is NICE in the business of prescribing or describing?

Peter Littlejohns: The Institute was there at the beginning to describe good practice and to provide guidance to the NHS on cost-effectiveness. This has had degrees of reinforcement by Government in various ways. From the Institute's point of view, its role is still to provide guidance. NICE is more into the descriptive than the prescriptive. This is the ethos NICE is trying to maintain.

Alan Williams: Is NICE's guidance viewed as mandatory?

Peter Littlejohns: Until recently, NICE's advice was guidance – and is still, at an individual level. Within the context of a professional-patient interaction, any professional is at liberty to deviate from that guidance. However, what has changed, subsequent to government directive, is that the reason for not adhering to NICE guidance on the grounds of cost does not exist.

Chris Heginbotham: In response to Jack Dowie's earlier comments, it is not a choice between simply saying either you deal with individuals through private medicine or you take a utilitarian approach to providing the best care you can for the majority. What we have to do is to try to provide the most responsive health care we can for individuals, subject to their needs, within the context of a properly resourced and thoughtful public provision. We have to think very carefully about what we can provide and how much we can afford to fund the different sorts of treatment. But within that, we clearly have to try to do the best for the individual. Rather than saying it has to be one or the other, we have to find a way of achieving a balance between the two factors.

Alan Williams: How, then, is NICE to deal with issues of distributive justice?

Chris Heginbotham: Clearly these are very difficult issues, and indeed we have not tackled them today. The issue, discussed in terms of thresholds for NICE appraisals, is very much a one-dimensional aspect of this, although maybe a number of criteria go into it. But that is just one step in the process: deciding on some criterion or criteria on which you will fund a particular drug or a particular device. Then this decision has to be put in a much wider political, social, ethical and organisational context.

From the perspective of having run a health authority for many years, it is difficult to make these decisions on the ground. When it comes to the crunch, the decisions you have to take have to do with the local, political, social and organisational context. While we try to take into account the opportunity costs, we do not, in NICE appraisals, take into account some of those other treatments foregone as a result of now having to fund a particular treatment. At the sharp end, however, you do have to take that into account and, at the end of the day, you do fudge through. However, we should try to fudge through better. We require much clearer distributive justice criteria. Difficult judgements are being made all the time. Anything that can help us to do it better would be welcome.

Adrian Towse: In NICE Appraisal Committee discussions at present, the question of who is benefiting from the QALYs must have an impact on the extent to which the Committee is prepared to go over a particular threshold. For example, if it is children who would primarily benefit, that is factored in the QALYs, but it might still be an additional consideration over and above the impact reflected in the QALYs. Another example, if it is only going to cost the NHS (say) £1 million to treat all of the people who will benefit but it has a high cost per QALY it may be accepted because of how horrible the disease is.

Referring back to Chris Heginbotham's point, there is some history here. It comes primarily out of the internal market, where individual health authorities had to make decisions about whether they were

prepared to pay for extra-contractual referrals for particular operations for particular patients. One way forward taken by the health authorities was to set up local priority forums to help them introduce some kind of consistency into the sorts of trade-offs they were having to make. We have been round these problems once before. There are routes forward and ways of involving the public in trying to get some understanding of the trade-offs that NICE is struggling with now.

Section 4

PUBLIC INVOLVEMENT

The discussion recorded here took place in March 2002. NICE has subsequently announced its establishment of a Citizens Council, consisting of 30 members of the public, to take part in its deliberations. At the time of this book's publication the first meeting of the Citizens Council was scheduled for three days from
21 November 2002.

Chapter 11

Public involvement in health care decision making

ROBIN CLARKE

This chapter looks at:

- definitions of ‘public involvement’;
- whether the public is able and willing to make trade-offs;
- case studies from Leicester and Coventry.

Introduction

Over the past five years, the Institute for Public Policy Research (IPPR) has been concentrating a good deal on how to involve the public in decision making. We have looked across various different forums, including local government, health and central government. One of the things we do at IPPR is to hold Annual Public Involvement Awards, which are chaired by Anna Coote and run jointly with the *Guardian* newspaper. We try to use them to reward those organisations that we think are doing innovative things, or that are taking a bit of a risk by opening themselves up to public involvement.

Reasonable expectations

I approach this topic neither as a health expert nor as an economics expert, and knowing precious little about QALYs. Drawing from my public involvement experience, I was asked to come along and give my perspective about what it is we can reasonably expect the public to contribute to debates about the treatments the NHS should provide. What type of mechanisms are out there, and what are the likely outcomes from having a more public-involvement approach?

I approach this from the perspective that public involvement is a good thing in itself, in civic republicanism, citizenship-type terms. But it is also a good thing in instrumental terms, as it can lead to better-quality decision making. However, I am not starry eyed about public involvement; I think bad approaches to public involvement are worse

than doing nothing at all. There are no guarantees that public involvement will lead us into a public policy wonderland.

What do we mean by ‘public involvement’?

Often we think that ‘public involvement’ refers to just going out and asking people their opinion on something. The things that spring to mind are the usual opinion surveys or focus groups – the sorts of things that newspapers pick up on. Over the last few years, however, there has been an explosion in new methods and approaches to public involvement.

The range of involvement

This is not the place to plunge into a detailed discussion of the differences between types of involvement methodologies. However, I want to illustrate in Figure 4 that ‘public involvement’ is a catch-all term for what are in actual fact a whole host of different approaches. It encompasses everything from the familiar opinion surveys and focus groups through to deliberative polls and citizens’ juries. It has a number-crunching quantitative side and a discursive qualitative element.

Figure 4 What do we mean by public involvement?

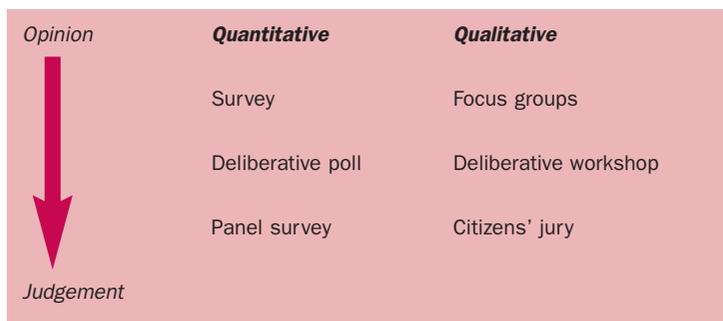
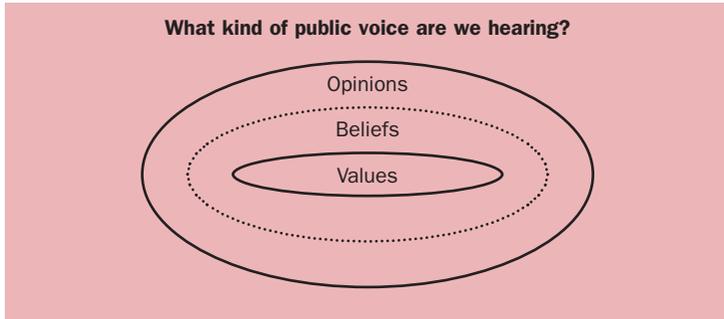


Figure 5 **Types of public voice**

People traditionally tend to associate public involvement with surveys and focus groups. These approaches are those towards the top of this diagram. However, newer and more deliberative approaches have been introduced in the last few years, to ensure that the public are able to have a meaningful input into increasingly complex decisions in which they have been invited to take part. These newer approaches are shown further down in the diagram.

There is no right or wrong method of public involvement. It is all about what you are looking for and what you need to get out of people. For example, do we want to be able to put a hard-and-fast percentage on something? Or are we after more depth? Do we want to find out what people think at the moment, regardless of how much or how little they know? Or do we want to find out what a more informed public would think, if they had access to more information and the time to look at it?

Opinion vs. judgement

In Figure 4, the arrow on the left-hand side shows 'opinion' at the top and 'judgement' at the bottom. Figure 5 looks in more detail at these two factors. When we talk to the public, we hear many different voices. The ones in which we are primarily interested are those articulating opinions and beliefs. There are people who say that

opinions derive from beliefs and beliefs from values, but I do not subscribe to that. I think that values often stand out as something separate and are often inconsistent with opinions and beliefs.

This illustrates that there are a number of different elements to the public voice that we often hear when we ask people for their views on something:

1. Values

These are the immovable core. They are views that are deeply held and are extremely unlikely to change. Opinions and beliefs do not necessarily derive from values. Values are personality- and culture-driven.

2. Beliefs

These are often passionately and firmly held. They are issues that people have thought deeply about and have come to an informed decision about over time. Beliefs are information-driven.

3. Opinions

In the outer ring, opinions are usually referred to as those views that are offered, but that by and large are perceived as free-floating. This line of thought is now being challenged. It is probably more correct to say that someone may well express opinions on issues that are new to them, but that whether they are free-floating or more anchored depends more on how they are elicited. Opinions that are sought through deliberative processes are more anchored and therefore are fast tracked beliefs. Therefore the two rings of opinions and beliefs can blur.

Typically, when talking about public involvement, people have focused on the outer ring – the opinions. These are the more free-floating views. In this category, if you ask someone the same question, they might say ‘x’ on one day and ‘y’ on the next. The beliefs, on the other hand, are what people feel they have thought more about; they are more information-driven. The quality of that information may

well vary, but the answers in this category will be more settled.

In public involvement terms, when it comes to dealing with complex issues, we are trying to fast track from the loose, uninformed opinions through to the more settled beliefs. We are asking people to look at information, to think about it, and then to come up with what they think.

The new sort of public involvement is focused on trying to get at the more grounded beliefs rather than the free-floating opinion.

Is the public able and willing to make trade-offs?

Members of the public are often willing to make trade-offs. They are able to do it, but whether that shows an actual, good quality ability, depends on one thing: remembering the three 'R's: reason, reflection, refinement. If you go out and ask members of the public their immediate or 'top of mind' opinion of rationing, they will say the first thing that comes into their heads, and it may not be what they think if they had a chance to engage with the issue. A lot of the newer public involvement involves trying to incorporate these three elements, as follows:

- 'reason' is taking the opportunity to access information, to internalise it and to discuss it;
- 'reflection' is taking the time to step away from the information, internalise it, think about what has been said, and to think about what other people have said;
- 'refinement' is the obvious last step – to be able to come back and say 'In the light of all of that, this is what I think, as a more informed person.'

Two case studies

1. Leicester

A few years ago, Leicester Health Authority had a service reconfiguration issue. There were three acute hospitals, and this was

not viewed as the most efficient way to deliver services. The health authority asked 'Why not go for two acute hospitals and one planned care and rehabilitation (PCR) site?'. Not surprisingly, personnel across the hospitals wanted to keep the glamour of being the acute site and no one wanted to be the PCR site.

The public did not like the idea either. There was a petition with over 100,000 signatures, saying 'Don't touch our health services'. The Health Authority decided to run a citizens' jury. They got 16 people together – ordinary members of the public – for four days and asked them the question: 'Do you feel there is a need for change and, if so, what would you do?'. They were given all the information; all the different sides of the debate were represented, from the health authority through to the 'Don't touch our hospitals' campaign.

At the end of the four days, these 16 people said 'Yes, unfortunately we do feel there is a need for change. The current system is not the best way of running things. We think that the best way is to have two acute hospitals and one PCR site, and this is how we would put it together.' The media covered the whole process of the jury. It was front page news and was also covered on television and radio. In the end, the opposition died away. People had bought into the process that those 16 people were doing a lot of hard work to try to come up with a more community-wide perspective.

2. Coventry

The second illustration comes from Coventry. I went there to do a 'top-of-mind' survey and also a more deliberative one. Both surveys related to health care spending priorities over a five-year period. In the top-of-mind survey, those questioned said 'We will spend more money on the glamorous acute services and primary care, and take it away from maternity services and mental health services'. In the deliberative one, where people had a chance to assess what the need was, respondents said 'We will put more money into maternity and mental health services'. The public do not always have a knee-jerk response, therefore.

Where does this leave decision makers? First of all, it is threatening. They may claim that they are elected to make these decisions; but we

Figure 6 **The public involvement balance sheet**

+	-
Better/public interest decisions	Drawn out decisions
Money saved	Resource intensive
Rebuild trust	Lack of leadership

would say that is a very passive notion of what it means to be a representative or a decision maker. A more active response is the right one to make. Second, these exercises are meant to supplement what is out there. They are not replacing it; they are not decision-making processes in their own right. They are adding to the decision-making process. If you like, they are decision-advising processes. Finally, I do not believe there is much evidence that the public actually do want to make the decisions themselves. All they want is an equal say at the top table where the decisions are being made.

Figure 6 is a public involvement ‘balance sheet’. There are pluses and minuses to public involvement. On the plus side, it can lead to better quality decisions based on fuller information. On the downside, it can lead to decision processes being drawn out.

On the plus side, it can lead to money saved. There are numerous examples of public involvement where people have said ‘No, don’t provide that service. It won’t be used.’ The service is not provided and the money is saved. But conducting a public involvement exercise can be resource intensive in terms of time and money.

A crisis of trust

The big picture is that we have a crisis of trust in all our institutions. Public involvement is just one method of starting to rebuild trust – if it is done right. The downside is that it can be perceived as displaying

a lack of leadership. Sometimes you cannot win. If you do go out and ask the public their opinion, you are accused of being indecisive and showing weak leadership. If you do not go out and ask for public opinion, you are accused of being arrogant and out of touch.

Chapter 12

The Citizens Council

PETER LITTLEJOHNS

This chapter looks at:

- what the Citizens Council is;
- how it will work.

Introduction

In the three years that NICE has been functioning, a number of issues have come to light that were not considered in the original set-up legislation. With such an innovative approach to improving the quality of health care services, it is inevitable that not all the challenges could be predicted, despite careful planning at the beginning.

A subjective process

The first issue is that despite all the scientific rigour underpinning NICE, there remains (and always will be) a degree of subjectivity in the final decision-making process. Within that context, we have used dialogue with stakeholders and consultative documents to try to tease out the areas of judgement and make the processes as transparent as possible. There is one stakeholder that has not come through very strongly with its views, however, and that is the public: the user and (through taxation) the purchaser of the service.

We address specific patient group concerns within the appraisal programme. However, by virtue of the way that the system has been established, the programme often targets a selected group of individual patients with a specific disorder. We have never had full input from the public – the potential patient – and that is something that needed to be addressed.

Value judgements

The second issue is that of values. This became of paramount

importance when we were considering the future guideline programme for NICE. One of the commissions from the Department of Health is a portfolio of guidance for the management of infertile patients. All of us are aware that there are many value judgements to be made in this area, as well as in resolving the technical issues around establishing the cost-effectiveness of different interventions in this field.

With that in mind, the Government and the Institute decided to take forward an approach that would bring the public into the decision-making process. In 'The NHS Plan'¹, the Government announced that a Citizens Council would be set up, as a way of trying to get public values into the decision-making process. We have spent more than a year working with various organisations and groups, in order to define very carefully how it should be initiated.

How will it work?

What we have decided to do, approved by the board of NICE in November 2001, is to go to the qualitative rather than quantitative side of public involvement – the citizens' jury approach. We will establish a Citizens Council consisting of up to 30 individuals, who will meet twice a year for a four-day period, and serve for three years. They would be representative of the public, and there are various processes to ensure that they represent the different groups within England and Wales.

A prerequisite is that they should not be part of any health care service (public or private), either in terms of the provision or management of health care or the supply of ancillary goods or services, including those lobbying on behalf of these groups or of patient groups. Those of you who followed the initial Oregon experiment (in which interventions were ranked in order of priority for funding) will be aware that lobbying by these groups at public meetings called to obtain public input into priority setting was a problem.

Running the sessions

At each session, the Council members will be presented with

evidence by expert witnesses. The meetings will take place in public. There are discussions still to take place about whether these will be televised, either live or subsequently. The subject will not be specific to an individual appraisal or a guideline. It will be defined by the Institute and its Advisory Committee (a board sub-committee), in terms of developing the generic social, ethical or moral values that the public would expect to see as part of that decision-making process, be it within the advisory committees for appraisals or guidelines.

Suggestions may come from Institute staff, ministers or the board, or may exceptionally be generated by the early stages of a technology appraisal or clinical guideline development. An external group will be used to generate a specification for the meeting based on the topic or topics identified, and to report to NICE on the proceedings. That information would be reviewed by the Board Committee and then fed back to the relevant decision-making groups within NICE.

These plans have gone through a number of discussions. The overall framework has been agreed by the NICE Board, but we have yet to establish the Citizens Council. We have employed a project worker who will be taking this forward, and it will be established over the next year. From our point of view it is a very challenging approach, but I think that it will address many of the issues that the authors of this book have aired.

REFERENCE

- 1 Department of Health. *The NHS Plan*. London: Department of Health, 2001.

Chapter 13

Discussion: Public involvement

Stirling Bryan: Robin's presentation was a sales pitch for public involvement that almost was not required, I would argue. Within this room, there is probably quite a substantial buy-in to the notion of public consultation and public involvement. The key issue is how to elicit values in the sense of willingness to see the NHS pay for one treatment rather than another. In doing so, we have to think about the frameworks within which we want those values to be used.

From a health economics perspective, we have much evidence of public surveys to elicit values. We do that as a necessary requirement, in terms of construction of QALYs. Now what we are talking about is seeking values on a QALY, in order to try and establish a threshold. There is buy-in to public consultation, but we need to focus on why we are doing the public consultation.

Robin Clarke: There is a horrible trade-off in public involvement. We can either go for real depth and find out what people really think about quite complex issues, and come up with something that makes common sense, where they have weighed up the different options – or you go for numbers. Numbers tend to lead you towards a lack of depth.

There was a stab at trying to marry the two, through a deliberative opinion poll approach; but no one can afford it. It will cost several hundred thousand pounds to do a proper deliberative opinion poll. Not surprisingly, the only people who have done it in this country are Channel 4. I do not think that it is a reality for the public sector – though it would be a nice thing to do.

Nancy Devlin: It would be quite useful to draw this discussion specifically towards the topic of today, which is how we go about using the involvement of the public to identify what an appropriate threshold might be or what the value of a QALY is.

One thing that I would like you to debate between you is what is the most appropriate way in which to get the public involved in ascertaining an appropriate threshold. Is the Citizens Council – the

use of a small number of the public deemed somehow to be representative – the best means of approaching this? Or is the best approach a more widespread use of members of the general public in some sort of sampling approach, perhaps using the kind of conjoint analysis methods Graham Loomes described in his paper, to ascertain willingness to pay? What is the best way forward? What is the best way of involving the public on the key issue that we have asked today?

Peter Littlejohns: My view as an individual – not as an Institute member – is that we want both. What we will get from the Citizens Council is a set of values (in the sense of ethical principles) related to the criteria that should be addressed when one takes the next step forward in identifying very specific research programmes that aim to define the NHS's willingness to pay for QALYs in particular instances. The comment from the floor earlier on was that the actual research base for QALYs is still not so strong that one would like to make national decisions on thresholds – certainly not decisions that all of us would feel happy with putting forward in the public arena.

I think it is a two-stage process, therefore. The Citizens Council will define the areas at which appraisal committees and clinical guidelines will start looking in more detail, and a research agenda will come out of that. It is an iterative process.

There is still a very active research agenda to take place within the world of QALYs. We need a process involving the public – and I think the Citizens Council is an example of that – where those questions that can be researched can be more defined.

Alan Williams: I want to challenge this proposition that you either have depth or you have numbers. We did interviews with more than 3000 members of the general public – a representative sample of the general public of England, Wales and Scotland. They were interviewed in their own homes for over an hour each. They were taken through a set of problems and asked to make choices between quite difficult things, being given all the time they needed to do it. We came out with a set of trade-offs that people were willing to make between, for example, the relief of pain, the improvement of mobility, and so on – things that they experience, could experience, could think about experiencing.

I do not agree, therefore, that the problem is that if you go into depth you will not get numbers. The problem with this particular issue will be that setting a value for a QALY is not part of people's everyday experience. It will be very difficult to find a format in which you can explain to them – in terms that they understand, in a context that is realistic for NICE – the implications of the number that they are being asked to give. That is the problem. It is not about method. It is about the detachment of this particular problem from people's everyday experience.

Robin Clarke: Going back to the first point about whether the Citizens Council is the most appropriate way of proceeding, the one thing that I would immediately want to pick up on is the idea that 30 people are representative. Thirty people cannot be representative of the wider country. What you get are 30 people who are a 'best fit'. You bring them into a room and you try to control for various elements, but in no way can you say that they are statistically representative.

The Citizens Council, if it is to have any wider validity and not operate in some sort of vacuum, needs to allow media access to it. It is the only way, if you have small numbers of people involved, to build up any legitimacy with the wider public. Otherwise, people start to ask 'Who are these people? Where have they come from? Who says they are deliberating on my behalf?'. I would say that is an essential element if you are to legitimise deliberative methods.

Having said that, for a central government approach to public involvement, it is fantastic. It is the most imaginative thing that central government has come out with. When you think that its bravest attempt at public involvement before this was the People's Panel, then a Citizens Council is a very positive way forward.

I take the point about 3000 people interviewed for an hour. You can get through quite a lot of in-depth material and get people to make difficult decisions. The one thing missing from that is that you are aggregating views. There is not a dialogue happening between different people, so that people are not listening to other people's perspectives and signing up to a common way forward at the end of it – which is what deliberative mechanisms can achieve. That is the

missing part of it. If you want that dialogue to happen between people, if you want them to feel as if they have taken part in a public debate about an issue, then you do need that deliberative method. The only one that also adds the numbers is the deliberative opinion poll method. As I said earlier, however, I do not think that is a realistic way forward for many organisations.

I do not personally think that it would be a problem, in a deliberative context, to get the public to discuss the value of a QALY threshold. The problem is how you make that a wider public discussion.

Anna Coote: There is the option of backing up findings from a deliberative session with polling.

Robin Clarke: Yes, but that points up the Coventry problem. If you have a quantitative poll sitting alongside a deliberative poll, you often get two very different public voices. The problem there is what do you do.

Anna Coote: It might depend on how you relate them to each other, but I will not get involved with that.

John Hutton: I would like to back up strongly what Alan Williams has just said about some of the uninformed comments regarding the basis of QALYs.

The history of it many years ago, as people have said, was less robust. It is possible now, however, to do the sort of in-depth work that gives some credence to the values produced, in the same way that Department of Transport work on willingness to pay has done in the road safety field.

There are all sorts of ways in which you can try to elicit views, which might inform what the average citizen's view of a QALY is, without asking direct questions. I think the small-group, in-depth discussion is an essential part of this, because you have to probe the other factors, apart from clinical effectiveness and things that might lead people to give a particular answer if you offer them a choice of a cancer treatment programme or a heart disease treatment programme.

But if you combine techniques imaginatively, you can get broad-sample representations of what choices based on limited information and experience might be, supplemented by in-depth investigation of key issues that you can identify from that. So we might go back to the old approach of learning from past decisions, which we talked about in the decision-making context, but actually using implicit decisions by people – albeit in perhaps a hypothetical context – which might help us to learn what the general citizenry feel about that.

My final comment is that all this seems to be a rather sad comment on democracy. Other countries have ways of democratically involving people at local level in their health services.

Adam Oliver: If we choose to go the QALY route, you have to be really confident that the methods you are using to elicit QALYs are actually eliciting the things that you want to elicit. We should remember that the core axioms or assumptions on which QALY measurements are based have been found – for many decades – to be systematically flawed, when tested in experimental settings.

Graham Loomes and Angela Robinson have done work trying to look at the reasons people use when they are trying to address these types of decision-making contexts. If you can get these qualitative explanations for the ways in which people are making their preferences, you can perhaps in the future develop that utility or QALY-elicitation technique, to take account of the ways in which people actually do make their decisions, and you can modify the techniques so that they are more scientifically accurate at getting what you are trying to get at.

Peter Littlejohns: I would like to pick up and reinforce John Hutton's point. There is no one way of doing this. We have to take the various approaches that we have. We have talked about patient involvement. We have responded to some of the patient organisations' concerns about how they get involved in the decision-making process. We have established a patients' research and support unit – initially within the guidelines – and it has moved across to the appraisal programme to encourage and sort out how those particular issues are addressed.

The value attached to that information is very important. How do you balance a research-based QALY approach with an individual patient who actually has the disease? The Institute will work on these issues. I think that we have to address both the individual/personal and the public approach in as transparent a way as possible. First of all, there is the move towards making the Citizens Council public. At the moment, a document that is out to consultation says the appeal process – the tail end of the appraisal – will be in the public domain. Once they are debated in public, these issues will be clearer to everybody, including those of us who are not health economists.

Robin Clarke: Oregon has been mentioned. Potentially, there were some very interesting points raised in the way they tried to find out what the public thought. One of the problems was that the public meetings or community hearings tended to be hijacked by familiar faces. That put a negative spin on what we would consider to be public involvement – which is trying to get beyond the regular, usual suspects that you see, to try to find out what those people think who do not usually take part in public involvement.

I feel that we are almost talking about two different things. A lot of people here are talking about research. Research is not about a dialogue. As I understand it, public discussion around QALYs is more dialogue-orientated. With research, we can pull together all the figures that we like, but the public will never feel any ownership of that, unless they have taken part in the discussion and heard other people's points of view. I would strongly push the NICE perspective, therefore, but to make sure that it is high-profile and popular.

Section 5

NEXT STEPS

Chapter 14

What research is required?

NANCY DEVLIN and ADRIAN TOWSE

This chapter looks at four key areas for research:

- generating a common currency;
- valuing QALYs and LYGs;
- factoring in criteria other than maximising health gain;
- local implementation.

Introduction

Cost-effectiveness analysis has tended to dominate the economic evaluation of health care, yet incremental cost-effectiveness ratios (ICERs) only facilitate rankings of health service options. Decisions by bodies such as NICE must go further than this by stating whether any one option is good value for money in absolute – rather than relative – terms. This can only be done by invoking an ICER threshold. The papers and discussions reported here demonstrate considerable uncertainty surrounding:

- the basis (and existence of) thresholds;
- how to discover the ‘right’ threshold(s);
- how (and which) factors other than cost-effectiveness should influence trade-offs against the threshold.

The need for consistency and transparency

Decisions informed by economic evaluation will be consistent, explicit and transparent only if the approach to the threshold is also consistent, explicit and transparent. We suggest four key areas for the research required to achieve this. They are:

- generating a common currency;
- valuing QALYs and life years gained (LYGs);

WHAT RESEARCH IS REQUIRED?

- factoring in criteria other than maximising health gain;
- local implementation.

Generating a common currency

Ranking across treatment areas and setting thresholds that can be used across the whole of the NHS requires a common currency for costs and outcomes. Resource use can be represented in terms of money – providing a consistent approach is taken to calculating incremental resource use and to the pricing of resources – but for outcomes, measurement has tended to rely on quality-adjusted life years (QALYs) or some other common currency.

Yet, are we confident that the extra QALYs associated with a switch from one treatment to another, if measured using different techniques (for example, using different generic health state descriptive systems, and assigning utilities to the states using different approaches) will give similar answers?

Validity of assumptions

More fundamentally, are the assumptions underlying how we measure QALYs in relation to how individuals make choices valid? How should NICE proceed when it is required to assess treatments whose outcomes cannot be captured as QALYs (for example, assisted reproduction technologies) or where it has to choose between treatments where ‘process utility’ (preferences about how, rather than which services are delivered) is the dominant consideration?

The use of willingness to pay (WTP) techniques to value health outcomes and health care processes will inevitably surface in NICE decision making, facilitating cost-benefit analysis, but will add to the complications NICE already faces in comparing disparate types of evidence. How widely WTP approaches could be used in health care as a basis for generating a common currency for comparison is currently unclear. Irrespective of the unit of measurement chosen to generate the common currency, the issue of an appropriate threshold remains, unless all treatments with a net present value can be accepted

i.e. unless budgets are unconstrained.

Valuing QALYs and LYGs

If we accept that cost per life year gained (LYG), or per QALY, will continue to be the most common method of economic evaluation of health care, then some means of valuing these outcomes is required. Such a valuation would implicitly convert an ICER decision rule into a quasi-net benefit criterion.* Options for valuing a QALY or life year include:

- eliciting individuals' willingness to pay for a 'statistical' QALY, using stated preference techniques commonly used in transport economics and also currently used in health economics to assess people's willingness to pay for 'process' aspects of health care (e.g. location, information, or waiting time) rather than those that are purely health outcome related.

Of particular interest is whether contextual factors, such as those discussed by Graham Loomes in Chapter 5, which are known to have an influence on willingness to pay, exert an influence over valuations pertaining to personal health and health care. We already have evidence that willingness to give up a treatment in order to release resources to enable a new treatment to be provided appears to require a higher threshold than is required when new money is available to fund the new treatment.

- using public panels or citizens' juries to consider valuations for QALYs and life years gained, and to consider the acceptability of decisions consequent upon those valuations.
- constructing shadow prices from existing budget constraints, i.e. the value of a QALY gained implicit in observed public policy decisions, controlling for factors other than cost-effectiveness. League tables of cost-effectiveness analyses and interventions can

* This would be 'quasi' cost-benefit analysis because the underlying method would remain grounded in an extra-welfarist ('decision makers'), rather than a welfarist, approach, i.e. society's valuation of a QALY is preceded by an assumption that QALYs are what are to be maximised.

WHAT RESEARCH IS REQUIRED?

be constructed to see whether there is an implicit willingness to pay in terms of what is and what is not used within the NHS.

However, a problem with using past decisions to 'reveal' the threshold to be used in future decisions is that this assumes that past choices are well informed, rational and consistent. It is also somewhat tautological, and is only very indirectly (via imperfect political processes) influenced by public opinion.

Arguably, some means of deriving values that directly involve the public will provide a more defensible basis for policy choices. In addition to the usual challenges involved in willingness-to-pay exercises, briefing lay participants on QALYs presents an additional complexity.

Public involvement

The public have to be involved at two different levels. Firstly, in the technical sense of providing data on the trade-offs they do make, in practice, or would wish to make, in theory, if they were making decisions about resource allocation in the NHS. Secondly, in the political sense of feeling that there is a legitimate process through which these decisions are being made. Stated preference techniques have been put forward for the former, and citizens' juries have been put forward for the latter. Yet they may give different answers. Thus a further potential avenue for enquiry is how to reconcile different results, should they arise.

Factoring in criteria other than maximising health gain

The way in which other factors important to policy choices (such as equity or aspects of 'process', such as waiting time or access to information) are traded off against the maximisation of QALYs requires more clarity. Society may value one QALY more than another depending on the age of the recipient, the disease they have, and whether the intervention is life saving or ameliorating. Society may value aspects of patient care that do not improve health outcomes – or give them a value over and above any improvement in health outcomes that they do deliver.

WHAT RESEARCH IS REQUIRED?

More agreement is required on the nature of any adjustment factors. For example, what are the relevant factors or ‘principles’, what exactly does equity mean in this context (what are the groups of people between which we hold preferences?), and what is the importance attached to the distribution of health gains between these groups? Again, the challenge lies in finding ways to involve the public in such deliberations.

Local implementation

Research is required on the local implementation of decisions and cost-effectiveness ‘in practice’. Decisions made nationally are handed down to local decision makers, who must decide what health services to forego in order to fund those deemed to be good value for money nationally.^{1,2}

While NICE guidance can overcome ‘postcode rationing’ for those services it considers, it cannot mandate what is given up locally in favour of them. National decisions could distort resource allocation if they shift funds from services that are not mandatory to fund to those that are. If what is foregone might have generated greater health improvements, the opportunity costs (and the implicit value of a QALY) may be higher than that which guided the national decision, and vice versa.

Efficacy of local solutions

NICE has argued that local interventions have not been assessed for cost-effectiveness to the same degree as NICE recommendations – if at all. In principle, however, local interventions could be more cost-effective than the NICE interventions that displace them. Of course, part of the problem is that local decision makers regard themselves as tightly constrained in terms of what they can give up. Often it is ‘growth’ money that NICE decisions pre-empt. The reality is that many existing activities will achieve poorer health outcomes per pound than money set aside to fund NICE decisions. Part of the problem is that NICE’s activity is weighted towards reviewing new interventions rather than existing activities that should, perhaps, be discontinued on grounds of lack of cost-effectiveness.

WHAT RESEARCH IS REQUIRED?

This tension between centralism and localism in decision making is likely to become particularly important in the future. NICE decisions are now mandatory. Budgets are increasingly being devolved to primary care trusts, which will have responsibility for 75 per cent of the NHS resource by 2004.³ Understanding variability in implementation and the means by which local priorities and preferences feed into this will be important if we are to avoid what Alan Williams described in Chapter 9 as 'postcode thresholds'.

REFERENCES

- 1 Madhok R, Taylor A, Soltani H & Ireland K. NICE and LICE: Lessons from a health district. *J Pub Health Med*, 2002; 24(1): 2-5.
- 2 Cookson R, McDaid D & Maynard, A. Wrong SIGN, NICE mess: is national guidance distorting the allocation of resources? *BMJ*, 2001; 323: 743-5.
- 3 Secretary of State. *Delivering the NHS Plan. Next steps on investment. Next steps on reform*. London: HMSO; 2002.

Section 6

APPENDICES

Appendix 1

Supporting analysis for Table 1

This appendix sets out the assumptions used in compiling Table 1 (page 28). The categorisation rankings set out in the final column of Table 1 are based on the following cell numbering set out in Table 3, below.

Table 4 sets out our analysis of the technologies appraised, including a column which indicates which category we have put the outcome into for inclusion in Table 1.

In addition, Tables 5 and 6 set out respectively:

- those technologies rejected for lack of evidence of clinical effectiveness;
- those technologies for which NICE published no specific cost-effectiveness details.

The assumption made to arrive at the rankings set out in Table 1, where QALY information is not available, is that life years are equated to QALYs. On this basis, the χ^2 is 9.80. The 5 per cent significance level for 2 degrees of freedom is 9.49 hence cost per QALY ranges are significant at the 5 per cent level. We also performed a sensitivity analysis assuming that a life year gained was only equivalent to 0.5 QALYs. On this basis the χ^2 was 9.21, which was therefore not significant at the 5 per cent level. We conclude that the results are sensitive to the assumptions made about the relationship between QALYs and life years gained.

Table 3

Cost per QALY £000	Accepted	Restricted	Rejected
< £20	1	4	7
20-30	2	5	8
> 30	3	6	9

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
3.	use of taxanes for ovarian cancer	1	no	£6500-£10,000 per LYG (life year gained)	1
5.	liquid-based cytology (LBC)	1	yes – not recommended pilots only	£1100 per LYG with a five-year interval (unlikely to exceed £10,000), £2500 per LYG with a three-year interval (unlikely to exceed £20,000)	7
6.	use of taxanes for breast cancer	1		£7000-£23,500 per LYG	1
9.	rosiglitazone	1	no	£563 per death-free or diabetic complication-free year gained	1
11.	implantable cardioverter defibrillators		yes	£26,000-£31,000 per LYG	5
12.	GP IIb/IIIa inhibitors	2	yes – unstable angina, on the basis of raised troponin levels	unstable angina or non-Q-wave MI: Up to £30,000 per death or AMI prevented at 30 days (£5000 for those with raised troponin)	4

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
			no – acute or elective percutaneous coronary intervention (PCI)	PCI: £7000-£11,500 per LYG	1
13.	methylphenidate (Ritalin, Equasym) for attention deficit/hyperactivity disorder (ADHD)	2	no	central estimate around £10,000-£15,000 per QALY gained, range £5000-£28,000	1
14.	interferon alfa-2b, ribavirin for Hepatitis C	2	no – for first six months' treatment	cost per QALY gained for six months' combination versus monotherapy: £7000 (interferon alpha naïve) or £3050 (relapsers)	1
			yes – after six months, treatment to be limited to genotype 1 patients	cost per QALY for extra six months' combination £5000-£36,000 (no benefit in genotypes 2 or 3)	5

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
15.	zanamivir (Relenza)	2	no – routine use yes – at risk adults	£38,000 per QALY gained for all adults £9300-£31,500 per QALY for at-risk adults	9 1
18.	laparoscopic surgery for inguinal hernia	2	yes – consider only for recurrent and bilateral	£50,000 per QALY gained for laparoscopic versus the preferred open surgery	6
19.	donepezil, galantamine, rivastigmine for Alzheimer's Disease (AD)	2	yes – treatment to be limited to those with mini-mental state examination score > 12	£0-£30,000 per QALY gained	4
20.	riluzole for motor neurone disease	2	no	£34,000-£43,500 per QALY gained	3
21.	pioglitazone for type 2 diabetes	2	no	£563 per death or diabetic complication-free year gained	1

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
22.	orlistat for obesity in adults	2	yes – data sheet (partially revised May 2001) says treatment should be discontinued after 12 weeks in six months the absence of a 5% weight loss. Guidance adds continuation beyond six months to be supported by weight loss of 10% or more	to achieve a cost of £20,000-£30,000 per QALY requires a weight reduction of 5% in each of the three months of treatment or a cumulative 10% reduction over six months	5

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
23.	temozolomide for brain cancer	3	yes – Karnofsky performance status of 70 or more yes – Karnofsky performance status of 70 or more	For glioblastoma multiforme (GBM), incremental cost per LYG relative to procarbazine is £35,000 For anaplastic astrocytoma (AA), the cost per LYG (versus placebo) is £35,000	6 6
25.	gemcitabine for pancreatic cancer	3	yes – Karnofsky performance status of 50 or more, no second line	cost per life year gained between approximately £7200 and £18,700 dependent on the 5-FU regimen used as comparator	4
26.	docetaxel, paclitaxel, gemcitabine and vinorelbine for non-small cell lung cancer (NSCLC)	3	no no	Cost per LYG of £2,250-£16,700 first line for newer combination therapies Just over £14,000 per LYG for docetaxel as a second line therapy compared to best supportive care	1 1

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
27.	cyclooxygenase (Cox) II selective inhibitors (celecoxib, rofecoxib, meloxicam and etodolac) for osteoarthritis and rheumatoid arthritis	2	yes – rejected for average-risk patients	independent non-JK cost per QALY estimates were above £30,000 for average risk OA and RA patients for celecoxib and rofecoxib compared with standard NSAIDs	9
28.	topotecan for the treatment of advanced ovarian cancer	3	no – recommended for high-risk patients	these treatments dominated for high-risk patients	1
29.	fludarabine for B-cell chronic lymphocytic leukaemia (CLL)	3	yes – not recommended in patients with ECOG performance status \leq 3	£32,500 per year of response	6
			no	£200-£2700 per year of remission versus CHOP	1

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
30.	taxanes for the treatment of breast cancer (re-appraisal)	3	yes – not recommended for first line treatment	£19,000 per life year gained for first line treatment	7
31.	sibutramine for the treatment of obesity	2	no – for second line therapy	£7000-£23,500 per LYG for second line taxane monotherapy	1
32.	beta-interferon and glatiramer acetate in MS	1	yes – assumed to be as per orlistat	£15,000-£30,000 per QALY	5
			yes – not recommended	£35,000-£104,000 per QALY	9

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
33.	irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer	4	yes – neither oxaliplatin-5FU/FA nor irinotecan-5FU/FA recommended as routine first line therapy	First line irinotecan-5FU/FA: £37,000 per PFLYG** (£29,000 per LYG) First line oxaliplatin-5FU/FA: £23,800-£67,900 per PFLYG	8 9
			no	second line irinotecan monotherapy: £17,000-£28,000 per LYG	2
			no	combination oxaliplatin in patients with unresectable metastases confined to the liver: £12,900-£39,400 per PFLYG	2

**PFL YG = progression-free life year gained

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
34.	trastuzumab for metastatic breast cancer overexpressing the HER2 protein	4	no – in combination	trastuzumab-paclitaxel combination: £37,500 per QALY trastuzumab monotherapy: £7500 per LYG (£19,000 per QALY)	3 1
35.	etanercept for juvenile idiopathic arthritis	4	no	£15,000-£30,000 per QALY gained	2
36.	etanercept and infliximab for rheumatoid arthritis	4	no	£27,000-35,000 per QALY gained	3
39.	nicotine replacement therapy and bupropion for smoking cessation	4	yes – only as part of abstinent – contingent treatment. Six-month gap between failed attempts at cessation	below £2000 per LYG. However if it is assumed that some smokers would have quit anyway it is more expensive but 'almost certainly cost-effective	1

Table 4 Analysis of NICE appraisals with QALY, LYG or related outcome measure (continued)

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness	Categorisation
40.	infliximab for Crohn's disease	4	yes – patients for whom surgery is inappropriate	£27,500 per QALY gained	5
41.	routine antenatal anti-D prophylaxis for RhD – negative women	4	no	£28,000 per QALY gained	2

Table 5 Lack of evidence of clinical effectiveness over relevant comparator

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness
1.	removal of wisdom teeth	1	yes – should be discontinued	not estimated
8.	digital hearing aid technology	1	yes – insufficient evidence to support the introduction of digital hearing aids	not estimated
16.	autologous cartilage transplantation (ACT)	2	yes – should only be performed as part of a properly structured clinical trial	not estimated
17.	laparoscopic surgery for colorectal cancer	2	yes – should only be undertaken as part of a randomised controlled clinical trial	not estimated
33.	ralitrexed for advanced colorectal cancer	4	yes – its use should be confined to appropriately designed clinical studies	not estimated
37.	rituximab for recurrent or refractory stage III or IV follicular non-Hodgkin's lymphoma	4	yes – not recommended for third and subsequent line therapy yes – use as last line therapy only in a case-series study	scale of clinical effectiveness not established scale of clinical effectiveness not established

Table 6 No cost-effectiveness outcomes stated

Guidance number	Technology appraised	Wave	Restricted?	Cost-effectiveness
2.	prostheses for primary hip replacement	1	yes – the benchmark should be a revision rate of 10% or less at 10 years	not estimated
4.	coronary artery stents	1	no	not estimated
7.	proton pump inhibitors (PPIs)	1	yes – omeprazole licensed for relief of symptoms associated with acid-related dyspepsia. NICE guidance: should not be routinely used in non-ulcer dyspepsia	not estimated
8.	analogue hearing aids	1	no	not estimated
10.	inhalers for under-fives	1	yes – choose on the basis of cost minimisation	not estimated
24.	debriding agents and specialist wound care clinics for difficult-to-heal surgical wounds	3	no	not estimated
38.	inhaler devices for routine treatment of chronic asthma in older children (aged 5-15 years)	4	no – but use lowest cost device (including dose) when more than one meets requirements	no robust cost-effectiveness or utility studies available. However, small improvements in asthma outcomes will be cost-effective

Appendix 2

What do the delegates think NICE's threshold should be?

Participants at the workshop were requested by Peter Littlejohns to anonymously record their personal view of the threshold NICE should apply to cost per QALY evidence in its decision making (see Table 7, overleaf). The majority of participants were health economists and, where appropriate were asked to record the number of years worked in health economics. Twenty-six responses were received.

Eighteen participants submitted threshold values. Five of these were recorded as a range rather than a single figure. Taking the mid-point of these five together with the other 13 point estimates, the average threshold was £29,000 per QALY. There was no relationship between the threshold chosen and the number of years worked as a health economist.

Eight of the 26 participants declined to specify a threshold value, offering a variety of explanations as set out below:

- 'whatever society will pay';
- 'cannot assess as no good info';
- 'do not use QALYs';
- 'no limit';
- 'none';
- 'ministers decide';
- 'formula linked to per capita GDP';
- 'do not know'.

Table 7 Cost-effectiveness thresholds selected by economists

Value of a QALY (£000)	Frequency*
5	1
15	1
17.5	1
20	4
25	2
30	1
31	1
32.5	1
36	1
40	2
50	3

* There were 18 responders

Appendix 3

List of attendees

Ron Akehurst, University of Sheffield

John Appleby, King's Fund

David Braunholtz, University of Birmingham

Stirling Bryan, University of Birmingham

Martin Buxton, Brunel University

Jeremy Chancellor, Innovus

Robin Clarke, Institute for Public Policy Research (speaker)

Edward Colgan, Dorset Health Authority

David Cook, University of Oxford (speaker)

Richard Cookson, University of East Anglia

Anna Coote, King's Fund (session chair)

Tony Culyer, University of York (session chair)

Elizabeth Davies, NICE

Nancy Devlin, City University (speaker)

Jack Dowie, London School of Hygiene and Tropical Medicine

Alastair Fischer, National Institute for Clinical Excellence

Bethan George, University of London

Chris Heginbotham, South Warwickshire General Hospitals NHS Trust

John Henderson, Department of Health

John Hutton, Medtap International

Peter Littlejohns, National Institute for Clinical Excellence (speaker and session chair)

Graham Loomes, University of East Anglia (speaker)

APPENDIX 3

Jo Lord, Imperial College

Ali McGuire, London School of Economics

Adam Oliver, London School of Economics

Christine Pond, Patients Association

Clive Pritchard, Office of Health Economics

Angela Robinson, University of East Anglia

James Raftery, University of Birmingham

Peter Sharplin, Aventis

Jon Sussex, Office of Health Economics

Keith Tolley, GlaxoSmithKline

Adrian Towse, Office of Health Economics (speaker)

Paul Trueman, Johnson & Johnson

Dominic Wake, AstraZeneca

Norma Weatherhogg, Patients Association

Nick Wells, Pfizer

Alan Williams, University of York (session chair)

Nick York, Department of Health

Recent OHE publications

Public Private Partnerships for Medicines and Vaccines Research and Development
by Hannah Kettler and Adrian Towse, 2002 (price £10.00)

Disability Adjusted Life Years (DALYs) for Decision-making? – An Overview of the Literature

by Julia Fox-Rushby, 2002 (price £10.00)

Interpreting and Addressing Inequalities in Health: from Black to Acheson to Blair to...?

by Robert Evans, 2002 (price £7.50)

The Life Cycle of Pharmaceuticals: a Cross-national Perspective

by Patricia Danzon and Jeong Kim, 2002 (price £10.00)

Health Economics: An Introduction to Economic Evaluation (2nd edition)

by Gisela Kobelt, 2002 (price £5.00)

The Links of Public Health and Economic Development

by Professor Jeffrey Sachs, 2001 (price £5.00)

Applied Econometrics for Health Economists – a Practical Guide

by Andrew Jones, 2001 (price £10.00)

Consolidation and Competition in the Pharmaceutical Industry

ed. Hannah Kettler, 2001 (price £10.00)

Don't Look Back? Voluntary and Charitable Finance of Hospitals in Britain, Past and Present

by John Mohan and Martin Gorsky, 2001 (price £10.00)

Capturing the Unexpected Benefits of Medical Research

ed. Clive Pritchard, 2001 (price £10.00)

The Economics of the Private Finance Initiative in the NHS

by Jon Sussex, 2001 (price £10.00)

Why Care about Health Inequality?

by Adam Oliver, 2001 (price £7.50)

Health Care Without Frontiers? The Development of a European Market in Health Services?

by Lyndsay Mountford, 2000 (price £10.00)

Productivity Costs: Principles and Practice in Economic Evaluation

by Clive Pritchard and Mark Sculpher, 2000 (price £10.00)

Improving Population Health in Industrialised Nations

ed. Jon Sussex, 2000 (price £10.00)

Recent King's Fund publications

Changing Relationships: Findings of the Patient Involvement Project, Rosemary Gillespie, Dominique Florin and Steve Gillam, 2002, £6, 32pp, ISBN 1 85717 468 2.

Claiming the Health Dividend: Unlocking the benefits of NHS spending, edited by Anna Coote, 2002, £10, 134pp, ISBN 1 85717 464 X.

Developing Intermediate Care: A guide for health and social services professionals, Jan Stevenson and Linda Spencer, 2002, £18, 134pp, ISBN 1 85717 466 6.

Every Voice Counts: Primary care organisations and public involvement, Will Anderson, Dominique Florin, Stephan Gillam and Lesley Mountford, 2002, £7.99, 80pp, ISBN 1 85717 460 7.

Five-Year Health Check, A review of Government health policy 1997–2002, edited by John Appleby and Anna Coote, 2002, £7.99, 138pp, ISBN 1 85717 463 1.

The Future of the NHS: A framework for debate, 2002, free, 22pp.

A Good Place to Learn? What young people think makes schools healthy, Kate Healey, 2002. free, 8pp.

Hidden Assets: Values and decision-making in the NHS, edited by Bill New and Julia Neuberger, 2002, £17, 228pp, ISBN 1 85717 458 5.

How Good is Your Service to Carers? A guide to checking quality standards for local carer support services, Roger Blunden, 2002, £5, 54pp, ISBN 1 85717 465 8.

Managing the Pressure: Emergency hospital admissions in London 1997–2001, Michael Damiani and Jennifer Dixon, 2002, £5.99, 28pp, ISBN 1 85717 461 5.

Old Habits Die Hard: Tackling age discrimination in health and social care, Emilie Roberts, Janice Robinson and Linda Seymour, 2002, £6.99, 42pp, ISBN 1 85717 462 3.

Out of the Maze: Reaching and supporting Londoners with severe mental health problems, Angela Greatley with Richard Ford, 2002. £10, 94pp, ISBN 1 85717 469 0. King's Fund and the Sainsbury Centre for Mental Health.

Psycho-Social Support for Dying People: What can primary care trusts do? Cathy Shipman, Steve Gillam and Ros Levenson, 2002, free, 56pp.

Public Health in the Balance: Getting it right for London, David Woodhead, Karen Jochelson and Ruth Tennant, 2002, free, 8pp.

Public Interest, Private Decisions: Health-related research in the UK, Anthony Harrison and Bill New, 2002, £17, 262pp, ISBN 1 85717 467 4.