



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/118329/>

Version: Published Version

Article:

Culyer, A. J. (Anthony J.) (2006) NICE's use of cost effectiveness as an exemplar of a deliberative process. *Health Economics, Policy and Law* , 1. pp. 299-318. ISSN: 1744-1331

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Perspective

NICE's use of cost effectiveness as an exemplar of a deliberative process

ANTHONY J. CULYER*

Institute for Work & Health, Toronto, Canada and University of York, UK

Abstract: This paper seeks to test 12 conjectures about the predicted use of deliberative processes by applying them to the technology assessment procedures used by the National Institute for Health and Clinical Excellence (NICE) in England and Wales. A deliberative process is one that elicits and combines evidence of different kinds and from different sources in order to develop guidance – in the present case, guidance for a health care system. A deliberative process entails the integration of three kinds of evidence: scientific context-free evidence about the general clinical potential of a technology, scientific context-sensitive evidence about particular evidence in realistic scenarios, and colloquial evidence to fit context-free scientific evidence into a context and to supply the best evidence short of scientific evidence to fill in any relevant gaps. It is shown that NICE's appraisals procedures and, in particular, its approach to cost effectiveness, entail both the weighing of each of these types of evidence and can be seen as rational responses to the 12 conjectures.

1. Introduction

The National Institute for Health and Clinical Excellence (NICE) represents one interesting possible solution to the problem of getting research-based evidence into practice – in this case research into the cost effectiveness of health care technologies (in quite a broad sense of ‘technology’). Not all evidence is provided by research and not all the relevant evidence is medical. Few organizations have tackled the complex challenge that this represents quite so boldly and imaginatively as it seems NICE has done. While it is premature to attempt any full appraisal of its success, and the present author has been too close to NICE's development to be truly objective,¹ this paper advances some broad conjectures about the circumstances under which ‘deliberative processes’, of the sort that characterize NICE's arrangements, may be chosen and hence begin to provide an explanatory account. These conjectures are presented in the

* E-mail: aculyer@iwh.on.ca.

¹ I was Vice Chair of NICE from its inception in April 1999 until October 2003. I have, in general, sought to locate much second-order descriptive detail in footnotes rather than blur the focus of text on many of the issues that had to be confronted.

next section. This is followed in Section 3 by a brief description of some of the principal characteristics of NICE's decision-making procedures regarding the assessment of health care technologies and their recommendation or otherwise to the National Health Service (NHS). These are related to each of the 12 conjectures. Section 4 addresses some characteristics of NICE's procedures that directly address the ambiguity of public policy statements. Many of these procedures involve value judgments, and some appear to flout the conventional canons for the conduct of cost-effectiveness analysis. This section also seeks to show how a deliberative process can help both to resolve ambiguity and to generate guidance that commands widespread assent. Section 5 elaborates and comments on NICE's solutions and Section 6 concludes.

2. Twelve conjectures about deliberative processes

I have chosen 'conjecture' rather than 'hypothesis' to indicate the essentially preliminary and tentative nature of the speculation that is the major concern of this section. The specific conjectures to be outlined can probably be underpinned by more fundamental features of which they are the implications. That task is for another occasion. The present purpose is to identify circumstances that are at least more generic than those in which NICE finds itself, so that NICE may be used as just one case study to test the *prima facie* plausibility of the conjectures.

The processes used by NICE in its decision making about health care technologies may be described as 'deliberative', and a systematic review and some initial theorizing about such processes in health care decisions (Lomas *et al.*, 2005; Culyer and Lomas, 2006) provide the basis for the conjectures to follow. A deliberative process is one that elicits and combines evidence of different kinds and from different sources. In particular, it combines context-free scientific evidence from bio-science and clinical trials (often in the form of narrative reviews, systematic reviews, and meta-analyses) with other forms of more context-sensitive scientific evidence (often in the form of applied decision theoretical procedures such as cost-effectiveness analysis or social surveys), and less systematic evidence referred to elsewhere as 'colloquial' (Lomas *et al.*, 2005) that may come in the shape of the 'experience' of those round the decisions-makers' table. Clinical scientific evidence is typically context-free in that it has been generated through processes that control for confounding variables in order to create suitable conditions for testing scientific hypotheses. This usually entails highly specific clinical contexts such as research centres of excellence and clinical work done by professionals of national and international repute. Context-sensitive scientific evidence is defined as evidence also collected to test hypotheses in systematic ways but ways that are more relevant to the context in which the technology (in the present case) is to be used. Evidence from a placebo-controlled trial is an example of context-free scientific evidence; evidence from a cost-effectiveness analysis of alternative drugs for treating

a particular disease in general practice in England and Wales is an example of context-sensitive scientific evidence. Colloquial evidence is neither scientific nor systematic but it is frequently all that it is possible to bring to bear on a particular issue and it frequently takes the form of expressions of opinion by experts (clinical, managerial, economic) based on their practical experience and professional judgment. The status of colloquial evidence is controversial. On the one hand, it is by definition neither scientific nor systematic (and it may be plain dangerous);² on the other, it is frequently all that is available on some critical aspect of a decision.

A deliberative process elicits evidence by virtue of the embodiment of the participants in deliberation and it combines evidence of both the two scientific kinds and the colloquial kind in a process of 'weighing up' and considering the contexts in which the guidance emanating from the deliberative process is to be used. Deliberative processes are said to increase the likelihood of what Daniels calls 'sound and acceptable decisions'. Daniels argues for processes that 'account for reasonableness' so that they have a moral authority over and above that which customarily attaches to market or bureaucratic processes (Daniels, 2000a, 200b), though the morality of the process is not an aspect to concern us here. Deliberative processes in health care stress the integration of technical analyses of clinical issues with (usually social) scientific analyses of the contexts in which decisions will be implemented, within an explicit decision-making model having clear criteria, and involving stakeholder and lay public consultation and even participation, in contrast to the more traditional top-down approach. Deliberative processes are not the same as consultative processes, though consultation will normally have a significant role to play in support of deliberation.

The following are conjectured to be circumstances that make the use of deliberative processes more likely:

- decisions have been delegated by a body with a democratic mandate to one without it
- evidence from more than one expert discipline is involved
- evidence from more than one profession is involved
- stakeholders have conflicting interests
- there are technical disputes to resolve and the evidence may be scientifically controversial
- evidence gathered in one context is to be applied in another
- there are issues of outcome, benefits, and costs that go beyond the conventional boundaries of medicine

2 A famous example is Dr Spock's advice to countless thousands of mothers: 'There are two disadvantages to a baby's sleeping on his back. If he vomits, he's more likely to choke on the vomitus. Also he tends to keep his head turned towards the same side, this may flatten the side of his head... I think it is preferable to accustom a baby to sleeping on his stomach from the start' (cited in Chalmers, 2003: 23). As Iain Chalmers has recently commented, reflecting on his early days as a medical practitioner, 'No doubt like millions of Spock's other readers, I passed on this apparently rational, theory-based and authoritative advice. We now know from the dramatic effects of the 'Back to Sleep' campaigns in several countries that the practice promulgated by well-intentioned experts like Spock led to tens of thousands of avoidable sudden infant deaths' (Chalmers, 2005: 229).

- there is substantial uncertainty about key values and risks that needs to be assessed and weighed
- there are other social and personal values not taken into account in the scientific evidence
- there are issues of equity and fairness
- there are issues of implementability and operational feasibility involving knowledge beyond that of the decision makers
- wide public and professional 'ownership' is desired

Scientific evidence on clinical efficacy is usually context-free in the sense that it is presented as the outcome of a quasi-laboratory style experiment. It is often summarized in the form of a narrative review, a systematic review, or a meta-analysis. Scientific evidence on context, including clinical effectiveness (cf. efficacy) under routine practice conditions and evidence on other contingent matters such as cost effectiveness and the social values attached to health outcomes, might also be gathered by controlled experiments, which in turn may also be summarized and synthesized. Colloquial evidence is often gathered through consultative processes, including social surveys, public meetings, and the hearing of witnesses, as well as directly from those participating in the deliberative process. The ultimate product of a deliberative process is guidance shaped by judgment – judgment about an effect of doing something, its size, the ways in which it is likely to be achieved, for whom, for how long, its cost in terms of the resources used that would otherwise have been employed in other ways to achieve other good things, and how worthwhile it is. The test, therefore, of a deliberative process is whether the resultant judgment is (or will be) more comprehensively 'evidence informed', better matched to the context of application, more efficiently implementable, and more widely acceptable to those affected by it.

3. The National Institute for Health and Clinical Excellence

NICE was given two main charges at its inception, though there were also others and further responsibilities have also been accumulated since then. The original two were:

- to identify cost-effective technologies and make recommendations for their use in the National Health Service (NHS) of England and Wales
- to create authoritative clinical guidelines to support cost-effective clinical practice in all health care settings³

³ More completely, the fundamental objective is to improve standards of patient care, and to reduce inequities in access to innovative treatment, by establishing a process which will: (1) identify those new treatments and products which are likely to have a significant impact on the NHS, or which for other reasons would benefit from the issue of national guidance at an early stage, (2) enable evidence of clinical and cost effectiveness to be brought together to inform a judgment on the value of the treatment relative to alternative uses of resources in the NHS, (3) result in the issue of guidance on whether the treatment can be recommended for routine use in the NHS (and if so under what conditions or for which groups of patients) together with a summary of the evidence on which the recommendation is based, (4) avoid any significant delays to those sponsoring the

The idea of 'technology' is broad: nearly as broad as 'ways of doing things'. It certainly and explicitly includes: pharmaceuticals, medical devices, diagnostic techniques, surgical procedures, other therapeutic technologies, health promotion, and workplace interventions for health and safety.

NICE's guidance is not binding on the NHS, though it carries much weight. It is the principal informational content underpinning 'Clinical Governance', a system of local management for accountability and audit of good clinical practice.⁴ The only statutory element in NICE's guidance is that any clinician wishing to follow it must be enabled to do so through appropriate resource provision by the local NHS – the commissioners (purchasers) and the trusts (hospitals and primary care providers) in question.

From the beginning, it was decided that NICE's procedures would be conducted with the highest degree of transparency possible and with much participation by 'stakeholders' (categorically defined as patients, informal caregivers, clinical and other professional caregivers, health care managers, manufacturers, researchers, and the public in general). NICE insisted on being located within the NHS rather than the Department of Health. It sought the respect of the overwhelming majority of the country's clinical and health service research community and the support of the Royal Colleges and other bastions of professional life. It was important to NICE that its guidance could not be dismissed as cranky, under-researched, or second rate. But it also had to be acceptable to the NHS's users and fair to the inventors and manufacturers of the various technologies that were used in a huge range of patient management pathways. It also had to be deemed 'do-able' by the managers. There had to be lots of opportunities for skeptics and any who might feel threatened to air their concerns and for NICE to respond appropriately.

Some of the ways in which NICE seems to be a model of deliberative process are:

- there are open Board meetings that take place bi-monthly around England and Wales (these are accompanied by public receptions and 'Question and Answer' sessions)
- minutes are published on the NICE web pages before confirmation
- there is a Partners Council⁵
- there is a Citizens Council⁶

innovation either in meeting any national or international regulatory requirements or in bringing the innovation to market in the UK.

⁴ Clinical governance is 'a framework through which NHS organisations are accountable for continually improving the quality of their services and safeguarding high standards of care by creating an environment in which excellence in clinical care will flourish' (Scally and Donaldson, 1998).

⁵ The Partners' Council has a statutory duty to meet once a year to review the NICE annual report. In practice, it meets more frequently as a source of advice and a forum for exchanging ideas and developing the future plans for the Institute. Its membership includes representatives from organizations with a special interest in its work such as patient groups, health professionals, NHS management, quality organizations, industry, and trade unions. Members are appointed by the Secretary of State for Health and the Welsh Assembly Government.

⁶ The Citizens' Council is a form of 'citizens' jury' that considers socially value-laden matters that are referred to it by the Institute's Board. Its members have no economic involvement in the health care system

- the membership of the Technology Appraisals Committee is set broadly⁷
- there are extensive consultation exercises throughout the appraisals process
- there is an appeals procedure⁸
- there are consultative processes about process⁹
- there are extensive liaisons with Royal Colleges,¹⁰ Independent Academic Centres,¹¹ and National Collaborating Centres¹²
- there is considerable joint working with NHS R&D¹³ and the National Coordinating Centre for Health Technology Assessment¹⁴

Thus, it came about that the process of technology appraisal was to be open, multi-disciplinary, multi-professional and multi-institutional, and that it would have what is sometimes somewhat unfortunately referred to as lay participation. It was heavily dependent upon people's willingness to serve *pro bono*

and are paid £150 per day plus expenses. It meets twice a year and adopts a deliberative approach (e.g. calls witnesses, commissions papers). It is managed at arm's length from NICE by Vision 21, a company specializing in research and community consultation (see Kelson, 2001, 2002 and Jarrett and PIU, 2004).

7 The Committee is a standing advisory committee of the Institute, having two branches, each of which duplicates the other by sharing the workload. Members are appointed for a three-year term. They are drawn from the NHS, patient and care-giving organizations, relevant academic disciplines and the pharmaceutical and medical devices industries. Names of Appraisal Committee members are posted on the Institute's website.

8 There are three grounds for appeal within the Institute's Appeals Process: that the Institute has failed to act fairly and in accordance with the Appraisal Procedure set out in its Guidance to Manufacturers and Sponsors; that it has prepared Guidance, which is perverse in the light of the evidence submitted; and that it has exceeded its legal powers.

9 An example is the process through which guidance was developed to inform stakeholders and analysts about the procedures for health technology assessment. Several committees were convened, with representation of experts from a variety of stakeholders. The outcome was NICE (2004). The members of the Economics Task Group were Mark Sculpher (Chair), Professor of Health Economics, University of York; John Brazier, Professor of Health Economics, University of Sheffield; Andrew Briggs, Public Health Career Scientist, University of Oxford; Martin Buxton, Academic Director of Health Economics, Brunel University; Ruth Carlyle, Information Materials Programme Manager, Macmillan Cancer Relief; Karl Claxton, Senior Lecturer in Economics, University of York; Francoise Cluzeau, Guidelines Technical Advisor, NICE; Michael Donaghy, Clinical Neurologist, Radcliffe Infirmary, Oxford; Dogan Fidan, Health Technology Analyst, NICE; Alastair Fischer, Health Technology Analyst, NICE; and Nick Wells, Senior Director, Outcomes Research, Pfizer.

10 The royal colleges are the principal professional associations of the English medical professions. They comprise: The Royal College of Anaesthetists, The Royal College of General Practitioners, The Royal College of Obstetricians and Gynaecologists, The Royal College of Paediatrics and Child Health, The Royal College of Pathologists, The Royal College of Psychiatrists, The Royal College of Radiologists, The Royal College of Surgeons of England, The Faculty of Public Health Medicine, The Faculty of Pharmaceutical Medicine, and The Faculty of Occupational Medicine.

11 NICE commissions seven independent academic Centres to review published evidence on the relevant technology when developing technology appraisals guidance. The Centres are: Health Economics Research Unit and Health Services Research Unit, University of Aberdeen; Liverpool Reviews & Implementation Group, University of Liverpool; Centre for Reviews and Dissemination, University of York; Peninsula Technology Assessment Group, Universities of Exeter and Plymouth; School of Health and Related Research, University of Sheffield; Southampton Health Technology Assessment Centre, University of Southampton; West Midlands HTA Collaboration, University of Birmingham.

12 NICE has created seven National Collaborating Centres within consortia of the royal colleges, professional bodies, and patient/carer organizations for developing clinical guidelines. They are: the National Collaborating Centres for Acute Care, Cancer, Chronic Conditions, Mental Health, Nursing and Supportive Care, Primary Care, and Women and Children's Health.

13 The main programme within NHS R&D of relevance here is Health Technology Assessment. This is a national programme of research the purpose of which is to generate high-quality research information on the costs, effectiveness, and broader impacts of health technologies.

14 The National Coordinating Centre for HTA (NCCHTA) is part of the Wessex Institute for Health Research and Development at the University of Southampton. It coordinates the HTA programme on behalf of the NHS R&D.

(that is for their travel and subsistence expenses alone). It was also plain from the outset that very large numbers of people would be involved and the Institute itself would be largely a virtual organization.

Each of these features corresponds to one or more of the conjectures of Section 2.

Decisions have been delegated by a body with a democratic mandate to one without it

This precisely describes NICE's relationship with Parliament via the Secretary of State (the 'minister'), select committees, and the floor of the House of Commons. In fact NICE has had delegated to it some of the most central value-laden issues involved in health care policy, not least of which are the definition of 'health' and 'health gain', the practical measurement of these entities, and the conceptualization and measurement of matters of equitable concern (largely undefined by the political leadership).

Evidence from more than one expert discipline and profession is involved

Decisions typically involve multiple disciplines and NICE's appraisals committees include statisticians, general practitioners, patient advocates, public health and consultant physicians, health economists, clinical pharmacists or pharmacologists, nurses, a consultant surgeon, NHS managers, a representative of the Association of British Healthcare Industries, a psychiatrist, members of Professions Allied to Medicine, and a paediatrician. In addition, subject specialists are usually present to provide expert testimony about the technologies being discussed.

Stakeholders have conflicting interests

The consultation process that generates the evidence before an Appraisal Committee affords opportunities for stakeholders to represent their interests, make submissions, engage in cost-effectiveness modelling exercises, and to understand the evidence being considered. NICE distinguishes between consultees and commentators. Consultees include manufacturers and other sponsors of the technology in question, national professional organizations, national patient organizations, the Department of Health and the Welsh Assembly Government, and relevant NHS organizations in England and local health boards in Wales. Consultees participate in the consultation on the draft scope, and they may make formal submissions and receive various interim documents (such as the Assessment Report, and the Appraisal Consultation Document). Consultee organizations representing patients, caregivers, and professionals can nominate clinical specialists and patient experts to present their personal views in person to the Appraisal Committee.

Commentators have more limited scope for participation. They are organizations that engage in the appraisal process but that are not asked to prepare a submission dossier and they receive the final report for information only. They

include manufacturers of comparator technologies, NHS Quality Improvement Scotland, relevant National Collaborating Centres for clinical guidelines, and other research groups where appropriate such as the Medical Research Council, as well as other official groups such as the NHS Confederation.

There are technical disputes to resolve and the evidence may be scientifically controversial

It is in the character of clinical trials and other evidential bases for assessing clinical effectiveness that the evidence will be disparate and that the conclusions of different studies will differ. Hierarchies of evidence ranked by quality are standard in the field of systematic reviewing because not all studies generate evidence of equal quality and there can be substantial dispute between experts over technical aspects of the interpretation of evidence. There is ample scope for technical disagreements between experts in health technology appraisal regarding, for example, the adequacy of control for confounding variables in the available trial literature, the interpretation of meta-analyses, or the comprehensiveness of the sensitivity analyses in guiding an appreciation of the consequences of particular analytical assumptions and empirical estimates, or the best way to model outcomes beyond the period for which data have been collected from clinical trials.

Evidence gathered in one context is to be applied in another

All bodies tasked, like NICE, with making formulary recommendations need to apply judgment in determining whether the context in which the evidence has been gathered (usually one or more centres of clinical and research excellence) affects those results such that their outcomes or costs might differ in other, less conducive environments. Similarly, results of trials conducted in other countries or health care systems may not be readily transferable. Making such judgments typically requires knowledge both of the scientific literature and of the 'reality' of clinical practice in the jurisdiction in question, which is often available only in the form of colloquial evidence – the judgment of practising physicians and managers engaged in the deliberative process.

Issues of outcome, benefits, and costs go beyond the conventional boundaries of medicine

NICE stipulates the scope of costs and benefits, and how they are to be compared. It recommends use of particular outcome measures and cost categories. The research, however, has rarely conformed fully or even substantially with these stipulations. While the significance of this problem may diminish over time as compliance increases, there will always be a residual scope for 'gap filling', some of which might be accomplished through technical means such as modelling but other elements may have to be dealt with through drawing colloquially on the practical experience and, no less importantly, values of those participating in the decision.

There is substantial uncertainty about key values and risks that needs to be assessed and weighed

Not all the scientific information available will have been subject to a systematic sensitivity testing procedure for the robustness of the results in response to alternative assumptions, different settings of care, or any one of a host of relevant matters about which there may be uncertainty. Uncertainty may sometimes exist even about the nature of the factors to be considered 'relevant'. The assessment of risk and the weight attached to it will almost always be matters for judgment and the circumstances of the time, and some aspects may be capable of being addressed only at the time of decision.

There are other social and personal values not taken into account in the scientific evidence

The existence of such other values might become evident either during the assessment of the evidence by expert advisory groups (like systematic reviewers), or through the process of consultation, or emerge during discussions at the time of decision. Both their existence and the weight to be placed on them are inherently matters for deliberation and probably only colloquial evidence will be available.

There are issues of equity and fairness

Issues of equity, such as the importance to be attached to costs or benefits that accrue disproportionately to some people than others (for example, geographical, socio-economic, or clinical subgroups) are rarely addressed in the background scientific research and, even when they are, determining their importance as distinct from their existence is quintessentially a matter for decision makers rather than scientists. There is frequently little scientific evidence available about what people (the general public, for example, or a target population) actually think about such issues or on the ways in which costs and benefits have an impact on different social groups.

There are issues of implementability and operational feasibility involving knowledge beyond that of the decision makers

Practical matters of implementability, managerial and financial consequences will rarely have been fully addressed, if at all, by the scientific research available to a decision-making group. To address and assess such matters in the context of a particular jurisdiction will normally require each possible recommendation to be assessed colloquially by those with appropriate practical knowledge.

Wide public and professional 'ownership' is desired

NICE has from the beginning been concerned that its recommendations have the confidence of not only the scientists and medical professional groups that advise it but also of users of the health care system, their caregivers at home

and elsewhere, and the public in general. It is difficult to identify any obvious excluded interest or stakeholder group.

4. Being explicit about the ethics of cost and effectiveness

There is no escaping the fundamental fact that the decisions and recommendations of a body such as NICE are laden with judgments of value. In the UK, successive ministers have decided that the prime objectives of the NHS are to maximize the health of the population and to do so in an equitable fashion. The former has been more specifically articulated than the latter. For example, successive ministers (both Labour and Conservative) in the UK have clearly set NHS managers the task of maximizing health (in official documents often referred to as maximizing 'health gain'). Thus:

The purpose of the NHS is to secure through the resources available the greatest possible improvement in the physical and mental health of the people of England [and]... aims to judge its results under three headings: equity, efficiency, and responsiveness. (Department of Health, 1996)¹⁵

The objective of the NHS Research and Development strategy is to ensure that the content and delivery of care in the NHS is based on high quality research relevant to improving the health of the nation. (Department of Health, 1991)

The knowledge produced by trials, overviews or technology assessment can be regarded as 'bullets' of effectiveness. Just like a bullet, they are of little value by themselves, and need to be loaded, aimed, and fired to hit their target. (Bandolier, 1994)

These three quotations say more than may at first appear. The first says that health maximization and equitable distribution are two prime aims of policy (not that they are the *only* aims).¹⁶ The second identifies evidence-based practice as important and commits the NHS to the generation of the appropriate evidence. The third reminds the NHS that mere information is never enough – not, at least, if the aim is to generate change in pursuit of the objectives outlined in the first quotation. Obvious implications of these broad instructions include the proposition that health care which does not contribute to health has no place in the NHS (even if people prefer it). That is quite a powerful implication. It also implies that cost-*ineffective* health care has no place in such a system, for

15 Under Labour administrations, it is also clear that arbitrary and avoidable unequal access to care is to be avoided. This has become known as 'post-code prescribing', though it is not in practice bounded by geography nor restricted to pharmaceuticals, so there is a presumption towards equality of *some* sort – the rhetoric loses precision at this point.

16 The clarity with which objectives are articulated varies greatly, of course. NICE's formal charges are to use the following criteria in its appraisals:

- Is the technology likely to result in a significant health benefit, taken across the NHS as a whole, if given to all patients for whom it is indicated?

providing cost-ineffective care would imply that resources which could have been put to the ministerial objective of maximum health were not being so put.

But the lack of specificity in the ministerial commands poses a problem for an organization such as NICE. How ought it to populate the general ideas with ones that are specific, relevant, and operational? What, precisely, is 'health' or 'health gain'? What is 'equitable'? What is the 'population'? What is it to 'maximize' something? How, more specifically still, are these ideas, each fraught with ethical overtones, to be implemented in a deliberative process whose outcome is to be a recommendation about the adoption of health care technologies? Here we have two ambitions, manifestly ethical in character, which we may summarily call 'efficiency' and 'equity'. They are (if we accept the legitimacy of the accountable minister) critically important starting points. Despite their ambiguity, they are also axiomatic, givens, primitive statements. And, given the authority (the accountable minister), they have a democratic basis which no mere academic, researcher, teacher, or even individual citizen cannot possibly lay claim to. These broad and imperfectly defined aims form the basis for the use of cost-effectiveness methods in NICE's decision-making processes and NICE has, perforce, had to inject potentially controversial content into them. All the ambiguities referred to have had to be demystified and made explicit in its approach to cost effectiveness.

5. NICE's resolution of ambiguities in its technology appraisals

NICE's guidance on the economic components of technology appraisal is set out in NICE (2004). Aspects of this guide that are worthy of note are conveniently addressed under the following headings:

- The context
- The Reference Case
- The scoping phase
- Perspective
- Outcome measurement (What costs and what benefits? How measured?)
- Distribution of costs and benefits (QALY bias and QALY weights)

The context

'Context' includes process characteristics that are not conventionally examined in economics. A distinction seems to be emerging in NICE's procedures between value judgments that may be conveniently decided upon in an algorithmic way and those that are better decided as the outcome of a more deliberative

- Is the technology likely to result in a significant impact on other health-related Government policies (for example, reduction in health inequalities)?
- Is the technology likely to have a significant impact on NHS resources (financial or other) if given to all patients for whom it is indicated?
- Is the Institute likely to be able to add value by issuing national guidance? For example, in the absence of such guidance is there likely to be significant controversy over the interpretation or significance of the available evidence on clinical and cost effectiveness?

process – one that is, in effect, a deliberative process conducted in the context of a specific policy decision rather than as an abstract exercise. Such a distinction exists between the treatment of the values embodied in the QALY (essentially algorithmic and only discussed in a NICE policy-making context if there is a feeling that the conventions are not representing the relevant values well enough) and the values embodied in interpersonal comparisons of QALYs accruing to specific groups of people who might be deemed ‘special cases’, as when, say, patient group representatives claim on the behalf of the people they represent that their claims deserve a higher priority than those of others.

The Reference Case

NICE has borrowed from the Washington Panel (Gold *et al.*, 1996) the idea of a ‘Reference Case’. This is explained by NICE thus:

The Institute has to make decisions across different technologies and disease areas. It is therefore important that analyses of clinical and cost effectiveness undertaken to inform the appraisal adopt a consistent approach. To facilitate this, the Institute has defined a ‘reference case’ that specifies the methods considered by the Institute to be the most appropriate for the Appraisal Committee’s purpose and consistent with an NHS objective of maximizing health gain from limited resources... This does not preclude additional analyses being presented where one or more aspects of methods differ from the reference case. However, these must be justified and clearly distinguished from the reference case. (NICE, 2004: 19)

This quotation neatly encapsulates three elements that have implications for the context of NICE’s work:

- it is constrained by the maximand (‘health gain’)
- it is constrained by a need for consistency in analytic methods (or as much as is necessary for its Appraisal Committee to be able to function effectively)
- it is pragmatic; not only in the two respects just mentioned but also in that exceptions to the Reference Case are permitted

The relevant details of the Reference Case are set out in Table 1. The scoping phase defines a framework for the subsequent appraisal and inescapably entails the making of explicit or implicit value judgments. Scoping precedes every technology appraisal and is in essence a deliberative process within a deliberative process. The framework determines the specific questions to be addressed so as to define the issues of interest (for example, target population, technology comparators) as clearly as possible and the questions that should be addressed by the Appraisals Committee when considering the clinical and cost effectiveness of the technology. Consultees and commentators are consulted during the scoping process. The Institute revises the scope in response to comments received and develops a final scoping document that describes the boundaries of the appraisal and the parameters that will be investigated. The scope is further developed into a protocol for the technology assessment.

Table 1. The NICE reference case

Element of health technology assessment	Reference case
Defining the decision problem	The scope developed by the Institute
Comparator	Alternative therapies routinely used in the NHS
Perspective on costs	National Health Service and Personal Social Service costs only
Perspective on outcomes	All health effects on individuals
Type of economic evaluation	Cost-effectiveness analysis
Synthesis of evidence on outcomes	Based on a systematic review
Measure of health benefits	Quality-adjusted life years (QALYs)
Description of health states for calculation of QALYs	Health states described using a standardized and validated generic instrument (like the EuroQol with UK weights)
Method of preference elicitation for health state valuation	Choice-based method, for example, time trade-off, standard gamble (not a rating scale)
Source of preference data	Representative sample of the public
Discount rate	An annual rate of 3.5% on both costs and health effects
Distributive equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit

The scoping exercise begins with a preliminary search of the literature, working with an Assessment Group. Potential consultees, commentators, and clinical specialists are identified and then consulted on the draft scope and they are also invited to a scoping workshop. Discussion at the scoping workshop is intended to ensure that all relevant issues have been considered and that the focus and boundaries of the appraisal have been clearly defined. The final scope for the appraisal is produced following the scoping workshop. The Assessment Group uses this scope to develop its assessment protocol.

The scope defines, amongst others, the following parameters of cost effectiveness:

- the clinical problem and the population(s) and any relevant subgroups for whom treatment with the technology is being appraised (choices here may be determined by both scientific judgments and colloquial ones – like the personal and social significance attached to particular clinical problems, for evidence NICE typically relies upon representative groups and experts with vicarious experience)
- the technology and the setting for its use (for example, hospital or community)
- the relevant comparator technologies (it is easy to 'rig' a comparison of technologies by suitable biased choices, or to render an analysis fairly useless by using evidence from trials using only a placebo comparator)
- the principal health outcome measures appropriate for the analysis (pragmatism is likely to rule here but choice of outcome measure – the denominator in an incremental cost-effectiveness ratio – is a critically important social value-judgment)

- the measures of costs to be assessed (interpreted as opportunity cost, this raises the same ethical issues as the choice of outcome measure, since cost for the NHS is ultimately ‘health gain forgone’)
- the time horizon over which benefits and costs will be assessed (both the period and the rate of discount entail value judgments)
- other considerations, for example, identification of patient subgroups for whom the technology might potentially be particularly clinically and cost effective (this implies that some patients might receive more favourable consideration than others, which is clearly value judgmental)
- special considerations and issues that are likely to affect the appraisal
- the extent and completeness of the evidence (while this may involve *scientific* value judgments – ‘is this research well-conceived and conducted?’ – it is less likely to involve *social* value judgments)
- consideration of the appropriateness of ‘routine’ assumptions embodied in NICE’s cost-effectiveness algorithms (for example, regarding the weights attached to incremental quality of life for different categories of beneficiary)

Perspective

The perspective from which an analysis is to be done plainly involves major ethical judgments, concerning, as it does, the scope of the costs and consequences that are deemed relevant. At NICE, many of these are embodied in the scoping phase, as we have seen. The specific guidance that NICE offers on perspective runs thus:

the perspective on outcomes should be all direct health effects whether for patients or, where relevant, other individuals (principally carers). The perspective adopted on costs should be that of the National Health Service and the Personal Social Services. If the inclusion of a wider set of costs or outcomes is expected to influence the results significantly, such analyses should be presented in addition to the reference case analysis.¹⁷

The recommended procedure is thus pragmatic in at least two ways:

- it locates the task as one of sub-optimizing within a frame defined by a budget and a mission statement
- it has a caveat enabling other elements to be adduced when judgment (ultimately that of NICE’s Board) suggests that it may be desirable

Some of the possibilities under the second bullet include any significant resource costs imposed outside the NHS, such as direct costs on patients or caregivers (for example, travel costs), or costs to other public sector organizations (but will not normally include productivity costs).

¹⁷ This perspective was substantially set for NICE in the two years prior to its establishment (1997–1998) in a Department of Health Expert Workshop on DH Guidelines for Pharmacoeconomic Studies, comprised of officials and academic health economists, chaired by me.

This represents a significant departure from the 'societal perspective' commonly advocated in the cost-effectiveness literature (e.g. Gold *et al.*, 1996). The NICE decision makers are very clear about their context and this is plainly an important determinant of how the public interest is interpreted.¹⁸ NICE does not try to simulate the workings of a perfect market (e.g. by postulating general welfare maximization based on individual preferences as its objective) and it deliberately curtails quite severely the scope of both the costs and the benefits to be considered. Its three grounds for doing so appear to be the principled one that the objective is not defined in terms of welfare but in terms of health, the conditional one that the overall resource commitment (and hence the opportunity cost felt elsewhere in the economy) as defined by the health care budget is a judgment reached elsewhere, and the pragmatic one that, in the event that ethically significant issues arise that might affect the decision but are outside the Reference Case, then they may be taken into account.

Outcome measurement

The quality adjusted life year (QALY) is a generic measure of health-related quality of life that takes into account both the quantity and the quality of life generated by interventions. It is plainly fraught with social value judgments. The version preferred by NICE is EQ-5D, a three-level, five-dimensional index where the dimensions are mobility, self-care, usual activity, pain/discomfort, and anxiety/depression, and each is scored on a three-point scale (1 – no problem, 2 – some problems, 3 – extreme problem). This generates 245 health states (3⁵ plus perfect health and dead) on a cardinal ('interval' – unique to a linear transform) scale.

Why the QALY? The following reasons were in *my* mind at the time:

- there was a general agreement amongst colleagues that we needed an outcome measure that related as closely as possible to the Secretaries of State's charge to NICE regarding 'health' (we effectively recognized the 'authority')

18 The most recent reasoning on this matter from NICE is in NICE (2004): 'The focus on cost-effectiveness analysis is justified by...the focus of the Institute on maximizing health gains from a fixed NHS/PSS budget. Given its widespread use, the QALY is considered to be the most appropriate generic measure of health benefit that reflects both mortality and health-related quality of life (HRQoL) effects. It is recognized that alternative measures exist (for example, the healthy year equivalent) but few economic evaluations have used these methods and their strengths and weaknesses are not fully understood. If the assumptions underlying QALYs (for example, constant proportional trade-off and additive independence between health states) are considered inappropriate in a particular case, then evidence to this effect should be produced and analyses using alternative measures may be presented as a non-reference case analysis...It is well established that different classification systems do not give consistent utility values to the same health states and hence results from the use of different systems cannot always be compared. Given the comparative nature of the Institute's work and the need for consistency across appraisals, the Institute would ideally wish that all appraisals used the same system. Currently, the most appropriate choice in the UK appears to be the EQ-5D. While it is widely used and simple to incorporate into studies, the EQ-5D may not be appropriate in all circumstances. Given the evolving state of the art in this area, the Institute believes it would be inappropriate to require the use of the EQ-5D to the exclusion of any other methods that meet its underlying criteria. Those submitting data should provide reasons for their choice of instrument. They should also indicate whether they have any evidence that will help the Committee to understand to what extent, and for what reason, their choice of instrument will have impacted on the valuation of the QALYs gained' (NICE, 2004, *passim*).

- I had a personal stake in the QALY from its earliest inception at York in 1971¹⁹ and in its subsequent development mainly by Alan Williams²⁰
- it was sufficiently close to common outcome measures used by clinicians in research (e.g. life-years or five-year survival rates) to be a familiar starting point
- it was an index rather than a profile
- it was generic – applicable to a wide range of technologies, thus facilitating comparison of relative cost effectiveness
- its theoretical properties and their consequences were (or were becoming) well-understood (properties such as constant proportional trade-off, risk neutrality over life-years, additive independence in health states)²¹
- the trade-offs embodied in it were derived from a representative sample of the UK population
- the ethical arguments adduced against it did not seem to me to be persuasive – particularly when it was compared with practical alternatives. These arguments were characterized chiefly by three weaknesses: (i) they attributed to QALYs claims of comprehensiveness that extended well beyond the scope of health-related quality of life-years (for example, to the value of a life itself); (ii) they confused the concept of a QALY with specific cases of it (and hence were usually objecting to incidental value assumptions, often the necessity of which had been largely uninvestigated until the QALY methodology exposed them as ethical issues to be resolved); (iii) they dealt in terms of heroic moral dilemmas (of the sort beloved by moral philosophers) rather than the routine and repeatedly arising, relatively small-scale, moral judgments that characterize most of the decisions about inclusion of health care technologies in benefit packages
- it had the attractive attribute of identifying and resolving, in a ready-made algorithm, a set of routine, frequently arising value judgmental issues concerning the dimensions in which ‘health’ was to be measured and the weights to be attached to each dimension (once one had, of course, accepted the essentially pragmatically crude utilitarian nature of the construct) – some of these come up under the distributional equity heading below:
- it was simple and low cost, and therefore had a good chance of becoming an ‘industry standard’

While other indexes had some of these virtues none, I thought, had all of them.

As ever, NICE is pragmatic. While endorsing the QALY as the Reference Case desideratum, it adds:

cost-benefit analysis may be particularly useful when non-health consequences are important in an evaluation. In such cases, willingness-to-pay methods may be used to value all consequences in monetary terms. Where such methods are used they should be fully described. (NICE, 2004: 22)

This illustrates a characteristic that was critical to my own thinking about the use of QALYs by NICE: from the beginning, their use was always context

19 See Culyer, Lavers, and Williams (1971).

20 This is, of course, no moral reason for promoting the QALY; merely an open admission of bias.

21 These and related issues are comprehensively discussed in Dolan (2000).

dependent. From the academic literature on QALYs it is easy to gain the impression both that their advocates think that they provide a complete and context-free answer to the complexity of issues bound up in measuring and assessing the outcomes of health care and that their use is as helpful in all decision settings as it may be in some. From NICE's beginning, the context for the use of QALYs was the Appraisals Committee, a multi-disciplinary, multi-professional, and multi-sectoral group that would always include patient and caregiver representatives and subject specialists. In that context the adequacy of QALYs, and of any other outcomes measures too of course, including whatever consideration of distributive fairness was felt to be needed, could receive full consideration, including any representations that might be made on behalf of those who would be living, as patients, caregivers, or medical professionals, with the consequences of the decision. The trick was always going to be to ensure that the special interests lined up behind particular groups of beneficiaries were never allowed to dominate the more generically represented general public. Making that assessment of adequacy and maintaining that balance were going to be critically important features if the Appraisals Committee's recommendations were to carry conviction. It is probably the case that *all* algorithms purporting to be decision related ought to be appraised, and preferably conceived, with a particular decision-making process in mind. At least, that is so if these methods, whether cost-effectiveness analysis in general or QALYs in particular, are to be seen as aids to thought rather than substitutes for it, in which case the thinkers in question need to have proper authority for making their decisions and a proper line of accountability.

Distribution of costs and benefits

The ethics associated with distributive justice arises unambiguously under this head. Earlier NICE guidance on equity made mention of the two great Aristotelian principles, which may be paraphrased as:

- *Horizontal equity – the equal treatment of those who are equal in an ethically relevant sense*
- *Vertical equity – the unequal treatment of those who are unequal in an ethically relevant sense*

There can be no doubt that NICE has been ready to apply both principles as may be appropriate. The trouble is, knowing what *is* appropriate and what the 'ethically relevant senses' are.

The reference case implies a particular ethical position regarding the comparison of health gained between individuals: *An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.* Some other characteristics of the individuals whose 'ethically relevant' sameness or difference might imply that they ought to be treated

either the same or differently might include (the list is not intended to be comprehensive):

- their pre- or post-treatment level of health (e.g. an incremental QALY is as socially valued when it accrues to someone with a lifetime of chronic incapacity as to someone who has hitherto been in abundant health)
- their current level of health (e.g. an incremental QALY is as socially valued when it accrues to someone who is currently very ill as to someone receiving preventive care who is perfectly fit)
- the size of the increment in health they may derive (e.g. the tenth incremental QALY someone receives is a socially valued as the first)
- age (e.g. a QALY for an aged person is as socially valued as one for a youngster)
- gender
- economic productivity
- lifestyle (e.g. a QALY for someone whose reckless style of life has brought about or exacerbated their ill-health is as socially valued as one for an impeccably clean-living type)

None of these assumptions is *inherent* in the idea of the QALY and alternative value assumptions may be made. But they are not readily made. NICE was in great difficulty in making a judgment on these matters. The QALY=QALY=QALY value assumption of the Reference Case is to be seen, pragmatically as always, as provisional – still a value judgment but one to be subject to scrutiny and special discussion. NICE was well aware that in assigning an equal weight to all QALYs in this way, it was inescapably making a value judgment.

NICE is adopting a pragmatic approach by testing the wider public's views on some of these trade-offs. The issue of age was put to the Citizens' Council (which recommended unitary weights). Others are being pursued through systematic research on the pioneering ethical principle that in serving the public it is useful to do empirical ethics – finding out what ethical views the public actually has and whether there is anything like a consensus. Simple-minded ethical talk about 'equality' in the health context is mere rhetoric unless one gets to be specific about equality of *what* and of *whom* and *for what*. There is a lot of unpacking that still needs to be done.

Despite its high dependence on professional consent, NICE has gone against economic opinion regarding cost effectiveness in a further respect. This is in its resolute refusal to declare a 'threshold' incremental cost-effectiveness ratio, which any technology must at least reach if it is to be recommended. On the face of it, this also flouts transparency leaving a murky region of discretion. There have been several attempts to infer what the threshold is, on the apparent assumption that NICE is dissembling (e.g. Devlin and Parkin, 2004; Towse and Pritchard, 2002). There appears to be a general concentration of estimates in the region of £20,000–£30,000 or even, according to Devlin and Parkin, higher. In April 2004, the Institute confirmed that interventions with a cost per QALY below £20,000 were likely to be recommended and the one explicit

announcement of a target threshold stipulates £36,000. It has been shown elsewhere (Culyer *et al.*, 2006) that NICE is not constitutionally placed to specify thresholds but must, rather like those who have speculated about the range in which a threshold might lie, search for the threshold that is implied by the existence of a fixed budget for the NHS (determined by Parliament), with the objective of maximizing health gain (also determined by Parliament), and relying on the range of technologies that are actually available for use. To date, NICE's 'searching' has focused mainly on new technologies rather than ones in current use, which may have cost-effectiveness ratios that are too high, though there are indications that NICE's searching will soon be more comprehensive and systematic.

6. Conclusions

This paper has sought to test a number of conjectures about the use of deliberative processes by applying them to the technology assessment procedures used by the National Institute for Health and Clinical Excellence in England and Wales. A deliberative process is one that elicits and combines evidence of different kinds and from different sources in order to develop guidance – in the present case, guidance for a health care system. A deliberative process entails the integration of three kinds of evidence:

- scientific context-free evidence about the general clinical potential of a technology
- scientific context-sensitive evidence about particular evidence in realistic scenarios
- colloquial evidence to fit context-free scientific evidence into a context and to supply the best evidence short of scientific evidence to fill in any relevant gaps.

It is shown that NICE's appraisals procedures and, in particular, its approach to cost effectiveness, entail both the weighing of each of these types of evidence and can be seen as rational responses to 12 conjectures about the circumstances under which a deliberative process is likely to be used. Each and every one of the conjectures applies in the case of NICE. Indeed, so closely does the NICE context match these circumstances that it seems almost inconceivable that its processes could have been otherwise. In that respect, it would appear to be exemplary.

References

Chalmers, I. (2003), 'Trying to do more good in policy and practice: the role of rigorous, transparent, up-to-date evaluations', *Annals of the American Academy of Political and Social Science*, 589: 22–40.

Chalmers, I. (2005), 'If evidence-informed policy works in practice, does it matter if it doesn't work in theory?', *Evidence and Policy*, 1: 227–242.

Culyer, A. J., R.J. Lavers and A. Williams (1971), 'Social indicators: health', *Social Trends*, 2: 31–42.

Culyer, A. J. and J. Lomas (2006), 'Deliberative processes and evidence-informed decision-making in health care – do they work and how might we know?' *Evidence and Policy*, 2: forthcoming.

Culyer, A. J., C. McCabe, A.H. Briggs, K. Claxton, M. Buxton, R.L. Akehurst, M. Sculpher and J. Brazier (2006), 'Searching for a threshold, not setting one; the role of the National Institute for Health and Clinical Excellence', *Journal of Health Services Research and Policy*, forthcoming.

Daniels, N. (2000a), 'Accountability for reasonableness in private and public health insurance', in A. Coulter and C. Ham (eds), *The Global Challenge of Health Care Rationing*, Buckingham: Open University Press, pp. 89–106.

Daniels, N. (2000b), 'Accountability for Reasonableness', *British Medical Journal*, 321: 1300–1301.

Department of Health (1991), *Research for Health*, London: Department of Health.

Department of Health (1996), *Priorities and Planning Guidance for the NHS 1997/98*, London: Department of Health.

Devlin, N. and D. Parkin (2004), 'Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis', *Health Economics*, 13: 437–452.

Dolan, P. (2000), 'The measurement of health-related quality of life for use in resource allocation decisions in health care', in A. J. Culyer and J. P. Newhouse (eds), *Handbook of Health Economics*, Amsterdam: Elsevier, pp. 1723–1760.

Gold, M. R., L.B. Russell, J.E. Siegel and M.C. Weinstein (eds) (1996), *Cost-Effectiveness in Health and Medicine*, New York: Oxford University Press.

Jarrett, L. and PIU (2004), *A Report on a Study to Evaluate Patient/Carer Membership of the first NICE Guideline Development Groups*, London: NICE.

Kelson, M. (2001), 'Patient Involvement in Clinical Guideline Development – where are we now?', *Journal of Clinical Governance*, 9: 169–174.

Kelson, M. (2002), 'The National Guidelines and Audit Patient Involvement Unit', *Journal of Clinical Excellence*, 4: 194–195.

Lomas, J., A.J. Culyer, C. McCutcheon, L. McAuley and S. Law (2005), *Conceptualizing and Combining Evidence for Health System Guidance*, Ottawa: Canadian Health Services Research Foundation.

Moore, A., H. McQuay and M. Gray (1994) *Bandolier: The First Twenty Issues*, Oxford: Bandolier supported by the NHS R&D Directorate.

NICE (2004), *Guide to the Methods of Technology Appraisal*, London: National Institute for Clinical Excellence.

Scally, G. and L. J. Donaldson (1998), 'Looking forward: clinical governance and the drive for quality improvement in the new NHS in England', *British Medical Journal*, 317: 61–65.

Towse, A. and C. Pritchard (2002), 'Does NICE have a threshold? An external view', in A. Towse, C. Pritchard, and N. Devlin (eds), *Cost-Effectiveness Thresholds: Economic and Ethical Issues*, London: Office of Health Economics, pp. 25–37.