



ELSEVIER

Contents lists available at ScienceDirect

## Health Policy

journal homepage: [www.elsevier.com/locate/healthpol](http://www.elsevier.com/locate/healthpol)



# Searching for cost effectiveness thresholds in the NHS

John Appleby<sup>a,b,\*</sup>, Nancy Devlin<sup>b,a</sup>, David Parkin<sup>b</sup>, Martin Buxton<sup>c</sup>, Kalipso Chalkidou<sup>d</sup>

<sup>a</sup> King's Fund, London, United Kingdom

<sup>b</sup> City Health Economics Centre, Department of Economics, City University, London, United Kingdom

<sup>c</sup> Health Economics Research Group, Brunel University, Uxbridge, United Kingdom

<sup>d</sup> Policy Consulting, NICE, London, United Kingdom

### ARTICLE INFO

*Article history:*  
Available online xxx

*Keywords:*  
Resource allocation  
Cost effectiveness threshold  
Cost utility analysis  
QALYs

### ABSTRACT

**Objectives:** The UK's National Institute of Health and Clinical Excellence (NICE) has an explicit cost-effectiveness threshold for deciding whether or not services are to be provided in the National Health Service (NHS), but there is currently little evidence to support the level at which it is set. This study examines whether it is possible to obtain such evidence by examining decision making elsewhere in the NHS. Its objectives are to set out a conceptual model linking NICE decision making based on explicit thresholds with the thresholds implicit in local decision making and to gauge the feasibility of (a) identifying those implicit local cost effectiveness thresholds and (b) using these to gauge the appropriateness of NICE's explicit threshold.

**Methods:** Structured interviews with senior staff, together with financial and public health information, from six NHS purchasers and 18 providers. A list of health care services introduced or discontinued in 2006/7 was constructed. Those that were in principle amenable to estimation of a cost-effectiveness ratio were examined.

**Results:** It was feasible to identify decisions and to estimate the cost-effectiveness of some. These were not necessarily 'marginal' services. Issues include: services that are dominated (or dominate); decisions about how, rather than what, services should be delivered; the lack of local cost effectiveness evidence; and considerations other than cost-effectiveness.

**Conclusions:** A definitive finding about the consistency or otherwise of NICE and NHS cost effectiveness thresholds would require very many decisions to be observed, combined with a detailed understanding of the local decision making processes.

© 2008 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Since its inception in 1999, the UK's National Institute for Health and Clinical Excellence (NICE) has provided evidence-based guidance for the NHS in England and Wales on the clinical and cost effectiveness of new and existing interventions. NICE is only one of many organisations that influence the use of resources in the NHS. Spending

is influenced by policy decisions made by the Department of Health – for example, to reduced waiting time targets. Other decisions about which services are provided, when, how and to whom, are made *inter alia* by Primary Care Trusts (PCTs), health care provider organisations such as Hospital Trusts and individual health professionals. Treatments recommended by NICE account for a very small proportion of NHS spending. Nevertheless, NICE decisions have an effect on resource use and it is important that those decisions use not only good information, but also good decision making criteria.

One accepted decision-making criterion in the NHS is the cost-effectiveness of the services that it provides. Whilst most NHS decisions do not require explicit

\* Corresponding author at: King's Fund, 11–13 Cavendish Square, London W1G 0AN, United Kingdom. Tel.: +44 20 7307 2540; fax: +44 20 7307 2807.

E-mail address: [j.appleby@kingsfund.org.uk](mailto:j.appleby@kingsfund.org.uk) (J. Appleby).

estimation of cost effectiveness, NICE is charged with a responsibility to be both explicit and transparent. It therefore has to state what is and what is not cost-effective. NICE's stated 'threshold' is based on the incremental cost-effectiveness ratio (ICER), expressed as an incremental cost per Quality Adjusted Life Year (QALY) gained.

However, the ICER level that NICE has adopted as the criterion for its decisions – a 'range' of £20,000 to £30,000 per QALY gained – has no firm basis in evidence or theory. It is based on a broad assertion that emerged from informal judgements made by NICE advisory committees. Cost effectiveness is not the sole driver of NICE decisions [1–5], but is *an* important consideration; so a defensible basis for judgements about what is acceptable value for money is important for NICE, the NHS, taxpayers and patients.

In a commentary on NICE [6], Alan Williams noted that a cost-effectiveness threshold can be established in two ways. We can *either* decide how much a QALY is worth, by finding its value to society *or* find the value of a QALY implied by budget-constrained decisions made in the NHS. These will coincide only under special circumstances that in practice are not met: the NHS budget is set to enable all health care to be provided whose marginal benefits, in terms of QALYs gained, are greater than or equal to their marginal costs; all budgets and their allocation are driven by QALY maximisation; and information on the costs and benefits of all possible health care is readily available. So, if NICE based its decisions on the social value of a QALY, they might be inconsistent with the budget constraints faced in practice by NHS organisations such as PCTs.<sup>1</sup>

NICE does not in fact use the social value approach. NICE's threshold is intended to reflect what is affordable given finite NHS resources and the demand on these resources from available health care services, management practices and population needs [7]. But if NICE adopts *any* threshold that is inconsistent with the threshold implied by PCT budgets, the same problem will arise: patients will be denied cost-effective services. If NICE's threshold is *above* that relevant to PCT budgets, implementation of NICE decisions will 'crowd out' more cost-effective services locally; if it is *below*, NICE will reject health care technologies that are cost-effective relative to others provided locally.

Williams noted that

"...it is extremely likely that the 'shadow price' of a QALY (i.e. the implicit value of a QALY as determined by the most cost effective intervention that each purchaser just cannot afford to buy) will vary from purchaser to purchaser. And it is widely believed that this 'shadow price' is much lower than the NICE benchmark of £30k. I think a major effort should be made to find out whether this belief is well founded." (p. 8) [6]

The study reported here assessed the feasibility of locating the implicit value of a QALY in the NHS by examining local NHS decisions. Specific objectives were to investi-

gate the feasibility of identifying services that reveal the threshold implied by local NHS decisions, estimating ICERs for those services and generating conclusions about the congruence of local and NICE judgements about value for money. Our focus was on services amenable to change, whether being introduced or being discontinued.

While this study is empirical, a conceptual framework is required to define a threshold concept that can be applied to decision making by both PCTs and NICE. This has the obvious difficulty that PCTs' thresholds are implicit, and NICE denies that it has a single threshold.

## 2. Cost effectiveness thresholds: developing a clearer conceptual framework

### 2.1. What is NICE's threshold?

NICE has always avoided the term 'threshold', instead expressing its cost-effectiveness criteria in terms of a "range of acceptable incremental cost-effectiveness ratios":

"There is no empirical basis for assigning a particular value (or values) to the cut-off between cost effectiveness and cost ineffectiveness. The consensus amongst the Institute's economic advisors is that the Institute should, generally, accept as cost effective those interventions with an incremental cost-effectiveness ratio of less than £20,000 per QALY and that there should be increasingly strong reasons for accepting as cost effective interventions with an incremental cost-effectiveness ratio of over £30,000 per QALY. These reasons include the degree of uncertainty surrounding the estimate of the incremental cost-effectiveness ratio and, where appropriate, reference to previous appraisals. The Institute and its advisory bodies will also wish to consider social value judgements including consideration of the nature of the condition, the particular patient population, and the intervention itself." [7]

While a reasonably clear statement of general principles, it lacks clarity about what happens *between* £20,000 and £30,000 and is vague about exactly what happens over £30,000. More recently the Chair of NICE clarified the latter in terms of process, though not of precise criteria:

"...NICE guidance should make explicit its reasons for recommending as cost effective those interventions with an incremental cost effectiveness ratio over £20,000 to £30,000." [9]

Other definitions exist, some of which add more precision and clarity [1], but they are often different in subtle but potentially important ways. Unfortunately, none of these definitions are wholly satisfactory, since they mix a precise, quantified criterion of a cost per QALY gained (CQG) range with an imprecise qualitative description of other factors affecting NICE decisions. Moreover, the way in which these other factors are combined with the CQG range in decision making is unclear (although a recent consultation issued by NICE in response to a review of topping up NHS care [10] has set out some circumstances where greater weight could be given to the benefits arising from treatments for

<sup>1</sup> For an interesting approach to the estimation of cost per life year based on PCT spending on different disease/programmes and population health outcomes see Marin et al. [8].



services up to and including service X, where  $\text{£CE}_X = \text{£X}$ , and accept that the NHS budget would be  $\text{£}\sum C_X$ . Services Y to N would not be provided. This threshold is shown in Fig. 1 as a bold line. The services immediately above and below it are those that are 'marginal'.

NICE states that its position is and has always been "that its threshold should represent the national shadow price of a QALY" [7]. The point at which the NHS budget of  $\text{£}\sum C_X$  is exhausted reveals the shadow price of a weighted QALY as lying between the CWQG of the last service funded ( $\text{£CE}_X$ ) and that of the next, most cost effective service *not* funded ( $\text{£CE}_Y$ ). In practice, NICE does not have access to the sort of data required for such a table: its threshold is effectively its best guess about the CWQG that *would* be revealed if it did.

#### 2.4. Using the model to guide the search for local NHS threshold(s)

We assume that local NHS decision making about services to be provided is conducted on an annual basis and that the existing threshold and any changes to the existing budget are known. This does not require precise annual recalculation of all data on all services to determine which will be provided – we assume that decision makers want to rank services and have in fact already achieved this to a great extent. They only need information about new services and *marginal* services – those which have a CWQG close to the existing threshold. Precisely how far 'close' is will depend on the size of any changes to the overall budget and the number and size of possible new services, since that will affect the position of the new threshold. A new service will be introduced if its CWQG is less than the threshold; an existing service will be excluded if its CWQG is greater than the threshold.

In practice, decisions are likely to concentrate not on marginal services but on new services and existing services suspected of being not being cost-effective. From an observer's perspective, the actual threshold is therefore not known with certainty, but each decision reduces the uncertainty about it. If decision makers behave according to the model and are consistent in their decision making, they will always include a service below the threshold and exclude one above it. The shaded entries in Fig. 1 represent decisions taken locally – the dark ones are included services and the light ones are excluded. The threshold lies between the lowest CWQG observed for an excluded service (disinvestment) and the highest observed for a newly included service (investment), with the range of uncertainty for the threshold shown by the arrow to the right.

To identify the implied PCT CWQG threshold would therefore require a large enough sample of decisions to identify a narrow enough range within which marginal services lie. The following describes a small-scale study to test the feasibility of such research.

### 3. Methods

Eight PCTs were identified as potential study sites. One declined to participate and one did not reply. The six PCTs in the study were not intended to be representative, geographically or in any other respect. Rather, they were either

accessible to the research team through professional networks or whose decision making was known to be led by individuals who would be interested in the study.

Structured interviews with the Directors of Public Health in each PCT were carried out between January and April 2007. An opportunistic sample of Finance Directors from NHS Trusts attending a course at City University also completed questionnaires. Interview scripts and questionnaires are available from the authors on request.

Interview notes and completed questionnaires were collated and a list of services in each of the three decision categories – Introduced, Discontinued and Deferred – was created. From this, a CQG was estimated for nine services, selected because they provide examples in each of the three decision categories and were clearly described. Services were excluded where a NICE technology appraisal decision applied, on the basis that to act on these, or not, was not within local discretion.

Cost effectiveness evidence that was considered as part of decision-making was recorded. Where evidence was not available, or not taken into account by decision makers, an ICER for affected services was imputed using published evidence on cost-effectiveness. There was no attempt to undertake an exhaustive search or conform to a strictly systematic procedure in the search for this evidence, as such a process better reflects the sort of evidence that PCTs might have easily found for themselves. We searched for papers via Pubmed (including Medline), the CRD website and Google.

### 4. Results

#### 4.1. Identification of services amenable to change

Completed questionnaires were returned by 17 Finance Directors, indicating specific services they had introduced or discontinued during the current financial year and the rationale for these decisions.

Most decisions were not relevant for our analysis. In almost every case, the reasons why providers discontinued services were that their unit cost exceeded the NHS tariff price or their local PCT had decided to commission the service elsewhere, or both. Similarly, most new services were related to entry into existing markets, for example replacing a current provider, or resulted from a merger with another provider.

Some services identified seemed genuinely to be marginal, such as a decision to replace methadone with a more costly but more effective alternative; and to introduce bariatric surgery where it had not previously been provided. However, no decisions were identified that were based on economic evaluation. Providers' decisions were based on considering the 'business case' – principally, financial evidence for existing services or projections of costs and revenue for new services. Health outcomes feature only peripherally in this, as the avoidance of commercial risk associated with an adverse outcome. Provider Trusts therefore seem to be responding to incentives to behave in a business-like way.

In contrast, it was easy to identify specific services that PCTs had introduced, discontinued or deferred during the

current financial year. The list of identified services was dominated by discontinuations and disinvestments, which may reflect the financial deficits most PCTs had at the time.

Discontinuations of services included a growing list of services that PCTs have explicitly decided to discontinue funding – such as tonsillectomies, grommets, tattoo removal, hysterectomy for heavy menstrual bleeding and surgery for mild to moderate varicose veins. These are not blanket exclusions, however, PCTs reserve the ability to make judgements over exceptional individual cases. There were also partial service discontinuations – for example, surgery for hip and knee replacement and cataract removal – where a decision was made about the severity of the condition, or clinical threshold, below which patients will not be offered treatment. Some PCTs are exploring patient prioritisation scoring tools such as New Zealand's CPAC [11,12] to assess patients' ability to benefit and as the basis for deciding what the explicit clinical threshold should be.

There were instances where published economic evaluations had clearly been consulted, but the general impression was that clinical and other non-economic considerations dominated decision making. In effect, PCTs focused on the factors constituting the 'weight' (W) in the CQOG metric.

Many of the PCT decisions identified were not appropriate for further analysis, including major service reconfigurations and minor changes to the way care was provided. They represented not a net addition or loss from the

NHS package of services, but changes in delivery patterns: changes in *how* or by *whom* rather than *which* services are to be provided. Most of these changes were perceived as being dominant in terms of cost-effectiveness, that is they reduce costs and either do not affect or improve health outcomes or 'soft' outcomes such as patient satisfaction.

The services identified as being amenable to change are summarised in Table 1. PCTs are numbered so that sets of decisions from any one PCT, and common decisions being made by PCTs, are evident. The following section reports the estimation of CQOG for a selection of these.

#### 4.2. Imputing the cost effectiveness threshold range

Table 2 lists the services chosen for further analysis.

In 5 of the services investigated it was not possible to estimate a CQOG. In 3 of these 5 cases – discontinuation of hysterectomy, cataract surgery for patients with good visual acuity, and cholecystectomy for non-symptomatic gallstones – this was because surgery in each case was dominated by alternatives. For partial discontinuation of routine hernia a CQOG could not be estimated because the nature of the restricted provision of services was unclear. For cognitive therapy for pain management, insufficient evidence was available to allow the estimation of the CQOG.

The information from decisions for which there was a CQOG gave a wide range within which the implied threshold could lie. We were not confident that we had identified

**Table 1**

Specific identified local changes to the range of NHS provision, 2006/7.

Introduced	Discontinued/'Managed access'	Deferred
New test for TB [1] Macugen [1]	Tonsillectomy [3,4,1,5,7] Grommets [3,4,1,5,7]	Chlamydia screening [6] Choosing Health strategies [6], e.g. smoking cessation.
Avastin [2] Bariatric surgery [7] Subutex (methadone replacement) [7]	Hysterectomy [4,1,5] Plastics [4,1,5] Removal of skin tags [4,5,2]	Retinal screening for diabetes [3] Bowel cancer screening [3] 2nd and 3rd line (non-tariff) drugs for cancer [2] PET/CTs for lymphoma [2] Tarceva for lung cancer [7]
Reduction in waiting times [1–7] Intensive stroke services: early rehabilitation [3] Children's speech and language therapy [3] 'Booze bus' ambulance service in city centres [7]	Tattoo removal [4] Varicose veins [4,1,5,7] Circumcision for non-medical reasons [2] Homeopathy [5] IVF – limit number of cycles to 2 [1] or 1 [2]	<i>Deferred discontinuation:</i> Community dental service annual oral health surveys for primary school children [3]
	Cochlear implants [1,5] Carpal tunnel syndrome [5] Knee wash out [5] Wisdom teeth extraction [5] Carotid endarterectomy for symptomatic carotid stenosis [8] Cognitive therapy for management of chronic pain [8] 'Managed access': Cataracts [2,3] Hip and knee replacement [3] Obesity restrictions on hips and knees [BMI] [2] Cholecystectomy for non-symptomatic gall stones [2] Inguinal/routine hernia repair [7] Orthodontic services [5,7]	

Note: 1–6 represent decisions taken by our sample of PCTs, 7 represents a decision taken by an NHS Trust (from the surveys of Trust finance directors), and 8 represents services selected from a consortia of PCTs' commissioning plans.

Note: 'Managed access' refers to a restriction on access, variably defined, for example, in terms of severity of illness or some patient characteristic related to clinical outcome.

**Table 2**

Nine NHS services, the nature of decision made in respect of them, and results from an investigation of the cost effectiveness of each.

Service area	Nature of decision	CQG
1. Hysterectomy for heavy menstrual bleeding	Discontinued	No relevant CQG as surgical option dominated by alternative pharmaceutical interventions
2. Surgery for mild/moderate Varicose veins	Discontinued	No relevant CQG as implication that 'mild/moderate' dominated by alternatives of conservative management and sclerotherapy. However, decision implies continuation for 'severe' varicose veins, at a reported CQG of £1,936
3. Cochlear implants	Introduced	Reported CQG: £10,341
4. 'Routine' inguinal hernia repair	Discontinued/'managed access'	Detailed nature of restriction unclear. Unable to estimate CPQG.
5. Cataract surgery for patients with good visual acuity	Discontinued	This intervention for 'good visual acuity' (presumed to be 6/12) dominated, therefore no CQG can be calculated
6. Cholecystectomy for non-symptomatic gallstones	Discontinued	Clinical consensus that intervention confers no clinical benefit. Intervention dominated and hence no CQG can be calculated
7. Chlamydia screening	Deferred	No clear evidence on CQG, 'widely believed to be very cost effective'
8. Carotid endarterectomy for asymptomatic carotid stenosis	Possible discontinuation/'managed access'	Weak evidence, but CQG could be in the range: £5,000 to £30,000
9. Cognitive therapy for management of chronic pain.	Possible discontinuation/'managed access'	UK evidence lacking; no reasonable estimate of CQG could be made

any truly marginal services, although it would have been hard to tell if we had. As a result, we cannot draw definitive conclusions about an NHS threshold.

**5. Discussion**

The concept of an implied cost-effectiveness threshold does not require PCT decision makers to have made a *detailed* analysis of the costs and benefits arising from decisions about what is to be included in and excluded from the package of commissioned services. It does however require that PCTs have made such decisions, and that these are not made randomly but are informed by costs and benefits. Our study suggests that they do make such decisions and these are quite easily identified. PCTs consider cost-effectiveness, but mainly on the basis of technical efficiency – removing ineffective procedures – or of dominance – cheaper procedures that are at least as effective. Affordability is also a concern, in terms of the total cost of a decision irrespective of its cost effectiveness. Evidence is considered, including cases where there is a trade-off between increased costs and increased benefits.

Cases such these enable us to identify the implied threshold. Unfortunately, their rarity means we cannot do so from our sample. Moreover, we doubt that simply obtaining a larger sample over a longer period would help. There is no guarantee that we would identify many more of the relevant marginal services, and in any case what is marginal changes over time. Instead, we require more sophisticated ways of researching NHS decisions.

For example, most PCT decisions were service reconfigurations, 'spend to save' schemes, and demand management initiatives. By their nature, it is difficult to define and describe these and to predict their incremental costs and effects. But they are important changes to NHS service delivery, with sometimes significant financial consequences, that have cost-effectiveness implications.

Similarly, waiting list initiatives have dominated elective surgery policies, and have cost-effectiveness implications that may be hard to tease out, but do exist.

Most likely, our sample of *formal* PCT decisions does not fully represent all *effective* PCT decisions, since they may have a topic selection process before looking in more detail at an intervention. It will also not fully account for the sort of decision making criteria that PCTs employ – in particular other factors affecting their judgements. This is important for our method, because this affects the weighting part of the CWQG required for a link with the NICE threshold. Our sample would need to include cases that had no special factors, in order to obtain a 'clean' CQG, with a weight of one; moreover it would be interesting to investigate whether or not the implied weights used by PCTs and by NICE are congruent. Making the correct comparison between PCT and NICE decision making – that is, looking at the process and criteria for NICE and PCT topic selection and the social value and other criteria PCTs use in decision making – would be worth investigating as part of the threshold issue.

Consideration of these issues is urgent. Even if NICE estimated its threshold correctly when it was first adopted, it is most probably incorrect now. Changes in the NHS budget and the range of health care interventions that is available will have altered the NHS shadow price of a QALY, affecting which services are marginal. Moreover, this shadow price will continue to change, necessitating a continuing update of the threshold. Up-to-date information on local decision making thresholds would be very valuable for such updating.

**Acknowledgements**

This project was funded by NICE R&D, within Framework Agreement 01/06/2004 between HAD and King's Fund.

We are grateful to members of our project steering group for valuable advice and feedback: Tony Culyer,

Peter Littlejohns, Alastair Fischer, Jo Lord and Jennifer Field (from NICE); Donald Franklin and Andrew Jackson (Department of Health); and Karl Claxton (University of York).

We wish to express our gratitude to NHS staff who participated in this study, including the NHS Finance Directors who provided us with information and useful insights on their decision-making and the PCT. Directors of Public Health (and other staff) who consented to being interviewed and providing us with documentation.

This article represents the personal views of Dr. Kalipso Chalkidou and should not be taken as necessarily representing the policies, practices, or the corporate position of her employer, the National Institute for Health and Clinical Excellence.

## References

- [1] Rawlins M, Culyer A. National Institute for Clinical Excellence and its value judgments. *BMJ* 2004;329:224–7.
- [2] Devlin N, Parkin D. Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics* 2004;13(5):437–52.
- [3] NICE. Social value judgements guideline. Principles for the development of NICE guidance. London; 2005.
- [4] Dakin H, Devlin N, Odeyemi I. Yes, no or 'yes, but': a multinomial model of NICE decisions. *Health Policy* 2006;77(3):352–67.
- [5] NICE. Appraising end of life medicines. <http://www.nice.org.uk/newsevents/infocus/EndOfLifeMedicinesConsultation.jsp>; November 5, 2008.
- [6] Williams A. What could be nicer than NICE? London: Office for Health Economics; 2004.
- [7] NICE. Guide to the methods of technology appraisal (reference N0515). April 2004. [http://www.nice.org.uk/niceMedia/pdf/TAP\\_Methods.pdf](http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf); 2004.
- [8] Martin S, Rice N, Smith PC. The link between health care spending and health outcomes for the new English primary care trusts. CHE Research Paper 42, University of York, York; 2008.
- [9] Rawlins M. Fifteen minutes with Professor Sir Michael Rawlins. *BMJ Career Focus*. <http://careers.bmj.com/careers/advice/view-article.html?id=1419>; 2006.
- [10] Richards M. Improving access to medicines for NHS patients: a report for the Secretary of State for Health by Professor Mike Richards. [http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_089927](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_089927); November 4, 2008.
- [11] Hadorn DC, Holmes AC. The New Zealand priority criteria project. Part 1: Overview. *BMJ* 1997;314(7074):131–4.
- [12] Derrett S, Devlin N, Hansen P, Herbison P. Prioritizing patients for elective surgery: a prospective study of clinical priority assessment criteria in New Zealand. *International Journal of Technology Assessment in Health Care* 2003;19(1):91–105.