National Institute for Health and Clinical Excellence

ITEM 8

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

SPECIAL HEALTH AUTHORITY

REPORT OF THRESHOLD WORKSHOP

The Board is invited to consider the report and agree further activities

Prof. Peter Littlejohns, Clinical and Public Health Director
Dr Sarah Garner, Associate Director, R&D

July 2009

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

SPECIAL HEALTH AUTHORITY

REPORT OF THRESHOLD WORKSHOP

1 Background

1.1 In their most recent report on NICE, the Health Select Committee expressed concern about the cost-effectiveness 'threshold' used by NICE and suggested:

- that the threshold used by NICE in its full assessments be reviewed
- that further research comparing thresholds used by PCTs and that used by NICE should be undertaken.
- that an independent body should determine the threshold used when making judgments of the value of technologies to the NHS.

1.2 As part of the Department of Health's response to the Health Select Committee report they announced that NICE would run a 'Threshold Workshop' to explore the issues raised by the Health Select Committee.

2 The workshop

2.1 The workshop was arranged for February 2009 but had to be cancelled due to inclement weather. The re-scheduled workshop was held on Monday 20th April 2009.

2.2 It was chaired by Professor Sir Michael Rawlins, Chairman of NICE, and its purpose was to provide an opportunity for an integrated discussion of NICE's 'threshold': Its focus was on exploring whether there is a need for the 'threshold' to be amended, and if so what methods would be available, and what would need to be done by NICE and by others for any possible improvements to be realised.

2.3 The workshop consisted of a series of speakers presenting the history and theory of the 'threshold', how the Technology Appraisal Advisory Committee uses it in making decisions, methodological options for calculating the threshold, how PCTS make decisions, and how the 'threshold' NICE uses compares with others in use in the public sector (see appendix 1 for programme).

2.4 In the afternoon the delegates worked in small groups addressing 3 questions: what factors should be considered in determining the need for a threshold review; what information is relevant; and whose responsibility is it to take these activities forward.

2.5 Fifty three people (see appendix 2) attended, drawn from NICE's stakeholders and Geoff Watts prepared a report (appendix 3).

3 Conclusions of the workshop

3.1 On the conceptual level, the threshold range used by NICE reflects, what the 'NHS can afford' i.e. it is the level of cost per QALY that new interventions need to achieve compared to those that they are likely to be displaced. It does not represent what 'society may wish to pay' for a QALY. It is directly related to the NHS budget which is not the responsibility of NICE.

3.2 There were conflicting arguments as to whether it was too high or too low. In addition there was no consensus on a methodological approach to clarify the position.

3.3 The current evidence suggests that the threshold is not dissimilar to that used in other public sector economic analyses.

3.4 There was sufficient theoretical basis to explain why a threshold range identified in 1999 could still be appropriate for 2009.

3.5 NICE is encouraged to work with the PCTS to understand their decision making processes and how NICE guidance effects local activity and decision making, and to explore opportunities for examining NICE's threshold in the context of local decision-making.

3.6 Programme budgeting could provide an appropriate way to gather further information on the cost-effectiveness 'threshold ranges' used in the NHS; however further methodological exploration is required as there is still controversy over its interpretation.

3.7 While the NHS is a national organisation the nature of autonomous local primary care trusts and their funding streams would mean that multiple 'thresholds' would always be present.

4 Next steps

4.1 The Board is invited to consider the report and agree further activities. It is proposed that:

4.2 On the basis of the current information it would be inappropriate for the Institute to change its current threshold range.

4.3 NICE should establish with PCTs an ongoing programme of work looking at the way that local PCTs make prioritisation decisions and the impact of NICE guidance on these decisions. This work would be undertaken by the R&D and Implementation directorates within NICE and a representative sample of PCTs, SHAs and the new PCT HTA liaison unit. The potential for disinvestment should be part of this programme.

4.4 NICE should encourage further research on methods to establish NICE's threshold. (It has been highlighted as a priority in the current MRC call for methodological research: http://www.mrc.ac.uk/Fundingopportunities/Calls/NICEdecisionmaking/index.htm).

4.5 Following this work NICE in conjunction with the Department of Health, researchers and research funding bodies should explore how a process for reviewing the threshold on a regular basis could be established.

Prof. Peter Littlejohns, Clinical and Public Health Director
Dr Sarah Garner, Associate Director, R&D

July 2009

APPENDIX 1

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Threshold Technical Workshop

20th April 2009, 9:30am – 5:00pm

NICE, MidCity Place, 71 High Holborn, London, WC1V 6NA

1. Welcome, background and aims of the day	Professor Sir Michael Rawlins Mr Kevin Barron MP	9.30 am
2. How is the threshold used by the Appraisal Committee?	Professor Andrew Stevens	9.40 am
3. Discussion	All	10.00 am
4. The concept of the threshold and NICE's approach	Professor Martin Buxton	10:20am
5. Discussion	All	10.40 am
Coffee break – 20 min		11.00 am
6. The link between health care spending and health outcomes: implications for the threshold	Professor Peter Smith	11.20 am
7. Searching for local cost-effectiveness thresholds	Professor John Appleby	11.40 am
8. Thresholds in non-NHS settings	Professor Graham Loomes	12:00
9. Discussion	All	12.20pm
Lunch – 60 min		1.00pm
10. NICE's previous investigations into the threshold (methods review and research)	Professor Peter Littlejohns	2:00pm
11. Determining the threshold – what factors should be considered? *Coffee to be served to small groups*	Small groups	2.15pm
12. Feedback and discussion	All	3.15pm
13. The implications for NICE	All	4.15pm

14. Conclusions and next steps	Professor Sir Michael Rawlins	4.45pm
15. Finish		5.00pm

Appendix 2

Threshold Workshop

Attendee List

1. John Appleby, Kings Fund (A)
2. Daphne Austin, West Midlands Specialised Services Agency (B)
3. Rachel Baker, University of Newcastle (C)
4. David Barnett, NICE (C)
5. Kevin Barron, Health Select Committee
6. Meindert Boysen, NICE (B)
7. David Brickwood, ABPI (B)
8. Ailsa Brown, NHS QIS (A)
9. Martin Buxton, Brunel University (D)
10. Sophia Christie, NHS Confederation (D)
11. Karl Claxton, University of York (A)
12. Angela Cooper, Medical Research Council (A)
13. Tim Crayford, Croydon PCT (C)
14. Tony Culyer, NICE (D)
15. Nancy Devlin, City University (C)

16. Andrew Dillon, NICE (B)
17. Cam Donaldson, University of Newcastle (D)
18. Nick Doyle, NICE (B)
19. Richard Edlin, University of Leeds (C)
20. Jane Fisher, MRC (B)
21. Alastair Fischer, NICE (C)
22. Sarah Garner, NICE (D)
23. Elisabeth George, NICE (A)
24. Jane Gizbert, NICE (C)
25. John Henderson, Department of Health (D)
26. Marcia Kelson, NICE (B)
27. Stefanie Kinsley, NICE (A)
28. Imran Khan, on behalf of Evan Harris MP (A)
29. Peter Littlejohns, NICE (A)
30. Graham Loomes, University of East Anglia (C)
31. Carole Longson, NICE (C)
32. Helen Mason, University of Newcastle (D)
33. Alan Maynard, University of York (D)
34. Chris McCabe, University of Leeds (B)
35. Barry McCormick, Department of Health (D)
36. Barbara McLaughlin, Chair of PIN (A)

37. Paul McManus, Barnsley PCT (D)
38. Bhash Naidoo, NICE (B)
39. Doug Naysmith, Health Select Committee (C)
40. Danny Palnoch, Department of Health (A)
41. David Parkin, City University (B)
42. Dani Preedy, NICE Appraisal Committee (C)
43. James Raftery (C)
44. Michael Rawlins, NICE (D)
45. Francis Ruiz, NICE (B)
46. Mark Sculpher, University of York / MRC scoping project (D)
47. Gabrielle Silber, ABPI (A)
48. Don Sinclair, South Central Priorities Support Unit (C)
49. Peter Smith, University of York
50. Andrew Stevens, NICE (A)
51. Adrian Towse, Office of Health Economics (C)
52. Michael Wallace, ABHI (A)
53. Geoff Watts, rapporteur

Appendix 3

Threshold workshop
Report of a technical meeting organised by NICE

Introduction

If the provision of advice was of itself the means of resolving dilemmas, NICE would have no difficulty in deciding whether or not the current level of its cost effectiveness threshold has been pitched appropriately. Since the £20-30K figure it uses first entered the public domain, all manner of critics and commentators have offered assurances that it's clearly been set too low. Or too high. Or that it's just about right. As NICE chairman Professor Michael Rawlins reminded his audience at the beginning of the meeting's introductory session, the threshold figure is not based on hard data. Nor indeed has it been used as a hard and fast cut off point. He instanced Glivec for chronic myeloid leukaemia at £48K per QALY. Hence the reason for organising this multi-disciplinary workshop. NICE needs to get a sense of what interested parties think is the course of action it should now be taking: whether to change the threshold and, if so, on what basis and how. Among the interested parties are, of course, PCTs – the bodies that have to implement NICE recommendations.

As NICE's clinical and public health director Professor Peter Littlejohns was to point out later in the meeting, although the threshold figure has not been altered, this is not because he and his colleagues have been asleep on the job. The obstacle to change is the lack of empirical evidence. NICE has sought professional and academic advice, undertaken or commissioned research, launched policy reviews, and canvassed the opinions of its Citizens Council. What it has not found is a mechanism for calculating an appropriate threshold figure on which all concerned can agree.

Following its own study of the issue in 2008, a Commons Health Select Committee reached the following conclusions: "The affordability of NICE guidance and the threshold it uses to decide whether a treatment is cost-effective is of serious concern. The threshold is not based on empirical research and is not directly related to the NHS budget. It seems to be higher than the threshold used by PCTS for treatments not assessed by NICE. Some witnesses, including patient organisations and pharmaceutical companies, thought NICE should be more generous in the cost per QALY threshold it uses, and should approve more products. On the other hand, some PCTs struggle to implement NICE guidance at the current threshold and other witnesses argued that a lower threshold should be used."

It was this report that triggered the setting up of the workshop. "We recommend that the thresholds used by NICE in its full assessments be reviewed; further research comparing thresholds used by PCTs and those used by NICE should be undertaken. An independent body should determine the threshold used when making judgements of the value of technologies to the NHS."

Kevin Barron MP, chairman of the Commons Health Committee, was among those taking part in the NICE meeting. As he pointed out, he and his colleagues had discussed NICE and its activities on a number of occasions. Most recently they'd been concerned by the extent to which fellow professionals appeared to have been criticising NICE decisions. This, he felt, sent out bad messages to all concerned. The other issue was, of course, the threshold, and the relationship between that used by NICE as opposed to those of PCTs. This has to be of concern to anyone with an interest in effective commissioning within the NHS.

Appraisal: some practicalities

Most of the morning session of the meeting was given over to a series of presentations and questions on various aspects of the threshold. Most of the afternoon was spent in small group and plenary discussion of the issues.

Professor Andrew Stevens, who has himself chaired many appraisal committee meetings, began proceedings by outlining the nature of the appraisal task as he sees it, and some of the difficulties. Each committee confronted with a new technology, he said, has to answer two questions: first, what is the technology's value to patients; and second, what is the consequence of adopting it for all patients? The need to answer both these questions arises out of the duty that NICE has to consider both the clinical *and* the cost-effectiveness of interventions, both of which are influenced by a variety of factors. The first question, he added, is much simpler to answer than the second.

Committees, he continued, work by:

- seeking to understand the condition, the patients and the technology, a process that may involve reading over 500 pages of evidence from manufactures, patient groups, clinicians and others
- considering the technology's benefits and harms using data from trials
- deciding the most plausible cost-effectiveness value, a process that relies on the economic model(s) provided for each appraisal: in effect a detailed thought experiments with multiple variables open to manipulation
- reviewing this in relation to NICE's threshold range
- refining the decision

The threshold itself is not an absolute cut off point. The process of deciding if a technology should be allowed is partly "cortical" and includes calculations, and a consideration of opportunity costs and of NICE's own previous decisions, themselves originally based on a knowledge of what the NHS actually spends. But there is also a "gut" component to decision making, and opportunity cost decisions can be as much influenced by this as by processes of reasoning. The committees' gut feelings also have to take account of the people unrepresented in the decision process: those whose NHS gains will be displaced.

Professor Stevens said that his original hope of being able to give the nod to a host of more effective and cheaper technologies had not been realised. And

the task had become harder as more of the technologies chosen for NICE's consideration - cancer drugs, for example - have come to lie within or above the £20-30K threshold range.

Appraisal committees are more flexible than they're often given credit for, and say yes to technologies that lie a little above the threshold more often than is appreciated. NICE's methods guide suggests factors that committees may wish to take into account. These include the certainty or otherwise of the ICER (committees, he added, try to reward good science), the innovative nature of the treatment, and the extent to which any substantial or unique benefits it offers may not have been adequately captured in the QALY measure.

In practice, decision making involves a great deal of what Professor Stevens described as "pragmatic thresholdry". Factors that might influence are many and various. Cancer tends to be viewed more sympathetically, as are children, or patients nearing the end of their lives or who are particularly disadvantaged by their condition. The severity of the illness too would be a consideration. Treatments for conditions that are time-limited, or affect only small numbers, or are the consequence of causes for which someone other than the patients themselves might be "blamed" also tend to given greater allowance. So too are interventions that clearly represent a great advance on what's already available. Even stakeholder persuasiveness cannot be ruled out as an influence.

The speaker repeated his earlier assertion that committees are not rigid in their decision-taking, and finished with a personal opinion. Leave the threshold where it is, he said.

The first issue raised in discussion was decision making at the PCT level – which is, arguably in some cases, far less rigorous than that within NICE. As Professor Stevens commented, the PCT decision bodies are small and understaffed, and face the added problem that they may even be dealing decisions on named patients – which makes the process more personal and even more arduous. In the latest NICE discussion on research priorities, top of almost everyone's list was a proper analysis of what PCTs actually do, their ICER range, and what in practice they choose to displace. He would advise PCTs to "get it together": to think of themselves as a drug-company board, albeit scattered across 150 subsidiary offices.

A commissioner in the audience pointed out that some PCTs *had* got it together. He then used the example of having to weigh the cost of a new and expensive drug for small numbers of people against, say, the cost of better end-of-life care for a substantially larger number. He added that in his experience, when you debate these life or death issues with the individuals directly affected, they do understand the commissioners' predicament. Someone else pointed out that PCTs simply do not have the information need to make sound decisions – and that if the NHS is soon to face more troubled times, this challenge will become even more pressing. The view was endorsed by another PCT member in the audience.

A representative of some of the patients who get involved with NICE drew attention to the assumption that patient organisations relish the conflict they sometimes have with NICE, and that they'd rather have a fight and get media attention than actually get their input right. Not true. Nonetheless, most patient organisations do not feel that the true impact medical conditions, especially of long-term ones, is properly taken into account. This would include, for example, the real cost to social services of letting somebody go blind. The key, she maintained, lies not so much in the threshold itself as in the assumptions underlying its application. Professor Stevens and Rawlins agreed that quality of life measurements do not necessarily capture everything.

Another delegate argued that one of the important functions of appraisal committees is to represent those who are otherwise neglected: the people who will bear the opportunity costs, but who can't be represented because their identity is unknown. The ethical principle involved is that everybody's health matters, whether you are known or unknown. The threshold is the proxy for this: the proxy for other people's health that could be lost.

Mindful of the "pragmatic thresholdry" included in Professor Stevens's presentation, someone waggishly suggested that it might be possible to devise a treatment that pressed all the buttons, and would therefore find acceptance at almost any price!

How NICE got where it is now – and where it goes next

The next speaker, Martin Buxton, professor of health economics at Brunel University and a member of the first NICE appraisal committee, offered a personal perspective on the recent history of the development of cost-effectiveness assessment. He distinguished four periods: pre-NICE; the "big bang" or birth and early development of NICE; the age of theoretical enlightenment (where we are now); and the future age of empirical underpinning (which we have yet to enter).

The pre-NICE era saw a variety of cost-effectiveness measures being used with no accepted view of how much it was worth spending, and a range of vague, poorly justified, and imperfect answers. ("Evidence suggests [i.e. I think] that society probably places a sufficient value on this effect to make this technology cost-effective" etc.)

With the advent of the "big bang" - the sudden emergence of NICE, and the rapid creation of an appraisal committee charged with making decisions on cost-effectiveness - a feeling began to emerge about what was a reasonable and defensible threshold. But NICE's public statements were "not without ambiguity" and made no reference to any specific financial threshold. In fact by studying what NICE did or did not recommend, health economists were able to infer the existence of a threshold at around the £30K mark. This was clarified in 2004 when Professor Rawlins and Professor Tony Culyer of the University of York wrote an article in the *BMJ* confirming that NICE was using not a single threshold figure, but a £10K range within which the probability of acceptance declines from bottom to top. They quoted £25-35K; the figures

currently in use are £20-30K.

At this point Professor Buxton posed a question: what is the underlying conceptual basis of such figures? He suggested two possibilities: that it could represent a judgement about social valuation; or that it could be a judgement designed to maximise health benefit within a given budget. The first approach carries with it an implication that the health system should undertake any activity that generates a QALY for less than the threshold value. The value itself might be affected by such matters as the national income, the health status of the population, the technological scope to improve it, and citizens' desire for health as opposed to other gains. In the second approach the threshold value is influenced by a politically-determined and fixed budget, requires the health system to disinvest in less cost-effective existing services, and may vary over time.

NICE, of course, has no responsibility to set the overall NHS budget, and is specifically required to recommend a cost-effective use of NHS resources. Approach one, Professor Buxton concluded, is therefore irrelevant to NICE as it currently exists in.

Pursuing his second approach, Professor Buxton then moved to the age of theoretical enlightenment, and wondered if there might be a technocratic solution to establishing a threshold value. In theory, he suggested, the answer is yes. However, to achieve this would require a knowledge of the cost-effectiveness of all interventions, present as well as new. In the real world, he went on, the cost-effectiveness of many technologies is not known and probably never will be. This led him to consider the role of NICE as a "threshold searcher". Accepting that the NHS budget is fixed, the aim must be to optimise the ways it is spent. True optimisation of spending may be unattainable, but the search for a threshold value encompasses the notion that new technologies are only accepted if their cost-effectiveness is greater than that of technologies they displace. This implies a requirement to consider *dis*investment as well as investment possibilities, which in turn depends on a political willingness to stop funding certain procedures. It is, he said, very difficult to stop doing things once you have started doing them. A process of searching also implies that the threshold will be "fuzzy" rather than sharply defined.

Moving finally to the age of empirical underpinning, Professor Buxton suggested that while we may have achieved conceptual clarity about an approach that seeks to maximise the QALYs gained within the NHS budget, this is neither a justification for the figure currently in use, nor a reason to preclude a search for more evidence on the threshold in terms of the opportunity cost of a QALY. In short, the system still needs firmer empirical underpinnings. He finished by reminding his audience that there is no logical reason why a threshold appropriate to the UK should be the same as those chosen by other countries.

In discussion, one delegate expressed surprise at how early on in NICE's existence the threshold appeared to have been established. She added that it

was very difficult to expect PCTs to find uniform thresholds when there was still so much geographical variation in the funds available to them. On the question of local disinvestment, another delegate suggested that the way to get big money savings might be to, say, stop prescribing statins to certain population groups or to change the intervention criteria for cataract surgery rather than to disinvest entirely in certain treatments. He thought the research and information most needed was probably on the point at which treatment should be undertaken rather than on individual treatments themselves.

The discussion then moved to the practicalities of disinvestment, and whether NICE had tried to generate any relevant information. Professor Rawlins remarked that, contrary to widespread belief, there are few ineffective drugs in the British National Formulary. Some are no longer recommended, but none is "completely off the wall". Neither did a review by the Cochrane Collaboration find much in current practice worth abandoning. Passing the ball back to PCTs he said we could save £40 million a year by getting people to give a 10mg rather than a 20mg maintenance dose of omeprazole. There are many opportunities for such actions in the NICE guidelines programme.

Professor Littlejohns added that NICE's optimal practice reviews had identified some 170 recommendations pulled out of its guidelines annually. Compiling the reviews this is not straightforward, he added. When a historical recommendation to stop doing something is identified, it's necessary to go back to the clinicians to see if it's still relevant, and also whether there will actually be savings. It's not as exciting as introducing innovation. But one of the messages of the day should, perhaps, be that this topic needs closer attention. Professor Rawlins added that it might be better to abandon the term "optimal practice review" and go for something more explicit – such as "disinvestment opportunities".

There was also some discussion of the extent to which appraisal committees should be bound by previous decisions. Professor Buxton pointed out that, in an ongoing system, previous decisions have to be tolerably consistent with future ones. If the questioner was asking whether he and his colleagues had got it wrong in the first place, and therefore that all previous decisions have gone off on the wrong course, he could only reply that the figures available at the time seemed the most reasonable they could come up with. He added that he would certainly worry if the thresholds introduced in 1999 and 2000 were completely wrong because that would mean that cost-effectiveness estimates had been not helping but hindering the system for 10 years! On which disturbing thought Professor Rawlins called for a coffee break.

Setting the threshold – but where?

The next speaker, Professor Peter Smith from the Centre for Heath Economics at the University of York, tackled the link between health spending and health outcomes, and their implications for the threshold. He and his colleagues have been using English Programme Budget data from PCTs: a valuable but often neglected source of information offering insights into what we get for the money we invest in the NHS. These data make is possible to link expenditure with indicators on 23 programmes of medical care - infectious

diseases, blood disorders, cancers and tumours, and so on - based on ICD 10. The expenditure is listed by category: pharmaceutical prescribing; inpatient and day case care; ambulatory and hospital care; primary care etc. From this it's possible to see at a glance that the three biggest areas of expenditure per capita are mental health disorders, circulation problems, and cancer. Reassuringly, plots of the data make it clear that more NHS spending does improve health outcomes – though with much variation from place to place.

He offered an example based on considering two programmes of care: for cancer and for circulatory disease. Imagine there is a trade-off between them. What should a PCT choose to do? Should it spend everything on circulatory disease or everything on cancer or divide its resources between the two? Assuming that health gains are all of a piece - a QALY is a QALY not matter how it's saved - the allocation between the two that maximises the total health gain is what that PCT should choose. This can be expressed graphically, or in the form of equations.

The data make it possible to calculate an implicit value of a life year saved, albeit with caveats. Thus for cancer, a one per cent increase in expenditure per head (£0.809) is associated with a 0.355 per cent reduction in lost life years (0.01919 days) – which, when scaled up, implies that one life years costs £15,387. Similarly for circulatory diseases: a one per cent increase in expenditure per head (£1.211) is associated with a 1.244 per cent reduction in life years lost (0.04431 days) – which implies that one year of life costs £9,974. The comparable figure for respiratory problems is £7,397, for gastrointestinal problems £21,538 and for diabetes £26,429. These findings seem to remain quite stable over several years, he said.

In broad terms, Professor Smith concluded, the numbers suggest that NICE is probably not completely out of line in using its current £20-30K per QALY. On the basis of this evidence Professor Smith could see nothing that would argue in favour of an increase in the current NICE threshold.

Local thresholds

The meeting passed straight to a further presentation, this time from Professor Nancy Devlin of the Office of Health Economics. She continued with the theme of local decision making.

She pointed out that there is no mechanism for linking NICE's threshold to the NHS budget. But if the two are supposed to be linked and the threshold was correct in 1999, it cannot be correct in 2009. Likewise if it was correct now, it cannot have been correct in 1999. In a project she and her colleagues were commissioned to undertake by NICE, they were trying to work out what the NICE threshold might be if it were to be derived from decisions made by PCTs and other NHS organisations at the local level. On what basis, they wanted to know, do these organisations make allocative decisions? And is the cost-effectiveness threshold used by NICE consistent with these PCT thresholds?

The work was at one point known as the "Williams Project" – and she quoted

the words of Professor Alan Williams that had prompted this name: "…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 if this belief is well-founded."

NICE's position, she went on, is that its threshold should "represent the national shadow price of a QALY". She then presented a conceptual model based on imagining that complete information was available on the cost-effectiveness of all current and potential NHS services, and that these were presented as a table from most cost-effective to least,. Imagine further that decision-makers worked their way down the table, adopting the services that were most cost-effective. They would eventually reach a point at which the NHS budget was exhausted. This would reveal the location of the threshold.

In the absence of such evidence, she said, NICE's threshold should represent its best guess about the measure of the shadow price. Her and her colleague's research has attempted to find evidence to support or refute that this is so. This was a feasibility study, so they collected data on investment and disinvestment decisions from just six PCTs. They also used questionnaire and discussions and talked to the staff of NHS trusts. They were seeking to fathom the basis for specific changes to a range of local NHS decisions over a set period.

They showed that it was feasible to establish CQG (cost per weighted QALY gained) evidence for at least some of the decisions scrutinised. They found a stark contrast between commissioners and providers: between the considerations driving decision-making by PCTs as opposed to NHS trusts. They found that cost-effectiveness evidence was rarely used in local NHS decision making, and that other factors (such as simple affordability) dominated. PCTs and the NHS may have implicit thresholds different to those of NICE because of local variations in costs and preferences, and the importance they place on non-cost-effectiveness factors. More encouragingly, PCTS did say they would be interested in having more access to evidence on cost-effectiveness, particularly to support decisions on disinvestment.

In summary, a definitive finding on the consistency or otherwise of the threshold figures used by NICE as opposed to other bodies would require the observation of many more decisions, and a more complete understanding of local decision making processes.

Looking beyond the NHS

The last presentation of the morning was by Professor Graham Loomes of the University of East Anglia. He reviewed the use of thresholds in *non*-NHS settings. This, he said, is a relevant consideration because the quality of our lives is affected not only by our health, but by all sorts of other factors from road traffic accidents to air pollution to crime. The various thresholds in use in different sectors don't necessarily have to be identical; but in so far as all are

about aspects of public policy, some degree of coherence is surely desirable.

As in the arena occupied by NICE, thresholds in non-NHS settings can be subjected to a variety of weightings - ranging from people's age and life expectancy through to the degree to which they themselves as opposed to others are responsible for whatever has befallen them. By way of example he quoted some of the values actually used in non-NHS settings. The Department for Transport, for example, puts a value of £1.5 million on preventing a fatality – a figure that works out at about £30-50K per year of life. DEFRA and the HSE both use a threshold figure of around £30K. As with NICE, there are imprecisions in the means by which these figures are reached, and no less reliance on factors such as gut instinct. All express hypothetical public preferences and commonly use 'willingness to pay' methodology and do not take 'opportunity cost' into account. That said, it would appear that the NICE threshold figure is not wildly out of kilter with those used in other areas of public policy making that depend on valuing human life.

Questions, questions...

The discussion of these three presentations, began with a comment on Professor Devlin's assertion that if the NICE threshold had been right 1999 it cannot be right now. Not so, said the questioner; if the size of the budget and inflation have moved in concert with spending on extra interventions in the NHS, the threshold might not have needed to change. Professor Devlin responded that such a coincidence seemed unlikely. Another delegate reminded the audience that changes in medical productivity would also affect matters.

A question (to Professor Smith) concerned the possibility that private sector use might have affected the NHS programme budgeting data he had studied. Professor Smith thought not. He had, he said, tested whether private-sector use made a significant impact on NHS expenditure or on health outcomes, but it did not turn up in either analysis. He suspected that private-sector use was heavily skewed towards a small number of PCTs, and that could be why he and his colleagues hadn't detected it.

Another questioner raised the issue of multiple thresholds and whether this approach should be explored. Professor Smith was not enthusiastic about their use, but felt there might be a case for adjusting thresholds, depending on the submission in question.

Professor Loomes was asked how much confidence he had that all public policy thresholds will eventually be brought into closer concordance. He admitted that people faced with some of the survey questions that underpin them already had difficulty answering, particularly if the differences they were asked to distinguish between were very small. There is also a tendency towards "reporting selection" in which new figures that appear to be very much out of line with existing figures are left to one side. His judgement, he continued, is that the true values would probably be lower than the values coming out of these surveys. He would not be surprised if figures in the region

of £10-£20K were more representative.

Professor Rawlins noted that the Treasury *Green Book* quotes NICE and the £30K figure quite extensively, and wondered if NICE's figure has influenced those who do the analyses. Professor Loomes thought it possible. In response to another question, Professor Smith commented that the NHS in its current form appears, at the margin, to be producing significant life savings at pretty modest costs. If he were in the health ministry, he said, he'd be using some of Professor Loomes's data to suggest that, relative to other ministries, health may in this respect be under funded!

Still thinking about the figures presented by Professor Loomes, another questioner wondered if any of the calculations distinguished between compensation for something lost as opposed to the value of a potential benefit. He was thinking of what applies in medicine where the licensing body will attach a higher negative value to adverse effects than they would a positive value to beneficial effects, because people do not expect to be harmed. Professor Loomes responded that he could not say if this applied in non-NHS fields; the data in these sectors were not sufficiently detailed for these purposes.

Mindful of Professor Devlin's suggestion that PCTs were operating for much of the time in what amounted to an information vacuum, the same questioner also wanted to know how Professor Smith explained these bodies' apparent efficiency. Professor Smith responded that there was no reason to assume that PCTs *are* necessarily being efficient in getting value for money. Back-of-the-envelope calculations in circulatory disease, for example, show the average spending of £9K per life-year. But there were big variations between PCTs with some spending up to £30K for an additional life-year. Others are under-spending. He did not think his data said anything about efficiency. His own view was that there is now quite a lot of healthcare which is good for you and that, by accident or design, PCTs are actually commissioning a lot of it. Professor Devlin repeated that PCTs are interested in making more use of cost-effectiveness data; but it is difficult to imagine them all replicating the sort of analytical process undertaken by NICE. With finance likely to grow tighter over the coming years, she imagined that PCTs would become even more interested in cost-effectiveness evidence, particularly to inform disinvestment decisions.

A question on the feasibility of measuring the outcomes of public health campaigns prompted Professor Smith to comment that a lot of the public-health fraternity still seemed locked into the false belief that virtually anything in the community is likely to be better value for money than conventional healthcare. He said it worried him that we get incredibly low costs per QALY in, for example, advice on smoking cessation. We are told that this is a terrific investment. Is it really true? He thought the public-health people should be doing more experiments to find out what really works. Hard data on this would be very helpful to PCTs.

All three of the panellists were challenged to say what research would be

most helpful in establishing whether or not the current threshed was about right. Professor Loomes said that he'd found the data presented by Professor Smith very reassuring. He'd been struck by the level of spending on mental health. He thought the EQ-5D was pretty hopeless on most things, but especially on the psychological dimension. Getting a better measure of this quality of life aspect would be valuable. He'd also like to see a penetrating account of the weaknesses of econometric analysis, and more studies of disinvestment. This last point was picked up and emphasised by Professor Smith who said he'd like to see a prospective follow through of some major piece NICE guidance to find out how PCTs had responded to it, and what they had decided to forgo. Professor Devlin also backed this suggestion, stressing the importance of trying to understand the true reasons for each decision. It was her suspicion that many were made more on grounds of political expediency than of cost-effectiveness. Another delegate wholeheartedly endorsed her view. A piece of research that took an expensive NICE recommendation and followed how it was put into effect it would be valuable to PCTs in improving local decision-making.

Following this, someone from a PCT suggested to Professor Devlin that NICE and PCTs were currently two worlds in collision. Neither knew what they could or should be doing together. Responding, Professor Devlin wondered if the two worlds were actually reconcilable when the whole point of having PCTs is that they tailor things to the needs of their local services. Is it ever going to be possible to eliminate postcode prescribing where the underlying variations in costs and values are different? One of her colleagues on the study, Professor John Appleby, added that although not all PCTs have a checklist of the criteria they use in making decisions, they are nonetheless pretty explicit about some of them. He added that many of their decisions were really at the margins of their budgets because the bulk of it is already allocated; the money flows in through the back door and out of the front according to a history of earlier decisions. A lot of new decisions were around relatively small proportions of the total budget, and much time was often spent on one or two exceptional cases.

One delegate commented that the saddest thing for him was that we have a public sector able to produce great improvements in health, safety, and the environment, but we do not have the nimbleness to fund significant programmes of work which would give quick answers to questions posed by bodies like NICE or by politicians. Perhaps what's needed is a research programme or research centre that would allow this sort of thing to be done.

Not everyone present was agreed on the right course of action. One economist said that he did not understand why we needed to know precisely what PCTs disinvested in, or why. This information wouldn't help NICE to set the right threshold.

Professor Devlin went on to stress the influence of government imperatives and waiting-time targets on PCT decision making: factors that have nothing to do with QALY-based decisions.

Mr Keith Barron closed the morning session by stressing the importance of public health measures, and the need to find quicker and more compelling method of assessing its impact and value.

Small group discussions

Following the lunch break, Professor Littlejohns introduced the small group work. Attendees were divided into four groups (**A, B, C & D**), each of which were to spend an hour or so discussing the same three questions:

- **(i)** What factors should be considered in determining the need for a threshold review (methodological e.g. linked to inflation, or ethical or political etc)?
- **(ii)** What information is relevant?
- **(iii)** Whose responsibility is it to take these activities forward?

Back in the plenary session, each group appointed a spokesman to feedback its conclusions. They summarised their deliberations as follows.

A (i) The group identified public confidence as a factor, along with the size of the NHS budget and changes in NHS productivity. They felt it would desirable to explore whether having more than one threshold might be beneficial. And they also discussed the frequency with which the threshold might be reviewed.

They added that this whole area is bedevilled by a lack of transparency (in pharmaceutical pricing, for example), so NICE need not feel too embarrassed by its own lack of methodological soundness!

A (ii) There was general support for extending the econometric approach. What was needed was a valuation of expenditure per ICD and also evaluations of the outcomes per ICD, both made while controlling for everything else. This would yield the marginal cost per QALY by ICD. There was less support for extending the "Williams project" approach, and less too for the approach outlined by Professor Devlin (to map actual PCT decisions on what is displaced by NICE). Also discussed, but again with rather less support, was a comprehensive programme-budgeting approach.

A (iii) The group agreed that it should not be the responsibility of politicians. More favoured was the idea of relying on a team assembled by NICE (and led by someone who was scientifically numerate). A second possibility was to choose a complete outsider. Perhaps the Select Committee's motive in suggesting a review by complete outsiders was that NICE had not been thought robust enough to do something that would give it a less easy life! Or at least that having outsiders would help to take some of the heat off NICE. (Professor Rawlins reassured his audience that he himself did not feel under pressure.)

B (i) The group started with the view that the threshold was about displacement, and therefore it was the NHS's responsibility to decide when a review was needed; in other words when the NHS had reason to believe that NICE's recommendations were not consistent with promoting the efficient use

of its resources. The NHS might for example be having great difficulty managing the implementation of new interventions within its available budget.

A second factor might be a change in the external environment, such as a major economic downturn which would be expected to reduce the budget and/or increase the demand for healthcare, so putting pressure on the capacity of the NHS simply to maintain the current provision. This too might lead the NHS to request a review.

A third situation might be when the NHS took the view that NICE's behaviour was not in line with stated policy; ie when a great many interventions above the threshold were being approved. The NHS might want NICE's behaviour changed to be more consistent with stated guidance.

B (ii) The group felt that a better understanding of the values that had been driving decisions, and whether they were appropriate, would be helpful. The definition of health, for example. And true statements of opportunity cost – something the NHS has historically been bad at - would be desirable.

B (iii) The decision should be made independently of NICE - perhaps by representatives of budget holders within the NHS. The NHS Confederation was one suggested candidate.

C (i) Group C felt that three things should be considered. First, money: the effects inflation, changing budgets, changing NHS productivity and the like. Second, whether we are in the "right place"; if the general feeling is that we are, a review is unnecessary. Third, politics. Is NICE being *required* to reconsider the threshold?

C (ii) The most important fact to establish is the unit of opportunity cost, and the ability to say every time a new technology is implemented what is being lost. There was a discussion of the use of differential thresholds, but no agreement among the group. It was pointed out that people have been trying to weight QALYs for years.

C (iii) Of the candidates considered as suitable to undertake a review, NICE itself was most favoured.

(The group also took a vote on whether the current threshold should be changed. Five were in favour of no change, three wanted it to go down, and two others abstained.)

D (i) The group identified NHS budget, price inflation, NHS productivity, and innovation - but projected forward. A disconnect between NICE and PCTs might also be seen as the signal for triggering a review.

D (ii) The view was expressed that a range of empirical work should be undertaken on, for example, programme budgeting - though recognising the importance of getting outcomes besides mortality into the mix. There was also a feeling that it was important for PCTs to become more familiar with what

NICE is trying to do.

D (iii) The group felt that it was the responsibility of the NHS to determine an appropriate threshold, and that this could be done either by PCTs, perhaps contracting it out, or by some collaboration between PCTs and NICE.

Final thoughts

A more general discussion followed the report back session. Someone suggested that maybe right now was the time for a review of the threshold. But in the light of some econometric work now in progress, there was also a case to be made for waiting a year or two.

Another topic that prompted much further discussion was the relationship between NICE and PCTs – one which may be become closer and more effective if plans to set up a new PCT support unit go ahead. This should, if nothing else, facilitate some kind of dialogue. But there were considerable areas of disagreement about the appropriate nature of the relationship, precisely who should be finding out what, and in how much detail. The difficulty of trying to convey what was happening to the general public also prompted some rather despairing reflections.

Drawing the meeting to a close, Professor Rawlins felt that NICE now had to raise the money to fund appropriate research. He added that his personal preference lay with letting NICE itself sort out the issue of threshold review. The hope must be, he added, that in a year or two when he (or his successor) was sitting in front of the Health Committee, it would be able possible to come up better answers than those available at the moment.

Appendix 4

What we were asked to consider
In what circumstances should NICE recommend interventions where the cost per QALY is above the threshold range of £20-30,000?

The conclusions we reached
Two of the 29 Council members attending the meeting took the view that there were no circumstances in which NICE appraisal committees should depart from the established threshold. These two members took no further part in the voting.

Of the remaining 27 Council members, the numbers who favoured taking account of each of a list of various possible circumstances were - in order of support - as follows:

- the treatment in question is life-saving 24
- the illness is a result of NHS negligence 23
- the intervention would prevent more harm in the future 23
- the patients are children 22
- the intervention will have a major impact on the patient's family 22
- the illness under consideration is extremely severe 21
- the intervention will encourage more scientific and technical innovation 21
- the illness is rare 20
- there are no alternative therapies available 19
- the intervention will have a major impact on society at large 16
- the patients concerned are socially disadvantaged 13
- the treatment is life extending 10
- the condition being tackled is time-limited 9
- the illness is a result of corporate negligence 2
- the stakeholders happen to be highly persuasive 0