A NICE Challenge Accepted

Alastair Fischer
Joanne Lord

St George's Hospital Medical School and NICE

Discussion Paper for the Health Economists Study Group,
Nottingham, July 2000

A NICE Challenge Accepted.

Introduction

In a recent editorial in *Health Economics*, Hutton and Maynard raise a number of issues of concern about the work that NICE is expected to do (1). We have summarised their concerns as follows.

1. There is a clear potential for conflict between clinical effectiveness and cost-effectiveness.
2. NICE will often have to make decisions on new technologies on the basis of limited information. Their optimism that cost-efficiency (or otherwise) can be inferred from a combination of cost-efficacy and modelling may be misplaced.
3. NICE might act as a further ("fourth") hurdle for manufacturers, slowing down the diffusion of worthwhile technologies without providing sufficient offsetting benefits in return.
4. NICE might be forced to make decisions on the basis of budgetary impact rather than cost effectiveness.
5. NICE might be unable to prevent "post-code prescribing" from sneaking in the back door after being banished through the front.
6. It was not clear how NICE would carry out economic evaluation…
7. …nor how it would see its guidance put into practice.

Background

NICE has three main functions:

1. appraisal of health care technologies,
2. commissioning clinical practice guidelines
3. supporting clinical audit.

In this paper, we concentrate on the first function (appraisal). In practice, the borderline between the appraisal of a single technology and the development of guidelines for a whole disease area is often blurred. Guidelines are generally broader than technology appraisals, covering a range of technologies and decision strategies for a defined patient group. For some appraisals it may be possible to compare the technology against a small number of alternatives, although often, as we mention in the next section, it may be necessary to consider broader issues.

A description of NICE's appraisal process may be found on its website www.nice.org.uk, under News; Press Releases for 1999; First Work Programme. Briefly, it works as follows:

The list of technologies to be appraised by NICE is set by the Department of Health and the National Assembly for Wales. Once the programme has been

agreed, NICE invites interested parties to submit evidence about the effectiveness and cost. This includes the manufacturers of the technology under discussion and all identified national patient and professional groups. NICE operates in close co-operation with the National Coordinating Centre for Health Technology Assessment, who commission a 'rapid' systematic review of the scientific evidence from one of a number of academic groups. This review also includes material from the company submissions.

All the evidence, including the HTA Report, evidence from the professional bodies and the patient groups, the company submissions and a NICE-secretariat overview is sent to the members of the Appraisal Committee. This Committee currently consists of 23 people, 15 of whom are senior academics in medical science, including one statistician and three health economists. Seven of the 23 are involved in primary care (two GPs, three in Primary Care Departments, one in Nursing, one in Pharmacy), three are NHS managers and two represent patient groups. (Details are on the NICE website.) The Committee also receives short written overviews from invited senior clinicians ("experts"), usually recommended by a professional body. These experts also answer questions put to them by the Committee. At the discretion of the Committee, patient representatives may also be invited to attend.

The Committee discusses the question. It considers the effectiveness of the technology, its cost-effectiveness, and other pertinent issues. The initial conclusions and recommendations of the Committee are documented in the form of the PAD (Provisional Assessment Determination). The PAD is sent out for a period of consultation. The consultees include all of those parties who are initially invited to make submissions (industry, patient and professional groups). The Committee revisits the topic, to consider consultees' comments and to make their final decision, two months after their initial meeting. Their conclusions and recommendations to the Institute are documented in the Final Assessment Determination (FAD). On the basis of the FAD, the Institute prepares its draft guidance document. This is sent out to the consultees, who have a period of ten days within which they can lodge an appeal. If there are no appeals the guidance is issued to the NHS and made public. If an appeal is lodged, the launch of the guidance is delayed.

The whole process takes almost a year, much of which involves refining the question, and gathering, sorting and analysis of scientific evidence.

Discussion

Let us now examine Hutton and Maynard's points.

There is a clear potential for conflict between clinical effectiveness and cost-effectiveness.

The Appraisals Committee has to steer a difficult path between these two concepts. This task is not made easier by the ongoing controversies about

methods of economic evaluation. These include continuing uncertainties around the concept and measurement of the QALY (2). In the appraisals conducted so far, the Committee has been presented QALY estimates, where sensible estimates are available. However, they have also been presented with information on the effects of the technology and the comparators on a range of other relevant clinical indicators and outcome measures. Thus, the method of evaluation employed by the Committee may more properly be called cost-consequence analysis, rather than a straightforward cost-utility analysis.

Even more difficult, however, is the need to consider equity on top of this. Such questions should be informed by both effectiveness and cost-effectiveness considerations, but these are only necessary but not sufficient conditions for making a decision. As we all know, there are no theorems to allow us to say where, as economists, we should stand on this efficiency versus equity plane.

A further complication arises because of the difficulty in defining the boundaries of each appraisal. If the boundary is drawn too closely, evidence of effectiveness and cost-effectiveness may still not be sufficient to reach appropriate conclusions. However, if too broad, the appraisal may become unmanageable – a 'guideline' rather than an 'appraisal'.

NICE will often have to make decisions on new technology on the basis of limited information. Their optimism that cost-efficiency (or otherwise) can be inferred from a combination of cost-efficacy and modelling may be misplaced.

NICE might act as a further ("fourth") hurdle for manufacturers, slowing down the diffusion of worthwhile technologies without providing sufficient offsetting benefits in return.

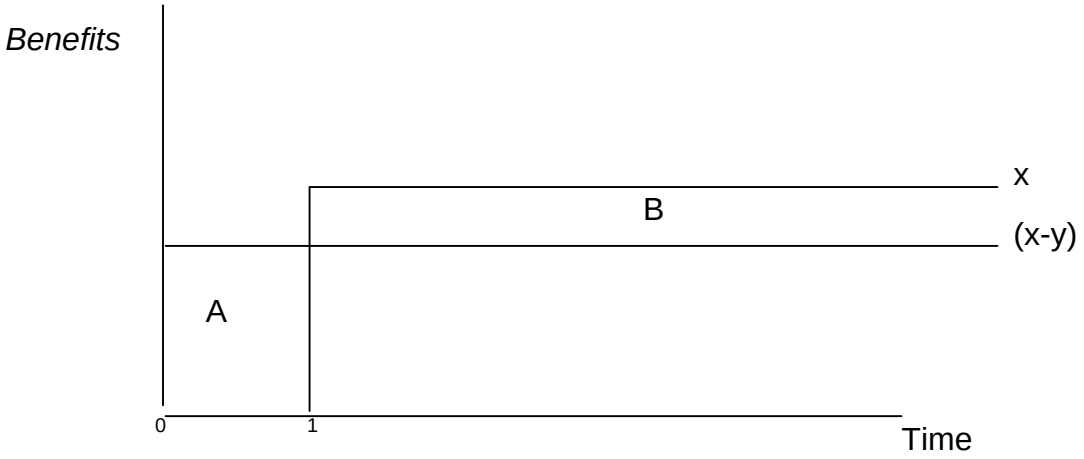
These two questions are related. *Any* body that tries to examine whether a new technology is worthwhile will often have very limited information to inform themselves. There is a trade-off between, on the one hand, waiting to get better information and, on the other hand, not waiting, but having to make decisions with poorer information. So if NICE acts quickly, but has to use modelling to breach the information gap, it will be criticised (as in the first quote above). But if NICE waits, it may have the effect of slowing the diffusion (as in the second quote). When did you stop beating your wife, Mr NICE?

Even so, the second quote needs addressing more closely, as there is another side to it. The quote is consistent with the following model.

Strategy 1: There is no NICE. Of all new technologies, which are introduced at time zero and diffuse to all players immediately, there are those that give benefits of x per year and others that give negative benefits of absolute value y each year. Thus, introduction of the new technologies gives a net benefit of $(x - y)$ each year forever.

Strategy 2: NICE exists. It takes a year to make a decision on all new techniques. It has a success rate of correctly predicting beneficial technologies of 100%. Thus, after benefits of zero in year 1, the ongoing benefits are *x* each year forever.

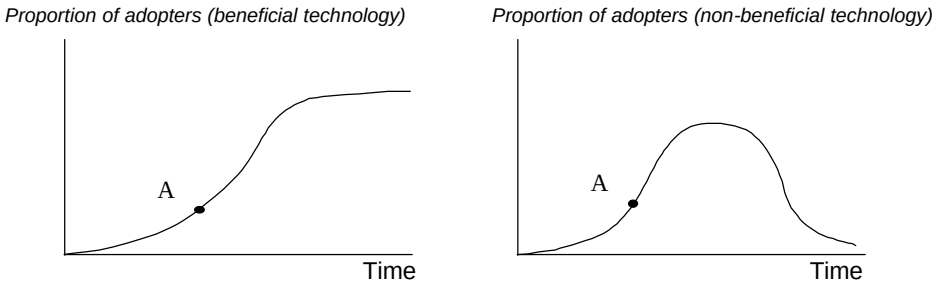
These things are shown in the diagram below.



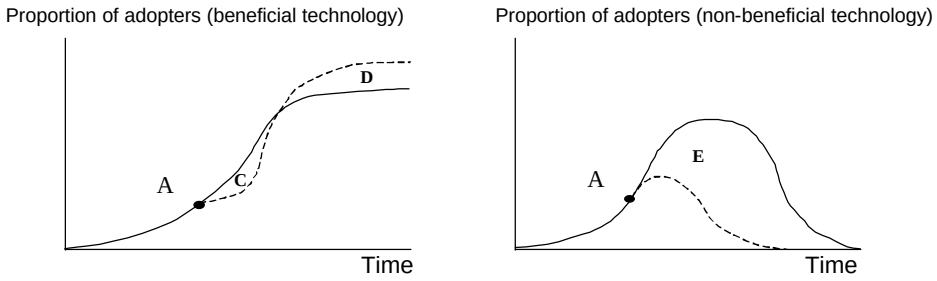
Without NICE, strategy 1 includes area A but excludes area B. With NICE, strategy 2 includes area B but excludes area A. When discounted at any positive rate, area B becomes finite and the net benefit from NICE involvement is given by (B – A).

However, a more realistic model is to recognise that diffusion is generally well under way when NICE begins to look at a technology. The question then is whether NICE speeds up that diffusion process or slows it down. The following model is probably more appropriate.

Strategy 1: Without NICE, the cumulative diffusion curves of new technologies are as given below. The first diagram shows technologies that are eventually shown to be beneficial, and the second, those that are not eventually beneficial.



Strategy 2: With NICE, acting at point A, the diffusion may be slowed a little at first, but if NICE is accurate at picking winners, the adoption rate will then pick up strongly in the case of beneficial technologies, but wither rapidly, or go to zero, in those that that are not beneficial.



If we assume that the savings are proportional to the proportion of adopters, then there may be savings for the beneficial technologies (when discounted D exceeds discounted C) as well as for the non-beneficial technologies (the large saving of discounted E). This compares with the previous model, in which there were no savings of D, so that net benefits occurred only when E exceeded C.

The caveat is that NICE may be risk-averse, and hold up technologies that it is unsure about. If the uncertainty is with respect to clinical effectiveness, NICE has every right to be conservative, and no-one is likely to quarrel with that conclusion. No-one wants to operate at p values greater than 0.05 lest another thalidomide situation emerges. However, on cost-effectiveness, it is another matter. The well-being of different patients cannot be substituted easily, but different pots of money can.

Let us explain. Suppose that there is a technology with a 60% chance of benefits of £1 million and a 40% chance of benefits of £(-1) million. The expected benefits are £0.2 million, but there is only a 60%, and not a 95%, chance of a positive result. So if a p value of failure of 0.05 is applied to the cost-effectiveness of the technology, the technology would be rejected. However, suppose that there are 100 such projects, all independent. Choosing all of them will (on average) give benefits of £60 million offset by disbenefits of £40 million. The nation would be better off (on average) by £20 million. Due to the law of large numbers, there is only about a 2 to 3% chance that the true net benefits would be less than zero. So on cost-effectiveness, over a large number of projects, NICE does not need to be so concerned about the uncertainty of its cost-effectiveness estimates as it does about those of clinical effectiveness.

Certainly, the *means* for the estimated net benefits for each such technology have to be positive. This is essentially a simplification of Claxton's position (3).

Though the government might reasonably ignore financial uncertainty, it should not ignore clinical uncertainty. With public investments the financial risks are distributed between many taxpayers, so the risk to each individual is negligible (4). Thus the government should behave as an expected-value decision maker, ignoring uncertainty in the assessment of financial risks. However, when risks associated with public investment decisions are borne by a limited number of private individuals, as with the adverse events of medical interventions, it is appropriate for the government to discount for risk as would these individuals (4).

We recognise that there may still be substantial publication bias in the estimates of both clinical and cost effectiveness, but that is another story entirely, which impinges on all of medical decision making, and is not specific in any way to NICE.

So the conclusion we reach in this section is a very strong one. We believe that there are good grounds for thinking that far from being a fourth hurdle, NICE could well help to speed up innovation that in the long run will prove beneficial. It will, of course, be a very useful fourth, fifth and sixth hurdle for technologies that are of a net disbenefit to society. In the long run, therefore, the existence of NICE should be in the interests of medical manufacturers as well as medical consumers. Products that are not of sufficient use to society will be replaced by products that are. Aggregate sales of manufacturers in long-run steady state should not decrease: if anything, they should rise marginally due to a healthier (and thus richer) population and greater trust in the products being supplied. (And as Hutton and Maynard also say, marketing costs should also reduce in this scenario.)

NICE might be forced to make decisions on the basis of budgetary impact rather than cost effectiveness.

So far, there is no evidence that we are aware of that this has happened. Of course, that is not to say that it might not happen downstream. NICE is required to tell the Government how much it thinks its recommendations will cost. So it should. But that is quite a different matter from cost-effectiveness, and is recognised to be so.

NICE might be unable to prevent "post-code prescribing" from sneaking in the back door after being banished through the front.

This is possible, though not inevitable. Here is the simplest possible model. There are two Health Authorities, A and B, and two technologies, I and II. Both

are subject to postcode prescribing. Authority A spends 1 on I and 0 on II; B spends 0 on I and 1 on II.

Health authority	Technology I	Technology II
A	1	0
B	0	1

Now suppose NICE “requires” all health authorities to adopt technology I. Then B will have to spend 1 on I and without an increase in budget, will have to spend 0 on II

Health authority	Technology I	Technology II
A	1	0
B	1	0

As a result, postcode prescribing has been banished in *both* technologies!

Although this is technically a counterexample that shows that Hutton and Maynard’s point is not correct in all situations, the argument loses some force as more technologies are added. Suppose we add in a third technology that both A and B spend one unit on. We thus start with an initial budget of 2 units, the same as the starting point for the two-technology case, but with Technology III added at the end.

Health authority	Technology I	Technology II	Technology III
A	1	0	1
B	0	1	1

If, as before, following NICE Guidance, B spends 1 on Technology I, then either B will have to spend 0 on Technology II (as before) or 0 on Technology III. In the first of these cases, postcode prescribing ends for all technologies, with Technology II being funded nowhere. In the second case, postcode prescribing transfers from Technology I to Technology III. (Keen observers will note that the theory described is isomorphous with the theory that says that in a 2-good, 2-consumer world, it is not possible for the goods to be complements.)

The point is that at the end of the day, the *net* effect will be less postcode prescribing overall than there was before it was abolished in the technology under discussion. Abolishing postcode prescribing is not like squeezing a balloon, only to find out it has re-emerged elsewhere, but more like smoothing a field by shovelling off the top of hillocks and distributing the earth elsewhere. On average, the field will be smoother, essentially by entropy, after the process than before.

There are also other ways of reducing postcode inequity. A more efficient way would be to tackle it directly, by ensuring that waiting lists are essentially the same over the whole country. (Postcode prescribing with a 0 or 1 treatment is simply a polar case of a variable length waiting list for each region.) For some treatments, this could be achieved by sending in "flying squads" of consultants, for others, by asking patients to go to another region for treatment. But this would re-centralise medical services, and downgrade the autonomy of the Health Authorities. (Perhaps this is what should be done, particularly if it is thought that the autonomy of the Health Authorities is not derived from a process of regional democratic election, other than perhaps in Scotland, Wales and Northern Ireland. But then again, perhaps not. Either way, it is beyond the remit of this paper to do more than flag this as an issue.)

It was not clear how NICE would carry out economic evaluation…..
……nor how it would see its guidance put into practice.

So far, the Appraisal Committee of NICE has used cost per QALY, alongside evidence on a relevant range of clinical indicators, to inform itself of cost effectiveness, but it has not been slavish about drawing a line in the sand, for all the usual reasons. First, it recognises that there is discussion about whether all QALYs are equal. Second, that there are measurement errors. Third, that there are other dimensions to the evaluation of outcome in health care. And of course, there are equity considerations that transcend all of the other discussion points about QALYs. It is therefore more that there is an area of clear cost-effectiveness, of a low cost per QALY, gradually eliding through a long "twilight zone" where cost-effectiveness is not clear, to an area of clear "cost-ineffectiveness", of very high cost per QALY. No-one has said, and we think no-one quite knows, exactly where these boundaries are placed. Just when does daylight become dusk?

Until we have firmer evidence of societal willingness-to-pay, and until we can be surer that estimated QALY gains adequately represent social preferences, it is not appropriate for NICE to be tied to a mechanistic decision-making rule of making cost per QALY equal at the margin.

On putting guidance into practice, NICE suggests audit strategies as part of its guidance. Further, a national R & D project will look at the effects of NICE guidance in the NHS as a whole. But it is early days yet, and that is still to occur.

In addition to the above issues, there are many other methodological and practical challenges facing NICE if it is to contribute to better clinical decision making. For example, Freemantle and Mason have inquired what level and quality of evidence NICE should require (2).

NICE does not always have the luxury of taking the pure, gold-standard approach of looking at RCTs alone. Where the evidence is of sufficient quantity and quality, of course, it would be remiss not to rely substantially, and perhaps entirely, on it. But what if there is little or no such evidence available?

Endpiece

This paper is clearly for discussion. In some cases, it proposes alternatives to those expressed in Hutton and Maynard's editorial, but it does not claim any definitive answers. The paper is also mostly concerned with the Big Questions. When it comes to detail, NICE is in the process of developing more detailed guidance on how a firm, a professional body or a patient group can respond to a NICE invitation to submit appraisal evidence.

The views expressed in this paper are those of the individuals concerned, and not official views of NICE. Apart from working for NICE, neither of the authors has any known conflict of interest. We believe strongly in full disclosure of interests for all those involved in debates such as this.

References

1. Hutton J and Maynard A. A nice challenge for health economics. *Health Economics* March 2000; 9(2): 89-93.
2. Freemantle N and Mason J. Not playing with a full DEC: why development and evaluation committee methods for appraising new drugs may be inadequate. *BMJ* 1999; 318: 1480-1482.
3. Claxton, K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 1999; 18: 341-64.
4. Arrow KJ, Lind RC. Uncertainty and the evaluation of public investment decisions. *American Economic Review.* June 1970; 60(3): 364-78.