The role of NICE technology appraisal in NHS rationing

Simon Walker*, Stephen Palmer, and Mark Sculpher
Centre for Health Economics, University of York, Heslington YO10 5DD, UK

Objective: This article examines the role of National Institute of Health and Clinical Excellence (NICE) technology appraisal in detail, focussing on the process itself and the methods used to establish cost-effective practices for the National Health Service (NHS).

Areas of agreement: Approaches to identifying both effective and cost-effective practices have become central to rationing decisions in the NHS. The establishment of the NICE, which produces guidance on what treatments should be provided by the NHS, represents the most visible approach to introducing economic considerations into these decisions.

Areas of controversy: The decisions over which activities will be displaced by NICE approved treatments are made at a local level, while the cost-effectiveness threshold used to evaluate technologies is set nationally. This may result in treatments being displaced which are more cost-effective than those being introduced.

Areas to develop research: The introduction of programmes looking at disinvestment opportunities to help aid local decision makers is a key step in improving the allocation of NHS resources and removing geographical inequalities.

Keywords: NICE/cost-effectiveness/rationing/economic evaluation/technology appraisal/priority setting

NICE’s origins

History of NHS decision-making

The National Health Service (NHS) was established with the aim of providing healthcare to all citizens, based on need, not the ability to pay. This has inevitably created tensions between increasing demands and limited funding. Since the early 1990s evidence-based healthcare has become a cornerstone in policies to identify effective, and in particular cost-effective, practices and to move away from decisions based on opinion or current practice to a greater use of
scientific research and evidence. The latest step came with the introduction of the National Institute of Health and Clinical Excellence (NICE) in 1999.

NICE was initially established in England and Wales to help the NHS meet three continuing objectives: (i) to improve continually the overall standards of care; (ii) to reduce unacceptable variation in clinical practice; and (iii) to ensure the best use of resources so that patients receive the greatest benefit. NICE has a wide remit involving several programmes of work including clinical guidelines, technology appraisal, public health interventions and implementation. All these programmes include economic analysis to some degree, however, it is the technology appraisal process for which it is most central. The focus of this article will be NICE’s technology appraisal activities.

Basic economic problem and the tragedy of the commons

Looking at why resource allocation problems arise leads us to the basic economic problem: resources are scarce but competing demands on them are infinite; and more specifically to the tragedy of the commons: ‘The NHS is a common resource. A patient acts rationally in seeking an expensive treatment that produces a benefit (even if small) because the cost falls almost entirely on others’. The same can be said of doctors who seek the best for their patients as the cost, in terms of fewer resources for other patients, falls almost entirely on other doctors’ patients. To ensure efficient resource allocation in healthcare, it is required that the health benefits of an intervention are greater than their opportunity cost, where the latter are the health benefits associated with interventions that are ‘squeezed out’ when new interventions that impose additional costs on the system are funded.

NICE has attempted to tackle these issues by publishing guidance about whether pharmaceuticals and other technologies should be provided by the NHS. NICE provides guidance to the NHS in England and Wales when requested to do so by the Department of Health (DoH) and the Welsh Assembly. The guidance is based on a technology appraisal, undertaken by the Institute, which synthesizes evidence on the clinical and cost-effectiveness of a technology in the context of its use in the NHS. The guidance indicates whether a particular technology, based on the balance of the current evidence, should be recommended as the cost-effective use of NHS resources. The process by which the Institute establishes this guidance is briefly summarized below.
The technology appraisal process

The process can be divided into three distinct phases: (i) scoping; (ii) assessment; and (iii) appraisal. Each of these is described below under the context of NICE’s multiple technology appraisal (MTA) process, which has been used for most appraisals. The overall process is underpinned by a number of key features including independence in the evaluation of the current evidence, transparency in the way the appraisals are conducted and inclusiveness such that stakeholders play an important role in the process.

Scoping

The scoping process aims to provide a framework for the appraisal by determining the specific questions to be addressed. Specifically, these include the technology (one or more) of interest, the patient population(s) for which they would be used and the relevant intervention(s) against which the technology will be compared. This process sets the boundaries for the assessment and the questions that should be addressed during the assessment and appraisal phases. The scope is revised in response to comments received from consultees (national groups representing patients and carers, bodies representing health professionals and manufacturers of the technology under review) and commentators (manufacturers of the products with which the technology is being compared, NHS Quality Improvement Scotland and research groups working in the area).

Assessment

The assessment phase comprises a systematic and independent evaluation of the evidence available on the technologies by an Assessment Group (an academic group commissioned by the NHS Health Technology Assessment Programme). The aim is to produce an independent and unbiased assessment of the clinical and cost-effectiveness of a technology following the framework established in the scope. The process can be split into two mutually dependent components: (i) a systematic review of the clinical and economic evidence which includes submissions made by the technology manufacturers; and (ii) an economic evaluation which attempts to synthesize this evidence in the context of the NHS. The methods of economic evaluation will be discussed later.
Appraisal

The appraisal process involves the Appraisal Committee (comprising a multi-disciplinary group of independent experts including healthcare professionals, patient representatives and academics) considering the outputs of the assessment phase within the context of additional information provided by consultees, commentators, clinical specialists and patient experts. After considering the evidence, the Appraisal Committee formulates an appraisal decision based on a range of factors including cost-effectiveness, strength of the clinical evidence, robustness of the economic evaluation and the degree of clinical need of the patients. The Committee’s decision is summarized in an Appraisal Consultation Document (ACD), which details the Institute’s preliminary views on the evidence for the technology of interest. Following circulation of the ACD, the formal stakeholders (patient/carer organizations, healthcare professional bodies and manufacturers) have one month to comment. After the consultation period, the Appraisal Committee meets again to consider the evidence alongside the comments received on the ACD. They then produce a Final Appraisal Determination (FAD) that is sent to stakeholders to consider whether they wish to appeal. If there are no appeals the FAD forms the basis of the NICE guidance to the NHS.

Appeals

All the nationally-based organizations involved in the appraisal process also have access to formal appeals on the grounds of process (due process), perversity (if the decision is considered perverse given the evidence) or powers (if NICE exceeds its power). Appeals are heard by a panel consisting of non-executive members of NICE and industry and patient representatives.

Methodological features

The overall purpose of the NICE appraisal process is to make decisions, which support an efficient use of NHS resources—that is, the maximization of population health from available resources. This has a series of methodological implications for the appraisal processes and NICE’s position on these is defined in a ‘reference case’ that specifies the methods which are considered most appropriate for the purpose of assessment. The key features of the reference case are discussed below.
Cost perspective

The costs to be considered for assessing the cost-effectiveness of an intervention should relate to resources that are under the control of the NHS and Personal Social Services (PSS). This involves quantifying the effect of the intervention on resource use in terms of physical units (e.g., days in hospital, number of visits to GP) and valuing those units in monetary terms using appropriate prices and unit costs.

Quality-adjusted life-years (QALYs)

The reference case states that the measure of health benefits that should be used is the QALY. QALYs are a generic (non-disease-specific) measure of health outcome, which simultaneously capture morbidity [health-related quality-of-life (HRQoL) gains] and mortality (survival duration gains) and combines the two into a single measure. The need for a generic measure reflects NICE’s remit to make consistent decisions across different technologies in a range of clinical areas.

To estimate QALYs, two types of data are required. Life expectancy is usually taken from the available clinical studies, although these may only report mortality or survival rates requiring assumptions if these are to be translated into estimates of life expectancy. The more difficult data required for QALY estimation are the weights (often referred to as utilities, preferences or values) to quality-adjust the life-years. Various methods exist to elicit these weights which, in terms of sophistication, range from the use of plausible ad hoc values posited by the authors of studies, to the use of preference-based instruments to elicit from individuals the weights that they would attach to particular states of health. The NICE reference case requires that patients’ HRQoL should be captured with a standardized and validated generic instrument, and that weights should be based on public preferences (from the UK population).

Cost-effectiveness analysis

Cost-effectiveness analysis (CEA) is a form of economic evaluation where both the costs and consequences of an intervention are considered simultaneously against other relevant comparators (e.g., best-alternative care). The comparative nature of these evaluations is key since it is not possible to establish cost-effectiveness without formal comparison against other ways of using these resources. As discussed previously, NICE specifies the QALY as its preferred measure of health benefits and resource use which is under the control of the NHS and...
PSS to measure costs for CEA. Below is a brief overview of how the results from a CEA are used to establish cost-effectiveness in the context of NICE technology appraisals.

Let A and B represent two alternative treatments. If intervention A is less costly and more effective, it is said to ‘dominate’ B. Similarly, if A was more costly and less effective, it would be dominated by B. Under either of these conditions, NICE would feel safe in concluding that the dominant option was the more cost-effective. In practice it is rare that the cost and outcomes lend themselves to the dominance rule, and it is usually the case that an intervention is more effective, but also more costly. The critical issue here is whether the additional (incremental) cost is worth paying for the incremental benefits. The decision rules developed to address this issue focus on the incremental cost-effectiveness ratio (ICER), which is defined as

\[
\text{ICER}_{AB} = \frac{\text{Costs}_A - \text{Costs}_B}{\text{QALYs}_A - \text{QALYs}_B}
\]

At this point the decision about whether an intervention is considered cost-effective hinges on the threshold ICER considered by the institute to represent an efficient use of NHS resources, such that treatment A will only be considered cost-effective if the ICER is lower than this threshold. In principle, the threshold value should be determined by the displaced intervention(s), representing the health benefits forgone by implementing treatment A (which could be in any clinical area). However, since NICE does not know which interventions will be displaced, as decisions on this are made at a local level, it uses an approximation with a threshold of around £20,000–£30,000 per QALY. The current threshold used by NICE has been the subject of intense scrutiny and a more detailed discussion of the threshold and methods for identifying the threshold can be found in Culyer et al. (2007).10

Systematic reviews

There are vast amounts of information available on treatments, to such an extent that the quantity can become unmanageable.11 For decision makers to make rational decisions, they should be able to take account of all the evidence that is relevant to the decision problem, in order to avoid the potential for bias that could be caused by selective consideration of the evidence base. Systematic reviewing allows the efficient integration of the evidence so that it can be used for decision-making. The techniques and benefits of systematic reviewing11 are well established and play an important role in the NICE technology appraisal process.5 The
importance of the process of assembling and synthesizing the evidence with regards to modelling is discussed further in the section below.

**Modelling and evidence synthesis**

While the randomized controlled trial (RCT) has developed an important role in the clinical evaluation of healthcare interventions, its potential limitations for decisions concerning the efficient use of NHS resources are widely recognized. Factors such as partial comparisons (where an RCT does not compare a treatment to all the comparators being considered by NICE) and short time horizon (where an RCT does not have sufficiently long period of follow-up to account for all treatment effects and resource use differences between trial arms which are relevant to CEA) often limit the value of RCTs as the sole basis of data for informing NICE guidance.

For NICE to make appraisal decisions it is necessary for clinical and cost-effectiveness to be considered over an appropriate time frame (i.e. the time over which costs and benefits may differ between treatments), to be relevant to UK patients and to compare all relevant treatment options for the patient population of interest. Since this information is unlikely to be available from a single source, additional methods are required to synthesize data from several RCT(s) and other sources (e.g. observational studies), as well as explicit assumptions, to provide a complete picture of clinical and cost-effectiveness. Decision analytic models offer a framework within which all relevant evidence can be synthesized and estimates of clinical and cost-effectiveness generated. Modelling is likely to be required when: (i) trial participants do not match the patient population of interest; (ii) intermediate outcome measures are used rather than the effect on HRQoL; (iii) data on all relevant comparators are not available from one trial; and (iv) costs and benefits of treatments are expected to differ beyond the trial follow-up. The process of evidence synthesis and decision analytic modelling is now seen as central to the process of HTA in general, and it plays a key role in the NICE appraisal process.

Given the range of data sources required for these models, the process of assembling evidence for the assessment phase needs to be systematic (i.e. evidence must be identified, quality assessed and where appropriate pooled using justifiable methods). These principles apply to all categories of evidence required to estimate clinical and cost-effectiveness, as they will all be typically drawn from a number of different sources. In addition, the process of combining data from disparate sources (and of differing quality) will inevitably lead to uncertainty surrounding the estimates applied to particular parameters.
Techniques have been developed (referred to as probabilistic sensitivity analyses) which allow consideration of the impact of uncertainty in these parameters to be quantified in terms of the degree of uncertainty surrounding NICE decisions (i.e. the probability that a given intervention, from among those being compared, is the most cost-effective). More importantly, the potential consequences of decision uncertainty can also be formally considered and methods have been developed to establish the value of obtaining additional information (e.g. by commissioning further research) aimed at reducing the current level of decision uncertainty.13

Review of NICE’s work

Raftery (2006)14 reviewed the guidance issued by NICE to the NHS between 1999 and April 2005. In total, NICE issued 86 guidance documents covering 117 technology or patient topics over this period. The guidance was categorized into four groups reflecting the decisions reached by NICE: (i) yes; (ii) yes with major restrictions; (iii) yes with minor restrictions; and (iv) no. The restrictions for pharmaceuticals were considered relative to their existing license and for non-pharmaceuticals relative to the size of the potential patient population. As summarized in Figure 1, NICE recommendations were no for 22 (19%) topics, yes for 27 (23%), yes with major restrictions for 38 (32%) and yes with minor restrictions for 30 (26%).

Raftery found that, of the negative recommendations, two-thirds were on the grounds of insufficient evidence and the rest were as a result of unacceptable cost-effectiveness. Where treatments were recommended with major restrictions it was generally based on evidence...
suggesting that while a technology was not cost-effective for all patients, particular subgroups could be identified that had cost-effectiveness estimates below the cost-effectiveness threshold. For those recommended with minor restrictions, they usually specified good clinical practice (e.g. use by specialist), but sometimes recommended the use of the lowest cost intervention (on the grounds that the effectiveness of the technologies were identical and hence the cheapest technology should be used first).

Dakin et al (2006)\textsuperscript{15} used a multinomial model to examine more formally the factors which the Appraisal Committee appear to take into account when making decisions. The model was based on three possible decisions: (i) recommend for routine use; (ii) recommend for restricted use; and (iii) not recommended. Their results suggested that interventions supported by more randomized trial evidence are more likely to be accepted (reflecting a hierarchy of evidence and a preference for clinical effectiveness data to be derived from RCTs). Higher ICERs increased the likelihood of rejection relative to restricted use but interestingly were not significant in the decision between routine and restricted use. Pharmaceuticals, interventions appraised early in NICE’s existence and those with more systematic reviews were also less likely to be rejected. The inclusion of a patient group submission as evidence also appeared to make a recommendation for routine use as opposed to restricted use more likely. A similar study was conducted by Devlin and Parkin (2004)\textsuperscript{16} based on a binomial model, where a treatment was either accepted or rejected. They found that in addition to cost-effectiveness, uncertainty around the ICER and the burden of disease also appeared to contribute to NICE’s decisions.

**Issues/future challenges for NICE**

**Methodological issues**

**Decision-making and thresholds**

Although CEA has become an established part of the resource allocation process in the UK and other countries, largely through pharmaceutical reimbursement mechanisms,\textsuperscript{17} it still has its detractors.\textsuperscript{18} The majority of these criticisms stem from the decision rules of CEA and on the threshold used to establish cost-effectiveness. In particular, it has been argued that the CEA does not consider the health gains forgone by reallocating resources from existing programmes to fund new programmes when a fixed threshold value is used.\textsuperscript{18} This criticism is particularly relevant to NICE, as decisions on how to fund NICE’s recommendations through disinvestments in other treatments are left to
the local NHS trusts. This raises the issue that the NICE recommendations may not necessarily increase the overall efficiency of resource allocation decisions in the NHS, since it is possible that more efficient interventions (i.e. those generating more QALYs per pound spent) could be displaced in favour of the new technologies approved by NICE whose provision is mandatory. This is certainly a key concern and it also has equity implications as it is possible that the most vulnerable groups (e.g. those with mental health problems) are those for whom interventions may be displaced in favour of less-efficient interventions which are supported by a stronger lobby. However, NICE is now considering the issues of disinvestments in existing services both through its technology assessment programme and its public health programme.19,20

The decentralization of resource decisions also leads to concerns for NICE as it may result in an inability for it to achieve one of its key goals: to reduce unacceptable variation in clinical practice.21 As discussed above, the NHS trusts make their own decisions on what to disinvest in and, while appraisals might ensure mandatory provision of approved technologies, it does nothing to ensure that technologies which have not been the subject of a NICE appraisal are provided equally geographically.

Process issues

Future research—the role of value of information for NICE
As described earlier there are methods that can be used as part of a CEA to evaluate the level and consequences associated with uncertainty surrounding NICE recommendations. These methods also enable a formal consideration of the value of future research through the value of information analysis. Within NICE’s methods guidance it suggests that it is helpful to present the contribution of uncertainty in parameters to decision uncertainty through value of information analysis. However, currently it is not in NICE’s remit to be able to do anything with such information, other than to use it to inform the current appraisal decision. This could actually lead to inefficiencies. For example, a new drug may appear cost-effective; however, there is uncertainty surrounding the decision (i.e. there is a probability that the drug may not actually be cost-effective). If NICE recommends the new technology, the manufacturer may no longer have an incentive to carry out further research to reduce the uncertainty (in fact it has very strong incentives not to do so). However, if NICE rejects, it is rejecting a drug which is currently expected to increase efficiency, so as to force future research. Clearly, both decisions have a cost associated with them.
However, what if NICE were able to say ‘yes’ but also to make this conditional on the conduction of future research? Medicare uses a similar system to this known as ‘Coverage with Evidence Development’ whereby companies seeking Medicare reimbursement for treatments are required to enrol in research as a condition for coverage.\(^\text{22}\) Such a system appears to solve the controversy as the decision would be better informed in the future but also expected benefits are not foregone today to ensure this.

### The single technology appraisal (STA) process

Clearly, one of the main aims in establishing NICE was to facilitate quicker access to cost-effective treatments. However, in 2005 the Health Committee of the House of Commons observed that NICE ‘acts too slowly’.\(^\text{23}\) In response, NICE and the DoH announced in November 2005 that NICE was launching a new, rapid process for assessing treatments which would co-exist with the MTA process. The STA process was to be used initially to produce faster guidance on life-saving drugs. This was to be achieved by asking for a single submission of evidence by the manufacturer, which would then be independently assessed by an Assessment Group. It was anticipated that this would result in guidance being published 6–15 months earlier than under the MTA process.

Assuming all other things remain the same, the earlier guidance can be released the better. It will increase the benefits to patients by allowing early uptake and should also prevent the spread of inefficient treatments. However, clearly not all things are equal and early guidance will most likely be based on a smaller evidence base making it potentially less reliable than guidance at a later date when evidence becomes more mature. There is also the concern that the shifting of the ‘onus of proof’ onto the manufacturer may result in lower quality and biased analyses. Miners et al. (2005)\(^\text{24}\) conducted a retrospective study of evidence submitted to NICE by the manufacturers of the relevant healthcare technologies and by the university-based assessment groups with the aim of assessing the results from economic evaluations from different types of organization. They found that the manufacturer’s estimates of the ICERs were generally lower than those produced by the Assessment Group (i.e. the treatments appeared more cost-effective). Similar results were reported by Bell et al. (2006).\(^\text{25}\) These results are, clearly, a potential cause for concern for the STA process where we are shifting the ‘onus of proof’ to the manufacturer who clearly has an interest in seeing the product approved and thus in underestimating the ICER where a decision is on the margin. This may undermine the independence of the NICE technology appraisal process, although it must be hoped that the review by the Assessment Group will maintain the...
independence of the process by informing the appraisal committee of any bias in the manufacturer’s submission. There are other concerns regarding the STA process as well, for example how to deal with commercial in confidence data (although NICE has been criticized before for its approach to this data as it lacks transparency, e.g. where all information is not made available to the consultees) and whether the process will actually lead to faster acceptance decisions.

**Lobbying—Patient groups, pharmaceutical companies and the media—how independent are NICE’s decisions?**

As part of the NICE’s appraisal process, patient groups are allowed to submit evidence which will be considered by the Appraisal Committee when formulating their decision. Patient groups share a common interest with pharmaceutical companies in promoting access to specific treatments that others will pay for (as long as they provide positive medical benefit) and this provides a route for companies to influence the perceptions of their drugs at a distance (between 50 and 82% of patient groups in the EU receive funding from pharmaceutical companies).

Dakin et al. (2006) showed that patient group submissions made a recommendation by NICE for routine rather than restricted use more likely. This could imply the ability of patient groups, and, possibly, indirectly pharmaceutical companies, to influence NICE, such that efficiency considerations may receive less emphasis than may be considered appropriate. This is particularly concerning since there is no lobby group for patients other than those relating to specific diseases, and therefore there is no group that considers simultaneously both the opportunity costs imposed on other patients whose treatment will be withdrawn and the benefits to those receiving the new treatment.

**An international perspective**

In 2003, the World Health Organization produced a review of the NICE technology appraisal programme. It noted that ‘Published NICE appraisals are already being used as international benchmarks—an obvious recognition of their credibility.’ This shows the high regard in which NICE is held internationally. Many countries now have similar tests for the cost-effectiveness of pharmaceuticals, although many have avoided processes similar to NICE’s MTA route (e.g. Canada and Scotland) and have instead focused on approaches more similar to the STA process.
The way forward

NICE now plays a key role in resource allocation decision in the NHS in England and Wales and this shift to evidence-based medicine must be considered to be a positive step forward for the NHS. Perhaps, the greatest test faced is to ensure that NICE’s decisions are not undermined or subverted by those who disagree with them as only through the wide acceptance of NICE’s appraisals can the NHS move towards better resource allocation.

References