BRIEFING PAPER

Pharmacoeconomics 2008; 26 (9): 781-798
1170-7690/08/0009-0781/$48.00/0

Exploring Uncertainty in Cost-Effectiveness Analysis

Karl Claxton

Centre for Health Economics, Department of Economics and NICE Decision Support Unit, University of York, Heslington, York, UK

Abstract

This paper describes the key principles of why an assessment of uncertainty and its consequences are critical for the types of decisions that a body such as the UK National Institute for Health and Clinical Excellence (NICE) has to make. In doing so, it poses the question of whether formal methods may be useful to NICE and its advisory committees in making such assessments. Broadly, these include the following: (i) should probabilistic sensitivity analysis continue to be recommended as a means to characterize parameter uncertainty; (ii) which methods should be used to represent other sources of uncertainty; (iii) when can computationally expensive models be justified and is computation expense a sufficient justification for failing to express uncertainty; (iv) which summary measures of uncertainty should be used to present the results to decision makers; and (v) should formal methods be recommended to inform the assessment of the need for evidence and the consequences of an uncertain decision for the UK NHS?

1. Background

When the UK National Institute for Health and Clinical Excellence (NICE) makes recommendations on the use of technologies in the UK NHS, information about cost effectiveness is critical. However, any assessment of effect and cost is uncertain and, thus, any decision based on cost effectiveness will also be uncertain. This uncertainty arises from a number of sources: not only the uncertainty in the estimates of parameters of the decision models commonly used to estimate costs and effects, but also a wide range of other sources of uncertainty, e.g. the potential bias or relevance of evidence and the assumptions required in extrapolating effects and costs over time. In the face of this uncertainty (only some of which is often explicitly characterized and presented), NICE must come to a view about whether the benefits of allowing access to an apparently cost-effective technology are greater than the potential consequences of an uncertain decision for the NHS.

A discussion of appropriate methods to characterize uncertainty requires first considering how decisions can be made based on evidence of cost effectiveness and why an assessment of uncertainty might be important for the types of decisions that a body such as NICE has to make.

1.1 Making Decisions Based on Cost Effectiveness

NICE appraisal requires an estimate of the cost effectiveness of the technology, often summarized as an incremental cost-effectiveness ratio (ICER), i.e. the additional cost for each additional QALY gained. This must be compared with some threshold for cost effectiveness, representing the health expected to be forgone elsewhere in the NHS because of the additional costs.[1,2]

For example, a threshold of £20 000 means that 1 QALY is expected to be lost for every £20 000 the NHS must find by curtailing other activities to accommodate the use of a more costly technology. If a

technology costs the NHS an additional £12 000 and provides 0.8 additional QALYs, the ICER of £15 000 per QALY gained (£12 000/0.8) is less than the threshold of £20 000 and the technology can be regarded as cost effective.

Alternatively, and entirely equivalently, this can be expressed as a positive net health benefit of 0.2 QALYs – the 0.8 QALYs gained more than offsets the 0.6 QALYs forgone elsewhere (£12 000/£20 000). This can also be expressed as positive net monetary benefit of £4000 – the monetary value of the health benefits of £16 000 ($0.8 \times £20\,000$) is greater than the monetary costs of £12 000.[3] The relationships between these entirely equivalent ways to express cost effectiveness are particularly important to appreciate when considering uncertainty. This is because, although ICERs are a useful and intuitive way to summarize cost effectiveness, they also have very unhelpful statistical properties.[4] Over recent years, almost all analyses of uncertainty in cost-effectiveness analysis have utilized measures of net health or net monetary benefit (of course, the results can and often are summarized using ICERs).

1.2 Variability, Heterogeneity and Uncertainty

It is useful to distinguish between variability, heterogeneity and uncertainty at the outset. Variability refers to the natural variation between patients in their response to treatment and the costs they incur – even when they have the same observed characteristics. Variability is irreducible – additional evidence can not reduce it. Heterogeneity refers to differences between patients who have different observed characteristics – for which, in principle, different guidance could be issued.[5] Uncertainty refers to the fact that we can never know for certain what the mean (expected) costs and effects would be if the treatment is provided for a particular population of patients, even if they have the same observed characteristics. Additional evidence can reduce uncertainty and provide more precise estimates. If the purpose of the NHS is to maximize health gains from its budget, then it is these expected effects and costs, and the uncertainty in estimating them, that are of primary interest.

2. Why Does Uncertainty Matter?

Although the importance of cost effectiveness is now well established, the reasons why uncertainty matters are less so. There are strong arguments for basing decisions about use of a technology on the expected incremental effects and costs rather than applying traditional rules of statistical significance to estimates of cost effectiveness. Indeed, previous guidance indicates that it is evidence about the expected effects and costs that is the primary basis for the guidance issued.[6] However, making decisions based only on expected effects and costs in no way implies that the uncertainty surrounding such a decision is unimportant. Indeed, an assessment of the implications of decision uncertainty is an essential part of any decision-making process that is concerned with improving health outcomes, given the resource constraint the NHS faces. There are three reasons why the uncertainty surrounding effects and costs matters: (i) to provide correct evaluation of expected effect and cost; (ii) to consider whether existing evidence is sufficient; and (iii) to assess the possible consequences of an uncertain decision for the NHS.

2.1 Evaluating Expected Effects, Costs and Net Benefit

When effects and costs are evaluated using a decision model in which there is a nonlinear relationship between inputs and outputs, e.g. Markov process,[7] the correct calculation of expected effects and costs will require the uncertainty around all the inputs (the parameters) to be expressed. This issue is illustrated in figure 1, where the value of parameter θ has a nonlinear relationship with net benefit (NB). Simply using the expected value of θ (θ_1) in the model would provide NB_1. However, the value of θ is uncertain and has a probability distribution. It is possible that θ will take a value of θ_2, generating NB_2 or a value of θ_3 generating NB_3. Both values are equally likely (they could represent the 95% confidence interval for θ) but generate very different values of NBs. When NB is evaluated over the possible values of θ, the expected NB [averaged over the possible values of θ or $E_\theta(NB,\theta)$] will clearly be less than NB_1. An analysis that only used the expected values of the parameters to estimate

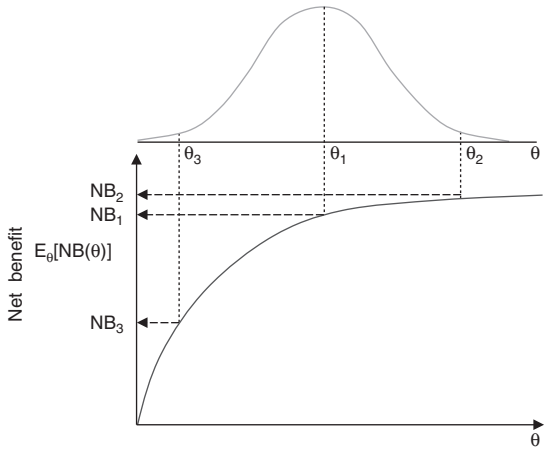

Fig. 1. Uncertainty and nonlinearity. The parameter θ has a nonlinear relationship with net benefit (NB). The value of parameter θ is uncertain and has a probability distribution. **E** = expected.

expected effects, costs and NB could be seriously misleading. Therefore, uncertainty matters, even if a decision maker only wishes to consider expected cost effectiveness in their decisions.

2.2 Is the Evidence Sufficient?

Simply basing decisions on expected cost effectiveness will ignore the question of whether the current evidence is a sufficient basis for guidance. It would fail to address the question of whether further research is needed to support NHS practice. The value of evidence or the health costs of uncertainty can be illustrated using a simple example in table I. Each row represents a realization of uncertainty, i.e. the net health benefit (QALYs) that results from taking a sample (a possible value) of each of the input parameters to the model. Therefore, each row can be thought of as representing one of the ways 'things could turn out' given our current uncertainty. The expected NB for treatment A and B is the average over all these possibilities (in this example, the range of potential values is simplified to only five possibilities).

On the basis of current evidence, we would conclude that treatment B was cost effective and on average we expect to gain an additional 1 QALY per patient treated compared with treatment A. However, this decision is uncertain and treatment B is not always the best choice (only three of five times), so the probability that B is cost effective is 0.6. For some realizations (two of five), treatment A would have been better. Therefore, a decision to approve B based on current evidence is associated with an error probability of 0.4. This is substantially greater than traditional benchmarks of statistical significance such as 0.05. Whether or not this level of uncertainty 'matters' depends on the consequences, i.e. what improvement in NB (or avoidance of harm) could have been achieved if this uncertainty had been resolved.

If uncertainty could be completely resolved (i.e. through complete evidence about effect and cost), then we would know the true value of NB before choosing between A and B (column 5 of table I). Of course we can not know in advance which of these values will be realized but on average (over the 5th column) we would achieve 13.6 rather than 13 QALYs. Therefore, the cost of uncertainty or the maximum value of more evidence is 0.6 QALYs or £12 000 per patient (at a threshold of £20 000 "per QALY") – more than half the value of the technology itself. It should be clear that the cost of uncertainty or the value of evidence is just as 'real' as access to a cost-effective treatment, as both are measured in terms of improved health outcomes for NHS patients. In principle, evidence can be just as, or even more, important than access to a cost-effective technology.[8]

2.3 Consequences of an Uncertain Decision for the UK NHS

Almost all decisions faced by NICE will impose some costs on the NHS which, if the guidance is reversed at a future date, cannot be recovered. For example, investment in equipment, the training of specialist staff or simply engaging in activities to

Table I. Why is evidence valuable?

How things could turn out	Net health benefit			Best we could do if we knew
	treatment A	treatment B	best choice	
Possibility 1	9	12	B	12
Possibility 2	12	10	A	12
Possibility 3	14	17	B	17
Possibility 4	11	10	A	11
Possibility 5	14	16	B	16
Average	**12**	**13**		**13.6**

ensure the guidance is implemented at a local level. These types of cost are often described as 'sunk' and are irreversible. When a decision is uncertain and there is a prospect that guidance may be reversed at some point in the future (e.g. if new evidence suggests that the technology is not cost effective or if another technological development makes it obsolete) then it may be better, in terms of NBs for the NHS, to wait until these uncertainties are resolved over time or until additional research is reported.[9,10]

There is another type of (opportunity) cost or irreversibility that is often more significant for the NHS and common to many NICE decisions. A decision to approve a technology will inevitably have an impact on the prospects of acquiring evidence to support its use in the future. This is because the incentives on manufacturers to conduct evaluative research (for the claimed indication) are removed once positive guidance or coverage has been granted. Furthermore, the clinical community is unlikely to regard further experimental research to be ethical and, even if ethical approval was granted, any randomized controlled trial is unlikely to recruit. This is particularly so when access to the technology is mandatory, as is the case with NICE guidance.

If the type of evidence the NHS needs can not be provided once positive guidance is granted, then the decision to issue guidance should take account of both the value of the technology (the expected NBs) and the value of the evidence that may be forgone (which can also be measured in terms of NB; see section 2.2). A technology may be regarded as cost effective when only considering the expected net health benefits of the technology itself. However, the value of the evidence (also expressed in terms of population health) that would be forgone for future patients (the opportunity cost of approval) may exceed the NBs of approval. In these circumstances, it would be better for the NHS if this evidence can be acquired by withholding approval, despite the technology having an ICER below the cost-effectiveness threshold. This can be achieved by NICE issuing an 'only in research' recommendation and reconsidering approval once sufficient evidence is available.[11,12]

2.4 Current and Developing Policy Environment

Since the 2004 *Guide to the Methods of Technology Appraisal*[13] was issued, a number of policy initiatives have occurred that pose more sharply the questions whether current evidence is sufficient to support NHS practice and what are the consequences of an uncertain decision? Importantly, NICE increasingly issues guidance on technologies close to launch when the evidence base is inevitably least mature. However, NICE does not have the remit to commission or fund research as part of its Technology Appraisal Programme, but it does make research recommendations, and can issue guidance conditional on additional evidence being provided, or restrict use within a clinical trial (an 'only in research' recommendation).

2.4.1 Research Recommendations

If evaluative research priorities are to be more closely related to the needs of the NHS and the guidance issued by NICE, as envisioned in the recent reforms to UK health research funding,[14] then a means to prioritize and effectively communicate the importance of additional evidence to those responsible for commissioning research will be required. To fulfil its remit, some form of assessment (whether informal or informed by explicit analysis) of the uncertainty in cost effectiveness, its consequences for the NHS and the importance of particular types of evidence, will be needed at some point during the NICE appraisal process.

2.4.2 Guidance with Evidence and Risk Sharing

In some circumstances, it might be possible to approve the use of a technology that appears to be cost effective based on existing evidence, while at the same time additional evidence is obtained from the manufacturer or by publicly funded research. In principle, this seems an efficient solution to the problem – some patients get early access to an apparently cost-effective technology but the evidence is also generated so that this decision can be reconsidered at a later appraisal when cost effectiveness as well as potential harm can be reassessed. This question of 'coverage or guidance with evidence' or 'risk sharing' seems to be envisaged within the recent reforms to research funding and proposed reforms to UK pharmaceutical pricing.[14-16]

If 'coverage with evidence' is to become a useful policy tool, an assessment of three questions will be needed: (i) is additional evidence needed; (ii) if so, what type of evidence is required; and (iii) will the type of evidence that can be gathered with concurrent approval be valuable and meet these needs? For example, if the key uncertainty is evidence about relative effects, then the type of observational registry data that are often envisaged will not be able to provide more reliable and precise estimates because a comparable control group will not be available. Early access will have been allowed, but the promise of the type of evidence needed will not be fulfilled. In these circumstances, early guidance would mean that the evidence base for future NHS practice will be undermined. Therefore, the question of how such assessments might be made by NICE is critical.

2.4.3 Only in Research

There are a number of circumstances in which it might be better for the NHS if approval of a technology that appears to be cost effective is withheld until the uncertainty surrounding the decision is resolved (i.e. an 'only in research' recommendation).[12] This includes situations in which (i) the NHS must make a significant and irreversible investment; (ii) new and potentially cost-effective technological development is imminent; and (iii) new and significant evidence is likely to become available. It also includes situations in which the type of evidence needed to reach a well informed decision about possible NHS use of the technology cannot be provided once it has been approved for NHS use. In all these circumstances, the decision to issue guidance ought to take account of both the value of the technology (the additional expected NBs) and the costs of the uncertain decisions – including the evidence that may be forgone and the likelihood that the investment required will ultimately be wasted.

2.5 Summary

All of the considerations described above require some assessment of the uncertainty surrounding a decision based on cost effectiveness and its consequences. Whether or not formal methods can assist in making these assessments is central to the development of appropriate methodological guidance. In particular:

- Which methods should be used to characterize parameter uncertainty?
- How should other sources of uncertainty be represented?
- How can the computational challenges be overcome?
- How should uncertainty in cost effectiveness be presented?
- How should the need for evidence and the consequences of an uncertain decision be assessed?

These key questions are discussed in turn in the following sections.

3. Characterizing Parameter Uncertainty

The 2004 *Methods Guide*[13] recommended that probabilistic sensitivity analysis (PSA)[17] should be conducted to reflect the combined implications of uncertainty in the parameters (inputs), and to quantify the uncertainty in cost effectiveness (outputs) and the decision to approve or reject a technology. This parameter uncertainty is only one source of uncertainty; nevertheless, it is an important component of decision uncertainty, so the question of which methods should be used to characterize it is important.

3.1 Probabilistic Sensitivity Analysis

The principles of PSA are very intuitive. Distributions are assigned to each of the model parameters. These are sampled (often using Monte Carlo simulation, which samples at random), and the output of the model (estimates of expected costs, effects and NB) is recorded for each set of samples from all the parameters. This process of sampling inputs and recording output is repeated (e.g. 10 000 times) so that the range of values that the parameters are likely to take are represented in the range of outputs.[18] As well as providing the correct estimate of expected effects, costs and NB in nonlinear models (see figure 1), the output of this process also provides the proportion of times (the probability) that each alternative is cost effective and all the information required to calculate the maximum value of additional evidence, as illustrated in table I and discussed in section 2.2.

3.1.1 Distributional Forms

It may appear that PSA introduces further assumptions about the choice of distribution to re-

present the uncertainty in the model inputs. However, the choice of distribution is not at all arbitrary if appropriately guided by the following three principles, as well as explicit judgments about the potential bias and relevance of the available evidence: (i) the nature of the parameter itself; (ii) the way the parameter was estimated, so that the distribution assigned reflects the statistical uncertainty (and any correlations) in its estimation; and (iii) the decision context, e.g. is it the average patient or the average of patients with a particular set of characteristics that is of interest?

Application of these general principles means that there will only be a very limited choice of appropriate distributional forms. For example, probability parameters are bounded by 0–1, so it would be inappropriate to specify a distribution that gave a probability to obtaining values outside this range. Where a probability is estimated from a proportion, the beta distribution is the natural choice. However, if the probability parameter is estimated from a logistic regression, then the parameters of interest are the coefficients on the log-odds scale, and multi-variate normality on this scale would be appropriate. For probabilities estimated from time-to-event data, the parameters would be the coefficients from a survival analysis estimated on the log hazard scale and, again, the appropriate assumption would be multi-variate normality on this scale.[8]

There is often insufficient evidence to apply standard statistical estimation and/or the evidence is limited in some other respect, e.g. potential bias and relevance. In these circumstances it becomes more apparent that the distributions assigned in PSA are really an explicit judgement (a reasonable belief often informed by evidence) about value of the parameter and its uncertainty.[19] Formal methods are available that can be used to elicit the views of experts about the likely value of a parameter and its distribution.[20] Of course, parameter uncertainty is only one source of uncertainty. Alternative but credible and plausible views of the quality and relevance of evidence as well as other assumptions are often possible. These other sources of uncertainty are discussed more fully in section 4.

3.1.2 Correlation

Model parameters may be correlated. For example, if a regression analysis is used to estimate model parameters, the relationship between them can be estimated from the co-variance matrix so that the correlations between regression coefficients can be included in PSA in the multi-variate normal case.[8] Similarly, methods of evidence synthesis[21,22] generate correlated outputs that can be fully captured in PSA by sampling from or directly using the output of such synthesis in the decision model. If there is no evidence from statistical analysis that parameters are correlated, it is generally not necessary to impose one. Of course, any logical or latent structural relationship between parameters should be reflected in the model structure rather than imposing correlation.

3.2 Deterministic Sensitivity Analysis

In deterministic sensitivity analysis (DSA), each parameter in turn is set at an extreme but plausible value and the decision is described as sensitive if a decision based on cost-effectiveness changes. Alternatively, a threshold value for a parameter is found at which the decision changes. Although DSA is very simple to understand and generally easy to implement, there are a number of problems.

- The estimates of expected effects, costs and NB will be biased in nonlinear models.
- What is an ‘extreme but plausible’ value for the parameter is not clear as there is an implicit assumption about the distribution of the parameter.
- Neither is it clear whether a threshold value for a parameter is likely or extremely unlikely to occur.

Therefore, any useful interpretation of one-way sensitivity analysis requires exactly the same information about the distribution of the parameter as PSA. The difference is that this information is not presented, so the judgments required remain implicit and opaque rather than explicit and transparent. Indeed, where parameters are based on statistical estimates (e.g. meta-analysis, evidence synthesis or regression analysis), then the information about the uncertainty and correlation in these estimates is in essence thrown away.[23]

A decision may not seem sensitive to any plausible parameter values in a series of one-way sensitivity analyses, but when taken in combination (parameters are simultaneously uncertain) there may be considerable uncertainty. Therefore, even if interpreted correctly, one-way sensitivity analysis will commonly (in the absence of correlation) underestimate uncertainty, making it particularly vulnerable to false claims that results are robust. Of course, a poorly and inadequately conducted PSA can be used to make the same false claims. The question is whether an analysis that is more explicit and transparent about distributions assigned to parameters, and that is directly interpretable, makes such claims easier to detect.

Best- and worst-case scenarios are very difficult to interpret correctly. The probability of all parameters taking extreme but plausible values simultaneously will be very small indeed, so if the decision is sensitive to such extremes it is not clear whether the decision should be regarded as uncertain or not. Multi-way sensitivity analysis with more than two or three parameters is very difficult to present. More importantly, it is generally even more difficult to interpret correctly and becomes impossible if some parameters are correlated, because it is impossible to locate a fixed (set of) parameter value(s) that can be regarded as 'extreme'.[23]

4. Representing Other Sources of Uncertainty

The uncertainties surrounding the parameters of the model are only one source of uncertainty. Other sources of uncertainty include the different types of scientific judgments that have to be made when constructing a model of any sort such as decision model, statistical model or even standard meta-analysis. These types of uncertainties have been classified in a number of different ways (methodological, model, clinical, etc.) but can be referred to collectively as structural uncertainties.[24]

4.1 Probabilistic Scenarios

It is important to recognize that the results of any type of model are correct if the necessary judgements and assumptions are acceptable. However, there are always other sets of judgments and assumptions that are possible, and some may be regarded as plausible and credible. In these circumstances, the 2004 *Methods Guide*[13] recommended that probabilistic scenarios[25] (a revision and rerun of the probabilistic model) should be presented, with the expected effects, costs and measures of uncertainty reported for each. It is then for the decision maker to take these into account in the light of other evidence (e.g. views of clinical experts).

There are broadly two possibilities. First, only one scenario is regarded as credible and all others can be disregarded – the parameter uncertainty captures all the uncertainty surrounding the decision. More commonly, one scenario may be the most credible but others are also possible and cannot be disregarded. In this case, there is uncertainty about effects and costs given a particular set of judgments and also uncertainty about which set of judgments might be realized. Parameter uncertainty no longer represents all the uncertainty surrounding the decision, and the decision maker must implicitly weigh these different scenarios to come to a view about effects, costs and overall uncertainty. When different scenarios suggest different decisions, the structural uncertainty clearly matters. However, even when this does not occur, it will affect decision uncertainty too, in ways that are difficult to assess implicitly and intuitively.[24] There are alternatives to this implicit weighting of probabilistic scenarios, which require the formal specification of probability distributions across scenarios or across structural parameters.

4.2 Model Averaging

The weighting of scenarios can be made explicit by assigning probabilities to represent how credible each is believed to be.[24,26] The weighted average of costs and effects from each scenario can easily be calculated but overall uncertainty and its consequences requires the simulated values from the scenarios to be merged, based on these probabilities, before calculating a new estimate of the error probabilities and value of information. This does not require additional simulation and is quick, easy to implement, and explicit. However, it does pose a number of questions:

- Who should be responsible for choosing the weights?

- If there is no consensus, how can differing views be combined?
- Which methods should be used to elicit probabilities (see section 4.4)?
- How many scenarios should be used to represent the possibilities?

Despite these practical issues, this approach would provide measures of overall decision uncertainty and its consequences. Importantly, it would do so even when the different scenarios did not suggest different decisions based on cost effectiveness. However, it remains difficult to assess which uncertainties matter most and what type of additional evidence might be most useful.

4.3 Parameterizing Structural Uncertainty

In almost all cases, structural uncertainty can be thought of as a missing and uncertain parameter in the model, i.e. different scenarios are special cases of a common 'meta-model' but with missing parameters taking extreme values.[24,27] For example, in a progressive disease where a treatment may or may not be disease modifying, two alternative scenarios of rebound following treatment failure could be constructed: (i) rebound is equal to initial improvement on treatment; or (ii) rebound back to where the patient would have progressed. This structural uncertainty is really an uncertain and missing parameter – the magnitude of rebound relative to disease progression. Rather than present two scenarios, this parameter could be included in the model despite there being no evidence available to estimate it.[28] The distribution assigned to it should represent an informed and credible view about its value and uncertainty. Similarly, differing views about the quality or relevance of trial evidence could be represented as scenarios where some trials are either included (they are relevant and unbiased) or excluded (they are irrelevant and so biased that the evidence they provide is worthless). The alternative to these extremes is to assign uncertain parameters (weights) for bias and relevance.[29] The latter approach has a number of advantages:

- it represents the source of uncertainty as a real and measurable parameter that is amenable to further investigation;
- it does not implicitly assign extreme and generally unlikely values (0 or 1) to the parameter;
- it indicates who might be best placed to provide informed and credible judgments;
- when there is more than one difference between scenarios it allows the most important source of uncertainty to be identified;
- the impact on decision uncertainty and the application of value of information methods allows consideration of whether further evidence about the parameter is needed.

4.4 Elicitation

These more formal approaches to structural uncertainty described in sections 4.2 and 4.3 require explicit judgments about the value of, and the uncertainty surrounding, a parameter for which no direct evidence exists.[28,30] Of course, similar issues are raised when selecting distributions for parameters in PSA but they are posed more sharply in this context. Any judgement made will need to be credible and defensible. Well established methods for this type of elicitation are available and documented.[20] However, they do take time and their use poses a number of detailed questions.

- Who should provide the judgements – which experts or decision makers?
- Which particular methods of elicitation should be used?
- How should the quality of judgments be calibrated (tested) and weighted?
- Should more than one expert be used? If so, then how should judgments be combined and if not combined, how should we choose which one is most credible?

5. Dealing with Computational Challenges

Some types of model structures are particularly computationally expensive, and expressing uncertainty using PSA becomes particularly burdensome. There are three key questions: (i) when can computationally expensive models be justified; (ii) can the computational expense be overcome; and (iii) if not, can computational expense justify a failure to adequately express uncertainty?

5.1 When Can Computationally Expensive Models be Justified?

The choice of model structure and complexity is always a trade-off between descriptive realism, computational burden and evidence requirements. The key issue is whether the model is sufficiently realistic to inform the decision, not whether it is a fine description of reality. There are two types of common model structure that can be particularly computationally expensive: patient-level simulation (PLS; which includes micro or discrete-event simulation) and dynamic models.

5.1.1 Patient-Level Simulation

PLS requires simulation to obtain a single estimate of expected effect and cost, so expressing uncertainty using PSA can become particularly burdensome. However, a single estimate of expected effect and cost based on a single set of (mean) values for the parameters will be biased because such complex models are inevitably nonlinear. Furthermore, uncertainty and its consequences will not be explored. PLS models have increasingly been used in NICE appraisals but only very rarely has PSA been conducted to capture uncertainty.[31] Therefore, the circumstances in which PLS may be justified need to be considered.

This type of model can be particularly useful for modelling time and state dependence. Although simple time dependence is easily handled in standard analysis, it becomes more difficult when the probability of, or the time until, future events depends on a patient-accumulated history, e.g. the sequence and time spent in different states. In some circumstances, this can be handled in standard Markov models and in some others by implementing semi-Markov processes using efficient programming platforms.[8] A model structure that is unable to take account of these issues may well lead to biased estimates of expected effects and costs. Therefore, the existence of time and state dependence itself is not necessarily a justification for adopting PLS, but clearly there are circumstances where such models are justified for this reason.

Other reasons given for using PLS include the ability to sample patient characteristics, integrate variability and treatment switching.

- Sampling patients with different observed characteristics provides the expected effects and costs averaged over this heterogeneous population. However, NICE needs separate estimates of effects and costs for identifiable groups of patients with different characteristics.[5] If NICE requires estimates for a number of combined patient groups, this can be based on the weighted average of the expected effects and costs for each group, using a standard analysis.
- These types of model necessarily sample variability. However, the primary interest for NICE is expected (population mean) costs and effects, not their variability. In addition, for most common model structures, variability will have no effect on expected effects and costs (e.g. in Markov models). There are some circumstances when expected effects and costs do depend on both the uncertainty and the variability. In these cases, both need to be expressed, but computationally less expensive methods can be used. In the calculation of expected effects and costs there are never circumstances when variability in effects and costs matters but the uncertainty does not.
- PLS is often used to switch patients between treatments to generate a sequence of treatments for each patient. The result of such analysis provides the effects and costs averaged over the many different sequences that might be experienced. However, NICE often wishes to issue guidance on which sequence is cost effective. This requires estimates of the effects and costs of each possible sequence or, when there are a vast number of possible sequences, the most common or clinically relevant ones. Identifying the most cost-effective sequence when there are a vast number of possibilities is challenging for all modelling approaches, but more so for those that are more computationally expensive.

5.1.2 Dynamic Models

Dynamic models are used to model infectious diseases, where the effective treatment of one individual can have an impact on others through reduced infection and herd immunity.[32] Where a technology is designed or likely to have a significant impact on an infectious disease and the impact of treatment on

the infection of others is a consideration, then models failing to account for this may underestimate benefits and overestimate costs. Dynamic models are often implemented using PLS and are computationally expensive. However, this is not always necessary, and other programming platforms can be used to solve the differential equations analytically or numerically.

5.2 Can Computational Expense be Overcome?

Clearly there are circumstances in which computationally expensive models may be justified. In these circumstances, computational burden can often be eased in broadly three ways: more efficient simulation, emulators and linear approximations.

5.2.1 More Efficient Simulation

Faster processing or more efficient simulation can ease this burden. Recently, it has been shown that substantial improvement can be made in the basic Monte Carlo approach to computing PSA in typical PLS models. This method is simple to apply and makes many otherwise intractable models amenable to PSA (for further information, see O'Hagan et al.[33]).

5.2.2 Emulators

An emulator is essentially a model of a model and can dramatically reduce the computational burden of any type of computationally expensive model. Not all emulators perform well on nonlinear models (e.g. simple linear regression), but there is a class of emulator (the Gaussian process) that can deal appropriately with the types of nonlinear models often used in appraisal. This approach is considerably more complex than PSA but the mathematical properties of these emulators are well established and have been demonstrated through their application in health technology assessment.[34]

5.2.3 Linear Approximations

Linear models provide unbiased estimates of expected effects and costs without fully expressing uncertainty or variability in the parameters. Therefore, if an unbiased linear approximation to a nonlinear model can be found, this will reduce the computational burden. However, it is not clear when this will work. In particular, linearization based on gradients may perform badly. A number of mathematical approximations are possible and have been used in health technology assessment and NICE appraisals.[35]

In summary, computationally expensive models may be justified in some circumstances, and methods are available to ease this burden. A failure to adequately express uncertainty means that the estimates of expected cost and effect will commonly be biased, and uncertainty and its consequences will be inadequately explored. Therefore, the computational expense of the chosen modelling approach and platform can not reasonably justify a failure to adequately express uncertainty.

6. Presenting Uncertainty in Estimates of Cost Effectiveness

The reasons why uncertainty matters for the NHS suggest that it might be useful for decision makers to have some information about (i) the probability that a decision based on expected effects and costs will be correct (p) and the probability it will turn out to be wrong – the error probability $(1 - p)$; and (ii) some indication of the consequences of error, i.e. what health and/or resource costs will be incurred by patients and the NHS if the decision turns out to be wrong. This information is useful because it contributes to an assessment of whether uncertainty matters. For example, a decision may be uncertain (very high error probability), but the consequences (costs to the NHS or NB forgone to the NHS patient population) might be limited. Conversely, a decision that has a low error probability (is more certain) might be more important if the costs and NB forgone are very large. Therefore, when considering alternative summary measures of uncertainty it is important to consider whether they can be interpreted in a way that might help to assess the consequences of uncertainty.

6.1 Confidence Ellipses and Scatter Plots

When there are only two alternatives (e.g. a single active treatment), a joint probability distribution can be represented in the incremental cost-effectiveness plane. This can be illustrated graphically in several equivalent ways, two of which are ellipses and scatter plots.[36] Scatter plots record each simu-

lated estimate of the expected additional effects and additional costs. This is illustrated in figure 2a using simulated output from a probabilistic model of two alternatives, B compared with A.

The threshold (λ) is represented by the dashed line, with slope of £20 000 per QALY and the circle indicating the expected incremental cost effectiveness of B compared with A. This is above the dashed line, so B is not regarded as cost effective (ICER = £61 680 per QALY). Consideration of the uncertainty surrounding a decision to reject B on these grounds requires an assessment of the proportion of points that lie below the dashed line (where B is cost effective). Even with only two alternatives to consider, there are a number of problems:

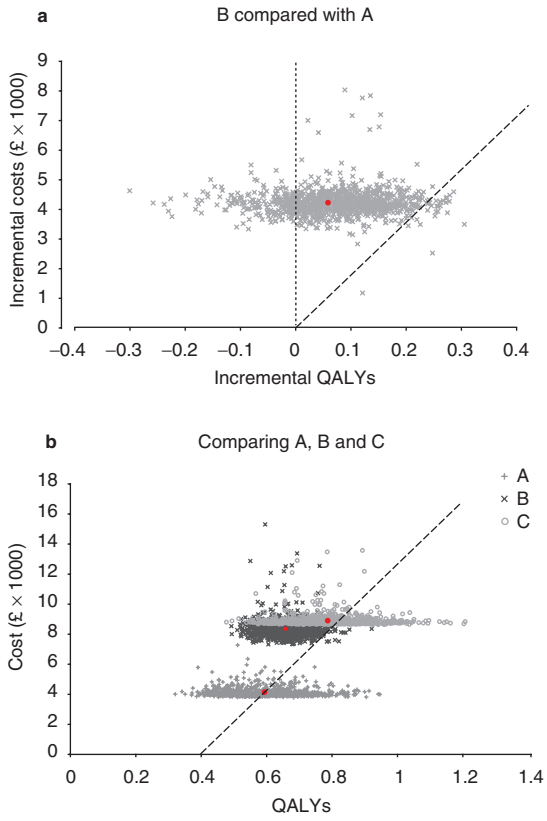

Fig. 2. Scatter plots of hypothetical treatment options: (**a**) B vs A (incremental cost effectiveness); (**b**) with multiple alternatives (effect-cost pairs). The dashed line represents the threshold (λ) of £20 000 per QALY (0.2 QALYs = £4000). Red circles indicate the expected incremental cost effectiveness.

- it is difficult to visualize alternative cost-effectiveness thresholds in this cost-effect space;
- when represented in two dimensions (or even in three), it is difficult to accurately assess the proportion of points that lie below the line to provide an intuitive estimate of the probability that B is cost effective compared with A.

When there is more than one active treatment, which is commonly the case, scatter plots or confidence ellipses become impossible to interpret correctly. Figure 2b illustrates a scatter plot when alternative C is also considered alongside A and B. Now the expected effect-cost pairs, rather than the increments, are plotted.

Intervention A remains cost effective at a threshold of £20 000 per QALY. But consideration of uncertainty requires some assessment of the proportion of points lying below the line for interventions B and C compared with A. This is difficult enough, but a critical piece of information is also missing: each of the cost-effect points for A, B and C will be correlated by the structure of the model itself. For example, if A has a higher than expected cost, then B or C might also have a higher than expected cost, etc. Without this information it is impossible to assess the probability that A is cost effective and the error probability associated with adopting A. This is also true for other proposed summary measures such as confidence ellipses, confidence intervals or the full distribution of NB of each alternative.

6.2 Cost-Effectiveness Acceptability Curves

Cost-effectiveness acceptability curves (CEACs) are constructed by recording the number of times each alternative has the highest NB (i.e. is cost effective) from the simulated output of a model. The probability (proportion of times) that each is cost effective is plotted for a range of possible cost-effectiveness thresholds. Figure 3 illustrates the CEAC associated with the scatter plot in figure 2b. At first sight, the CEAC seems easier to interpret since the probability that A, B or C is cost effective (p) and the associated error probability (1 − p) can simply be read off for any particular threshold.[37]

However, it is important to note that the alternative with highest probability of being cost effective may not be the most cost-effective alternative (hav-

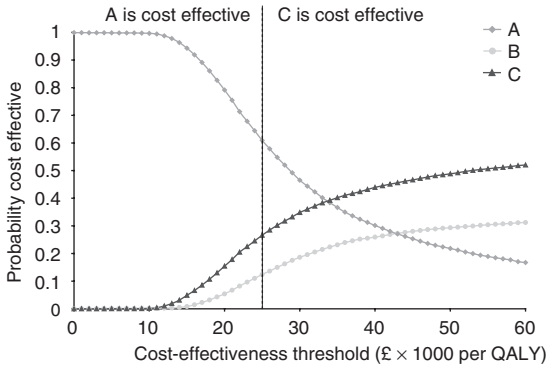

Fig. 3. Cost-effectiveness acceptability curve for hypothetical treatment options A, B and C. The dashed line represents the cost-effectiveness frontier.

ing the highest expected net-benefit) – this is illustrated in figure 3 for thresholds between £24 628 and £34 000 per QALY. This may seem counter intuitive, but is easily illustrated in table II, which is very similar to table I. B is cost effective and has the highest expected NB, but the probability that it is cost effective is only 0.4 (two of five), which is a lower probability than A (0.6; three of five). The reason is that when B is better than A it is 'much better', but when A is better than B it is only 'a little bit better'.

Therefore, presenting a CEAC alone is not enough. It is important to indicate which of the alternatives is expected to be cost effective as well as its probability. This is indicated by the dashed line on figure 3, or by the cost-effectiveness acceptability frontier (CEAF), which is a plot of the probability that the most cost-effective alternative (with the highest expected NB) is cost effective.[38]

The type of CEAC in figure 3 and the associated CEAF have significant advantages over scatter plots and other alternatives – they provide the probability that a decision based on expected costs and effects will be correct (p) and the error probability (1 − p). However, they do not provide information about the differences in the NBs of the uncertain alternatives, so it is difficult to assess whether the uncertainty 'matters'.

Table II. Probability and cost effectiveness

How things could turn out	Net health benefit			Best we could do if we knew
	treatment A	treatment B	best choice	
Possibility 1	10	14	B	14
Possibility 2	12	11	A	12
Possibility 3	11	16	B	16
Possibility 4	15	14	A	15
Possibility 5	12	10	A	12
Average	**12**	**13**		**13.8**

6.3 Other Ways to Present Uncertainty

NICE is primarily interested in thresholds between £20 000 and £30 000 per QALY. A simple alternative to the CEAC and CEAF is to report expected effects, costs, ICERs, the expected NB (expressed in health or money terms) and probabilities for each alternative, as well as the error probability for a decision based on expected cost effectiveness. This is illustrated in table III.

This indicates that A is cost effective at a threshold of £20 000 per QALY, but with a probability of error >0.2. This uncertainty is primarily in the choice between A and C – there is little chance that B will be cost effective. At a threshold of £30 000 per QALY, C is expected to be cost effective, but this decision is more uncertain (>0.6 error probability). Again the uncertainty is primarily between A and C. The difference in expected NB between the uncertain alternatives also gives some indication of the consequences of error. At £30 000 per QALY, the decision is more uncertain and the consequences might be more serious (the difference in expected NB between C and either A or B is greater) than at £20 000 per QALY (between A and either C or B).

The information in table III can also be presented graphically as in figure 4. This visual representation can be particularly useful for large numbers of alternatives;[39] indicating which appears to be cost effective, the associated uncertainty, how this uncertainty is 'distributed' over the other alternatives, and the consequences of uncertainty.

The consequences of an uncertain decision can not be directly calculated from table III or figure 4 as they only report the expected NB of the uncertain alternatives. However, this type of information is indicative of the relative importance of uncertainty in many common circumstances.

Table III. Presenting uncertainty and cost effectiveness

Option	Cost	Effect	ICER	λ = £20 000 per QALY			λ = £30 000 per QALY		
	(£)	(QALY)	(£/QALY)	NB (£)	probability	p(error)	NB (£)	probability	p(error)
A	4147	0.593		7722	0.792	0.208	13 656	0.465	
B	8363	0.658	ED	4794	0.054		11 373	0.186	
C	8907	0.787	24 628[a]	6827	0.154		14 695	0.348	0.652

a Option C versus option A.

ED = extended dominance; **ICER** = incremental cost-effectiveness ratio; **NB** = net benefit.

7. Assessing Need for Evidence and the Consequences of an Uncertain Decision

Addressing why uncertainty surrounding expected costs and effects matters includes asking whether existing evidence is sufficient and assessing the possible consequences of an uncertain decision for the NHS. Whether more explicit methods might help NICE make these assessments is considered below.

7.1 Is the Evidence Sufficient?

The *Methods Guide* of 2004[13] suggested that research priorities can be identified on the basis of evidence gaps identified by systematic review of effects and cost-effectiveness analysis. It suggested that these may be best prioritized by considering the value of additional information in reducing the degree of decision uncertainty. It also recommended that the value of information associated with particular groups of parameters can be used to identify where such research should be focused.

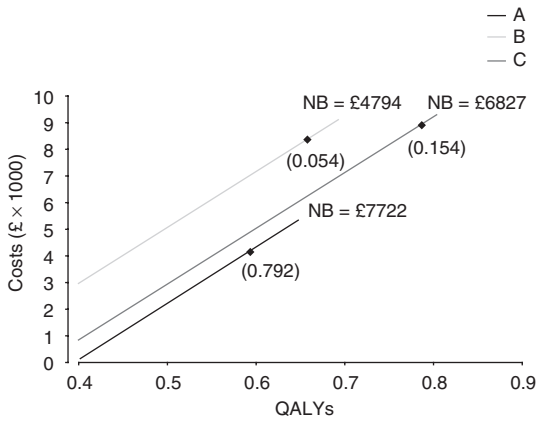

Fig. 4. Uncertainty and cost effectiveness for hypothetical treatment options A, B and C. The numbers in brackets are the probabilities of cost effectiveness for a threshold (λ) of £20 000 per QALY (0.2 QALYs = £4000). **NB** = net benefit.

7.1.1 Expected Value of Information

The expected value of information (EVI) combines both the probability of error and the consequences of error in terms of the NBs forgone. It can be expressed in health or monetary values or can simply be regarded as a metric of the importance of uncertainty.[8,35,40] The principles of EVI have already been illustrated in table I and discussed in section 2.2. As demonstrated, the additional expected net health benefits offered by the technology was 1 QALY but the expected value of resolving the uncertainty would contribute 0.6 QALY, or £12 000 per patient (at a threshold of £20 000 per QALY). This represents the value of resolving all the uncertainty for an individual patient. It also represents the maximum value of acquiring additional evidence, i.e. the expected value of perfect information (EVPI). Of course, any additional research will never resolve all uncertainties but EVPI does place an upper bound on the value of additional evidence. An informal indication of the importance of uncertainty can be made by comparing the per-patient value of access to the technology (the expected additional NB) to the maximum value of additional evidence about its use.

7.1.2 Population Expected Value of Perfect Information

Additional evidence can be used to guide the treatment of all other current and future NHS patients. Therefore, the maximum value of evidence to the NHS as a whole requires estimates of this current and future patient population (where the population EVPI is the discounted sum). This requires a judgment to be made about the time over which additional evidence that can be acquired in the near future is likely to be useful and relevant. Generally, fixed time horizons of 10, 15 and 20 years have common-

ly been used in the health literature as well as the environmental risk and engineering literature.[41] There is some empirical evidence that suggests that clinical information may be valuable for much longer (a half-life of 45 years).[42] However, any fixed time horizon is really a proxy for a complex and uncertain process of future changes, all of which impact on cost effectiveness and the future value of evidence.[43] In health, some future changes can be anticipated (a new technology will be launched, a trial that is recruiting will report or a branded drug will go generic), and differing judgments about time horizons for different appraisals might be appropriate.

As well as a simple metric of the relative importance of uncertainty across different appraisals, the population EVPI can be compared to the expected cost of additional research, which includes the NB forgone if conducting research requires delaying approval of a technology that appears to be cost effective based on current evidence. If these expected opportunity costs of research exceed the population EVPI (maximum benefits) then the research is not worthwhile – the resources could generate more health improvement by being used elsewhere in the NHS, and guidance should be based on current estimates of expected cost effectiveness.

7.1.3 Computation

The per-patient EVPI is easily calculated from the simulated output of a probabilistic model (see table I and section 3), which may include a formal assessment of other sources of uncertainty (see section 4). Therefore, the time and effort required to calculate EVPI is negligible once all the evidence has been assembled and judgments have been made in constructing a model, synthesizing evidence and assigning probability distributions to all the sources of uncertainty. Calculation of the population EVPI is also trivial given that estimates of incidence and prevalence are already required in the evidence review and the appropriate discount rate is detailed in NICE guidance. However, a judgment about the time horizon is required. The uncertainty in such judgments can be incorporated in the analysis and will influence the 'expected' population EVPI.[43]

7.2 What Type of Evidence is Needed?

If further research is potentially worthwhile, it would be useful to have an indication of what type of additional evidence might be most valuable. This can inform the decision of whether guidance with evidence, or an 'only in research' recommendation might be more appropriate. There are a number of alternative measures that can be considered.

7.2.1 Expected Value of Perfect Parameter Information

The analysis of the value of information associated with different (groups of) parameters is, in principle, conducted in a very similar way to the EVPI for the decision as a whole. The EVPI for a parameter or group of parameters (EVPPI) is simply the difference between the expected NB when their uncertainty is resolved (and a different decision can be made) and the expected NB given existing uncertainty.[8,35,44]

EVPPIs can be used as a simple metric of the relative importance (sensitivity) of different types of parameters and sources of uncertainty in contributing to the overall EVPI.[45] As a simple measure of sensitivity it has a number of advantages: (i) it combines both the importance of the parameter (how strongly it is related to differences in NB) and its uncertainty; (ii) it is directly related to whether the uncertainty matters (whether the decision changes for different possible values); and (iii) it does not require a linear relationship between inputs and outputs. In addition, it can be expressed in health or monetary values and either per patient, or for the population of NHS patients.

When population EVPPI is expressed in monetary terms it can be directly compared with the expected opportunity costs of the type of research that might be needed to provide the evidence. This is important because some uncertainties are relatively cheap to resolve (in terms of time and resource) compared with others (e.g. an observational study to link a clinical endpoint to quality of life compared with an RCT of long-term relative treatment effect). Which source of uncertainty is most important requires a comparison of these benefits and opportunity costs.

7.2.2 Computation

Evaluating EVPPI does come at a computational cost. For linear models, each estimate of EVPPI requires some additional computation (the manipulation of the simulated values rather than repeated simulations). When the model itself is not computationally expensive it is a generally manageable expense. However, if the model is nonlinear, EVPPI may require many repeated runs of the same probabilistic model, which can become prohibitively expensive given the time constraints of NICE appraisal. Therefore the computational expense of EVPPI needs to be justified. For example, estimates of EVPPI may be justified if the analysis of population EVPI suggests that additional evidence might be required (population EVPI is greater than the expected opportunity costs). If justified, there are a number of ways to reduce computation expense:[44]

- It is more efficient and more informative to first consider a limited number of groups of parameters, informed by the types of research required, e.g. RCT, survey of QALYs, or an observational epidemiological study. If there is substantial EVPPI associated with a particular group, only then conduct additional analysis to explore which particular source of uncertainty within the group matters most.
- The number of simulations to conduct and where to take them (inner and outer loops) can be made more efficient.
- In addition, computational burden can be reduced by faster processing and more efficient sampling; finding mathematical linear approximations to the model itself or to nonlinear relationships within it; and using emulators (see section 5).

There are other considerations that also impact on computational expense. First, EVPPI can be evaluated for a group of correlated parameters (often without additional computation expense). However, the EVPPI of each, individually, will not capture the relationship to others and may under/overestimate the value of more evidence. When certain groups of parameters are correlated, the type of repeated sampling can be required even in a linear model.[35] Second, when considering the order in which research studies might best be conducted, measures of the EVPPI of a sequence of studies is possible but generally requires more computation.[46]

7.2.3 Other Measures of Sensitivity and Importance

Other measures of the importance of parameters are also available. For example, there is a large body of literature outside health on the use of conditional PSA.[46] This can provide a variety of different measures of sensitivity, e.g. the contribution that a parameter makes to the overall uncertainty in model outputs (contribution to variance in expected NB). Unfortunately, these approaches are often just as computationally demanding as EVPPI, are not always related to whether the uncertainty matters (whether the decision changes) and do not provide a measure of value that can be compared with the opportunity costs of research. They offer no real advantage when parameters are correlated.

Much simpler approaches are also possible. Once parameter values and associated expected NBs have been simulated and recorded, it might be tempting to summarize the correlation between them. However, simple correlation can be high, even when the contribution to variance in NB is low. Another approach would be to use analysis of co-variance methods (ANCOVA), which can summarize the proportion of the variance in the output 'explained' by variation in the input parameters.[8] However, in its simple form, this assumes a linear relationship between inputs and outputs. Therefore it offers little advantage in terms of computation (i.e. if the model is linear, EVPPI is not particularly expensive anyway). They also have some disadvantages: they are not directly related to whether the uncertainty matters (whether the decision changes) and do not provide a measure of value that can be compared with the opportunity costs of research.

These simpler methods might provide some indication of where uncertainty matters in a nonlinear model – even though they would be biased. The bias will depend, among other things, on how nonlinear the model really is. However, evaluating EVPPI as if the model was linear would only require similar computation, but have the advantages described above. It is not clear whether a biased estimate of sensitivity or EVPPI is better than none at all, or how much nonlinearity or correlation and therefore potential bias would render them of no value. This question is very difficult to answer, as the appropriate comparison ought to be between the quality of

the assessment of the need for evidence with and without such imperfect information. But this is a common question for NICE when considering imperfect evidence of cost effectiveness.

7.3 Assessing the Consequences of an Uncertain Decision for the NHS

Almost all decisions faced by NICE will impose some costs on the NHS which, if the guidance is reversed at a future date, cannot be recovered. Importantly, a decision to approve a technology will inevitably have an impact on the prospects of acquiring evidence to support its use in the future.

7.3.1 Costs of Changing Clinical Practice

The sunk or irreversible costs of changing clinical practice may include investment in equipment and the training of specialist staff or simply engaging in activities to ensure the guidance is implemented at a local level. The importance of these costs for early adoption of a technology requires, first, some assessment of the probability that guidance will change (at some point in time), and, second, the value of the sunk cost forgone (over time). If the sunk costs expected to be forgone exceed the expected additional net health benefits of immediate adoption, then it may be worth withholding approval until more secure guidance can be issued.[9,10]

The uncertainty surrounding cost effectiveness gives some indication of how vulnerable the guidance may be to new evidence. In addition, judgment about whether research is likely to be conducted and whether the range of possible results is likely to change guidance is also needed, which can be based on estimates of predicted posteriors and the expected value of sample information (EVSI).[35,40] It is also possible to evaluate the maximum sunk costs that the NHS can afford while still adopting the technology,[6] or integrating estimates of sunk costs to consider the overall NBs of the decisions to adopt now (no research), adopt and conduct research, or conduct research and withhold adoption until it reports.[9] However, research and new evidence is only one source of change that might lead to a change in guidance. Others include changes in technologies, prices or evidence from other sources.[43] Attempts have been made to explicitly model these types of future changes or use methods from financial economics (real options) to evaluate this complex and uncertain prospect.[10]

7.3.2 Value of Evidence Forgone

All decisions are uncertain, particularly when they are made close to launch of a product, and a decision to approve a technology will inevitably have an impact on the prospects of acquiring evidence to inform judgements about its use in the future. The upper bound on the value of evidence that may be forgone by issuing positive guidance is the population EVPI. This can be compared with the additional expected NBs of early access to the technology. If the benefits of early access exceed the maximum value of evidence forgone, immediate positive guidance is worthwhile (even though some evidence will be forgone). Where the maximum (evidence) costs exceed the benefits of early access, a judgment will be necessary about when and what type of evidence might become available by issuing an 'only in research' recommendation. Of course, these judgements can, in principle, be explicitly incorporated in the analysis and can be linked to the value of particular types of evidence that might be forgone through estimates of EVPPI and EVSI for particular research designs.[35,47,48] It may not be possible or desirable to make explicit in formal analysis all the judgments required for this type of assessment. However, the question is whether the type of explicit analysis that can be made available is a useful starting point for deliberations around 'only in research' recommendations?

8. Conclusions

This paper has described the key principles of why an assessment of uncertainty and its consequences is critical to the decisions that NICE must make. In doing so, it poses the question of which formal methods may be useful in helping NICE and its advisory committees make such assessments. In coming to a view about whether more explicit methods of assessment are required, it should be recognized that no formal analysis can capture everything that might be important to the NHS. The question is 'do they capture enough to be useful to the deliberations that take place in the appraisal process?' For example, few would argue that cost-effectiveness analysis can capture everything of importance, but it

is generally accepted that if it is based on well conducted analysis it captures enough to be at least a useful starting point for deliberation. The same principles ought to guide the choice of methods for the characterization and presentation of uncertainty and its consequences.

Acknowledgements

This paper was initially prepared as a briefing paper for NICE as part of the process of updating the Institute's 2004 *Guide to the Methods of Technology Appraisal*. The work was funded by NICE through its Decision Support Unit (DSU), which is based at the universities of Sheffield, Leicester, York, Leeds and at the London School of Hygiene and Tropical Medicine.

The author has no conflicts of interest that are directly related to the contents of this article.

The author thanks members of the DSU who commented on the briefing document that forms the basis of this paper as well as Iain Chalmers, Alex Sutton, Alan Brennan, Louise Longworth and Carole Longson, who provided helpful comments on earlier drafts of this paper. All errors and omissions are the responsibility of the author.

References

1. Culyer AJ, McCabe C, Briggs A, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. J Health Serv Res Pol 2007; 12: 56-8
2. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics 2008; 26 (9): 733-44
3. Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. Int J Technol Assess Health Care 1991; 7 (1): 12-21
4. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998; 18: S68-80
5. Sculpher MJ. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics 2008; 26 (9): 799-806
6. Claxton K. The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. J Health Econ 1999; 18: 342-64
7. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998; 13: 397-409
8. Briggs A, Claxton K, Sculpher MJ. Decision analytic modelling for the evaluation of health technologies. Oxford: Oxford University Press, 2006
9. Eckermann S, Willan A. Expected value of information and decision making in HTA. Health Econ 2007; 16: 195-209
10. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. J Health Econ 2000; 19: 755-66
11. Griffin S, Claxton K, Palmer S, et al. Dangerous omissions: the consequences of ignoring decision uncertainty. Med Dec Making 2007; 27 (1): E8
12. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. J R Soc Med 2007; 100 (10): 453-60
13. National Institute for Health and Clinical Excellence. NICE guide to the methods of health technology appraisal. London: NICE, 2004
14. Cooksey D. A review of UK health research funding. London: Stationery Office, 2006
15. Office of Fair Trading. The pharmaceutical price regulation scheme: an OFT market study. London: Office of Fair Trading, 2007
16. Claxton K, Briggs A, Buxton MJ, et al. Value-based pricing for NHS drugs: an opportunity not to be missed? BMJ 2008; 336: 251-4
17. Claxton K, Sculpher MJ, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005; 14: 339-47
18. Thompson K, Graham J. Going beyond the single number: using probabilistic risk assessment to improve risk management. Human Ecologic Risk Assess 1996; 2: 1008-34
19. O'Hagan A, Luce BR. A primer on Bayesian statistics in health economics and outcomes research. Bethesda (MD): Bayesian initiative in health economic and outcomes research, MEDTAP International, 2003
20. O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain judgements: eliciting expert probabilities. Chichester: John Wiley and Sons, 2006
21. Sutton A, Ades AE, Cooper N, et al., on behalf of the NICE Decision Support Unit. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008; 26 (9): 753-67
22. Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. JRSS (A) 2005; 16 (1): 5-35
23. Ades AE, Claxton K, Sculpher MJ. Evidence synthesis, parameter correlation, and probabilistic sensitivity analysis. Health Econ 2006; 15: 373-81
24. Bojke L, Claxton K, Palmer S, et al. Defining and characterising structural uncertainty in decision analytic models [CHE research paper 9]. York: Centre for Health Economics, University of York, 2006
25. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. Pharmacoeconomics 2000; 17 (5): 501-13
26. Draper D. Assessment and propagation of model uncertainty. J Royal Stat Soc 1995; 57: 45-97
27. Chatfield C. Model uncertainty, data mining and statistical inference (with discussion). J Royal Stat Soc 1995; 158: 419-66
28. Bojke L, Claxton K, Bravo Vergel Y. Using expert elicitation to resolve issues of structural uncertainty in decision analytic models. Med Dec Making 2007; 27 (1): E8
29. White IR, Pocock SJ, Wang D. Eliciting and using expert opinions about dropout bias in randomised controlled trials. Clin Trials 2007; 4 (2): 125-39
30. Leal J, Wordsworth S, Legood R, et al. Eliciting expert opinion for economic models. Value Health 2007; 10 (3): 195-203
31. Griffin S, Claxton K, Hawkins N, et al. Probabilistic analysis and computationally expensive models: necessary and required? Value Health 2006; 9: 244-52
32. Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. Stat Med 1999; 18 (23): 3263-82
33. O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. Health Econ 2007; 16 (10): 1009-23
34. Stevenson M, Oakley J, Chilcott J. Gaussian process modelling in conjunction with individual patient simulation modelling: a case study describing the calculation of cost-effectiveness

ratios for the treatment of established osteoporosis. Med Dec Making 2004; 24: 89-100

35. Ades AE, Lu G, Claxton K. Expected value of sample information in medical decision modelling. Med Dec Making 2004; 24: 207-27
36. Van Hout BA, Al MJ, Gordon GS, et al. Costs, effects and c/e-ratios alongside a clinical trial. Health Econ 1994; 3: 309-19
37. Fenwick E, Claxton K, Sculpher MJ. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 2001; 10: 779-89
38. Fenwick E, O'Brien B, Briggs AH. Cost-effectiveness acceptability curves: facts, fallacies and frequently asked questions. Health Econ 2004; 13: 405-15
39. Colbourn T, Asseburg C, Bojke L, et al. Preventive strategies for group B streptococcal and other bacterial infections in early infancy: cost effectiveness and value of information analyses. BMJ 2007; 335: 655-62
40. Pratt J, Raiffa H, Schlaifer R. Statistical decision theory. Cambridge (MA): MIT Press, 1995
41. Yokota F, Thompson KM. Value of information literature analysis: a review of applications in health risk management. Med Dec Making 2004; 24: 287-98
42. Poynard T, Munteanu M, Ratziu V, et al. Truth survival in clinical research: an evidence-based requiem? Ann Intern Med 2002; 136: 888-95
43. Philips Z, Claxton K, Palmer P. The half-life of truth: what are appropriate time horizons for research decisions? Med Dec Making 2008; 28: 287-99
44. Brennan A, Kharroubi S, O'Hagan A, et al. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. Med Dec Making 2007; 27: 448-70
45. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. Med Dec Making 1998; 18: 95-109
46. Griffin S, Claxton K. Interpreting the expected value of perfect information about parameters. Med Dec Making 2007; 27 (1): E8
47. Brennan A, Kharroubi S. Expected value of sample information for Weibull survival data. Health Econ 2007: 16: 1205-25
48. Conti S, Claxton K. Dimensions of design space: a decision theoretic approach to optimal research design [CHE research paper 38]. York: Centre for Health Economics, University of York, 2008

Correspondence: Professor *Karl Claxton*, Centre for Health Economics and Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK.
E-mail: kpc1@york.ac.uk