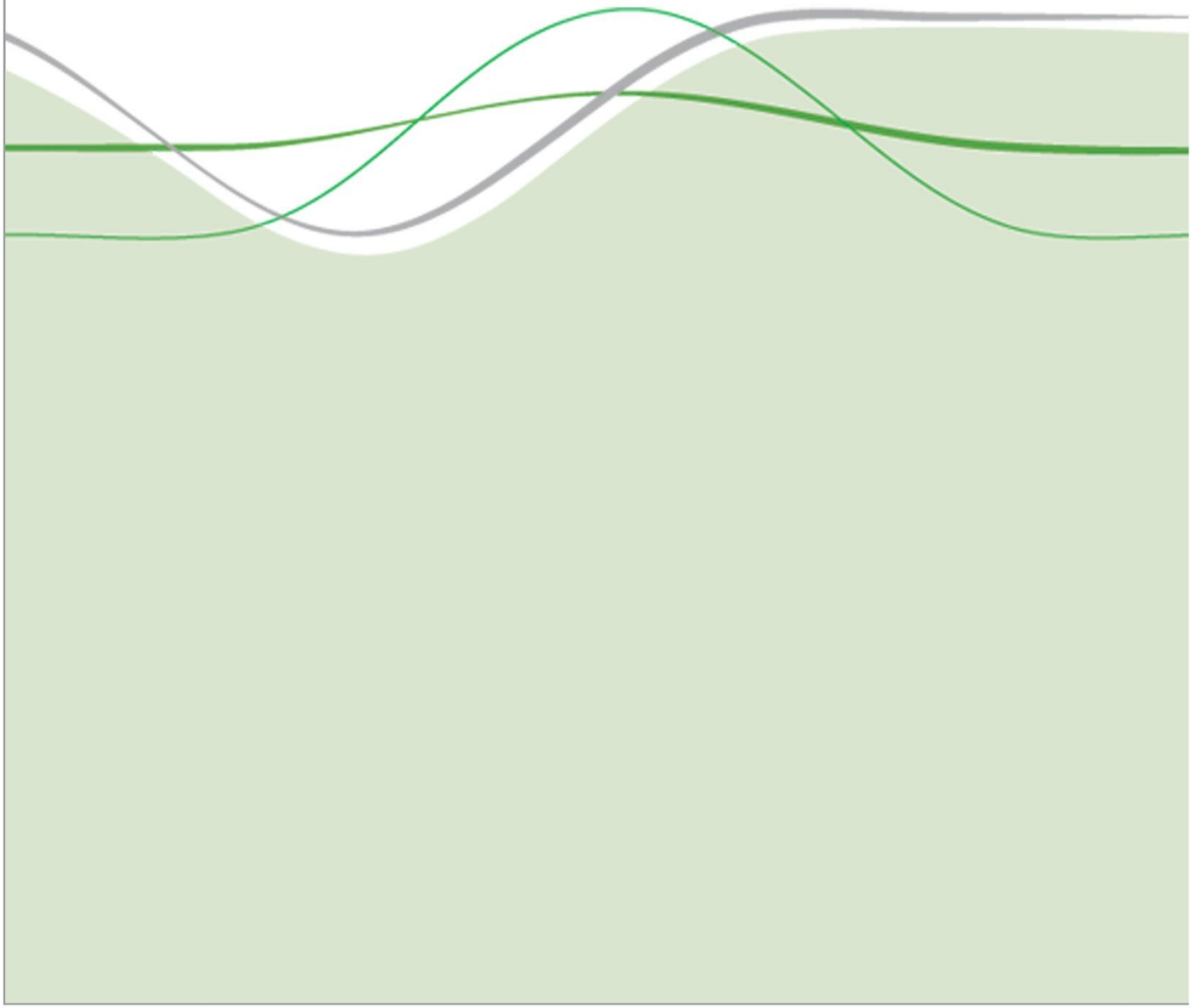
Office of Health Economics Research

Research Paper 16/05

Uncertainty and Risk in HTA Decision Making

November 2016

Paul Barnsley, Patricia Cubi-Molla, Alastair Fischer and Adrian Towse



About OHE Research Papers

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication.

Any views expressed are those of the authors and do not necessarily reflect the views or approval of OHE, its Editorial or Policy Committee, or its sponsors.

Once a version of the Research Paper's content is published in a peer reviewed journal, that supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Funding and acknowledgements

The work was commissioned from OHE by the Pharmaceutical Oncology Initiative.

Acknowledgements

Most of the work on this paper was undertaken in 2014 by Paul Barnsley with help and review from Adrian Towse. Following comments from an OHE Editorial Committee reviewer it was revised in the second half of 2015 and the first half of 2016 by Patricia Cubi-Molla with help from Alastair Fischer. The final review and edit was undertaken by Adrian Towse.

Uncertainty and Risk in HTA Decision Making

Paul Barnsley, Patricia Cubi-Molla, Alastair Fischer and Adrian Towse

November 2016

Research Paper 16/05

For further information please contact:

Nancy Devlin

ndevlin@ohe.org

The Office of Health Economics
(a company limited by guarantee of registered number 09848965)

Southside, 105 Victoria Street

London SW1E 6QT

United Kingdom

Tel: +44 207 747 8858

Table of Contents

Executive Summary 1
1. Introduction 4
2. Terminology: what is uncertainty and what is risk? 5
2.1. Uncertainty aversion 5
2.2. Risk aversion 5
2.3. Variance and potential bias 6
3. Uncertainty in the Rule Books and in Practice 7
3.1. Policy documents and uncertainty 7
3.2. Uncertainty in NICE Technology Appraisals 8
4. Considerations for a certainty premium in HTA as a result of risk aversion 9
4.1. Uncertainty about expected health outcomes 12
4.1.1. Health gains 12
4.1.2. Patients treated 14
4.1.3. Aggregate health production 14
4.1.4. Uncertainty about the health production function 15
4.1.5. Budget impact 17
4.2. Uncertainty about the threshold (or the optimal allocation of the health budget) 17
4.3. Loss aversion 19
4.4. Consequences of excessive risk aversion 19
5. Uncertainty due to estimating expected outcomes: variance and potential biases 20
5.1. Uncertainty due to sampling variability: Uncertainty in a dynamic setting 20
5.2. Bias and ambiguity 23
6. Recommendations for decision making in HTA 26
7. References 27

EXECUTIVE SUMMARY

The quality of decision-making in key public sector bodies dealing with resource allocation is a major determinant of their efficiency. One of the most difficult and contentious areas of decision-making is the way that uncertainty is dealt with. Uncertainty is an unavoidable part of all decision-making, and in particular, of the way that choices are made in HTA.

This report is concerned with uncertainty as it affects the cost effectiveness aspects of HTA. A review of the policy documents governing cost benefit analysis and cost effectiveness analysis in respect of HTA in England and Wales found a significant amount of concern about the handling of uncertainty over decision outcomes, but no detailed advice as to how it should be incorporated into the decision making process. There are good reasons for this lack of detail, which later sections of this report make clear.

In the interviews we conducted with former NICE decision makers, we found that they were very concerned about uncertainty when assessing technologies, but that they differed as to their views of which kinds of uncertainty were of most significance and how to weigh them against other important considerations. Former decision makers noted that NICE Committees could benefit from additional guidance on when and how uncertainty around ICERs should influence their decisions.

Uncertainty has a number of definitions, and a number of components, which we set out below.

Uncertainty aversion and Risk aversion

The first interpretation of uncertainty derives from risk aversion, which reflects diminishing marginal utility. This is the assumption that people appear to value a unit of money more highly if they own very few £ in total, compared with if they own many £. This means that they do not regard a 50:50 chance of gaining (say) £1000 or nothing, as the same as the certain receipt of £500. Most people would be willing to accept a certain figure of less than £500 rather than the gamble, even though the expected (in the sense of "average") monetary outcome is the same.

The obvious question then is "should NICE Appraisal Committees be risk averse and so look for a certainty equivalence?" In other words, should they adjust the threshold ICER downwards or reduce their probability of approving the technology at the best estimate of the ICER in order to take account of risk aversion?

We explore several factors that contribute to risk aversion in relation to decision making about using new technologies in the NHS:

- We as individuals may value an extra QALY less than a QALY we might lose. There is evidence that we have very slight risk aversion in relation to our health, but not at a level that would justify it being reflected in NICE decision making. Most of us have a strong preference for health, want more and do not want to lose any;
- We as a society may place a lower value on delivering additional QALYs for the same patient even though the patient does not. There is evidence that disease severity matters to the public. When a person receives an intervention that improves their health, their disease severity is being reduced, and each additional part of a QALY they receive from the intervention therefore has less social value than the previous part. Again, the effect is likely to be very small and can be ignored in practice.

- For the same aggregate community QALY gain, decision makers may favour an intervention giving a large QALY gain for a small number of patients over another giving a small QALY gain for a large number of patients. However, the analysis in this report argues that this is not an issue related to risk aversion.
- Decision makers may be more averse to approving interventions when the baseline health of the population is already high, especially when there is greater uncertainty about the likely QALY gain. But this is unwarranted. There is no reason to think we may start to lose interest in getting additional health "produced" for the population.
- More plausibly it may get harder to produce additional health gain. Each additional £1m we invest in the NHS produces less additional health gain than the previous £1m. This is an empirical issue with contradictory evidence. Moreover, most if not all NHS projects evaluated by NICE are small in relation to the total NHS budget. In any case, it would not be right to compensate for any effect through risk aversion on the part of the Committee leading to an implicit or explicit arbitrary reduction in the ICER.
- The budget impact of adopting a new technology will vary. If we have a fixed health budget, then money converts into QALYs, and there is no new case for risk aversion. If there is new, non-NHS, money for NICE decisions, then Treasury rules on investment appraisal, which recommend a very small adjustment for risk aversion, would apply. This is not currently the case, however.
- If there is uncertainty about the size of the NICE cost effectiveness threshold, this should be dealt with (if at all) by changing the threshold itself and not by adjusting the best estimate of the ICER of the technology being appraised.
- Risk averse behaviour can be encouraged by our perceptions of the probability of health loss; we weight losses significantly more heavily than gains, and we overestimate the likelihood of very unlikely events whilst underestimating the probability of more likely events. However, there is no rationale for loss aversion which involves misapplying probabilities.

Thus we find limited reasons for a NICE Appraisal Committee to be risk averse. This is the case providing society does not value QALYs differently according to disease severity, population- or budget-impact. There is a case for a very small amount of risk aversion but in policy terms is best ignored. However, this is only part of the broader case for "uncertainty aversion" which we outline below.

We have been considering a world where the different states of nature (outcomes and probabilities of achieving them) are known in order to consider the case for being risk averse in relation to those expected outcomes.

Uncertainty aversion in relation to variance and potential bias

The second view of uncertainty aversion that we consider is uncertainty described in terms of variability in the probabilities and outcomes associated with people receiving each treatment option, as they are estimated from observing a sample of people. That introduces two potential problems, variance and the possibility of bias. This is the main way that a NICE Appraisal Committee considers uncertainty in the context of economic evaluation.

Uncertainty in the form of variance arises because information is sampled. At the effectiveness stage of a health technology appraisal, the higher the variance of

estimated effectiveness, the wider the confidence interval of the estimated mean effect and the lower the probability that the intervention is deemed effective, where effectiveness requires meeting a confidence interval (e.g. 95%). This is not usually contentious, except occasionally for subgroup analysis, and for rare but serious adverse events. There may also be biases that have unknown directions and/or unknown sizes. Here uncertainty remains, because unlike random errors, biases (by definition) do not cancel out as sample size increases. However, the extent of the uncertainty cannot be determined if we do not know what biases there are, their direction or magnitude.

At the cost effectiveness stage, decision theory is used, which assumes that the variance of the incremental cost effectiveness ratio (ICER) can almost entirely be ignored. This is because, implicitly, the NHS can be regarded as an insurer who can pool the risk from many intervention decisions. This has the effect of allowing it to act as if it is risk neutral. Thus ICER uncertainty should form very little part in the decision process except to estimate whether more information is required (either collected alongside the use of an intervention, or as additional evidence to inform the adoption decision) to improve the efficiency of decision making. This occurs by reducing the uncertainty of the estimate of the ICER in such a way that it reduces the likelihood of making the (*ex post*) wrong decision. A value of information (VOI) approach can help determine the costs and benefits of additional research. In this context, we note that since the main part of the work on this paper was done, NICE has introduced reforms in its processes to handle the new arrangements for the Cancer Drugs Fund. The new arrangements include Managed Access Agreements between NHS England and pharmaceutical companies, setting out the terms of a drug's entry into the CDF and the means by which data will be collected to resolve any uncertainty relating to a drug's clinical- and cost-effectiveness. This introduces a formal mechanism to enable NICE to recommend adoption by the NHS whilst additional evidence is collected (sometimes called a "coverage with evidence development" or "only with research" recommendation).

In conclusion, the NICE Methods manual states:

> *"In particular, advisory bodies will be more cautious about recommending a technology when they are less certain about the ICERs presented in the cost-effectiveness analysis."*

Our assessment is that:

1. Members of the NICE Appraisal Committees and Guidelines Committees should take a risk neutral approach to handling uncertainty;
2. Sampling variance may be ignored in the consideration of cost effectiveness except insofar as it helps assess the potential for efficiently collecting more information (i.e. the Committee might suggest adopting the technology on condition of additional information collection, or reject use until the information is forthcoming);
3. The Committee should look for the most plausible ICER. The choice of the relevant ICER should reflect concern about potential biases arising from key modelling assumptions or data sources. Again, the potential for efficiently collecting more information should be explored.

1. INTRODUCTION

Decision makers in HTA are faced with uncertainty as to the real world outcomes associated with their choices. A review of policy documents, and interviews conducted with former members of Appraisal Committees suggest that HTA committees tend, ceteris paribus, to prefer certainty of outcome and are willing to select options with a lower expected value in order to reduce the variation between best and worst case outcomes. These can reflect both clinical uncertainty and concern about the assumptions made in the cost-effectiveness modelling. (Williams, et al., 2008)

In this report, we will consider whether this preference for certainty is justifiable on policy grounds. Since HTA, whether viewed through a welfarist or extra-welfarist lens,[1] attempts to maximise something other than a standard financial Return on Investment (ROI), approaches to risk imported from a purely financial context will often be inappropriate. However, other approaches to thinking about uncertainty in relation to biomedical evidence may well be relevant.

Our focus will be on consideration of approaches to thinking about decision uncertainty in relation to its application to the Cost Effectiveness Analysis methodology applied by the National Institute for Health and Care Excellence (NICE) in England and Wales. This comprises: (i) estimation of Incremental Cost Effectiveness Ratios (ICERs), calculated on a cost per Quality Adjusted Life Year (QALY) basis; (ii) comparison to a threshold based on an assumed opportunity cost of a QALY; with (iii) ad hoc adjustments to the ICER for non-QALY factors, including the uncertainty of the ICER; (iv) reaching a yes/no/restricted decision for use within a given patient population.

The report does not cover the way that uncertainty, at the stage of evaluating the effectiveness of an intervention, is handled by HTA bodies. This methodology follows a well-trodden path that has been relatively uncontentious in comparison with the second, cost-effectiveness, stage of the analysis. In summary, most HTA evaluations involve one or more adequately-powered randomised controlled trials (RCTs). The analysis is normally carried out using a frequentist methodology. The proportion of 'false-positive' results (accepting a treatment intervention that is less effective than the most usual existing treatment) is reduced to an acceptably-low level by the use of confidence intervals. These are set customarily at 95%. It is not the role of the cost effectiveness stage of the analysis to act as a further such filter.

The general approach to the validity of preferences over uncertainty as to a technology's effectiveness and cost are potentially applicable to other HTA systems, not only those which seek to allocate a "fixed" health care budget. The notion of a "decision" can be broadened to encompass incremental pricing decisions in therapeutic added value based pricing systems such as France's.

However, before considering any of these things, we define our terms and summarise the approaches of NICE and the UK Treasury.

[1] For reasons of readability we will tend to make use of welfarist terminology to describe the maximand of HTA decision making. The argument made are, we believe, consistent with both the broad extra-welfarist perspective including non-goods characteristics and non-utilitarian valuation of capabilities and the narrower version of extra-welfarism which views health (however defined) as the maximand of HTA policy. For "welfare" read "flourishing" or "health" for these broad and narrow approaches to extra-welfarism respectively.

2. TERMINOLOGY: WHAT IS UNCERTAINTY AND WHAT IS RISK?

2.1. Uncertainty aversion

Uncertainty matters because we have a preference for certainty – this is, for an outcome which is consistent across all states of the world, or for outcomes which differ little between states of the world. We can call this "uncertainty aversion".[2] We use "uncertainty aversion", "preference for certainty" and "certainty premium" as neutral descriptors of an observed preference for predictable outcomes over unpredictable ones.[3] However, there are different types of uncertainty, and so of uncertainty aversion, which require different responses.

2.2. Risk aversion

If we know the possible outcomes and the probability of them occurring (we shall refer to this as 'knowing the true state of nature'), we may then experience a particular type of uncertainty aversion that is termed risk aversion and derives from the theory of risk aversion used in economics and finance. This is used extensively in the business world, and is perhaps the view best known to economists in general.

Risk aversion manifests itself when people take out insurance policies on large items such as their houses or cars. The theory explaining such behaviour is based on the assumption that if someone only had £1 left, he or she would value it more than they would value an additional £1 if they already had a £1m. By extension, it would mean that someone with £500,000 would not risk gambling it by tossing a coin, where they received £1 million if the coin came down tails and zero £ if it came down heads.

On the other hand, if they were risk neutral, they would be indifferent between the certain £500,000 and the gamble described. In this situation all probabilities and outcomes are known with certainty. However, for the person making a decision about whether to insure or not a house, the likelihood that the house might catch fire or be burgled is not known. It might be possible to come up with an estimate of the likelihood based on observation of incidence in a large population of houses with similar relevant attributes. But there remains an unknown element. However, it is the degree of risk

[2] We note that this is not the standard usage of this term, which has traditionally referred only to a preference for lotteries with known probabilities over those with unknown probabilities, and sometimes broadened to include a preference for probabilistic uncertainty over Knightian uncertainty – see Schmeidler (1989) and Epstein (1999). These references use "uncertainty averse" in situations where, for example, subjects display a preference for a known payoff distribution over an unknown payoff distribution, even though the known distribution cannot be shown to have a higher expected value. We acknowledge the significance of this phenomenon, but for practical reasons wish to use the term "uncertainty aversion" more broadly in this paper so as to be able to distinguish the economic phenomenon of "risk aversion", arising from diminishing marginal utility of the payoff variable to an otherwise expected utility maximising agent, from a more general dislike of risks. It also enables us to avoid tautological justifications for certainty preference ("the decision maker's aversion to risks is justified because they are risk averse") and so distinguish between a positive description of choices ("the committee is uncertainty averse") and a normative justification or criticism of them (such as "the committee's uncertainty aversion arises because they are (in)appropriately risk averse").

[3] Formally, we might imagine uncertainty averse preferences, as we use the term, being modelled via a value function for which the second derivative with respect to the uncertain variables is negative.

aversion rather than uncertainty about the exact probability or outcome that will determine the decision whether or not to take out insurance.

We need to separate financial risk aversion as it affects individuals from issues relating to health effects and the NHS in two important ways. The first is because a decision maker with many and diverse health intervention decisions to make should take a perspective on financial risk that reflects this. The second is because we are dealing in health effects as well as money to pay for them, and it is the uncertainty around health effects that we are most interested in, as in many instances, health costs are known.

To continue on this second point - risk aversion, in the way the term is used in economic theory, does not refer to an aversion to risks. A risk averse individual, according to economic theory, is assumed to be an expected utility maximiser, completely indifferent to the distribution of payoffs across different states of the world, and indifferent between accepting and not accepting fair gambles across arbitrarily large payoff values, provided that the payoffs in question are denominated in terms of *the value the individual ultimately obtains from them*.

Risk aversion, then, is not a property of an individual's preferences across risk but an *emergent* property of a relationship between final payoffs and the intermediate goods in which those payoffs happen to be denominated: we are risk averse only when there are, and only because of, diminishing marginal returns to the currency in which the risk is expressed.

Since money is an intermediate, not final, object, one that translates into utility-increasing consumption at a diminishing rate (because high utility purchases are prioritised), financial risks will tend to generate rational risk averse behaviour emergent from the underlying diminishing marginal utility of money. Other goods – such as health - may or may not display similar diminishing marginal returns, and therefore similar preferences in relation to their distribution across states. There is no reason to expect this to be true for all goods or for the rates at which marginal utility diminishes in different goods to bear any relation to each other.

2.3. Variance and potential bias

In health care, uncertainty arises in very diverse situations such as not knowing the exact number of years a patient is expected to live after the treatment; or the increase in quality of life that a patient can expect from a treatment; or ignoring which health resources we will be foregoing when the money spent is diverted to the coverage of the new treatment. However, we may be able to estimate the likelihood or probability of any particular increase in quality of life that a patient can expect from a treatment. The most usual ways of doing this are:

(i) Sampling from a source of adequate and relevant information. We can collect data from a cohort of patients, and construct a probability distribution. This is usually done using the "relative frequency" interpretation of probability, although in principle a Bayesian approach can be used. This can help us estimate the most likely value of the increase in quality of life, as well as construct confidence intervals around the estimate. If so, although the increase in quality of life for a patient is a random event, the decision maker *knows the probability* of an increase of a particular size happening;

(ii) Modelling. In more complicated scenarios, the right mechanism for estimating a probability is not so straightforward. For instance, overall survival cannot be based

on a "wait and see" data-generating mechanism as in the previous example, since this would delay decision making by months or years. In this case, different models of survival analysis have been developed to estimate the probability distribution that most likely describes the reality. The decision maker could find these models reliable if there were general agreement about the validity of both the mechanism used to generate probabilities and the required assumptions or prior beliefs based on the data available.

In these cases we may or may not have a degree of uncertainty about the potential states of nature and the probability of them occurring.

If we do *not* know the possible outcomes and/or the probability of them occurring then a second form of uncertainty aversion occurs. We shall refer to this situation as 'not knowing the true state of nature'. The errors may be random ones, which give rise to variance, which may be reduced by increasing sample size. According to this interpretation, uncertainty in the form of variance arises because information is sampled. Additionally, the precise forms of the equations used to model outcomes based on the clinical evidence are not known. The errors may be systematic ones, such as not having the correct form of the equation, in which case a bias is introduced.

3. UNCERTAINTY IN THE RULE BOOKS AND IN PRACTICE

3.1. Policy documents and uncertainty

The current NICE Guide to Methods of Technology Appraisal (2013) refers to uncertainty in the second bullet point of section 6.3.3. It states that:

> *"Above a most plausible ICER of £20,000 per QALY gained, judgements about the acceptability of the technology as an effective use of NHS resources will specifically take into account [inter alia]… the degree of certainty around the ICER."*
>
> *"In particular, the Committee will be more cautious about recommending a technology when they are less certain about the ICERs presented."*

We can contrast this with the advice given in relation to risk aversion (in its technical sense) by the UK Treasury's "Green Book" on Appraisal and Evaluation in Central Government (UK Treasury, 2011).The Green Book notes that:

> *"A decision-maker who is risk averse cares about this potential variability in outcomes, and is willing to pay a sum in exchange for certainty (or willing to put up with variability on receipt of compensation). This compensation is the cost of variability, and should be included in appraisal when it is considered appropriate."*

As we will discuss in detail below, there is an important distinction between variability in outcomes denominated in QALYs and those denominated in monetary terms.[4] Dealing with monetary risks, the Green Book adopts a broad definition of "risk aversion" – any concern about variability in outcomes, for any reason – but in practice proposes a metric

[4] Strictly, there is a further distinction between projects which produce changes in income and those which produce changes in individuals' endowments of other goods, even where the value of those goods is denominated in monetary terms. We discuss this issue in more detail below.

for the appropriate certainty premium which will tend to yield very low levels of risk aversion in practice.[5,6]

The approach suggested by the Green Book involves some relatively strong assumptions about diminishing marginal utility of income, but produces relatively plausible estimates for financial risk aversion: it suggests that beneficiaries of a project which gives them either £200, £0 or -£100 with equal probability would, assuming an average starting income of £20,000, give up approximately 50 pence in order to receive a certain payoff of equal expected value (£33 vs approximately £32.50) instead.

The NICE Methods Guide is not as explicit in its approach to uncertainty. There is, in our view, an unstated assumption, in both current and proposed versions of the Guide, that ICER uncertainty justifies some degree of uncertainty aversion, such that more uncertain ICERs should be tested against a lower maximum ICER threshold than more certain values.

3.2. Uncertainty in NICE Technology Appraisals

In addition to reviewing the policy documents which guide decision making, the authors conducted interviews with two former NICE Appraisal Committee members, focussing on their attitudes to uncertainty and their understanding of risk-related terminology. This sample of interviewees is small and was selected on the basis of convenience. Given, however, that committee members tend to learn from each other and tend to come to a relatively homogeneous group view, these observations will help us understand committee members' attitudes to, and concerns about, uncertainty.

Key observations from the interviewees were as follows:

- "`Risk aversion' is a psychological framework and describes people who are particularly averse to taking risks… For example, one of the problems with the appraisals of beta interferons is the uncertainty over how long the benefits would persist beyond the life of the clinical trial"
- "There is always uncertainty in cost-effectiveness analysis. There is the risk that those uncertainties are such that you might say "yes" to a product or a technology which would have the consequence of depriving other people of cost-effective care (opportunity cost)"
- "You have to make a judgement about whether the most plausible ICER is beyond what you're comfortable with because of the opportunity cost"
- "It is very important to appreciate that there is a lot of judgement that has to take place…Perhaps HTA bodies are not good at expressing those judgements – as a decision-maker, explaining why you arrived at the judgement is often very difficult"
- "If you are risk neutral, you ignore uncertainty if there is an equal chance of the world being better than you thought as being worse than you thought (by the same amount)… if you are making lots of decisions, on balance, your best estimates will

[5] It states that "the fraction of income worth paying for certainty (C) is approximated by the expression: C= - var(y) / 2y* where y is the net additional income resulting from the proposal, and y* is the total expected income or benefits of those impacted by the proposal (…) Given the size of national income relative to the scale of most individual projects, the cost of variability for projects that benefit the community as a whole is usually negligible."

[6] Arrow & Lind (1970) showed that risk involved in government-funded projects could not only be spread among many projects, but that in funding such a project, the risk to individual taxpayers would also be minuscule.

come out correct, however, where HTA is concerned, you are not necessarily making lots of decisions and some are rather big decisions"

- "Uncertainty is sometimes ignored when [committee members] are happy about the way the decision is going and might figure more highly in the discussions when more members of the Committee are unhappy about the way the decision is going"
- "NICE's guidance to Committees should be much more explicit in how they are advised to handle uncertainty".

These comments indicate concerns about combining the tasks of arriving at the most plausible ICER and estimating the degree of uncertainty around that value. There is an understandable focus on the harms associated with *ex post* incorrect approvals, which logically should be mirrored with corresponding concerns in relation to incorrect rejections. Risk aversion is treated as both a description of and justification for uncertainty averse preferences.

The views stated by former decision makers can be cross checked with the revealed published reasoning behind a number of Technology Appraisal (TA) decisions, which we identified selectively via a search of the NICE decision database for the terms "uncertainty" and "risk".

For illustrative purposes, because the cause of the uncertainty is quite different, we show here two examples from NICE appraisals.

1. In TA174 for Rituximab, a leukaemia drug, the Committee noted that:

   "…the results became very sensitive to the difference between the utility values used for the progression-free survival health state and the progressed health state. The Committee considered the lack of appropriate utility data contributed to substantial uncertainty in the economic modelling"

Here the problem highlighted by the Committee is that it is not very clear whether utility is very well measured by the instruments used, so creating uncertainty (because of the potential for bias) about the results of the economic model. To reduce uncertainty would require the collection of a more relevant set of utility data.

2. In TA244 for Roflumilast, the Committee stated:

   "The Committee noted that the relative risk of 0.84 for reduction in exacerbations for roflumilast…was associated with considerable uncertainty and that many of the results …had confidence intervals for the relative risks that crossed one, indicating a statistically non-significant result. It concluded that further data were needed so the Committee could be more certain about the cost effectiveness of roflumilast."

In this second example, the relative risk for reduction was associated with considerable uncertainty. Although it is clear what it is measured, here the problem is that the degree of uncertainty around the mean is quite large and the committee are uncomfortable about that. This, in principle, could be solved by a bigger study which could be expected to reduce the confidence limits around the mean.

4. CONSIDERATIONS FOR A CERTAINTY PREMIUM IN HTA AS A RESULT OF RISK AVERSION

If the decision maker is indifferent to uncertainty, then the existence of randomness should not affect the optimal decision, because only the expected values are

contemplated. For instance, if the decision maker were indifferent to uncertainty around QALYs, then she would find the following scenarios as equivalent:

Scenario 1: a gain of 1 QALY with probability 1

Scenario 2: a gain of 0 QALYs with probability 0.5, and a gain of 2 QALYs with probability 0.5

If the decision-maker is not indifferent between scenarios 1 and 2, she will most likely be risk averse, that is, she will prefer scenario 1. This is equivalent to saying that the extra QALY that might be obtained is less valuable than the QALY that might be lost. From a classical economic viewpoint (expected utility theory), this attitude would be explained in terms of **diminishing marginal returns to health** (or concavity of the payoff as a function of health, regardless what we understand by "payoff"). Note that this reasoning can be extended to other parameters involved in the decision making process that are different from health, for instance the cost of the intervention or the health budget. Note also that in this example the alternative outcomes and their probability of occurring (the alternative states of nature) are known. This enables us to focus on response to this uncertainty in terms of known risks of alternative outcomes. In the remainder of this paper we examine which attitudes to uncertainty are theoretically justifiable and whether we can reach any overall conclusions as to the appropriate size and structure of the certainty premium in HTA decision making. Finally, note that from another viewpoint (prospect theory) we could attribute this mentality to a different psychological valuation of losses and gains by the decision maker, also referred as **loss aversion**. This point will be also discussed further.

In the sections which follow we will consider whether observed risk aversion when the true state of nature is known by HTA decision makers can be justified as part of an attempt to maximise some measure of societal welfare, or as reflecting societal preferences. Therefore, in order to illustrate our argument, and establish an easier connection between the microeconomic foundation of "risk aversion when the true state of nature is known" and HTA decision-making, we base our discussion on an explicit extra-welfarist policy model.

Any welfarist social value of a given QALY distribution is capable of being written as a weighted sum of the value to individuals of their QALY allocation. Kaplow and Shavell (2001) also show that any extra-welfarist social welfare function[7] can be written as the sum of its welfarist component and an additional extra-welfarist term. Finally, extra-welfarism which focuses narrowly on QALY-denominated health can be simulated using a weighted QALY maximising social welfare function.[8,9]

We follow an explicit extra-welfarist policy model suggested by Meltzer & Smith (2012) as a representation of the generalized model presented in several papers as Claxton et al. (2010), Griffin et al. (2008) and Gravelle et al. (2007). We contemplate a two-step

[7] In the broad sense of seeking to maximise some wider measure of flourishing of which utility is a subset.

[8] The preferences of extra-welfarists who argue for the maximisation of a non-QALY person-weighted health measure can be modelled in a welfarist approach by using an additional, additively separable element to represent the value of non-QALY health, to which Kaplow and Shavell's result applies by analogy.

[9] An extra-welfarist might assert that individuals are risk neutral across life years but not across quality adjustments, meaning that QALYs are not a measure of welfare but rather an extra-welfarist proxy for health. This position would still require an independent philosophical justification for the diminishing marginal value of individual QALYs.

decision making procedure. First, a "legitimate authority" defines the objective social function for the resource allocation and fixes the budget for health care. The social welfare function will depend on health benefits and the net consumption benefits to society in areas other than health. A second policy maker will choose the optimal resource allocation for health, trying to maximise the social function under the budget constraints fixed by the first policy maker. This model fits well the UK health policy context, where the first and second policy makers could be the Cabinet and the Department of Health or NICE, respectively. This is also in line with a series of papers (for example McCabe, et al. (2008) who argue that it is not NICE's responsibility to determine the ICER threshold (and equivalently the health budget) but to seek it out and use it for "maximising health gain from limited resources" (NICE, 2004).

We assume *n* independent potential health programmes, with a single technology *i* available at an additional cost c_i (every technology is compared to standard care which has a zero cost). The cost of such technology has to be covered by the health budget (*B*). We assume that the total policy budget (*M*) is also fixed, and (*M* – *B*) is the budget for the total production of non-health-related goods. In order to simplify, the model takes the total investment in non-health goods as a proxy for the total benefit derived from these goods (*x*). We assume that the health benefits of programme *i,* h_i, are equal for each of the π_i individuals who benefit from the treatment. The decision variables, $\lambda_i \epsilon [0,1]$ with $h = \sum_{i=1}^{n} \pi_i \lambda_i h_i$ being the total amount of health benefits from the *n* programmes and $c = \sum_{i=1}^{n} \pi_i \lambda_i c_i$ the total cost, reflect the amount of treatment *i* that will be purchased.[10] Finally, let $f(h,c)$ be the classical health production function, with the total cost (*c*) as input, and the aggregate health *h* as output. We write the model as:

$$\left\{ \begin{array}{c} Max\, W(h,x) \\ subject\ to\ f(h,c) \geq 0 \\ c \leq B \\ B + x \leq M \end{array} \right\}$$

This model establishes the societal decision rule:

$$v\left(h_i - \frac{c_i}{k}\right) \geq 0,$$

where *k* represents the optimal cost-effectiveness threshold (the shadow value associated to the budget constraint), and v reflects the marginal value/social value of health (with respect to non-health sector), that is, $v = W_h/W_x$, with $W_h = \partial W/\partial h$ and $W_x = \partial W/\partial x$.

Note that $\frac{c_i}{k}$ represents the opportunity cost of the intervention in terms of health foregone, since the cost of the new treatment has to be financed through the health budget. Thus, a new technology will be adopted if the social value of the net health benefits (health gain from the treatment h_i minus health foregone) is non-negative.

Note also that a reduced form of the model (with $W(h) = h$) reflecting simply the impact of a decision in the health care sector would be (also following Meltzer & Smith (2012)):

[10] Note that we could also consider $\lambda_i \epsilon \{0,1\}$ for indivisible treatments and yes/no decisions.

$$\left\{ \begin{array}{c} Max\ h = \sum_{i=1}^{n} \pi_i \lambda_i h_i \\ subject\ to \sum_{i=1}^{n} \pi_i \lambda_i c_i \leq B \\ \lambda_i \epsilon [0,1] \end{array} \right\}$$

This formulation is frequently used in the framework of using mathematical programming to inform decisions (e.g. Epstein et al. (2007)). The societal decision rule derived from the new formulation would be:

$$h_i - \frac{c_i}{k} \geq 0$$

Or equivalently:

$$\frac{c_i}{h_i} \leq k$$

The formula above indicates that the treatment's cost-effectiveness ratio has to be no greater than the threshold adopted by the decision maker (again optimally equal to the shadow price of the budget constraint); that is, the classical CEA decision rule.

4.1. Uncertainty about expected health outcomes

Accordingly within our model, we can consider the expected number of QALYs in each treatment (h_i), the aggregate number of QALYs (h) and the health production function$f(h,c)$, and the number of patients involved in every treatment (π_i).

In the following sections we explore the potential justification for diminishing returns to health gains (QALYs), as well as the rationale for presuming the existence of diminishing marginal returns to the number of patients treated and diminishing marginal returns to the aggregate health production.

4.1.1. Health gains

We explore the potential justification for different diminishing returns to QALYs at two different levels:

- The individual utility of health shows diminishing returns to health. That is, following the example above, the individual herself does not view two QALYs as being exactly twice as valuable as one;
- The societal utility (or wellbeing) produced by health gains shows diminishing returns to health. That is, society views the extra QALY that the patient (or a particular patient cohort) may obtain as less valuable than the penultimate QALY she (or they) gained.

One challenge with the first bullet point is at a technical level. It arises directly from the assumptions necessary to arrive at ratio-scale, aggregated estimates of the QALY. QALY recipients are assumed to be risk neutral across their remaining life years, and, in welfarist formulations of the HTA decision problem, across QALYs as well.[11] So individual welfare is assumed, as a direct result of the decision to use QALYs as a valid measure of total health, not to display diminishing marginal returns to additional QALYs. Referring to the model, for a technology *i* such that $\pi_i = 1$, this type of uncertainty would enter by introducing net health gains as $u(h_i)$, with $u'(h_i) > 0$ and $u''(h_i) < 0$, rather than simply h_i.

[11] See Pliskin, Shepard, and Weinstein (1980).

However, the QALY assumes risk neutrality across remaining life years: it assumes that measures of individual utility will, *ceteris paribus*, increase linearly with additional QALYs. In other words: by definition, QALYs represent the individual utility associated with a health gain. In this respect, if we are using QALYs, we have assumed away the problem of uncertainty about individual expected QALYs. Patients are assumed to be risk neutral.

If we were to relax the above assumption, necessary for the calculation and aggregation of QALYs, that individuals are risk neutral across remaining life years (i.e. we now reject the view that there are no diminishing marginal returns to QALYs on an individual level), we find that some level of risk aversion may be optimal. It is possible that the marginal QALYs subject to uncertainty would generally be small when considered on a per-patient basis. If so, the degree of diminishing marginal value to the individual patient would be small in practice unless there is extremely rapid diminution of the value of additional life years to an individual.[12] Barnsley (2013) uses an analysis of differences between standard gamble and time trade-off quality of life adjustments, which should be equal if the risk neutrality condition holds, to estimate the actual degree of risk aversion across remaining life years displayed by a small UK sample. Based on these figures, we can estimate the certainty premium for a patient with 40 QALYs remaining in relation to an expected 0.1 QALY gain instead of participating in a lottery as less than 0.00001%.

The second bullet point justification for uncertainty aversion relates to whether society has less interest in delivering the second QALY to the patient than it does in delivering the first, *even though the patient themselves values them equally*. In the model, we would find, for every patient, weighted QALYs of $g(h_i)$, with $g' > 0$ and $g'' < 0$, where $g(.)$ reflects the social value of the QALYs gained by a particular individual. The only rational basis society has for departing from the patient's own ranking of the outcomes is prioritarian: society values QALYs accruing to the "badly off" more than those accruing to the "well off" in an endowment deemed relevant by society for the allocation of health resources. Where the relevant endowment is "health" (often referred to as disease severity), the only relevant difference between the certain first QALY and the uncertain second is therefore the simple fact that the same patient receiving a second QALY is exactly one QALY healthier than the patient receiving the first. This implies that the proper level of societal risk aversion for QALY-denominated risks is given by the additional value society places on a QALY accruing to a patient who expects one less lifetime QALY than an otherwise identical patient.[13] Results derived by Brazier et al. (2013) suggest that the additional social value of a QALY in this situation might be in the order of 0.035 QALYs, implying that the expected value of the QALY from an uncertain medicine (with 50% probability) is 0.0175 QALYs below that of the certain QALY. However, this analysis is misleading: treating larger uncertain health gains as accruing to marginally less disadvantaged patients because of the disadvantage-reducing effects of the treatment itself also directly implies that a treatment providing individual patients with certain health benefits of two QALYs is less than twice as valuable to society as one providing one QALY of certain benefits. Since implementing diminishing marginal returns

[12] This effect might be more pronounced in the case of regulatory consideration, where negative realisations of uncertainty might lead to significant losses in patient health. Eichler et al. (2013) show that regulatory agencies demand a more than fourfold ratio of potential benefits to potential harms in these circumstances, though it is not clear that this level of uncertainty aversion is justified.

[13] This appears to be consistent both with the priority given to "end of life" patients in NICE's current methods guide and the proposed higher threshold to be applied to patients with high "QALY shortfall" measures in the (discontinued) NICE draft proposals to implement value-based assessment.

to individual patient gains is administratively complex and has been rejected by both NICE and the UK Department of Health (2004) despite some societal preference evidence in its favour,[14] consistency demands that larger uncertain gains be treated equivalently, meaning risk aversion should be ignored as a matter of practice, even if a slight degree of risk aversion is defensible in theory.

Note that the previous discussion can be easily extended to societal preferences for prioritising particular cohorts of patients (rather than individual patients) based on their health endowments – for example lifetime QALY expectations. However, the idea of prioritising patients' cohorts based on other endowments (e.g. income) cannot be discussed with the same argumentation, since the social value for a QALY in these cases is somewhat exogenous to the QALY itself. In this case, the expected social value of any number of additional QALYs will not be decreasing at the margin unless, for some reason, the expected *distribution* of the additional QALYs arising from the next intervention is less socially valuable than the distribution of the preceding QALYs. That is, each additional QALY coming from the patient cohort j (or h_j in the model) is just as valuable to the *patient* as the last, but QALYs may have diminishing marginal utility at the societal level if that patient cohort j receiving the QALYs is not the prioritised group – for instance, $\partial W/\partial h_j < \partial W/\partial h_i$ even if the health gain for both cohorts i and j are the same.

4.1.2. Patients treated

A special case arises where uncertainty in HTA relates not to the size of each patient's health gains, but only as to the *number* of patients who will receive a health gain of known size. For instance, if the drugs under consideration provided either one QALY to 100 patients with certainty or one QALY to either zero or 200 patients with equal likelihood then no argument from diminishing marginal disadvantage would apply, since the patient populations are relevantly identical. Nor would arguments about diminishing marginal utility of QALYs apply, since the health gain is a certain QALY.

In particular, looking at the model we observe that the societal decision rule does not change in situations where only the number of beneficiaries (π_i) is in question.

4.1.3. Aggregate health production

In this context we do not focus on valuation of QALYs at individual level but on the aggregate. Rephrasing the example at the beginning of this section, now more accordingly to the new perspective, we would say the decision maker would be indifferent to uncertainty around the total QALYs if she found a gain of 100 QALY with probability 1 equivalent to a 50% chance of only 50 QALYs and a 50% chance of 150 QALYs. In case of diminishing returns to aggregate QALYs, additional QALYs are less valuable and uncertainty matters. The extra 50 QALYs that might be obtained are less valuable than the 50 QALYs that might be lost. This statement could be a simple implication of assuming QALYs have decreasing marginal utility for individuals. However, we have argued against such a view and uncertainty in the aggregate health production can matter, regardless the assumptions made at the individual or patient cohort levels. If risk neutrality across remaining life years is assumed, then the statement "additional QALYs are less valuable" would make intuitive sense only in two cases. In the first case, we would be simply talking about society getting "satiated" of QALYs, in the sense that even if no exchange is needed, they would not want to buy new QALYs. (The only

[14] See Linley and Hughes (2013)

condition in the model would be the concavity of W in the health space h, or $W_{hh} < 0$). But this idea only makes sense in the scenario where every citizen is living up to the maximum number of potential QALYs, which is not realistic. Whilst the evidence is that we do not necessarily value the 40th QALY for an individual patient, as a society, at the same level as the first QALY, there is no reason for thinking that that social valuations would incorporate risk aversion to the total number of QALYs at stake across all patients.

Alternatively, in a second scenario, society may be willing to give up smaller and smaller amounts of other (non-health) goods than what is necessary to obtain additional QALYs (i.e. we have decreasing marginal rates of substitution between *h* and *x* or W concave in *(h,x)* or $W_{hh}<0$ and $W_{hh}W_{xx}>W_{hx}^2$). Note that assuming decreasing marginal rates of substitution is equivalent to stating that the v in the model (the marginal value/social value of health with respect to non-health sector) changes its value depending on the current h, say $\bar{h}$, rather than being a constant across every initial allocation of health at the society level. However, when incorporating this fact into the societal decision rule, $v(\bar{h})\left(h_i - \frac{c_i}{k}\right) \geq 0$, we observe that uncertainty in the total health gain h , with h_i and c_i fixed, does not affect the decision rule, as long as $v(\bar{h}) > 0$, that is, as long as the society is still willing to trade health and non-health goods, whatever the rate of substitution is.

Finally, note that the fact that society is willing to give up smaller and smaller amounts of other things than what is necessary in return for additional QALYs, is related to the idea that for buying these additional QALYs it may be necessary to sacrifice increasing amounts of other things in order to buy them. We consider this point next.

4.1.4. Uncertainty about the health production function

A decision maker may show risk aversion if she assumes that each additional dollar is buying fewer QALYs than the last one. In other words, she has a belief that there are diminishing marginal returns to general health expenditure (including the health expenditure not devoted to specifically identified technologies subject to HTA). Mirroring the model, a risk averse decision-maker would assume the health production function in its implicit form $h = \tilde{f}(c)$ to be concave, with a more general interpretation of *c* as the total expenditure on health as the input and the aggregate general health outcomes achieved *h* as output. This view implies that, under a fixed health budget, since additional expenditure by a health system is of decreasing value, increasing *reductions* in expenditure elsewhere as a result of approval of new technologies involve the sacrifice QALYs at an increasing rate. That is, for example, if we spend more in a new technology with a 50:50 chance of gaining or losing, in terms of the impact of the gain it would take it away from somewhere where we would have gotten more QALYs. This assumes that all unweighted QALYs are the same but the opportunity cost of achieving them, in terms of QALYs foregone, is rising.

Intuitively, assuming some rationality in the organisation of the health care system, we would expect health care expenditures to display diminishing marginal returns over sufficiently large changes in expenditure: were health spending to be reduced by 80%, say, we would expect the services retained to be of above-average effectiveness. Similarly, a fivefold increase in health spending would be likely to lead to the approval of relatively unproductive treatments and mobilise resources and technologies which are relatively unsuited to additional health production. This means we are interested in finding out whether the degree of uncertainty around total cost for individual HTA decisions observed in practice is sufficiently large to generate observable differences in the marginal effectiveness of health care expenditure.

Claxton et al. (2015) investigate this issue by looking at cross sectional differences in expenditure between different Primary Care Trusts ("PCTs"), local health care commissioning bodies in England and Wales. On p261 they note:

> *"To test this [diminishing marginal returns] hypothesis we used the expenditure model for each of the big four programmes to divide the 152 PCTs into two groups: those whose predicted spend is greater than the average predicted spend in that programme (ceteris paribus), and those whose predicted spend is smaller than the average predicted spend (ceteris paribus). We then re-estimated our outcome model for each of these two groups of PCTs …For all four programmes, the coefficient on the expenditure variable is larger (in an absolute sense) for the `high spend` PCTs than for the `low spend' PCTs. This result contradicts our hypothesis that `high spenders' will have a lower elasticity than 'low spenders'."*

That is: having separately estimated marginal productivities for a sample partitioned into high and low spending localities, they find that high spending localities appear to be more productive at the margin, implying *increasing* marginal productivity.

Claxton et al. then repeat this analysis, this time partitioning the sample into low and high spenders based on an alternative criterion: whether the PCT in question is over or under its "target allocation" calculated by the Department of Health (2004) on a needs-based formula.[15] This approach yields the opposite result, showing areas with higher levels of funding (relative to need) to have a "marginal" cost per life year more than 75% higher than less well-funded areas.

This inconsistent evidence in relation to marginal productivity is in some ways not surprising, since Claxton et al also find evidence that PCTs are not attempting to maximise QALYs at the margin, with costs per QALY varying widely across different disease areas.[16] If decision makers in the NHS are not operating within the ranges of expenditure we are considering as QALY maximisers at the margin, there is little reason to expect the cost per QALY to be increasing in overall expenditure, since the traditional rationale for diminishing returns, based on the most productive investments being adopted first (and disinvested from last) does not hold where decision makers use a different measure of productivity.

Overall, we conclude that diminishing marginal returns across the levels of cost per QALY uncertainty observed in practice in individual HTA decisions are likely to be extremely small and potentially even negative. Therefore it is reasonable to assume constant returns, and that the degree of cost uncertainty is likely to be extremely small relative to any curvature of the health system's marginal production function. Note that this is a separate question from what the mean opportunity cost should be and the uncertainty about the mean of what we are giving up. This is about, at the margin, do we get more from an additional unit of money spent at different levels of expenditure? If there were evidence that a NICE decision could have a non-marginal impact on expenditure then it would be more appropriate to adjust the cost-effectiveness threshold or the timing of implementation, rather than seek to compensate through risk aversion on the part of the Committee leading to an implicit or explicit arbitrary reduction in the ICER.

[15] See p263.
[16] See Table 31 on p76 of Claxton et al (2015).

4.1.5. Budget impact

We have argued above that cost uncertainty in the HTA context is in fact QALY uncertainty where there is a fixed health care budget, and that any issue of risk aversion in relation to budget impact is therefore about the issues we consider in relation to QALYs in 4.1.1 to 4.1.4 above, notably 4.1.4, our discussion about the nature of the health production function.

If the health budget is not fixed, i.e. the health care budget constraint is not binding at the relevant margin, either because additional expenditure will be covered by the Treasury or because funds will remain unspent at the end of the accounting period, they might then conclude that cost uncertainty (though not health gain uncertainty) in HTA displaces funds rather than QALYs. We are back into a world where approving new technologies leads to additional NHS spending, which leads ultimately either to an increase in taxes and lower personal consumption, or to a reduction in other government spending (e.g. on education or defence). In this context, we can characterise HTA uncertainty which potentially displaces general government expenditure as in line with the Green Book approach: "the cost of variability for projects that benefit the community as a whole is usually negligible." In other words, diminishing marginal value of money implies risk aversion, but at a very low rate since uncertainty in relation to individual decisions will be small relative to national income and the large number of risky decisions made by government means that portfolio variance will be small.

This alternative approach suggests that the appropriate level of risk aversion arising from cost uncertainty where the health budget is not fixed is both small and an increasing function of a technology's budget impact, i.e. only large decisions justify some level of risk aversion. Note that the risk aversion here arises from society's attitude to the *non-health* resources required to transfer into health care to fund the new treatment and not to the uncertainty about the lost output from the *health* resources used to fund the new treatment discussed in 4.1.4 above.

Overall, in line with other authors such as Culyer et al. (2007), we believe that the health care budget is best treated as fixed on average for consideration of uncertainty, even if it may display some flexibility at certain margins. This means that questions of financial risk aversion do not arise. Assuming a fixed health care budget, (B) in our model, in which all funds are ultimately spent to produce health gains, cost uncertainty can itself be denominated in QALYs – unanticipated additional costs will displace additional QALY increasing expenditure from the NHS, while unexpected cost reductions will have the opposite effect. As a result, any uncertainty faced by decision makers can ultimately be rephrased as uncertainty in relation to the state-distribution of QALYs.

4.2. Uncertainty about the threshold (or the optimal allocation of the health budget)

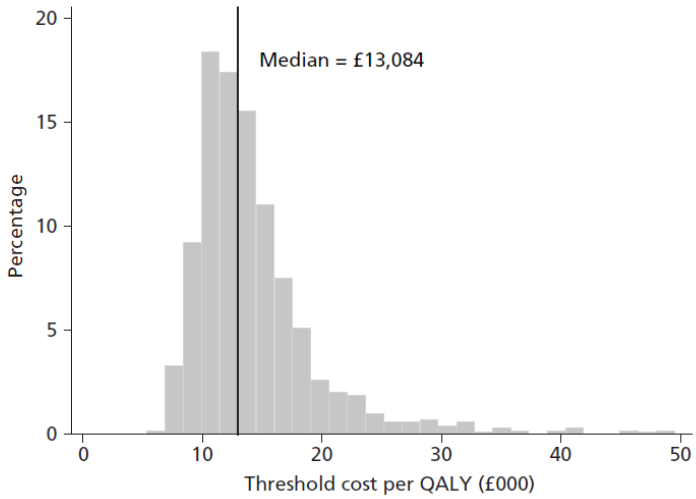
The threshold in the model is represented by the inverse of the Lagrange multiplier for the health budget constraint $c \leq B$. Thus the threshold is not taken as exogenous but it has to be identified by the decision maker throughout the maximisation search. This connection between the threshold and the health budget allocation has been highlighted for instance by Culyer et al. (2007):

> *"NICE is neither a threshold-taker nor a threshold-maker. NICE is, in effect, a threshold-searcher, where the threshold is logically implied by the combination of the technologies that are available and the budget, but is not readily visible."*

Errors around the identification of the threshold (i.e. differences between the estimated threshold and the actual one) can be recognised as a potential source of uncertainty. The issue of uncertainty around the threshold was more explicitly raised by Claxton et al. (2010). More recently, Claxton et al. (2015) show that a £10,000 overestimate of the true value of the threshold (for instance, assuming it is £20,000 when it should be £10,000) has more healthcare consequences than an underestimate (for instance, assuming it is £20,000 when it should be £30,000). This result, as Claxton et al. make clear, assumes that the distribution of possible values is symmetrical. In other words, if the distribution of possible values for the threshold is symmetrical, the larger the uncertainty about the threshold – irrespective of the ICER of the technology – the lower the threshold that should be applied, because otherwise more QALYs would be expected to be lost.

However, as discussed in Barnsley et al. (2013), there is no reason why the distribution of possible values for the threshold should be symmetrical, in which case this conclusion that a lower threshold should be applied may not hold. Technically, the expected size of errors conditional on their realisation is likely to be inversely proportional to their likelihood of occurring. The following example (Figure 1) is taken from Claxton et al. (2015). The issue we are considering is the distribution of the threshold. Expected errors around the Claxton et al. threshold estimate are not equally sized, since the distribution is clearly skewed to the right. Thus, there is a small probability that the threshold may have been substantially underestimated but a much larger probability that the threshold may have been slightly overestimated and no chance that it has been substantially overestimated. The potential significance of the underestimate is quite substantial: what matters is not the relationship that is identified in the diagram in the Claxton et al. (2015) paper, but rather what the distribution is in practice. We conclude, following Barnsley et al. (2013), that whilst uncertainty about the threshold makes the theoretical case for using a lower threshold, the effect is reduced by the positive skew of potential errors and, arguably, can therefore be ignored.

Figure 1: Histogram of simulation of undiscounted threshold (all 23 PBCs)



Source: Claxton et al. (2015).

4.3. Loss aversion

Loss aversion is the name given by Tversky and Kahneman (1991) to the experimentally observed decision making heuristic which departs from risk neutral decision making in the following ways:

- Decision makers weight losses relative to some perceived status quo significantly more heavily than gains.[17]
- Decision makers systematically overestimate the probability of very unlikely events, while underestimating the probability of more common occurrences.

As a result, loss aversion generates risk aversion-like behaviour in response to low probability events which lead to losses from an existing valuable endowment. In reference to our model, assuming loss aversion will make us depart from expected utility theory. Thus the social welfare function, if we wished to incorporate loss aversion, should be written as a combination of the value attached to different outcomes as suggested in the first bullet point above, and weighted by the subjective probability linked to every potential outcome, as suggested in the second bullet point.

Since loss aversion appears to be a relatively universal phenomenon we would expect that it is likely to influence the attitudes to risk of HTA decision makers. However, there is no rational reason for using it in decision making in HTA. Since both limbs of the loss aversion heuristic lead decision makers to overvalue certainty relative to the value maximising optimum, decision makers ought to be aware of, and resist, any tendency towards loss averse behaviour in evaluating uncertain prospects.

This is easiest to see in relation to the second aspect of loss aversion outlined above: obviously it is preferable for decision makers to evaluate the likelihood of events based on accurate, rather than arbitrarily rescaled, probability values. Analogously, the disproportionate response to losses which are the other key aspect of loss aversion lead to decisions which are *ex post* suboptimal. Setting aside issues of diminishing marginal value, which we have dealt with in detail above, decision makers should not place any special value on the loss of a QALY by an unidentified patient than on the gain of a QALY by some other patient. They should maximise the expected value of their decisions from a utility expected theory approach, and weighting losses relative to an arbitrary baseline is inconsistent with this approach. Since the current endowment level of health in a society is only ever transitory and imperfectly known to decision makers, there can be no rational basis for fixating on a particular aggregate level of health as the baseline from which gains are to be measured.

4.4. Consequences of excessive risk aversion

To the extent technologies with uncertain evidence are incorrectly[18] rejected due to decision makers' instinctively risk averse preferences, the potential harm caused by those rejections may not be randomly distributed throughout the community: among the technologies for which it has hardest to provide certain evidence are those which treat very small patient populations, including the beneficiaries of some orphan drugs.

[17] The phenomenon of overvaluing a good which one already possesses is also known as the "endowment effect". For our purposes, the endowment effect and the first limb of loss aversion are equivalent concepts.

[18] As we note above, we view a rejection as incorrect where it reduces expected societal health, welfare or flourishing, meaning we do not accept the strong form extra-welfarist view that decisions are "correct" provided that they are properly reached by a properly-appointed body.

There are two issues here. Firstly, it would be inappropriate to reject technologies on the basis of uncertainty when that uncertainty is difficult or impossible to reduce, which may arise with limited patient populations available to participate in trials. Secondly, there is a value judgement that could be made that where a population suffers from a large burden of illness and unmet need for treatment, it should receive more, rather than less, priority in resource allocation.

Similarly, uncertainty may be large for treatments which show significant health gains early in clinical trials, leading to patient crossover from comparator arms of the trial, or for late stage treatments where trial length is unavoidably limited by short patient lifespan.

These patient groups – orphan and ultra-orphan populations, late stage terminal patients are of particular concern in HTA decision making.

In addition, since the uncertainty involved in technologies targeting them is a largely unavoidable feature of the patient population, the arguments for delaying approval to incentivise further research do not arise in these situations. Even if it were the case that data on technology effectiveness was generated more rapidly via controlled trials than through use by a wider patient population, this does not justify delaying approval for technologies with inherently uncertain health benefits due to small patient population size.

In considering the optimal level of uncertainty aversion, decision makers should be sensitive not only to the theoretical arguments set out above, but also to the practical possibilities for uncertainty reduction and the identities and existing burdens of those who will be harmed by a rejection on the basis of uncertainty.

5. UNCERTAINTY DUE TO ESTIMATING EXPECTED OUTCOMES: VARIANCE AND POTENTIAL BIASES

So far, in this paper we have modelled uncertainty as a known probability distribution around potential health and cost outcomes, while ignoring the sources of uncertainty and the forms it can take as part of the HTA process. In this section we consider more explicitly how uncertainty arises and can be dealt with in the context of HTA.

5.1. Uncertainty due to sampling variability: Uncertainty in a dynamic setting

The main place where sampling variability is important in NICE appraisals is where a committee must decide on whether to wait for the collection of more information or to make a positive decision to adopt a technology only if further information on its effectiveness is collected, rather than make a simple yes/no decision. The variance may arise from clinical effects or from costs.[19]

[19] Costs are uncertain and almost always are positively skewed: the average cost of some very few patients with the largest costs is a great deal higher than the overall average, so there is a long right-hand tail to the distribution of costs. This is problematic for small samples, where a single large observation may make a great difference to the average cost, but it is difficult to decide whether it should be regarded as an outlier. However, with larger samples this effect tends to dissipate. Because of the Central Limit Theorem, the distribution of the mean cost tends towards the normal distribution as the sample size increases. If the costs come from the same randomised trial as the health outcomes, it means that substantial uncertainty will arise only in the form of variance, as the randomisation process will reduce bias to a very small component. Having relatively

Apart from a decision about waiting for more information or providing more information along with a "yes" decision, we conclude that risk aversion should be either irrelevant or of marginal significance to HTA decision making and conclude that decision makers may currently respond too negatively to the risk component of uncertainty.

The approaches to uncertainty outlined above treat the level of uncertainty faced by a decision maker as given. However, a different source of uncertainty can be triggered by a dynamic interpretation of the decision analysis process. If today's decisions cause the variance of some estimates over time to shrink or enlarge, then we can address uncertainty in a way that is giving us some information. For instance, a policy rule which favours saying "no" when a technology is too uncertain may generate incentives to produce additional information aimed at narrowing the estimates of key coefficients, and this additional information can leave decision makers with a new, higher expected value, decision problem. This implies that investing in uncertainty reduction can be valuable, even where decision makers are uncertainty neutral or even uncertainty loving when faced with a given level of uncertainty.

Although the estimated variance of a sampled variable is used to determine VOI, a decision to wait for further information rather than make a "yes" or "no" decision on current information alone is not based on risk aversion, but rather, on attempting to optimise society's health benefits from a given health-care budget. To illustrate this, we first consider a decision-making body that is faced with some promising but very preliminary data. It is known that the final data will be available in two weeks' time, and that there is a much higher probability that the correct decision will be made after the new data has been considered. Committees faced with such a situation will usually postpone their decision for the two weeks. That is to say, the information has a value. That value can be quantified using a decision-theory approach.

We can suppose that the estimated ICER of the intervention before the new information becomes available is £27,000 per QALY and the threshold is £30,000 per QALY. If VOI is greater than £3,000, it will be optimal to wait for new information. The committee is not being risk averse when it decides to wait rather than approving the intervention immediately. To indicate why this is the case, let us assume that for a separate intervention, the initial estimate of cost effectiveness is £34,000 per QALY and VOI is over £4,000. The choice is between an immediate rejection or to wait for more information. If the committee decides to wait, it is because it optimises societal health (on average) to do so, and not because it is risk loving.

Thus although variance has been used in the decision process, an optimal decision will not be based on the decision-maker's attitude to risk.

Whether decision-makers in NICE committees act optimally with respect to the theory as just outlined or whether they do so because they are risk-averse when faced with an uncertain ICER of £27,000 is not known. In a situation such as this, committee members with little or no theoretical knowledge are likely to be guided within the committee by other members who have been chosen because of their theoretical knowledge and practical experience of the statistics and economics.

large trials (of some hundreds of patients in each arm) means that the small-sample problem alluded to above has little influence. An exception can arise for a trial that is highly clustered if there are large cost differences between clusters, but in general this will not be problematic if, as expected, costs of different patients within each cluster vary sufficiently so as to dwarf the difference of between-cluster costs.

A Value of Information (VOI) framework can be used (Griffin, et al., 2011) to show that, if the rate of uncertainty reduction is higher following a "no" decision than following a "yes", decision makers should optimally reject products with small positive expected benefits in order to enable a subsequent, better-informed decision with a higher expected value. Griffin et al. argue that approval will reduce the rate at which uncertainty-reducing information is generated for five reasons:

(i) The adoption of a technology removes incentives on the manufacturer to conduct further research;

(ii) The required information can only come from clinical trials;

(iii) These clinical trials can only be conducted in the health system subject to the HTA body's jurisdiction;

(iv) The early diffusion of a technology means that future clinical trials are less likely to be supported or regarded as ethical by the clinical community, even when public funds are made available for such research; and

(v) Patients are unlikely to enrol in clinical trials once they have unrestricted access to the new technology.[20]

We note that (i) is, at least in theory, avoidable via approval combined with commissioning of research. To the extent that a company is incentivised by the costs of delayed approval to incur the cost of additional research, that research could also have been incentivised by a side payment no greater than the costs of delay to the company, or commissioned by the HTA agency or some other part of the payer's organisation, directly, possibly with the requirement that the research be funded by the company. Where such Coasian bargaining is possible, the costs of the delay to the company and/or the patient population are a pure dead weight loss.[21]

Assumptions (ii) and (iii) do not account, respectively, for the possibility that observational research may provide information or that transferable research may be conducted in jurisdictions outside the HTA body's remit.

In relation to (iv) and (v) we make two points. First: it is an odd approach to ethics to regard it as wrong to withhold health-improving technology from *some* patients in order to gather information in the context of a clinical trial, whilst accepting that it is ethically acceptable to withhold the same health-improving technology expected to be cost-effective from *all* patients in order to allow a trial to be conducted.

Second: at the heart of the Griffin et al (2011) approach to VOI in HTA is the assumption that uncertainty will be resolved at a faster rate prior to approval. The authors understate the ability of approval to generate clinical effectiveness and real world cost data at far higher rates than is possible with pre-approval clinical trials. If we ultimately decide that widespread use of a medicine is a more effective way of resolving uncertainty about its costs and benefits, then the VOI analysis presented by Griffin et al. remains relevant, but the calculations are changed.

[20] Griffin et al (2011), p.214-215.

[21] Note that, theoretically, under a predictable maximum ICER threshold, companies will by incentivised to price up to the maximum allowable amount and appropriate the entire surplus associated with a new technology so the dead weight losses in question will accrue to the company, taking the form of a disincentive to research.

One obstacle to a "coverage with evidence development" approach might be a situation of imperfect reversibility. In other words, even if the new evidence is obtained, it may be difficult to change a decision for a variety of reasons – e.g. because of the investments that have been made, whether they are political or physical. The critical point is that these are real effects: it is not about a concern about uncertainty; it is about the possibility of getting additional information that would change the decision. Irreversibilities change the potential value of additional information by adding to the cost of changing a decision, and therefore matter for uncertainty aversion.

5.2. Bias and ambiguity

In HTA, 'observable' variation in the form of sample variance usually contributes a relatively small amount to total uncertainty. Bias and ambiguity between them usually contribute far more. Ambiguity, in this context, may be measured by the difference between the estimated distribution and the true distribution of an estimate of mean outcome. Because this difference reduces with sample size, it requires only that RCTs are sufficiently large in order to allow us to ignore this problem when considering effectiveness. HTA trials, particularly for drugs, generally achieve this size. It is the role of potential bias that takes precedence in most instances. When considering cost effectiveness, however, assumptions about the form of predictive equations involve prediction errors that do not shrink as sample sizes increase. These equations could be regarded either as biased or ambiguous. (Berger & Bosetti, 2016)

We know biases will exist but we do not always know their origin or if they are present or absent in any particular case. Occasionally in healthcare, the direction of a bias is known (e.g. publication bias tends to over-represent trials with an effect that is larger than the true average effect and is in the 'expected' direction). If the effect of a bias (its direction and size) is known, then it can be compensated for, and after the compensation has taken place, nothing more needs to be done about it. But mostly it does not happen like that, because neither direction nor size of a bias is known. If the direction is known but not the size, too great a compensation can make things worse, albeit in the opposite direction. Furthermore, compensating in this way is subjective and constitutes a bias in its own right. Bias should be treated as one would treat an irreducible level of variance: that is, a fixed (but unknown) level.

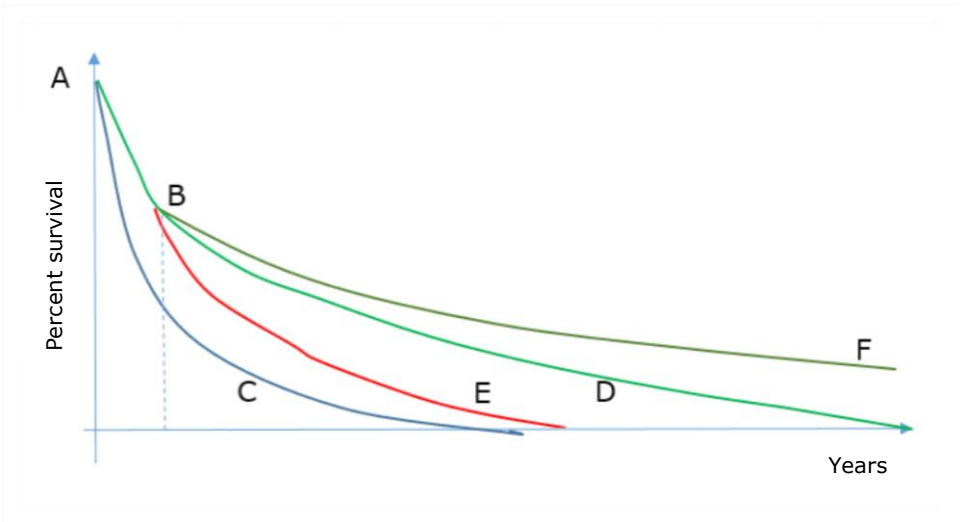
When committees look at the ICER under such conditions, the estimated variance within the ICER refers to sampling variation, which can be ignored because this is risk that can be pooled for an institution dealing with a large number of projects. However, we usually have no evidence that tells us about the size of the bias. Whatever size it is, it is not related to the sampling variation, because if the sample were twice the size, the sampling variation of the estimated mean effect would decline, whereas the bias (by definition) would remain unchanged. Large biases generally do not come from the RCTs, provided that they have been properly designed and carried out. Possible biases usually arise from economic modelling of eventual effect size, and that means making assumptions about the effect of a treatment beyond the end of a trial follow-up period. If, for example, data in a clinical trial for quality of life improvement and life expectancy for a treatment for a chronic condition only exist for (say) the 12 months beyond the end of treatment, extrapolation to the end of life depends on assumptions made about the continuation of an effect. Such assumptions may give rise to very large errors which cannot be improved by making the size of the trial larger. Knowing the relationship between a surrogate endpoint and overall survival may help, though not by much unless the correlation between the surrogate and the actual survival is very high.

Bias, or what we can regard as a *systematic* deviation from the true value of a variable, is thus pivotal in the consideration of uncertainty. If biases did not exist, uncertainty in the ICER could be ignored except for consideration of the case for additional evidence collection. If biases are recognised to exist, it does not help to look at the variance of the ICER, because the variance of the ICER is a recognition of random error, and is not bias.[22]

Given a properly conducted RCT (whose aim is to avoid the effect of confounders and biases), the main sources of bias will be from assumptions made regarding cost effectiveness. We illustrate using the example of trying to estimate overall survival from a disease with and without treatment. The normal way of capturing this is through a Kaplan-Meier survival curve (see Figure 2).

Suppose that an RCT aims to find the survival advantage of treatment X compared with not treating with a comparator. The treatment plus follow up is assumed to end at time B. With treatment with the comparator, survival over time follows the curve AC (in blue). The curve will usually be well-known from past observation. For the treatment arm, the true survival curve (if we knew it) would be given by ABD (in green). However, the only evidence we have is the line AB. By modelling, we would suggest (say) a worst possible extrapolation of ABE and a best possible extrapolation of ABF. If the extrapolation is anything other than ABD, there will be bias. If all possible paths between BE and BF show enough benefit for X to be cost effective, the bias does not matter: the drug should be recommended as being cost effective. The reverse occurs when no possible path between BE and BF is cost effective, and X should not be recommended. However, if ABE is not cost effective but ABF is cost effective, then the choice we make impacts on the cost effectiveness of X. If over all possible cancers and all possible drugs, the bias of our extrapolation around its true survival line was in essence random, then the bias could be treated as a variance and ignored as the NHS would 'gain on the swings as much as it lost on the roundabouts'. However, the survival bias might be systematically negative too often (or alternatively, too often positive) for bias to be ignored.

Figure 2: Illustrative Kaplan-Meier survival curve



[22] This also means that the cost effectiveness acceptability curve (CEAC), whose slope depends on the *variance* of the ICER, is of little or no use in decision making.

Overall, therefore, the effect of sampling variance on an ICER can be ignored but potential biases arising from modelling assumptions can only be somewhat ignored. Sensitivity analysis can be carried out and it might be possible to ignore a bias altogether if the size of the bias, within limits that are regarded as plausible, does not affect an adoption decision.

A NICE Appraisal Committee may be clear about the existence of a potential bias but may be unduly prudent. Too great a level of conservatism will generally be suboptimal.

We have focussed here on modelling assumptions extrapolating clinical trial effects. However, there are other sources of possible bias:

- NICE technology appraisals do not usually take the sampling variance of its utility measurements for quality-of-life estimation into account. Utility measurements are sometimes taken from the RCT or RCTs that apply to the effectiveness section of an appraisal. The sample variation that is involved is an uncertainty that will not have been taken into account. A Committee will generally not be aware of the omission of this uncertainty from the modelling analysis. Often, utility values do not come from the same RCT as is used to determine effectiveness. Sometimes they are for patients that are of different degrees of disease severity or from different countries, and occasionally the patients' quality of life refers to people who have a similar but not the same disease as is being appraised. The instrument used to estimate QoL may not be the same from one appraisal to another. These differences represent potential biases.
- More generally, we do not know which group of people is the 'best' at estimating and valuing the patient's QoL. It might be the general public, the clinician, the patient or the carer. The choice of which group is 'best' can be seen as a social value judgement. Whether risk, bias or social value judgement, however, such an issue none would not become apparent to the relevant committee members. To the extent, however, that the same preferences are used from appraisal to appraisal (and this has been the general public's preferences since NICE's inception) any source of bias would have been applied to all appraisals. There is a level playing field for all appraisals. We conclude that this form of bias can probably be ignored.
- NICE takes a number of things for granted. Most of them are requirements imposed on it by governments. This includes the level of the cost effectiveness threshold (though NICE has had some say about what this level should be); NICE looks to the Treasury "Green Book" about how government bodies should do economic evaluation, and NICE in general adheres to these recommendations. This includes the setting of discount rates for future costs and benefits (though the Green Book has been open to different interpretations on this matter: if one of these interpretations is the 'true' one, then deviations to an ICER threshold by taking a different interpretation must be a bias).

In all, healthcare researchers and decision-makers alike try to avoid bias by means of RCTs, and avoid "false positives" by the use of 95% confidence intervals. At the cost effectiveness stage, decision-makers should try to avoid the further conservatism of taking risk into account again. However, many sources of uncertainty in the form of biases of unknown size or even direction are introduced at this stage of decision-making. Decision-making in such a world is no longer crisp and clear-cut, but overly-cautious behaviour by those making decisions should be avoided. First, risk should be ignored; second, biases that are likely to be in common to all appraisals can probably be ignored,

and other biases that may be in different directions for different appraisals may cancel out. Biases should of course not be ignored in their entirety but should not generally be judged by their worst-case effects.

6. RECOMMENDATIONS FOR DECISION MAKING IN HTA

We suggest in our final section some recommendations which we hope might improve some of the weaker points in handling uncertainty in the decision-making process.

We do so, recognising that more evidence is required on the following:

- The rate at which marginal returns to general NHS expenditure diminish and the rate at which valuable information is generated prior to and following technology approval.
- The potential costs of reversing a decision. This is because where a decision is likely to be costly to reverse, option value arguments imply some level of uncertainty aversion and the value of collecting additional evidence prior to making a positive decision increases.
- The importance of uncertainty in NICE decision-making. This could be examined by looking at the proportion of appraisals where a NICE decision depended on the way that uncertainty was viewed by an Appraisal Committee.

Decision makers are likely to be instinctively loss averse, weighting negative realisations of uncertainty too heavily relative to positive realisations and overestimating the probability of unlikely events. These instincts will tend to lead to improper treatment of uncertainty. We suggest that NICE modify its posture towards uncertainty. It should take virtually no regard of the variance of the ICER, but it should be aware that the uncertainty with respect to biases in the estimation of the ICER cannot be ignored in the same way. These biases give rise to uncertainty aversion, whose effects are similar to those of risk aversion on individuals. We have shown in sections 2 and 3 that individuals are willing to trade-off risk against the mean number of QALYs, and show that in these circumstances, the effect is likely to be quite small and can probably be ignored.

We therefore suggest that NICE adopt an uncertainty neutral posture: that it remove uncertainty from the list of factors which determine a technology's value in its Methods Guide and provide decision makers with appropriate guidance so as to ensure that they assess a technology based only on what they consider to be the best estimates of parameter values, ignoring variance about those values, conventional measures of statistical significance and sensitivity analysis. The potential value of collecting additional information to reduce uncertainty, in terms of the impact on decision making, is important to address. Our recommendation would need to be revised if evidence suggested the effect of uncertainty aversion was not as small as our analysis suggests. Finally we note that it is not only NICE that may require to take a different approach to handling uncertainty. Other NHS bodies allocating resources between competing uses also need to be knowledgeable about how to optimise improvements in societal health within a given health budget.

7. REFERENCES

Arrow, K. J. & Lind , R. C., 1970. Uncertainty and the Evaluation of Public Investment Decisions. *American Economic Review,* Volume 60, pp. 364-78.

Barnsley, P. D., 2013. *Subjective wellbeing, mean reversion and risk ,* Sydney: University of Sydney.

Barnsley, P., Towse, A., Karlsberg Schaffer, S. & Sussex, J., 2013. *Critique of CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold,* London: Office of Health Economics..

Berger, L. & Bosetti, V., 2016. Ellsberg Re-revisited: An Experiment Disentangling Model Uncertainty and Risk Aversion. *Research in Agricultural and applied economics, http://purl.umn.edu/236239.*

Brazier, J. et al., 2013. *Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey ,* Sheffield/York: EEPRU.

Claxton, K. et al., 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment,* 19(14), pp. 1-542.

Claxton, K., Walker, S., Palmer, S. & Sculpher, M., 2010. *Appropriate perspectives for health care decisions,* York: Centre for Health Economics..

Culyer, A. et al., 2007. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *Journal of Health Services Research & Policy,* pp. 56-58.

Department of Health, 2004. *Policy Appraisal and Health,* s.l.: Department of Health.

Eichler, H. G. et al., 2013. The risks or risk aversion in drug regulation. *Nature Reviews Drug Discovery,* Volume 12, p. 907–916.

Epstein, D. M., Chalabi, Z., Claxton, K. & Sculpher, M., 2007. Efficiency, Equity, and Budgetary Policies: Informing Decisions Using Mathematical Programming. *Medical Decision Making,* pp. 128-137.

Epstein, L. G., 1999. A definition of uncertainty aversion. *The Review of Economic Studies,* 66 (3), pp. 579-608.

Gravelle, H. et al., 2007. Discounting in economic evaluations: stepping forward towards optimal decision rules. *Health Economics,* 16(3), pp. 307-17.

Griffin, S. C., Claxton, K. P., Palmer, S. J. & Sculpher, M. J., 2011. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Economics,* pp. 212-224.

Griffin, S., Claxton, K. & Sculpher, M., 2008. Decision analysis for resource allocation in health care. *Journal of Health Services Research & Policy,* 13(Suppl 3), pp. 23-30.

Kaplow, L. & Shavell, S., 2001. Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle. *Journal of Political Economy,* 109(2), pp. 281-286.

Linley, W. G. & Hughes, D. A., 2013. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health economics,* 22(8), pp. 948-964.

McCabe, C., Claxton, K. & Culyer, A. J., 2008. The NICE Cost-Effectiveness Threshold. What it is and what that means. *Pharmacoeconomics,* 26(9), pp. 733-744.

Meltzer, D. O. & Smith, P. C., 2012. Theoretical issues relevant to the economic evaluation of health technologies. In: *Handbook of Health Economics, Vol. 2.* s.l.:Elsevier, pp. 433-469.

NICE, 2004. *Guide to the methods of technology appraisal,* London: NICE.

Pliskin , J. S., Shepard, D. S. & Weinstein, M. C., 1980. Utility functions for life years and health status. *Operations research ,* 28(1), pp. 206-224.

Schmeidler, D., 1989. Subjective probability and expected utility without additivity. *Econometrica,* 57(3), pp. 571-587.

Tversky, A. & Kahneman, D., 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics,* pp. 1039-1061.

UK Treasury, 2011. *The Green Book: appraisal and evaluation in central government..* London: TSO.

Williams, I., McIver, S., Moore, D. & Bryan, S., 2008. The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technology Assessment -Southampton ,* 12(7).