Dealing with uncertainty in the economic evaluation of health care technologies

Susan Claire Griffin, MSc.

This thesis was submitted for the qualification of PhD by publications at the University of York

The research was conducted in the Centre for Health Economics

November 2010

Abstract

This collection of papers aims to addresses why, and how, uncertainty should be incorporated within economic evaluations of health care technologies and how the results can be used to inform decision making. Section one provides a brief overview of the methodological and policy background of using economic evaluation to inform decisions about health care technologies. With the context for utilising economic evaluation to inform reimbursement decisions established, sources of uncertainty are considered at each stage of the evaluation. Section two discusses the identification of relevant evidence to inform estimates of the effectiveness of health care technologies. The key uncertainty addressed is how to assess whether the available evidence can be analysed to provide an unbiased estimate of the treatment effect. As multiple sources of evidence may exist, Section three explains how they can be combined to sum up the available evidence, and in Section four the techniques for characterising uncertainty within that combined analysis are examined. Sections five and six demonstrate why a characterisation of uncertainty is crucial in order to inform decisions about the need for further research and the consequences of making reimbursement decisions under uncertainty. Methods to establish the opportunity cost of uncertainty and sufficiency of the evidence base are reviewed in Section five with a view to informing decisions to acquire further research. In Section six the irreversible aspects of reimbursement decisions, as well as the interdependence of decisions about reimbursement and about further research is illustrated. These factors spell out why an economic evaluation that does not deal with uncertainty will be inadequate for informing reimbursement decisions. Finally, Section seven concludes the thesis and summarises the contribution of the presented papers and areas for further research.

Contents

Page number

Acknowledgements iv

Relevant publications and declaration of candidate's contribution v

Integrative chapter

1. Methods and policy background 1

2. Identifying evidence of treatment effects 3

    2.1 Impact of method of data collection on potential for bias 3

    2.2 Methods of analysis to control for selection bias 4

    2.3 Methods to correct for missing outcomes and data 4

    2.4 Utilising observational data on treatment effects 5

3. Combining multiple sources of evidence 7

    3.1 Multiple sources of evidence providing the same information 7

    3.2 Multiple sources of evidence providing a range of information 9

4. Characterising uncertainty 11

    4.1 Sensitivity analysis for parameter uncertainty 11

    4.2 Sensitivity analysis for other sources of uncertainty 12

    4.3 The implications of model design for PSA 13

5. Informing the need for further research 14

    5.1 The value of perfect information 14

    5.2 The value of sample information 16

    5.3 Realising the benefits of research 16

6. The impact of uncertainty on reimbursement decisions 17

    6.1 Incorporating investment and reversal costs 17

    6.2 The impact of reimbursement on the prospects for research 18

    6.3 Linking reimbursement with uncertainty 18

7. Summary 20

Appendix A. Publications 23

Glossary 75

References 76

Acknowledgements

I would like to dedicate this thesis to James, who makes me smile every day.

To my parents, Robert and Margaret Griffin, I would like to say thank you for supporting me throughout my education.

The included research papers were developed while working alongside esteemed colleagues from a range of institutions, and I would like to thank all of my co-authors for making this collection of papers possible.

Carrying out the research was made pleasurable thanks to the supportive, friendly and rich intellectual environment at the Centre for Health Economics, University of York. I would like to thank everyone who has accompanied me throughout my development as a researcher.

I am grateful to Linda and Ness for proof reading the manuscript.

A special thanks to Karl Claxton and Mark Sculpher for prompting this enterprise and providing invaluable support, encouragement and advice.

Relevant publications and declaration of candidate's contribution

1. Griffin S, Barber J, Manca A, et al. Cost effectiveness of clinically appropriate decisions on alternative treatments for angina pectoris: prospective observational study. British Medical Journal 2007; 334: 624[1]

Contribution of the candidate: Analysis of primary trial data by calculating costs and performing the cost-effectiveness analysis and responsibility for preparation of manuscript, technical report and subsequent revisions.

2. Griffin S, Bojke L, Main C and Palmer S. Incorporating direct and indirect evidence using Bayesian methods: an applied case study in ovarian cancer. Value in Health 2006; 9(2): 123-131[2]

Contribution of the candidate: Conception of the research question and was responsible for conducting the analysis, preparation of the manuscript and subsequent revisions.

3. Griffin S, Claxton K, Hawkins N and Sculpher M. Probabilistic analysis and computationally expensive models: necessary and required? Value in Health 2006; 9(4): 244-252[3]

Contribution of the candidate: Conception of the research question and was responsible for conducting the literature review, preparation of the manuscript and subsequent revisions.

4. Griffin S, Welton N, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs. Medical Decision Making 2010; 30: 155-162[4]

Contribution of the candidate: Involved in conception of the research question and was responsible for developing the methodological framework, preparation of the manuscript and subsequent revisions.

5. Griffin S, Claxton K, Palmer S and Sculpher M. Dangerous Omissions: the consequences of ignoring decision uncertainty. Health Economics 2010; DOI 10.1002/hec.1586[5]

Contribution of the candidate: Involved in conception of the research question and was responsible for developing the methodological framework, preparation of the manuscript and subsequent revisions.

The accompanying integrative chapter is the sole work of the candidate.

1. Methods and policy background

Within the health care sector there exist numerous competing health care interventions and programmes, of which only a proportion can be provided with the resources available. The challenge is in identifying which interventions to provide given the resource constraints, and this can be informed by an assessment of the costs and benefits attributable to each intervention.[6 7] Decisions about allocating collective resources need to be made in a justifiable and auditable manner, so any methods need to make explicit the underlying scientific and social value judgements. Those judgements, amongst other things, determine which costs are considered relevant, what are the benefits and how should they be measured and the limits of the available resources. The benefits of health care interventions can be measured in terms of the amount of health they generate. The use of health as the outcome of interest when evaluating returns from investments would seem consistent with the aim of providing collectively funded health care[8 9] but represents a departure from neoclassical economics in which benefits would always be measured in terms of individuals' subjective utilities.[10] However, this departure is necessary if it is not considered possible to appropriately value health outcomes in the same way as benefits derived from the consumption of goods and services.[11] For the resource constraint to be endogenous some knowledge of the social welfare function is required in order to value health care relative to other activities.[12 13] However, if one accepts the existence of a legitimate social process for determining the budget allocated to health care, the relevant resources can be regarded as fixed and determined exogenously.[14] With this value judgement the allocation problem takes the form of a constrained optimisation to maximise health gains subject to a budget constraint, and the relevant costs are those that fall on the budget constraint. The most cost-effective use of available resources would be the one that provides the greatest health benefits.

The costs and health outcomes of interventions can be estimated by referring to available evidence that describe the course of the disease, health outcomes and health service resource use by individuals who receive them. By identifying all possible interventions for all potential recipients, calculating for each the expected health outcomes and cost burden and comparing this to the total amount of resources available, it would be possible to solve the whole allocation problem simultaneously using mathematical programming techniques.[15-17] However, the informational requirements of this global approach make it impractical. To make the problem manageable it may be necessary to consider the choice between all relevant treatment options for a specific indication. Should one of the mutually exclusive options be found to offer an improvement in health gains but at a greater cost than the currently funded alternative, the potential health outcomes must be compared to those lost from interventions displaced by reallocating resources to fund the new intervention.[18 19] Identifying and evaluating targets for disinvestment may be difficult in practice,[20] and instead an estimation of the shadow price of the budget constraint may provide a threshold for assessing cost-effectiveness.[21]

These comparisons necessitate the use of a summary index measure of health outcomes that is common across all interventions in all disease areas. In practice this is often estimated by weighting survival according to some preference-based measure of health-

related quality of life (HRQL) to generate quality adjusted life years (QALYs).[22-24] The evidence and methods of calculation must distinguish the health gains that are attributable to the use of different interventions. In practice the process of identifying the available evidence and the manner in which it is interpreted and used to calculate expected costs and health outcomes will be subject to uncertainty. Uncertainty in estimates of cost-effectiveness leads to the possibility that the wrong set of interventions is funded such that population health is not maximised, and hence there is an opportunity cost to uncertainty in terms of health forgone. The uncertainty in evidence can be characterised by describing the range and likelihood of alternative plausible values and assumptions supported by the evidence, and from this computing a distribution of costs and health outcomes for each treatment under evaluation. Using this distribution it is possible to address questions about whether the current evidence base is sufficient to support the choice between alternative treatments and whether further research is valuable. Methods for calculating the value of eliminating decision uncertainty are well-established.[25-27] This provides an upper bound for the value of additional research to reduce that uncertainty, which can be compared to the costs of conducting that research.[28] On the basis of the available evidence, the decision maker can maximize expected health by reimbursing the treatments with the highest expected net benefits. There may be an error probability associated with those decisions, but any other allocation of resources would be associated with a lower expected level of health. However, reducing the likelihood of wrong decisions would also increase the expected level of health.[29 30] This can be achieved by acquiring more evidence, whether actively or passively, to reduce the decision uncertainty. When accounting for uncertainty the costs of reversing decisions in light of new evidence should be incorporated in the analysis.[31] While population health benefits can be maximised by optimising reimbursement decisions to fund interventions and to fund further research, those two actions are not mutually independent.[5]

The value of economic evaluation in supporting reimbursement decisions is evidenced by the proliferation of national institutions with the remit to grant approval for the use of particular health interventions within health services worldwide on the basis of a review of available evidence and an assessment of cost-effectiveness.[32 33] Thus advancing the methods for economic evaluation can have a direct impact on uncertain decisions about resource allocation in health care systems around the world.

2. Identifying evidence of treatment effects

To identify from the wealth of available information the sources of evidence on which a particular assessment of cost-effectiveness should be based, it is necessary to specify criteria for determining the relevance of evidence. Estimation of treatment effects, that is the change in health outcomes or resource use that would result from the use of an intervention, requires evidence on the causal effects of interventions.[34] The desired information is what the health outcomes and resource use would be with the intervention and what would have happened without the intervention. However, it is impossible directly to observe the counterfactual as, at any given point in time, we can only observe an individual either with or without the intervention, and must also rely on past observations to predict the experiences of future patients. Thus the relevance of any source of evidence for estimating a treatment effect depends on its ability to approximate the counterfactual in the studied population and how generalisable are the estimates of treatment effect to the target population.[34] The ability to identify a causal relationship between interventions and outcomes (by approximating the counterfactual) is often referred to as internal validity.[35] The extent to which the results in the studied population hold true for the target population is often referred to as external validity.[36] An estimate that lacks validity can be described as biased because it differs systematically from the true value it is intended to describe. Unbiased estimates will be uncertain due to sampling error (unless the full population is observed), and any bias is an additional source of uncertainty in its extent and magnitude. If the internal or external validity of a source of evidence is judged to be too low it could be said to be irrelevant to the decision problem.

2.1 Impact of method of data collection on potential for bias

Randomised controlled trials (RCTs) are designed to minimise bias in identifying the causal effects of interventions. They do this by allocating patients to different interventions in a way that guarantees no systematic differences in the baseline characteristics of individuals within the treated and control groups.[37] This allows average treatment effects to be estimated by a simple comparison of the mean outcome in each group. Treatment effects calculated in this way have high internal validity as they describe with minimum bias the effects on the outcome in the study participants. However the rigours of designing such a study to enable the detection of a clear treatment effect (signal) can be restrictive. The units of observation (e.g. patients), the treatments provided, the observations made on the units and the setting in which the study is conducted may differ to those pertaining to the decision problem that the study seeks to inform, leading to problems with generalisability.[38] For example, strict inclusion and exclusion criteria may mean that the patients recruited differ systematically in their characteristics compared to the wider population or the individual patient for whom treatment is intended.[39 40] The restrictive nature of randomised trials means that for some interventions or patient populations of interest, it is not feasible to design a randomised trial due to reasons of practicality, cost or ethics.[41] In these circumstances an alternative source of evidence are quasi-experimental studies in which treatment is not assigned by randomisation.[35] However, when assignment to treatment selected, whether by patients, doctors or researchers, or when randomisation is unsuccessful, a simple unadjusted comparison of outcomes between treatment groups would be subject to selection bias. This is due to confounding by patient characteristics

that determine both whether a patient would receive an intervention and their post-intervention costs and health gains.

2.2. Methods of analysis to control for selection bias

There exists a range of methods to recover unbiased estimates of treatment effects from non-randomised studies by controlling for potential confounding. Where treatment assignment is determined by characteristics of individuals observed by the analyst (selection on observables), then regression analysis and methods of matching, including propensity scores, can be utilised.[42] In a regression framework it might be sufficient to condition on the set of confounding characteristics when estimating the relationship between the intervention and outcomes of interest. If the specification of the model conforms to the underlying data generating process this will produce an unbiased estimate of the treatment effect. An alternative method that does not rely on the parametric assumptions underpinning the regression approach is the method of matching.[43] The outcomes of treated individuals are compared to a matched comparison group for whom observed pre-treatment characteristics match those of the treated group up to some chosen degree of proximity. If suitably matched individuals can be identified, then their average outcomes provide the counterfactual for the average outcome for treated individuals. A number of methods have been proposed for defining a suitably matched group of individuals and a leading method is propensity score matching.[44] Where selection into treatment is based on unobservable characteristics of individuals, regression analysis is unsuitable and the method of instrumental variables (IV) can theoretically be used to make inference about the causal effect of an intervention.[45] An instrument is a covariate that is correlated with treatment assignment but is not independently correlated with outcome (i.e. other than through its influence on treatment assignment). Randomisation is therefore the perfect instrument, and a suitable instrumental variable is one that operates by the same mechanism. By utilising independent information in the instrument to determine treatment assignment, the method seeks to infer the underlying treatment effect. However, in practice it is difficult to find a suitable instrument and so examples of economic analyses that utilise IV are rare. Another approach is to model the selection mechanism directly. That is, if the characteristic determining selection was not observed, it may be possible to specify a model to predict its value in order to incorporate it within subsequent analyses to control for sample selection bias.[46]

2.3 Methods to correct for missing outcomes and data

The choice of outcome measures can be determined at the design stage for an RCT, but quasi-experimental studies conducted in 'naturalistic settings' may rely on data that are collected during routine care. Previously these were not likely to include preference-based measures of HRQL, although this may change with increased emphasis on patient reported outcome measures tied to reimbursement mechanisms.[47] In the absence of a preference-based measure of HRQL it may be possible to predict this from available clinical, disease-specific measures.[48-50] Prediction models for this purpose have been referred to as 'cross-walks' or 'mapping algorithms'.[48] The internal validity of such an approach relies on an assumption that the observed measures capture the same information as would have been

measured by the generic preference-based measure. A judgement about external validity is also required if a mapping algorithm is to be applied in a different patient population to that in which it was developed. The predicted values are more uncertain than directly observed measures; this is in part captured in the standard error of the prediction equation but can also stem from structural uncertainty as to the most appropriate form of equation.[51]

Another consequence of a more naturalistic setting may be an increased likelihood of missing observations compared to the more closely regulated RCT setting. Unless the data are missing completely at random (MCAR), a simple complete case analysis will be biased. This is due to systematic differences between the observed values and the values of the missing observations such that the summary statistics of the complete observations are not expected to equal those of the overall sample.[46] If the value of the missing observations is dependent on other observed values, the data can be said to be missing at random (MAR). While the assumption of MCAR can be proven inappropriate, no test can verify the assumption of MAR, so this relies on judgement about the mechanism for missingness. If data are MAR, regression analysis can be used to predict the missing values in order to correct for bias.[52 53] Using a single imputed value for each missing observation would underestimate the uncertainty and sampling error by giving imputed values equal weight to observed values even though their true value is predicted with uncertainty. To correct this Rubin proposed multiple imputation of missing values and corresponding rules for the calculation of summary statistics that account for the additional uncertainty introduced by the imputation procedure.[52 54] If data are missing not at random (MNAR), the probability of an observation being missing depends on its true value and the mechanism of missingness must be modelled to correct for bias.

2.4. Utilising observational data on treatment effect

Paper 1[1] illustrates how observational data collected in a naturalistic setting can be used in an economic evaluation by correcting for the potential biases described above while characterising the remaining uncertainty. The study describes the primary analysis of observational data, missing data and the prediction of HRQL data where it has not been measured directly. The ACRE study[55] design sought to address the question of whether medical management, coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI) might be cost-effective treatments for coronary heart disease in a patient population who, due to the characteristics of their heart disease, were excluded from previously conducted randomised trials, and for whom those same characteristics indicated that the results from available randomised trials may not be pertinent. Despite their exclusion from the randomised trials of efficacy, published clinical guidelines indicated that CABG and PCI were clinically appropriate for some of these patients, and correspondingly they received them in routine UK practice.[55] However, there was considerable uncertainty as to whether the magnitude of any health gains offered by these procedures, especially when provided after a waiting time exceeding that observed in previous trials, justified the considerable additional upfront expense relative to ongoing medical management. The judgement about which evidence was relevant was that the external bias from relying on a dissimilar patient population able to be recruited to RCTs

exceeded the potential for internal bias from analysing data from a quasi-experimental study design.

The ACRE study included a generic preference-based measure of HRQL at six years, but not at baseline or one year when only disease specific questionnaires were available. In order to estimate QALYs a regression analysis was specified to predict EQ-5D score[56] on the basis of angina, coronary heart disease measures of quality of life and other covariates. This generated a mapping algorithm that could be described as self-contained because it was estimated within-study, using information from the same population in whom it would be used to generate predictions. This meant that internal validity could be assessed and external validity was not of concern; however, the use of multiple covariates could make it difficult to re-use for another study.[50] Missing data were addressed using multiple imputation on the assumption that they were MAR.[57] Selection bias was addressed by identifying a matched set of patients based on their assessed clinical appropriateness for each of the procedures.[58] To control for potential remaining confounding factors regression analysis was used to estimate the differences in within-trial costs and QALYs attributable to each treatment. Uncertainty was characterised by using seemingly unrelated regression to capture the correlation between costs and health outcomes.[59]

The importance of study design in establishing the quality of evidence may vary with the characteristics of the intervention or the nature of the parameter to be informed.[60] The analytical frameworks reflected in many of the existing guidelines for economic evaluation were developed with the focus on health care technologies and, in particular, pharmaceutical interventions, provided within the health care sector.[32 61] The approach demonstrated in Paper 1[1] may have increased value as national reimbursement agencies take on responsibility for evaluating public health interventions which are less likely to be evaluated using RCTs, as are non-pharmaceutical interventions such as devices and diagnostics.[42 62] A limitation of the analysis presented in Paper 1[1] is the reliance on a single data source. Rarely will a single source of data be all that is available to inform an economic evaluation.[63] The next section discusses how an economic evaluation may incorporate multiple sources of evidence.

3. Combining multiple sources of evidence

A scientific approach to utilising available evidence can be used to support or justify societal level decisions about resource allocation.[64] The evidence base to inform an economic evaluation of health care interventions can be large. In order to evaluate the cost-effectiveness of a particular technology it must be compared to all feasible alternatives and so the evidence base may include data collected on a number of interventions. The cost and health differences resulting from the use of alternative treatment options may extend across a patient's lifetime, or perhaps even multiple lifetimes if inter-generational effects are considered. Thus relevant evidence may include data on the long-term prognosis of patients and information on the duration of treatment effects. A range of study designs may be relevant depending on the nature of the parameter to be informed. For example, RCTs might be the preferred source of data to inform estimates of treatment efficacy, while the cost of providing an intervention in practice may be derived from administrative data sources and evidence for the HRQL associated with a particular health state may be validly sourced from an observational study design. In order to identify the range of available evidence in a transparent and replicable manner, techniques for systematic review are used.[65 66] To conduct a systematic review a set of desirable criteria for included studies are defined against which those identified in a literature review can be assessed, first by determining whether they should be included in the review and then by assessing the quality of those eligible for inclusion. The issues addressed in Section two exemplify how the inclusion criteria and quality assessment might be considered for studies to inform estimates of treatment effects. The resulting set of identified evidence can then be brought to bear on the decision problem.

3.1 Multiple sources of evidence providing the same information

Methods for meta-analysis enable the synthesis of multiple sources of evidence on single parameters or even multiple parameters.[67-70] The results of a series of studies are combined to provide a single overall estimate that describes the combined weight of evidence, with the contribution of each individual study weighted according to its precision. In its simplest form a number of controlled studies making the same pairwise comparison between interventions are combined. However, where two or more active treatment options exist it is not unusual to observe a series of RCTs that include different sets of comparators. If all relevant treatments have not been compared directly within a single trial, then the choice between treatments must be based on an indirect comparison.[71] Under these circumstances techniques for network meta-analysis enable multiple treatment comparisons by synthesising a series of studies linked by common comparators.[72-74]

The fundamental assumption underlying all meta-analyses relates to the exchangeability of the relative treatment effects measured in different studies. In a fixed-effects meta-analysis the assumption is that the same relative treatment effect would be observed, subject to stochastic variation, if the same comparison was made in any of the trial populations. In a random-effects meta-analysis the population value of the relative treatment effect is assumed to vary between studies but to be drawn from a common distribution. Techniques for meta-regression can be used to assess whether study-level covariates

explain systematic differences in the relative treatment effect measured in different trials, but such methods are often underpowered in practice by the small number of the studies to be pooled.[75] The assumption of exchangeability must be correct in order for the pooled estimate to have internal validity. Problems with validity can arise when the availability of study results is correlated with the value of those results such that the synthesised studies are not a representative sample of all of the studies conducted. As covered in Section two, data that are not MCAR must be addressed and techniques such as funnel plots, trim and fill, regression methods and selection models are among those that have been used to identify and address the issue of missing studies (publication bias) in the context of meta-analysis.[76]

It can be more difficult to test the assumption of exchangeability in a network of trials compared to a set of trials that each make the same comparison. It is perhaps for this reason that many analysts and policy makers would choose to focus solely on direct data when both indirect and direct data are available.[77 78] However such an approach could lead to the exclusion of many relevant studies. For example, if two treatments are developed in parallel and each subject to a series of placebo controlled trials, a single trial comparing each head to head may represent only a small portion of the available evidence. Excluding relevant data could lead to the value of a further data collection being overestimated. Good quality indirect evidence can provide more information and alter estimates of decision uncertainty even where comparable direct evidence is available.[79 80] For this reason the exclusion of indirect data should not be the default position. Instead, if there is uncertainty about whether the trials providing indirect data are exchangeable with those providing direct data, the analyst must judge whether the benefits of including the additional data outweigh the consequences of potentially introducing bias. With more than two treatments under comparison, the same trial can provide both direct and indirect evidence and hence indirect evidence cannot be excluded from the analysis. If the assumption of exchangeability is appropriate, the trial estimates will be consistent and there will be a transitive relationship between the true underlying treatment effects (measured on some appropriate scale). This can be used to derive tests for consistency.[81] Inconsistency within the network will lead to the indirect comparison increasing uncertainty and may raise concerns about the validity of the underlying assumptions.[73 74] When the quantity of direct evidence is large, indirect evidence, which is automatically down-weighted in the analysis by having a larger variance, will have less impact on the pooled statistics compared to a situation in which less direct evidence is available. Thus the value of expanding a network to include more indirect evidence will vary with the particular decision problem. As identified studies become more distant from the comparison of interest, the evidence they provide is more indirect and has less impact on the pooled statistics. This leads to diminishing returns in terms of increased accuracy from widening the network of included studies, which must be balanced against the opportunity cost of increased analysis time.

Paper 2[2] demonstrates the method of a mixed treatment comparison for synthesising data from three RCTs, each evaluating a different pair from a set of three comparator treatments. Each RCT provided a unique pairwise comparison in a complete network, meaning, firstly, that indirect data could not be excluded from the analysis; and, secondly, that consistency could be assessed within the set of trials. However, a test for consistency

would not identify which of the three trials might not be exchangeable with the others as removing any one trial from the synthesis would eliminate inconsistency. One of the trials was different in that it was stopped early due to recruitment problems and consequently had shorter follow-up and was never published in a peer-reviewed journal. Presenting the results of pooling different subsets of the trials in a series of scenarios allows the decision maker to decide informally which set of trials may be most relevant to the decision problem. A more formal approach would be to weight the trials in the synthesis not only according to sample size, but also according to some assessment of quality or relevance, thus incorporating that additional uncertainty in the final pooled estimates. Extending evidence synthesis methods to include studies of differing quality or networks of evidence without a common comparator is practicable when using a Bayesian approach.[82 83] Such methods could have allowed the economic analysis in Paper 1[1] to draw on evidence from both observational and RCT designs.[84] However, this is an area of ongoing methodological work and consensus about how to determine appropriate weights has not yet been achieved.

3.2 Multiple sources of evidence providing a range of information

The trial data synthesised in Paper 2[2] provided estimates of progression free and overall survival from ovarian cancer. These do not fully describe the outcomes of interest to patients, who may be interested in overall survival and HRQL, and do not describe the outcomes of interest for decision makers who may wish to compare the health gains achievable with treatments for ovarian cancer to those generated in other disease areas. For these reasons the meta-analysis represented only a portion of the relevant evidence base, as additional information was also required on long-term prognosis, costs and HRQL. The need to incorporate all relevant evidence, to compare all appropriate comparators over a suitable time horizon, and to bring together multiple sources of evidence across a range of parameters, can be facilitated with the use of a decision analytic model.[63 85-87] Such models specify mathematical relationships between the various data sources to provide a quantitative assessment of cost-effectiveness that directly addresses the decision problem. They provide an explicit framework that should make clear the assumptions and logical relationships between inputs that determine the final outputs.[88] Paper 2[2] utilised a simple cohort model that relied on an assumption of exponentially distributed survival times to convert hazards to mean survival, weighted by HRQL, and incorporated the costs of treatment and adverse events.[89] In general, the appropriate design of such models depends on the decision problem to be addressed and the characteristics of the available data.[90-92] A model framework can be developed based on knowledge of the disease area and intervention, the data appropriately analysed for that framework and finally the outputs can be assessed in terms of validity to see whether modifications are required.[93] Keeping models as simple as possible (i.e. being parsimonious in design) is a desirable property as it aids communication and transparency.[92] The choice of more complex modelling approaches that increase analysis time should be justified in terms of increased accuracy in results.[94] Models must be understandable to both technical experts and policy makers if they are to be useful and used in informing resource allocation decisions.[95]

Some assessment of model validity can be made on the basis of how accurately a decision model represents the source data. With a single data source it may be possible to split the data into estimation and validation samples,[96] but this requires large samples and is infeasible in the context of a decision model that relies on aggregate level data. However, it is a model's predictive ability that is of interest when informing policy choices.[95] If an economic evaluation incorporates all available evidence or extrapolates beyond available data, there is no external data source against which to validate the resulting estimates of costs and health outcomes, even though such validation seems desirable.[97 98] The face validity or credibility of the model predictions can be assessed based on prior beliefs concerning plausible outcomes.[96] While certain models may be ruled out as invalid, it is, in general, not possible to identify a 'correct' model, only ones that are believed to be useful.[99] The choice between alternative modelling frameworks for the same decision problem relies in part on value judgements by the analyst and is an area of ongoing methodological development.[100] The way in which uncertainty in these choices and the relative precision of the multiple sources of evidence can be reflected in decision models is the subject of Section four.

4. Characterising uncertainty

There are numerous sources of uncertainty in cost-effectiveness analysis.[101] When the inputs to decision models are informed by sample data the underlying population values are estimated with imprecision and the evidence supports a range of plausible values with varying degrees of likelihood. Where direct data are unavailable to inform a particular parameter, a range of possible values can be obtained by eliciting beliefs from persons considered to have some capability for estimating relevant information.[102 103] Each different parameter value would imply a different estimate of the expected costs and health outcomes for the interventions under comparison, and this has been referred to as parameter uncertainty. The manner in which the multiple sources of evidence are identified and synthesised can be informed by the demands of the decision problem, the normative judgements underlying the analysis and finally the characteristics of the identified data. However, while particular methods may be ruled out as inappropriate a single best approach may not be apparent. This will, in part, be due to ignorance or a lack of understanding of the collection of physical mechanisms and processes to which the data relate, but will also stem from differences in value judgements, for example about the appropriate measure of HRQL. The existence of alternative plausible characterisations of the mathematical and logical relationships between parameters and alternative definitions of the set of relevant evidence leads to modelling or structural uncertainty in estimates of cost-effectiveness.[104]

4.1 Sensitivity analysis for parameter uncertainty

Uncertainty can be propagated through to the results of a decision model, thus describing a range of values of costs and health outcomes that could be estimated from the available (uncertain) evidence. This provides information as to how the results of the analysis would change for different assumptions and input values, and is termed sensitivity analysis. A distinction can be made between value sensitivity and decision sensitivity. Value sensitivity refers to the change in magnitude of the estimated costs and health outcomes as a result of a change in the value of the inputs. Decision sensitivity focuses on changes in the health care intervention that is regarded as most cost-effective following changes in the value of the inputs.[27] In this regard, if the most cost-effective alternative is the same for all possible input characterisations, the model results could be said to be robust or insensitive to those changes. Providing information only on value sensitivity, that is the range of possible values for costs and health outcomes, could be regarded as inadequate for informing decisions.[27] Paper 2[2] illustrated this fact by describing an example where reduced uncertainty in the estimated treatment effects and consequently the range of the model outputs was accompanied by increased decision uncertainty.

One-way and multi-way sensitivity analyses have been used to show the impact of varying parameter values within some plausible range. Individual parameters or groups of parameters are set to alternative or extreme values and the corresponding model outputs recorded. Tornado diagrams or spiderplots are among the common methods used to illustrate the resulting value sensitivity in model outputs.[103 105] While these may indicate parameter values for which the decision could change, they fail to quantify the likelihood of

any particular combination of parameter values. They fail to capture the variable support provided by the available evidence for alternative possible true values of uncertain empirical quantities. Another disadvantage of such 'one-at-a-time' or local sensitivity analyses is that the remaining parameters are held at their base case values and so important combinations can be missed.[106] When multiple deterministic sensitivity analyses have been conducted it can be difficult to summarise the combined impact of all the uncertainty.[85] An alternative approach for assessing parameter uncertainty is to describe the range of possible values supported by the data by assigning probability distributions to parameters. The selection of distributions should be chosen to reflect the characteristics of the parameter and of the data generating process, and should capture any dependence or correlation between parameters. The combined uncertainty from all input parameters can then be reflected in the model results using probabilistic sensitivity analysis (PSA).[27]

In Paper 1[1] bootstrapping the seemingly unrelated regression was used to conduct the PSA,[107] but Paper 2[2] utilised a more common approach for conducting PSA within decision analytic models by means of second order Monte Carlo simulation.[108 109] In a second order Monte Carlo simulation sets of random values are repeatedly drawn from all of the distributions assigned to input parameters and the corresponding model outputs recorded. This is a direct empirical approach for describing the distribution of model outputs as a function of variation in the model inputs. Sufficient simulations must be carried out to adequately describe that distribution and to obtain stable mean estimates that would not vary significantly were the analysis to be repeated. In practice this may require thousands of simulations. Fewer simulations may be required with more directed sampling of the inputs, for example with Latin Hypercube Sampling.[110] PSA is required to obtain unbiased mean estimates of costs and health outcomes for non-linear models where the expected value of the outputs cannot be derived from the expected value of the inputs.[85] The resulting distribution of costs and health outcomes characterises the degree of parameter uncertainty in the cost-effectiveness results. The decision uncertainty can then be summarised, for example by presenting the results in the form of cost-effectiveness acceptability curves that describe the probability of each alternative being the most cost-effective.[111]

4.2 Sensitivity analysis for other sources of uncertainty

In some cases, differences between alternative assumptions and judgements can be expressed in the form of an additional parameter in the model and thereby incorporated in the PSA.[104] When structural uncertainty is not parameterised within the model the decision uncertainty encapsulated by the probabilistic outputs will be incomplete.[104] Structural or modelling uncertainty can be addressed in one-way sensitivity analysis by reanalysing the decision model for each of set of plausible assumptions (scenarios). This characterisation of uncertainty may be appropriate when assessing sensitivity to value judgements, for example to the choice of discount rate. However, if more than one modelling approach is considered reasonable and each produces a different estimate of expected costs and health outcomes, it may be appropriate to use more sophisticated methods to characterise this uncertainty.[104] Methods for Bayesian model averaging would allow representation of this structural uncertainty by combining the outputs of the

alternative models, generating results that could incorporate both parameter and structural uncertainty.[112] To utilise this approach it is necessary to construct and analyse each appropriate model structure, thereby increasing analysis time. It is also necessary to estimate the appropriate weight to assign to each approach.[113] The increase in analysis time may be minimal with the choice between alternative regression analyses, but could increase if the construction of more than one type of decision analytic model is required. The representation of structural uncertainty within economic evaluation is an area of ongoing methodological work.

4.3 The implications of model design for PSA

When PSA is conducted by means of bootstrapping or Monte Carlo simulation, the results of the decision model are repeatedly calculated. For some models a lengthy analysis time is required for PSA and, in these computationally expensive models, PSA has in the past been regarded as unnecessary.[3 88 114] A particular example is models that rely on a first order Monte Carlo simulation in order to represent variability among patients as a way of estimating expected costs and effects. Stochastic variation between patients does not inform decisions about cost-effectiveness and introduces statistical noise in the calculation of expected costs and health outcomes. It may need to be included as a nuisance parameter where there is a non-linear relationship between a patient characteristic and the model outputs in order to obtain unbiased results. One way to achieve this is by using a patient level simulation or an individual sampling model, but sufficient simulations must be run to overcome the noise and to provide a stable estimate of costs and health outcomes. This use of first order Monte Carlo simulations to represent variability can make a second order Monte Carlo simulation 'computationally expensive' in terms of analysis time.[115] The total number of simulations required can be several orders of magnitude greater than with a cohort model that excludes stochastic variation by describing the experience of the average patient.

In order to have impact on reimbursement decisions the results of an economic evaluation should be considered valid and be available in a timely fashion.[116] Omitting a full characterisation of uncertainty in order to characterise variability may have seemed a necessary trade off to some but Paper 3[3] illustrated a range of alternative modelling approaches that retain the characterisation of variability while reducing analysis time to allow for a characterisation of uncertainty. Among the methods discussed are direct analytical solutions, alternative modelling frameworks and the use of response surface models which utilise statistical models to describe changes in the model outputs as a function of changes in some or all of the input values based on a smaller sample of simulations. However, these approaches are still not commonly applied and may prove more difficult to implement in more complex models.[106 117] The use of approximations would also introduce another element of modelling uncertainty that is absent when the distribution of model outputs is empirically sampled.[118] Ongoing methodological development to reduce the analysis time for computationally expensive models[119] alongside improvements in computing capabilities should pave the way for a characterisation of parameter uncertainty to be routinely included in economic evaluations. The benefits and importance of this are addressed in Sections five and six.

5. Informing the need for further research

When decisions are made on the basis of uncertain information there is a possibility that subsequent evidence will indicate that the interventions reimbursed did not in fact represent good value for money. The difference between the realised population health gains and those that could have been achieved by reimbursing the 'true' optimal set of health interventions represents the opportunity cost of uncertainty. This opportunity cost is a function of the probability of error and the magnitude of the health outcomes forgone if the wrong decision is made.[29] Further evidence is expected to reduce decision uncertainty, the probability of error and the expected value of this opportunity cost.[120] A characterisation of uncertainty in the estimated costs and health outcomes is therefore relevant when we wish to address questions about the sufficiency of the current evidence base, and whether health gains could be achieved by allocating resources to fund further research.[28] Economic evaluation can inform an assessment about whether current evidence is sufficient by quantifying the opportunity cost of uncertainty and the anticipated costs of acquiring further evidence.

For this reason it is important to distinguish between variability and uncertainty. Variability (aleatory uncertainty) arises from natural stochasticity and is irreducible through further research, as is heterogeneity in individual patient's responses to treatments. Uncertainty (epistemic) stems from incomplete knowledge and is, in principle, reducible through further measurement.[121] For example, conducting an additional trial would be expected to increase the precision of the pooled estimate in a meta-analysis. Similarly, further data collection can be used to validate modelling assumptions, reducing modelling uncertainty.[96] Where the expected health gains from acquiring additional information exceed the anticipated costs, a decision maker should, in principle, be willing to pay to obtain further evidence. The limited availability of research funds means that it is not possible to fund every research study where the expected benefits of the additional information exceed the costs of conducting the research. Hence methods are required to prioritise among alternative research designs, and utilising the results of a probabilistic model could provide a means to achieve this.[26 28 122-126 127]

5.1 The value of perfect information

The value of an error-free choice can be estimated by supposing that we can select the optimal intervention in every possible future state of the world. Comparing these maximum health gains that assume perfect knowledge to those expected under uncertainty provides the expected value of perfect information (EVPI), which represents the upper bound on the value of further evidence. This quantity is readily estimated from the results of a PSA as the distributions assigned to parameters describe the range of possible true values and the associated distribution of expected costs and health outcomes is calculated. As information is non-rival, value is generated in every subsequent decision that would utilise the additional evidence. Therefore, an estimation of the number of decisions is required to estimate the total value of research or population EVPI.[128] In order for further research to be worthwhile it is necessary that this exceeds some minimal cost of research.

Consequently, for some decision problems, the current evidence may be deemed sufficient on the basis of low or zero estimated population EVPI.[85 129]

Where the EVPI indicates that further research is potentially worthwhile, further questions can be addressed to determine what type of research may be valuable. In order to establish which of the model inputs are associated with decision uncertainty it is possible to estimate EVPI for individual parameters or groups of parameters.[120 130] The value of perfect information about a parameter provides another measure of sensitivity for judging the relative importance of uncertainty in the model inputs and can help determine more precisely what should be measured in future research studies.[27] It is calculated in a similar way as overall EVPI by supposing that we can select the most cost-effective alternative for every possible value of the parameter(s) in question, but with the uncertainty unaltered in the remaining inputs. This is readily obtained from the results of a PSA in linear models but requires an additional level of Monte Carlo simulation for non-linear models and consequently increased analysis time.[131 132]

The study design appropriate for providing additional information will vary according to the characteristics of the parameter in question, as will the costs of carrying out the research. A lengthy research design will be able to benefit fewer patients, and a more costly research design will have lower net benefits as it will displace either treatments that would have generated health gains or alternative research studies that could have generated health gains. Paper 4[4] describes methods for evaluating the value of information for a range of different research designs. Those previously described in the literature focussed on estimating the value of further information for particular parameters or groups of parameters within the decision model.[27 31 133-135] A prioritisation of research based on the value of alternative 'one-off' research designs might suggest that the study with the greatest individual expected health gains net of research cost be funded. However, this assessment would be incomplete. In general, the relative value of any particular research design should be evaluated against the full range of alternative research designs.[122 127 136] Paper 4[4] reviews existing approaches and extends the methods to evaluate a sequence of research to incorporate the impact that additional research on one parameter will have on the expected value of further information about the remaining parameters. The strategic value of a sequence of research is obtained when information that can be obtained relatively cheaply about one parameter indicates that a previously valuable more costly research design is unnecessary. In these circumstances the first study may not warrant prioritisation on its own merit. However, the value of a sequential research design is largely based on avoiding research designs with high opportunity costs, meaning there is a clear need to estimate accurately the costs of the proposed research designs and this may require an assessment of the optimal sample size as well as the research design.[136 137] This would require estimation of the value of sample information.[130]

5.2 The value of sample information

In practice, further evidence will reduce rather than eliminate uncertainty. The principle underlying the estimation of the expected value of sample information (EVSI) is the same as that for EVPI. The results of the intended future study must be predicted in order to calculate the expected health gains from a decision that incorporates that additional information.[130] This value is compared to the expected health gains from a decision based only on current evidence to estimate EVSI. The expected cost of the proposed research design is subtracted from the EVSI to provide the estimated net benefit of sample (ENBS) information. A positive ENBS provides a sufficient condition for conducting further research. However, when the results of further studies are predicted by means of Monte Carlo simulation the required analysis time can be very high unless the underlying model is very simple. Applications demonstrating the calculation of ENBS typically feature linear models with few parameters or are restricted to comparisons of two interventions where the new study is assumed to update the estimated incremental net benefits directly.[31 136] While methodological work is continually suggesting ways to reduce analysis time for more complex scenarios,[138] currently the widespread application of methods to prioritise between research designs, such as those discussed in Paper 4,[4] may rely on EVPI, meaning they act only as a guide to decision making.[125]

5.3 Realising the benefits of research

The health benefits of research are realised when they are translated into treatment decisions. Decisions about reimbursement and research should be made on the same basis, as both have the same objective to improve population health gains and both place demands on the resources available within the health care system.[29 139] The use of economic evaluation to inform research designs can ensure that the information collected directly addresses the decision uncertainty. However, some caution should be exercised when using PSA to inform the need for further research as this could give undue weight to uncertainty that is easily characterised by means of probability distributions at the expense of other sources of uncertainty.[99] The methods for valuing research designs described above assume full implementation of the health care intervention identified as most cost-effective. If in practice implementation is variable, this should be reflected in the calculations and requires some assessment of how further evidence would change clinical decisions.[127 140] Alternatively, the decision maker may wish to invest in strategies to alter clinical practice.[141] The impact of investment costs when reimbursement decisions are made on the basis of uncertain evidence is considered further in Section six which discusses the consequences of uncertain decisions.

6. The impact of uncertainty on reimbursement decisions

The evidence available to inform any decision is not static. As new evidence emerges economic evaluations can be updated and reimbursement decisions reconsidered.[142 143] If the decision to reimburse a health care intervention could be changed instantaneously and without cost as new information arrives, then the probability of executing those changes would be inconsequential.[30] However, this may not be the case in practice, and revising decisions may be impossible (irreversible) or costly. When there are resource implications associated with implementing and reversing decisions, a characterisation of uncertainty is required in order to estimate the consequences of that uncertain decision. Increasingly reimbursement decisions about new interventions are extending beyond simple approval or rejection, to consider a range of policy options that incorporate by design some response to decision uncertainty. For example the use of risk-sharing or patient access schemes may make the price paid for the technology conditional on the realised health gains.[144] Alternatively a decision may be linked to evidence collection by limiting reimbursement to patients included in research trials or by making reimbursement conditional on the provision of further information.[145-147] To inform the optimal use of these kinds of reimbursement decisions, a characterisation of uncertainty is always relevant and essential.

6.1 Incorporating investment and reversal costs

Investment costs that are sunk cannot to be recovered if a decision is changed. These may, for example, include the costs of training and persuading clinicians to deliver a new intervention. Such costs are relevant to, and should be incorporated when, assessing whether a new health care intervention is cost-effective. Typically this has been achieved through annuitisation, where up-front investment costs are converted into annual equivalent outlays across the useful life span of the asset in question.[148] For a piece of capital equipment the useful lifespan is simply the time until a replacement for the machine is required. The equivalent annual outlay should take account of the resale value (recoverable cost), reflecting the depreciation of the asset and the opportunity cost of the funds tied up in the asset.[148] For training costs the relevant life span is the time until the treatment decision is altered to one for which the training is irrelevant. The additional health gains expected by switching to a new intervention must be accumulated in sufficient patients to outweigh any up-front investment costs in order for an intervention to be regarded as cost-effective. Reversal costs are incurred only if the treatment decision is altered, and may include the dissemination of information on the policy change or political costs if it is accompanied by a loss of credibility. For reimbursement decisions informed by economic evaluations, changes would be precipitated by the arrival of new evidence.[128 142] Some assessment of the prospect that further evidence would alter the optimal treatment decision, such as that provided by PSA, is therefore required in order to appropriately incorporate investment and reversal costs in estimates of cost-effectiveness. This should be accompanied by some estimate of the timing of those changes.[5 128] By deferring a decision until uncertainty is reduced, investment costs can be avoided or mitigated. This describes the option value of delaying a decision, which will vary according to the degree of uncertainty and irreversibility associated with the particular decision problem.[30] This option value should be compared to the opportunity costs of delaying access to potentially

cost-effective health care technologies in terms of the health forgone by patients who, in the meantime, receive an intervention expected to be less cost-effective.[29 149 150]

6.2 The impact of reimbursement on the prospects for research

The option value of delay could be used to inform a decision to await the arrival of new evidence. Alternatively a decision maker with control over both reimbursement and research decisions could use an economic evaluation to inform simultaneous decisions about both.[31 149 151] Whether decision makers take an active or a passive role in the generation of new evidence, the impact of reimbursement on the incentives for further research must be considered. Once approval for reimbursement is granted manufacturers are no longer likely to retain an incentive to provide further evidence for the technology in question. Furthermore, randomising new patients to ongoing or new trials may prove difficult or be deemed unethical once an intervention is recommended for widespread use. While additional trials may still be feasible outside the jurisdiction if other countries have not yet approved the intervention, that evidence may not be able to reduce decision uncertainty if differences in patient population or health care systems limit generalisability.[152 153] A negative reimbursement decision may also limit the feasibility of further research, for example if additional evidence is desired on rare adverse events. Thus reimbursement decisions may result in the forgoing of further research. Only with estimates of the decision uncertainty can this potential opportunity cost, that is the potential value of information forgone, be evaluated. Paper 5[5] demonstrated these issues and proposed an amendment to existing decision rules based on expected cost-effectiveness that would incorporate this opportunity cost of value of information forgone.

6.3 Linking reimbursement with uncertainty

Incorporating the option value of delay in reimbursement decisions means that interventions associated with greater uncertainty and irreversibility are less likely to be approved for reimbursement on the basis of current available evidence than under a decision rule that ignores this opportunity cost.[30] Taking account of the impact of the value of information forgone would further decrease the likelihood of immediate approval for technologies associated with greater uncertainty where the ability to resolve that uncertainty was impaired by the decision to reimburse.[5] By incorporating uncertainty in reimbursement decisions a different optimal set of health interventions is identified compared to decision rules based on expected costs and health outcomes. The importance of incorporating uncertainty is increased further when decisions about whether to approve a health care technology for reimbursement extend beyond a simple approval or rejection. Decisions about reimbursement are increasingly made jointly with a decision about pricing (e.g. patient access schemes) that aims to mitigate the impact of decision uncertainty.[144] These schemes reduce the effective price of a technology, making it more likely to appear cost-effective. A characterisation of uncertainty is required in order to assess the extent of the effective price reduction, and allow any savings to be compared to the potential transaction costs of implementing such schemes.[154] Paper 5[5] showed how price adjustments can impact not only on the value of immediate access to a new treatment, but also on the value of further research, further altering the balance between

immediate approval and deferral. Another aim of risk-sharing schemes is to allow the collection of further data alongside use to support the reimbursement decision. The type of data that can be collected while a technology is approved for widespread use may be restricted in practice for the reasons addressed above. Consequently the feasibility of such schemes to reduce decision uncertainty may depend on the type of evidence required, the assessment of which was discussed in Section five. The need for randomised evidence could imply that technologies should be approved for use only within the context of a research study.[145] Even with such schemes the completion of further research is not guaranteed,[155 156] and the methods proposed in Paper 5[5] allow for the incorporation of uncertainty as to when and whether further evidence will become available.

7. Summary

The papers presented within this thesis cover a range of issues relevant to the characterisation of decision uncertainty in the economic evaluation of health care interventions. Together, the first four Sections of this Integrative Chapter provide an overview of how uncertainty can be incorporated in economic evaluations. This requires consideration and judgement about how the available evidence actually relates to the decision problem. Section two considered the problem of identifying relevant evidence on the causal effects of interventions. Details of analytical methods to address concerns about internal validity, by correcting for selection bias in observational data sources and addressing missing data, were described, in addition to methods for translating clinical outcomes into those required for economic evaluation. Paper 1 showed how these methods could be combined and used jointly to perform cost-effectiveness analysis when concerns about external validity meant that data from RCTs was considered irrelevant and available data sources failed to include all outcomes of interest for economic evaluation. The analysis simultaneously corrected for multiple potential sources of bias, and estimated outcomes on a scale that was appropriate for informing reimbursement decisions. The approach exemplified in Paper 1 is applicable to a wide range of health care technologies, demonstrating that the use of economic evaluation need not be limited to decisions about pharmaceutical or clinical interventions or about patient populations who are represented in RCTs.

Section three described how the weight of evidence could be assessed by combining information from a range of sources. Paper 2 showed how techniques for network meta-analysis could be applied to a series of studies that did not feature a common comparator in order to provide a single summary estimate of cost-effectiveness. The results demonstrated how different judgements about which sources of evidence are relevant could impact on estimates of cost-effectiveness. By enabling all available studies to be synthesised simultaneously, the methods demonstrated in Paper 2 allowed the question of the relevance of available studies to be based on an assessment of their relative merits, and not driven or constrained by the method of analysis. This can ensure that economic evaluations reflect the judgements of the decision makers they seek to inform.

Section four provided a technical description of how uncertainty and the weight of evidence can be characterised within decision analytic models. The focus was on how to summarise the decision uncertainty, and in particular the use of PSA for reflecting the combined impact of uncertainty in the model parameters. To this end Paper 3 illustrated some amendments to decision models that would mean that a full assessment of decision uncertainty was not curtailed by available analysis time. The methods demonstrated in Paper 3 can help ensure that the information provided by economic evaluations is driven by the needs of decision making and the requirements of the decision problem, and not constrained by the choice of modelling approach.

The importance of assessing decision uncertainty was discussed in Sections five and six. Section five showed how an assessment of decision uncertainty could be used to inform the design of further research to reduce that uncertainty, thereby improving treatment decisions and expected health outcomes for future patients. Paper 4 advanced existing methods by showing how the value of a sequential research design could be estimated. Utilising the methods described in Paper 4 to evaluate a more comprehensive range of potential research designs would ensure that the appropriate opportunity costs are reflected in economic evaluations that seek to inform the prioritisation of further research.

Finally in Section six, the various consequences of decision making under uncertainty were considered. The first was that the optimal reimbursement decision would change over time with the arrival of new evidence. When this was combined with investment or reversal costs it was illustrated that a characterisation of uncertainty was essential to identifying the optimal set of health interventions to reimburse. Paper 5 demonstrated how reimbursement decisions can impact on the prospects for further research, and that the value of information forgone is an irreversible aspect of reimbursement decisions. The methods presented in Paper 5 showed how this opportunity cost could be estimated and how it would alter decisions about which treatments should be adopted in order to maximise population health gains. Paper 5 provided a clear set of arguments as to why failing to incorporate the impact of decision uncertainty in reimbursement decisions would lead to health gains not being maximised, and showed that this was the case even when economic evaluations are used to inform decision makers whose remit is limited to approval or rejection of technologies for reimbursement. The framework presented in Paper 5 can be used to illustrate the impact of price adjustments on both assessments of cost-effectiveness and the need for further research, and can incorporate uncertainty in when and whether further research will report. The strength of the framework is that it is relevant to a wide range of decision problems, particularly when economic evaluations are used to inform reimbursement decisions made jointly with decisions about pricing or further evidence collection.

It is not yet possible to fully characterise all uncertainty within an economic evaluation or, for example, to routinely implement some of the more complex but theoretically feasible analyses discussed in this thesis. Nevertheless an economic evaluation that incorporates a characterisation of known or agreed uncertainties can provide a useful aid to decision making. Uncertainty is inherent in decisions about resource allocation regardless of how they are made, but by conducting a formal analysis the range of sources of uncertainty can be made explicit, and their impact on subsequent decisions made transparent. This can help ensure that decision makers are held accountable for their choices concerning the allocation of collective resources by the people on whose behalf they are made, by making clear the reasoning and judgements underlying those decisions. It can also help to ensure that the response to uncertainty is consistent across decisions made at different times or for different groups. Continued testing and development of the methods for characterising uncertainty in economic evaluations can ensure that the analyses provide information that is useful, timely and relevant to the decision problems they seek to inform.

Appendix A. Publications

Cite this article as: BMJ, doi:10.1136/bmj.39129.442164.55 (published 5 March 2007)

BMJ

RESEARCH

Cost effectiveness of clinically appropriate decisions on alternative treatments for angina pectoris: prospective observational study

S C Griffin, research fellow,[1] J A Barber, lecturer in medical statistics,[2] A Manca, Wellcome Trust training fellow in health services research,[1] M J Sculpher, professor of health economics,[1] S G Thompson, professor of medical statistics,[3] M J Buxton, professor of health economics,[4] H Hemingway, professor of clinical epidemiology[5]

[1]Centre for Health Economics, University of York, York
[2]UCLH/UCL Biomedical Research Centre, University College London Hospitals NHS Trust, London, and Department of Statistical Science, University College London
[3]Medical Research Council Biostatistics Unit, Cambridge
[4]Health Economics Research Group, Brunel University, London
[5]Department of Epidemiology and Public Health, University College London Medical School, London WC1E 6BT

Correspondence to: H Hemingway
h.hemingway@ucl.ac.uk

doi: 10.1136/bmj.39129.442164.55

ABSTRACT

Objective To assess whether revascularisation that is considered to be clinically appropriate is also cost effective.

Design Prospective observational study comparing cost effectiveness of coronary artery bypass grafting, percutaneous coronary intervention, or medical management within groups of patients rated as appropriate for revascularisation.

Setting Three tertiary care centres in London.

Participants Consecutive, unselected patients rated as clinically appropriate (using a nine member Delphi panel) to receive coronary artery bypass grafting only (n=815); percutaneous coronary intervention only (n=385); or both revascularisation procedures (n=520).

Main outcome measure Cost per quality adjusted life year gained over six year follow-up, calculated with a National Health Service cost perspective and discounted at 3.5%/year.

Results Coronary artery bypass grafting cost £22 000 (€33 000; $43 000) per quality adjusted life year gained compared with percutaneous coronary intervention among patients appropriate for coronary artery bypass grafting only (59% probability of being cost effective at a cost effectiveness threshold of £30 000 per quality adjusted life year) and £19 000 per quality adjusted life year gained compared with medical management among those appropriate for both types of revascularisation (probability of being cost effective 63%). In none of the three appropriateness groups was percutaneous coronary intervention cost effective at a threshold of £30 000 per quality adjusted life year. Among patients rated appropriate for percutaneous coronary intervention only, the cost per quality adjusted life year gained for percutaneous coronary intervention compared with medical management was £47 000, exceeding usual cost effectiveness thresholds; in these patients, medical management was most likely to be cost effective (probability 54%).

Conclusions Among patients judged clinically appropriate for coronary revascularisation, coronary artery bypass grafting seemed cost effective but percutaneous coronary intervention did not. Cost effectiveness analysis based on observational data suggests that the clinical benefit of percutaneous coronary intervention may not be sufficient to justify its cost.

INTRODUCTION

Guidelines based on clinical appropriateness criteria (optimising net benefits to health) are widely used to inform decisions about practice but are insufficient grounds for allocating healthcare resources. Although consensus exists that cost effectiveness analysis is needed to maximise the health gains achieved from a limited budget, how closely formally measured clinical appropriateness accords with cost effectiveness is not known. Population rates of coronary revascularisation, particularly percutaneous coronary intervention,[1] have increased rapidly, but many influential trials report no cost data,[2] analyse costs but not in relation to effectiveness,[3] or report cost effectiveness but not in terms of quality adjusted life years.[4] No three way randomised comparisons of the cost effectiveness of medical management, percutaneous management, and coronary artery bypass grafting exist.[2] Most importantly, as clinical equipoise is seen as an ethical prerequisite of randomisation, patients considered clinically more suitable for one procedure than another are, generally, not included in trials.

We sought, therefore, to identify the cost effectiveness of treatments rated as clinically appropriate by a multidisciplinary panel. We studied three alternative management strategies for coronary disease: coronary artery bypass grafting, percutaneous management, and medical management. We used the RAND appropriateness method,[5] which has been shown to be a prognostically valid method to determine the clinical suitability of unselected patients to have revascularisation,[6 7] but without consideration of costs. A cost effectiveness analysis based on quality adjusted life years is suitable for comparing treatments that are expected to affect mortality, morbidity, or both and is the most suitable basis to inform decisions about the provision or reimbursement of healthcare technologies.[8]

METHODS

Participants

The appropriateness of coronary revascularisation (ACRE) cohort consists of consecutive patients, recruited without exclusion criteria, who had coronary angiography between 15 April 1996 and 14 April 1997 at three hospitals of one National Health Service trust in London.[5] We identified 4121 patients and followed them for six years. Most patients had chronic coronary disease at baseline; only 6% were recruited during an admission with acute myocardial infarction. This economic evaluation focuses on the subgroup of 1740 patients rated as appropriate to have bypass surgery, percutaneous management, or both. Of these, we excluded 20 patients because they died without having revascularisation and no record of their intended management plan existed. All patients gave consent.[5] A detailed technical report of the methods used for this six year analysis is available (www.york.ac.uk/inst/che/staff/griffin.htm).

Clinical appropriateness ratings

Before recruitment of patients to the ACRE study, a nine member expert panel rated separately the clinical appropriateness of bypass surgery and percutaneous management in hypothetical patients for 985 specific clinical indications (based on the RAND-Delphi technique[9]). Appropriateness was defined by clinical judgment, based on available evidence, that doing a procedure would be associated with more benefit than harm. Appropriateness ratings were assigned to ACRE participants on the basis of their individual clinical characteristics at the time of angiography to identify those patients who would benefit clinically from revascularisation. Details of these ratings have been reported previously.[10] The ratings did not take account of patient preferences or cost considerations.

NHS resource use and costs

We adopted the cost perspective of the NHS and included the costs of bypass surgery, percutaneous management and angiography procedures (including hospital stay), drugs, admissions for chest pain, general practitioner visits, outpatient appointments, and visits to the emergency department. The occurrence of hospital admissions, their reasons, and lengths of stay came from the NHS-wide clearing service. Data on drugs at baseline and at one year and six year follow-ups came from hospital case notes, general practitioners' and patients' questionnaires (response rate 85% at baseline, 77% at one and six years), and case notes after admissions for chest pain. Frequency of attendance in the previous year at a general practice, outpatient department, or casualty came from the patients' questionnaire. Unit costs came from previously published studies and published pricing lists for the United Kingdom.[3 11 12] Costs are reported in UK sterling (£), updated to 2003/4 prices and discounted at 3.5%/year.[8 11]

Outcomes and quality adjusted life years

More than 99% of the study sample were flagged with the UK Office of National Statistics, which notified us of dates of death. We ascertained acute non-fatal myocardial infarction through the NHS-wide clearing service and classified it according to recent criteria.[13] As part of the six year questionnaire, patients completed the EQ-5D health related quality of life instrument, from which we derived utility scores.[14 15] Utilities represent quality weights for the calculation of quality adjusted survival; 1 corresponds to the highest degree of quality of life, and 0 is equivalent to dead. As we did not collect EQ-5D data at baseline and one year follow-up, we estimated utilities for these time points by using other variables on the patients' questionnaire to predict EQ-5D scores in a regression model. The variables used included the patient rated severity of angina symptoms,[16] shortness of breath, and demographic factors. We calculated an estimate of quality adjusted survival for each patient by weighting their survival according to their quality of life. We discounted quality adjusted survival at a rate of 3.5% a year.[8]

Statistical methods

We analysed patients in three groups on the basis of their being rated clinically appropriate for bypass surgery only (that is, not for percutaneous management), for both procedures, or for percutaneous management only (that is, not for bypass surgery). Within the three groups, we compared those who had bypass surgery, those who had percutaneous management, and those who had neither type of revascularisation (medical management). These actual management groups are defined as the treatment received within one year of index angiography. For patients who died within one year without having a procedure, we used the intended management recorded at baseline as a proxy for actual management.

We used regression analyses with interaction terms to estimate the effect of actual management, by appropriateness category, on cost effectiveness (total costs and quality adjusted survival), presence of angina, and mortality at six years' follow-up. We adjusted analyses for the potential confounders in table 1, and additionally included baseline utility for adjustment of quality adjusted life years.[17] Odds ratios for presence of angina at six years came from multiple logistic regression and hazard ratios for death from Cox regression. Regression of life years (cumulative survival) used an ordinary least squares approach. We analysed cost effectiveness by using seemingly unrelated regression,[18] a multivariate regression technique that accounts for the potential correlation between costs and quality adjusted survival.

To compare the three management strategies, we used standard cost effectiveness decision rules.[19] We calculated the incremental cost effectiveness ratio comparing two management strategies, which represents the cost per quality adjusted life year gained by moving to a more costly, more effective method of management. To reflect uncertainty, we derived cost

Table 1 | Baseline characteristics by appropriateness category. Values are numbers (percentages) unless stated otherwise

Baseline covariate (No missing)	Appropriate for CABG only (n=815)	Appropriate for both (n=520)	Appropriate for PCI only (n=385)
Mean (SD) age (years)	63 (9)	59 (9)	59 (9)
Male	671 (82)	403 (78)	292 (76)
Ethnic group (141):			
White	623 (83)	396 (84)	303 (85)
Other	130 (17)	76 (16)	51 (14)
Previous CABG	60 (7)	68 (13)	29 (8)
Previous PCI	31 (4)	62 (12)	45 (12)
Previous MI	371 (46)	268 (52)	207 (54)
Heart failure	116 (14)	48 (9)	44 (11)
Previous stroke	76 (9)	34 (7)	29 (8)
Diabetes	130 (16)	69 (13)	52 (14)
No of diseased vessels:			
≤1	18 (2)	213 (41)	282 (73)
2	106 (13)	263 (51)	79 (21)
3 or left main stem	691 (85)	44 (8)	24 (6)
Diffuse disease	199 (24)	71 (14)	28 (7)
Left ventricular function (329):			
Normal	456 (67)	324 (82)	251 (78)
Impaired	220 (33)	71 (18)	69 (22)
Mean (SD) Parsonnet operative risk score[37]	7.3 (5.4)	5.2 (4.4)	5.8 (5.3)
CCS score (362):			
0 (no angina)	38 (6)	22 (5)	14 (5)
I-II (mild angina)	179 (27)	97 (24)	76 (25)
III-IV (severe angina)	437 (67)	282 (70)	213 (70)

CABG=coronary artery bypass graft; CCS=Canadian Cardiovascular Society; PCI=percutaneous coronary intervention; MI=myocardial infarction.

effectiveness acceptability curves showing the probability that each treatment is cost effective for a range of threshold amounts that the NHS would be willing to pay per quality adjusted life year.[20] We have used a threshold of £30 000 (€45 000; $58 000) per quality adjusted life year in presenting the cost effectiveness results, on the assumption that the maximum incremental cost effectiveness ratio acceptable to the NHS lies between £20 000 and £40 000 per quality adjusted life year.[21]

We imputed missing data on length of stay and patient reported resource use (general practitioner visits, outpatient attendances, and visits to casualty) with simple ordinary least squares. In adjusted analyses, and when obtaining utilities, we used multiple imputation with chained equations.[22 23] We created five imputation datasets to allow retention of between imputation variance in estimating standard errors.[24]

We used univariate sensitivity analyses to investigate assumptions about the need to adjust for the definition of actual management, the inclusion of patient reported resource use, the exclusion of hospital admissions for reasons unrelated to chest pain, and the differential timing of the one year patients' questionnaire (one year from revascularisation for percutaneous management and bypass surgery or one year from index angiography for medical management).

RESULTS

Baseline characteristics

Of the 1720 patients in the economic analysis, 815 (47%) were rated as appropriate for bypass surgery only, 520 (30%) were rated as appropriate for both procedures, and 385 (22%) were rated as appropriate for percutaneous management only (table 1). The prevalence of current smokers was similar across the three appropriateness groups (10%, 10%, and 13%). The severity of angina (Canadian Cardiovascular Society class) was also similar. Those rated as appropriate for bypass surgery tended to have a higher number of diseased vessels, higher prevalence of impaired left ventricular function, and higher operative risk scores. Single vessel disease was more common among patients rated as appropriate for percutaneous management only.

Clinical outcomes

Over the six year follow-up, 44% (335/754) of patients initially treated with medical management and 26% (93/364) of those initially treated with percutaneous management went on to have additional revascularisation procedures; of those who initially had bypass surgery, further revascularisation was needed for only 4% (25/602) of patients. Angina was present in 55% (560/1020) of patients at six years. Among patients rated as appropriate for bypass surgery, adjusted analyses showed a significantly raised odds of angina for those who had percutaneous management or medical management compared with those who had bypass surgery (table 2). A suggestion of similarly raised odds was apparent in patients suitable for percutaneous management who had medical management compared with those who had percutaneous management. Overall, 16% (277/1720) of patients died during follow-up. Adjusted analyses in the group appropriate for bypass surgery showed a raised risk of death for those who had medical management compared with bypass surgery or percutaneous management. Among the group rated appropriate for percutaneous management and bypass surgery, we found some evidence that the risk of death in patients who had percutaneous management or medical management was almost twice that in those initially treated with bypass surgery (table 2).

Costs

The costs of treating patients with medical management remained considerably lower than those for percutaneous management or bypass surgery (fig 1). In year one, the costs of medical management were 9-12% of bypass surgery costs and those for percutaneous management were 43-50% of bypass surgery costs. By year six, these ratios rose to 43-50% for medical management and 78-82% for percutaneous management, primarily owing to the need for additional revascularisation procedures (table 2). Approximately 74% of percutaneous interventions involved one or more stents. The average unadjusted costs of each initial treatment strategy were, at baseline, £12 500 for bypass surgery, £5800 for percutaneous

Table 2 | Six year clinical outcomes by appropriateness category and actual management. Values are numbers (percentages) unless stated otherwise

	Appropriate for CABG only			Appropriate for both			Appropriate for PCI only		
	Received CABG (n=408)	Received PCI (n=54)	Received MM (n=353)	Received CABG (n=149)	Received PCI (n=173)	Received MM (n=198)	Received CABG (n=45)	Received PCI (n=137)	Received MM (n=203)
Revascularisation	10 (2)	20 (37)	193 (55)	9 (6)	47 (27)	83 (42)	6 (13)	26 (19)	59 (29)
Chest pain admission	80 (20)	20 (37)	88 (25)	58 (39)	73 (42)	82 (41)	14 (31)	45 (33)	70 (34)
Non-fatal MI	19 (5)	2 (4)	23 (7)	15 (10)	19 (11)	16 (8)	3 (7)	7 (5)	10 (5)
Angina at 6 years:	(n=238)100 (42)	(n=34)27 (79)	(n=195)99 (51)	(n=89)52 (58)	(n=102)61 (60)	(n=119)82 (69)	(n=26)11 (42)	(n=95)49 (52)	(n=122)79 (65)
Adjusted odds ratio* (95% CI)	1	2.90 (1.03 to 8.14)	1.63 (1.09 to 2.42)	1	1.17 (0.69 to 1.98)	1.60 (0.89 to 2.87)	0.91 (0.39 to 2.13)	1	1.65 (0.95 to 2.87)
Deaths:	68 (17)	7 (13)	76 (22)	18 (12)	28 (16)	34 (17)	9 (20)	9 (7)	28 (14)
Adjusted hazard ratio*† (95% CI)	1	0.58 (0.25 to 1.40)	1.41 (1.01 to 1.98)	1	2.00 (1.08 to 3.69)	1.70 (0.95 to 3.03)	2.10 (0.81 to 5.44)	1	1.33 (0.62 to 2.87)

CABG=coronary artery bypass graft; MI=myocardial infarction; MM=medical management; PCI=percutaneous coronary intervention.
*Adjusted for age, sex, ethnic group, Canadian Cardiovascular Society score, left ventricular function, previous stroke, MI, previous CABG, previous PCI, diabetes, diffuse disease, diseased vessels, heart failure, and Parsonnet score; for all adjusted analyses, missing data have been imputed by multiple imputation[23] and results are a summary of those from five imputation datasets.[24]
†Overall hazard ratio presented despite some evidence of non-proportional hazards over time.

management, and £1400 for medical management; these rose to £16 200, £12 100, and £9200 at six years. The adjusted analysis made little difference to the estimated cost differences (table 3).

Cost effectiveness

Average predicted baseline utility was 0.55, considerably lower than the UK population norm of 0.80 for people aged 55 to 64.[25] The average utility score among those alive at six years improved to 0.65. Because of deaths, the average length of survival in the study was around five years. Adjusted analyses by appropriateness category showed no significant differences in discounted mean survival duration (that is, life expectancy) across treatment groups. For the groups rated appropriate for bypass surgery only or percutaneous management only, the clinically appropriate treatment had the highest mean quality adjusted life years. In the group rated appropriate for either type of revascularisation, bypass surgery had the highest mean quality adjusted life years. Adjusted analyses showed significant differences in quality adjusted survival comparing medical management with bypass surgery in patients appropriate for bypass surgery only or for percutaneous management and bypass surgery (table 3).

Among patients rated as appropriate for bypass surgery only, the incremental analysis showed that percutaneous management had an incremental cost effectiveness ratio of £11 000 per quality adjusted life year compared with medical management and bypass surgery had an incremental cost effectiveness ratio of £22 000 per quality adjusted life year compared with percutaneous management (table 3). Thus bypass surgery is the most cost effective procedure for patients rated as appropriate for bypass surgery. The probability that bypass surgery is the most cost effective treatment strategy was 59% at a willingness to pay of £30 000 per quality adjusted life year, compared with 40% for percutaneous management (fig 2).

In patients rated as appropriate for both procedures, percutaneous management was ruled out on grounds

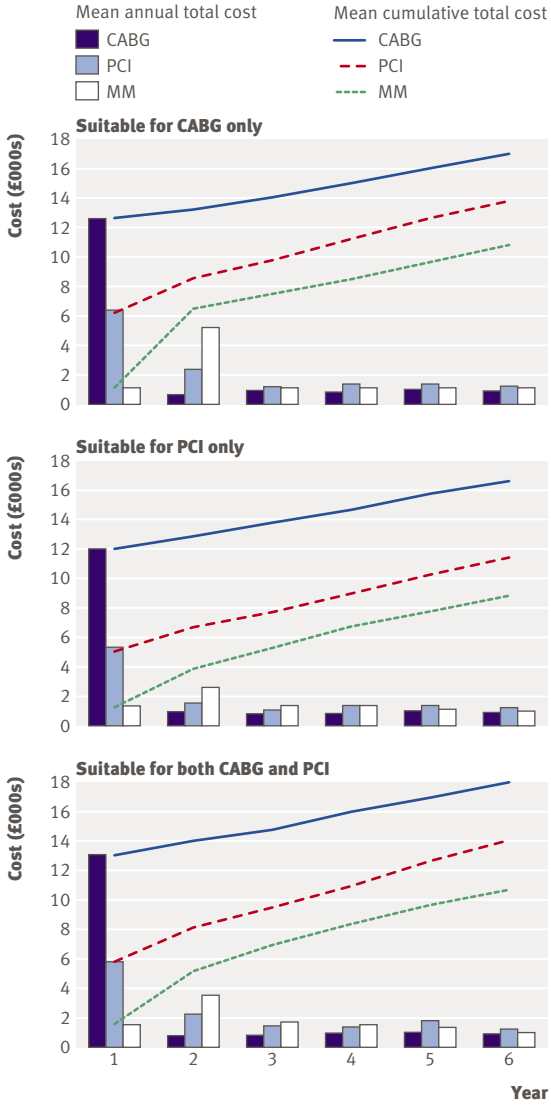

Fig 1 | Unadjusted total cost: mean annual and cumulative cost by actual management and appropriateness category. Costs presented in 2003/4 UK sterling, discounted at 3.5% a year. CABG=coronary artery bypass surgery; MM=medical management; PCI=percutaneous coronary intervention

Table 3 | Effectiveness and economic measures by appropriateness category and actual management. Values are mean (SD) unless stated otherwise

	Appropriate for CABG only (n=815)			Appropriate for both (n=520)			Appropriate for PCI only (n=385)		
	Received CABG (n=408)	Received PCI (n=54)	Received MM (n=353)	Received CABG (n=149)	Received PCI (n=173)	Received MM (n=198)	Received CABG (n=45)	Received PCI (n=137)	Received MM (n=203)
Utility at baseline*	(n=281) 0.54 (0.23)	(n=35) 0.48 (0.22)	(n=262) 0.60 (0.22)	(n=94) 0.45 (0.22)	(n=120) 0.50 (0.22)	(n=145) 0.54 (0.22)	(n=33) 0.56 (0.24)	(n=96) 0.57 (0.23)	(n=148) 0.61 (0.21)
Utility at 6 years*	(n=264) 0.69 (0.29)	(n=35) 0.61 (0.36)	(n=219) 0.67 (0.31)	(n=100) 0.66 (0.31)	(n=108) 0.65 (0.30)	(n=131) 0.61 (0.30)	(n=28) 0.69 (0.28)	(n=100) 0.65 (0.29)	(n=129) 0.66 (0.29)
Life years†:	4.95 (1.47)	4.95 (1.57)	4.94 (1.32)	5.14 (1.19)	5.07 (1.27)	5.08 (1.14)	5.07 (1.15)	5.31 (0.95)	5.20 (0.95)
Adjusted MD (95% CI)	0‡	0.03 (−0.32 to 0.39)	−0.03 (−0.21 to 0.15)	0‡	−0.17 (−0.45 to 0.11)	−0.09 (−0.36 to 0.17)	−0.13 (−0.55 to 0.29)	0‡	0.06 (−0.21 to 0.33)
QALYs†:	(n=317) 3.29 (1.55)	(n=40) 3.01 (1.54)	(n=293) 3.02 (1.53)	(n=114) 3.13 (1.37)	(n=127) 2.93 (1.65)	(n=164) 2.83 (1.39)	(n=40) 3.08 (1.59)	(n=111) 3.31 (1.47)	(n=161) 3.15 (1.43)
Adjusted MD (95% CI)	0‡	−0.15 (−0.51 to 0.20)	−0.40 (−0.58 to −0.22)	0‡	−0.24 (−0.52 to 0.04)	−0.39 (−0.70 to −0.09)	−0.07 (−0.50 to 0.37)	0‡	−0.06 (−0.36 to 0.24)
Total cost (£)†:	16 980 (7879)	13 875 (7815)	10 850 (7220)	17 859 (6940)	14 007 (10 453)	10 690 (7888)	16 541 (5571)	11 493 (6468)	8775 (7364)
Adjusted MD (95% CI)	0‡	−3230 (−5417 to −1044)	−5870 (−6961 to −4779)	0‡	−3820 (−5510 to −2130)	−7255 (−8875 to −5636)	4947 (2359 to 7534)	0‡	−2847 (−4510 to −1184)
ICERs (£ per QALY) (from adjusted values)	22 000 (*v* PCI)	11 000 (*v* MM)	–	19 000 (*v* MM)	ED	–	D	47 000 (*v* MM)	–

CABG=coronary artery bypass graft; PCI=percutaneous coronary intervention; D=dominated; ED=ruled out by extended dominance; ICER=incremental cost effectiveness ratio (cost per QALY gained calculated in comparison with next relevant, less costly alternative)[19]; MD=mean difference; MM=medical management; QALY=quality adjusted life year.
*Includes values from prediction model as well as observed utility (for six year values only).
†Discounted at rate of 3.5% a year.
‡Reference category.
Adjustments are for age, sex, ethnic group, Canadian Cardiovascular Society score, left ventricular function, previous stroke, myocardial infarction, previous CABG, previous PCI, diabetes, diffuse disease, diseased vessels, heart failure, and Parsonnet score; adjusted analysis of QALYs and total cost are from seemingly unrelated regression and include an additional adjustment for baseline utility; for all analyses missing data have been imputed by multiple imputation[23]; results are a summary of those from five imputation datasets.[24]

of "extended dominance"—at any cost effectiveness threshold, either medical management or bypass surgery was a more cost effective option (fig 2). The incremental cost effectiveness ratio for bypass surgery compared with medical management was estimated as £19 000 per quality adjusted life year (table 3). The probability that bypass surgery, percutaneous management, and medical management are the most cost effective forms of management was estimated at 63%, 22%, and 15% respectively (fig 2).

Among patients appropriate for percutaneous management only, percutaneous management dominated bypass surgery (that is, it was less costly and more effective than bypass surgery) and had an estimated incremental cost effectiveness ratio of £47 000 per quality adjusted life year compared with medical management (table 3). This incremental cost effectiveness ratio is above the maximum usually considered acceptable by the NHS, making medical management the most cost effective treatment strategy in patients rated as appropriate for percutaneous management. The probability that percutaneous management is most cost effective was estimated at 36%, compared with 54% for medical management (fig 2).

Sensitivity analysis

The cost effectiveness results were robust to adjustment for the timing of the one year questionnaire, the exclusion of patient reported cost data, and the inclusion of other hospital admissions. Altering the definition of actual management to treatment received within 15 months of angiography did not affect the study results. However, among the group rated as clinically appropriate for bypass surgery, altering the definition of actual management to treatment received within nine months caused the incremental cost effectiveness ratio for bypass surgery compared with percutaneous management to increase from £22 000 to £33 000.

DISCUSSION

Among consecutive patients, all of whom were judged to be clinically appropriate for revascularisation, the analysis suggests that bypass surgery within 12 months was cost effective in relation to a standard UK threshold but that percutaneous management was not. These findings challenge clinical practice and healthcare policy, which has evolved on a basis of evidence of effectiveness from clinical trials, largely in isolation from considerations of cost effectiveness. As angina pectoris has a high incidence and prognostic burden in the general population, where many patients are not evaluated for revascularisation,[26] these findings are important for public health.

Strengths and limitations

This study has several strengths. Firstly, we defined clinical appropriateness by using an explicit method that has been shown to be highly reliable and prognostically valid.[5 27] We used an expert panel to provide an independent measure of appropriateness, before recruitment of patients. Based on all forms of evidence—meta-analyses,[28 29] trials, observational studies, and clinical experience—the panel's judgments aimed to articulate clinical appropriateness, without consideration of cost.

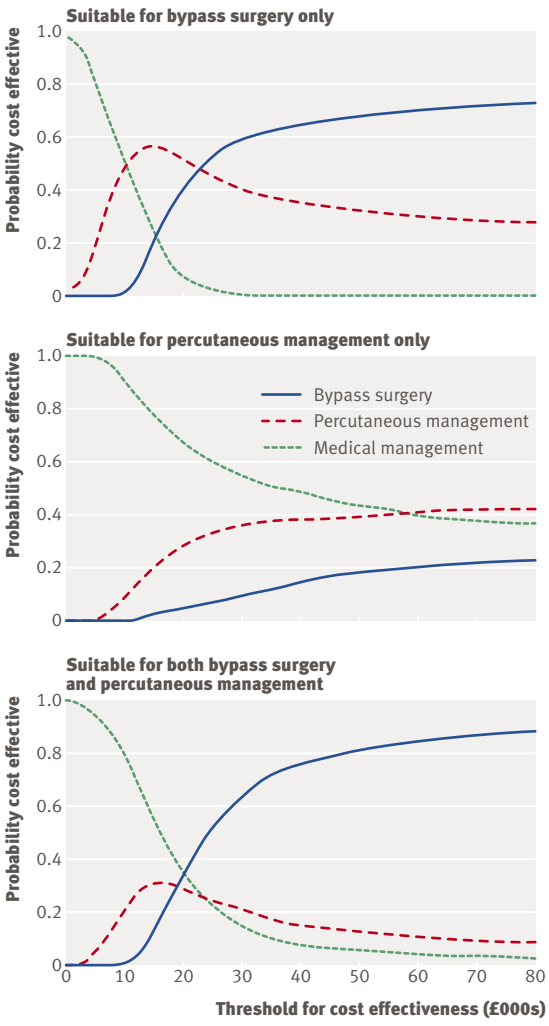

Fig 2 | Cost effectiveness acceptability curves

Secondly, more than 90% of unselected consecutive patients were matched to an appropriateness rating, allowing our results to represent a real world view of cost effectiveness; patients selected into clinical trials may not be representative of unselected patients in terms of baseline severity of disease, health related quality of life, use of resources, and prognosis.[30] The case fatality we observed in this population (approximately 3% a year) was comparable to that seen in large, less selected, primary care populations.[26] Our analysis was made possible by the ubiquitous phenomenon that not all patients in whom revascularisation is deemed appropriate will actually receive it. We have shown that clinical appropriateness ratings in a broad unselected population accord with evidence of clinical effectiveness from trials.

Thirdly, we were able to make comparisons between three alternative management strategies that may never be simultaneously investigated in cost effectiveness analysis alongside randomised trials. This is both a strength and limitation; observational studies may be the only study design to answer the research question, but they come at a cost of confounding. Patients who go on to receive bypass surgery may have been destined to have better outcomes than those who do not, and the results may therefore be "confounded by indication." We sought to redress this both by design (patients who are judged to be suitable candidates for revascularisation are by definition more similar than those who are not) and by analysis (by using multivariate regression analysis to adjust for the potential confounding effects of baseline clinical and demographic characteristics).

A second limitation of our analysis, inherent in the need for long term follow-up studies, is that we do not know whether percutaneous management judged clinically appropriate according to the most recent criteria remains not cost effective. Although this awaits empirical testing, several lines of evidence indicate that this may be the case. The cost of the percutaneous management procedure has increased with drug eluting stents,[31] and percutaneous management remains associated with higher costs resulting from subsequent admission to hospital. Meanwhile, an increasing number of drugs for secondary prevention have been shown to improve outcomes in chronic coronary disease.

Association of clinical appropriateness with better outcomes

Patients who had bypass surgery were least likely to have angina present at six years. This confirms findings from randomised trials,[2-32] but in a broader, unselected population. The treatment rated as clinically appropriate corresponds with the greatest number of quality adjusted life years, although this result was statistically significant only for the comparison of medical management with clinically appropriate bypass surgery. Throughout the six year follow-up of this unselected patient group, quality of life remained lower than expected from age specific population norms. The low utility scores, despite intervention, reflect the findings of the bypass angioplasty revascularisation investigation (BARI) trial, in which angina was also found to have a substantial negative impact on quality of life.[33]

Lack of cost effectiveness of percutaneous coronary intervention

Our analysis indicates that clinically appropriate percutaneous management within 12 months was not cost effective. Despite the large increase in numbers of percutaneous coronary intervention procedures seen in many countries, we do not find this result surprising: the high costs of percutaneous management and the need for subsequent procedures,[3 32 34] absence of mortality benefit,[35] and absence of a marked gain in quality of life have all been separately reported in trials. Our contribution is to estimate jointly the cost and outcome of percutaneous management that is considered clinically appropriate, in comparison with medical management and bypass surgery. The use of percutaneous management has increased rapidly, on the sole basis of clinical criteria without consideration of the economic consequences. As a result, the funds invested in percutaneous management could potentially be invested in more cost effective treatments that would provide greater benefit to NHS patients.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Guidelines based solely on clinical appropriateness are widely used to inform decisions about practice; whether care based on appropriateness criteria is also cost effective is unknown

No three way randomised comparisons of the cost effectiveness of medical management, percutaneous coronary intervention, and coronary artery bypass grafting have been done

Evidence of cost effectiveness in "real world," routine practice settings is also lacking

WHAT THIS STUDY ADDS

Coronary artery bypass grafting seems both clinically beneficial and cost effective for patients judged clinically appropriate for revascularisation

The clinical benefit of percutaneous coronary intervention, however, does not seem to be sufficient to justify its additional cost

Cost effectiveness of bypass surgery

Bypass surgery within 12 months was the most cost effective strategy among patients rated as appropriate to have bypass surgery if the maximum incremental cost effectiveness ratio the NHS is willing to accept is around £30 000 per quality adjusted life year. However, if we define actual management as treatment received within nine months (which may be viewed as short given a mean UK waiting time of 6.5 months at the time of the study[36]) the incremental cost effectiveness ratio just exceeds £30 000 per quality adjusted life year. We found that the relative cost differences between bypass surgery and percutaneous management reduced over the follow-up period, but the absolute difference remained significant at six years. Medical management is consistently the least costly form of management; the low treatment costs are not fully offset by high admission rates or costs of late procedures. Previous trials have indicated a greater degree of "catch up." The BARI trial showed an increase in percutaneous management costs from 65% to 98% of bypass surgery costs after 12 years of follow-up; most of the gain was in the first five years.[33] Future work could extend the cost effectiveness model over a lifetime horizon, in which interventions with large "up-front" costs, such as bypass surgery, may seem more cost effective. Willingness to pay thresholds differ markedly across countries with the greatest numbers of people with chronic coronary disease—India, China, Russia, United States—and similar study designs could be used to inform national policies.

Conclusion

This cost effectiveness analysis in a real world setting offers a challenge to physicians, providers, and payers to show that the management of coronary disease currently offered, however clinically beneficial, is also cost effective. This was the case for bypass surgery within 12 months, but not for percutaneous management, for which the additional benefit was too small to justify the additional cost over the consistently less costly strategy of medical management.

SCG and JAB contributed equally to this paper.

Contributors: All authors made substantial contributions to conception and design (health economic: MJS, MJB, SCG, and AM; statistical: SGT and JAB; clinical epidemiology: HH) and analysis (SCG and JAB). All authors contributed to interpretation of data, drafting the article or revising it critically for important intellectual content, and final approval of the version to be published. HH, JAB, MJS, MJB, and SGT obtained funding. HH is the principal investigator on the ACRE study and is the guarantor. Adam Timmis contributed to the conception and design (clinical cardiology) and interpretation of data. Angela Crook and Roger Stafford prepared the data. Sue Philpott and Natalie Fitzpatrick were the project coordinators.

Funding: British Heart Foundation project grant. The British Heart Foundation had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. HH and MJS hold public health career scientist awards from the Department of Health. During the course of the ACRE cost effectiveness analysis study, AM became recipient of a Wellcome Trust training fellowship in health services research.

Competing interests: None declared.

Ethical approval: The five local ethics committees: West Essex, South Essex, Redbridge and Waltham Forest, East London and the City, and Barking and Havering.

1 Togni M, Balmer F, Pfiffner D, Maier W, Zeiher AM, Meier B, et al. Percutaneous coronary interventions in Europe 1992-2001. *Eur Heart J* 2004;25:1208-13.
2 Hueb W, Soares PR, Gersh BJ, César LAM, Luz PL, Puig LB, et al. The medicine, angioplasty, or surgery study (MASS-II): a randomized, controlled clinical trial of three therapeutic strategies for multivessel coronary artery disease. *J Am Coll Cardiol* 2004;43:1743-51.
3 Sculpher MJ, Smith DH, Clayton T, Henderson R, Buxton MJ, Pocock SJ, et al. Coronary angioplasty versus medical therapy for angina. *Eur Heart J* 2002;23:1291-300.
4 Claude J, Schindler C, Kuster GM, Schwenkglenks M, Szucs T, Buser P, et al. Cost-effectiveness of invasive versus medical management of elderly patients with chronic symptomatic coronary artery disease: findings of the randomized trial of invasive versus medical therapy in elderly patients with chronic angina (TIME). *Eur Heart J* 2004;25:2195-203.
5 Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53-63.
6 Hemingway H, Crook A, Feder G, Banjeree S, Dawson J, Magee P, et al. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. *N Engl J Med* 2001;344:645-54.
7 Kravitz RL, Laouri M, Kahan JP, Guzy P, Sherman T, Hilborne L, et al. Validity of criteria used for detecting underuse of coronary revascularization. *JAMA* 1995;274:632-8.
8 National Institute for Clinical Excellence. *Guide to the methods of technology appraisal*. London: NICE, 2004.
9 Bernstein SJ, Laouri M, Hilborne LH, Leape LL, Kahan JO, Park RE, et al. *Coronary angiography: a literature review and ratings of appropriateness and necessity*. Santa Monica, CA: RAND, 1992.
10 Hemingway H, Crook AM, Dawson JR, Edelman J, Edmondson S, Feder G, et al. Rating the appropriateness of coronary angiography, coronary angioplasty and coronary artery bypass grafting: the ACRE study. *J Public Health Med* 1999;21:421-9.
11 Netten A, Curtis L. *Unit costs of health and social care*. Canterbury: Personal Social Services Research Unit, University of Kent, 2004.
12 British Medical Association, Royal Pharmaceutical Society of Great Britain. *British national formulary*. London: BMA, RPS, 2004. (No 48.)
13 Joint European Society of Cardiology/American College of Cardiology Committee. Myocardial infarction redefined—a consensus document of the Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. *Eur Heart J* 2000;21:1502-13.
14 Kind P. The EuroQoL instrument: an index of health-related quality of life. In: Spilker B, ed. *Quality of life and pharmacoeconomics in clinical trials*. 2nd ed. Philadelphia: Lippincott-Raven, 1996.
15 Dolan PG, Kind P, Williams A. *A social tariff for EuroQol: results from a UK general population survey*. York: Centre for Health Economics, University of York, 1995.
16 Campeau L. Grading of angina pectoris. *Circulation* 1976;1976:522-3.
17 Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;14:487-96.
18 Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;13:461-75.

19 Johannesson M, Weinstein S. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993;12:459-67.
20 Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;10:779-89.
21 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ* 2001;323:1300-3.
22 Royston P. Multiple imputation of missing values. *Stata Journal* 2004;4:227-41.
23 Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18:681-94.
24 Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, 1987.
25 Kind P, Hardman G, Macran S. *Centre for Health Economics discussion paper 172: UK population norms for EQ-5D*. York: Centre for Health Economics, University of York, 1999.
26 Hemingway H, McCallum A, Shipley M, Manderbacka K, Martikainen P, Keskimäki I. Incidence and prognostic implications of stable angina pectoris among women and men. *JAMA* 2006;295:1404-11.
27 Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;338:1888-95.
28 Pocock SJ, Henderson RA, Rickards AF, Hampton JR, King SB 3rd, Hamm CW, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet* 1995;346:1184-9.
29 Yusuf S, Zucker D, Peduzzi P, Fisher LD, Takaro T, Kennedy JW, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the coronary artery bypass graft surgery trialists collaboration. *Lancet* 1994;344:563-71.
30 Krumholz HM, Peterson ED, Ayanian JZ, Chin MH, DeBusk RF, Goldman L, et al. Report of the National Heart, Lung, and Blood Institute Working Group on Outcomes Research in Cardiovascular Disease. *Circulation* 2005;111:3078-86.
31 Kaiser C, Rocca B-L, Buser P, Bonetti PO, Osswald S, Linka A, et al. Incremental cost-effectiveness of drug-eluting stents compared with a third-generation bare-metal stent in a real-world setting: randomised Basel stent kosten effektivitäts trial (BASKET). *Lancet* 2005;366:921-9.
32 Bypass Angioplasty Revascularization Investigation (BARI) Investigators. Five-year clinical and functional outcome comparing bypass surgery and angioplasty in patients with multivessel coronary disease: a multicenter randomized trial. *JAMA* 1997;277:715-21.
33 Hlatky MA, Boothroyd DB, Melsop KA, Brooks MM, Mark DB, Pitt B, et al. Medical costs and quality of life 10 to 12 years after randomization to angioplasty or bypass surgery for multivessel coronary artery disease. *Circulation* 2004;110:1960-6.
34 Hill RA, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, et al. Coronary artery stents: a rapid systematic review and economic evaluation. *Health Technol Assess* 2004;8:1-256.
35 Bucher H, Hengstler P, Schindler C, Guyatt GH. Percutaneous transluminal coronary angioplasty versus medical treatment for non-acute coronary heart disease: meta-analysis of randomised controlled trials. *BMJ* 2000;321:73-7.
36 Hospital Episode Statistics. Main operations: summary, England, N HS Hospitals, 1998-99. www.hesonline.nhs.uk.
37 Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:I-3-I-12.

Accepted: 23 January 2007

Volume 9 • Number 2 • 2006
VALUE IN HEALTH

Incorporating Direct and Indirect Evidence Using Bayesian Methods: An Applied Case Study in Ovarian Cancer

Susan Griffin, MSc,[1] Laura Bojke, MSc,[1] Caroline Main, MSc,[2] Stephen Palmer, MSc[1]

[1]Centre for Health Economics, University of York, York, UK; [2]Centre for Reviews and Dissemination, University of York, York, UK

ABSTRACT

Objective: To demonstrate the application of a Bayesian mixed treatment comparison (MTC) model to synthesize data from clinical trials to inform decisions based on all relevant evidence.

Methods: The value of an MTC model is demonstrated using a probabilistic decision-analytic model developed to assess the cost-effectiveness of second-line chemotherapy in ovarian cancer. Three clinical trials were found that each made a different pair-wise comparison of three treatments of interest in the overall patient population. As no common comparator existed between the three trials, an MTC model was used to assess the combined weight of evidence on survival from all three trials simultaneously. This analysis was compared to an alternative approach that combined two of the trials to make the same comparison of all three treatments using a common comparator, and an informal approach that did not synthesize the available evidence.

Results: By including all three trials using an MTC model, the credible intervals around estimated overall survival were reduced compared with making the same comparison using only two trials and a common comparator. Nevertheless, the survival estimates from the MTC model result in greater uncertainty around the optimal treatment strategy at a cost-effectiveness threshold of £30,000 per quality-adjusted life-year.

Conclusions: MTC models can be used to combine more data than would typically be included in a traditional meta-analysis that relies on a common comparator. They can formally quantify the combined uncertainty from all available evidence, and can be conducted using the same analytical approaches as standard meta-analyses.

Keywords: Bayesian, cost-effectiveness, evidence synthesis.

Introduction

Economic evaluation of health-care technologies is increasingly recommended for informing allocation of health-care resources in several countries around the world [1]. For each disease area under consideration, there may exist several competing, mutually exclusive treatment options. A decision that recommends one or more of these treatments as preferable to the rest should be based on a simultaneous comparison of all the relevant alternatives. Where this decision is made on grounds of cost-effectiveness, the comparison needs to produce a cardinal ranking of the alternatives. In such cases, informal approaches that produce an ordinal ranking of the alternatives will be insufficient. For almost every set of technologies considered, it will be necessary to combine information on costs and effects from several sources, and modeling techniques will invariably be employed. To aid decision-makers, decision modeling should be conducted within a clear analytical framework that is consistent, transparent and able to take account of all available evidence.

A common feature of models comparing three or more treatment options is the existence of several clinical trials that each compare a different combination of the relevant comparators. This network of evidence requires that the analyst use methods of evidence synthesis to derive the relative treatment effects of the relevant comparators in a systematic and explicit framework. There are many examples in the literature of meta-analyses that synthesize trials comparing various treatments relative to a common comparator, often placebo. Nevertheless, some networks may not feature a common comparator between all trials, particularly where trials compare treatments to an active control. This article presents an application of a mixed treatment comparison (MTC) model to synthesize three different pair-wise comparisons of three treatments of interest [2,3], and discusses why this is the most appropriate method of synthesizing these types of data.

These methods were applied as part of a Technology Assessment Report (TAR) for the National Institute for Health and Clinical Excellence (NICE) in the UK. The process of decision analytic modeling is now

Address correspondence to: Susan Griffin, Alcuin A Block, University of York, Heslington, York YO10 5DD, UK. E-mail: scg3@york.ac.uk

10.1111/j.1524-4733.2006.00090.x

seen as central to the process of health technology assessment in general, and it plays a key role in the NICE appraisal process [4]. To inform treatment decisions for advanced second-line ovarian cancer, two previous TARs had separately examined the cost-effectiveness of topotecan and pegylated liposomal doxorubicin hydrochloride (PLDH) [5,6]. An update was commissioned because the previous assessments did not provide the simultaneous direct comparison of topotecan, paclitaxel and PLDH necessary to inform the decision.

This article explores alternative approaches to estimating the relative effectiveness of topotecan, paclitaxel and PLDH for input into a cost-effectiveness analysis. First we discuss the problems with an informal "naive" approach to the simultaneous comparison before applying two different formal approaches. The first formal approach is based on comparing relative treatment effects to a common comparator. This approach employs direct evidence on treatments where they are compared with a common comparator, and infers indirect comparisons between uncommon treatments from different trials. The second formal approach is based on an MTC model for combining direct and indirect evidence simultaneously, and does not require a common comparator between all trials. Nevertheless, this approach does rely on the network being "complete," that is, that every trial have a treatment in common with at least one other trial. Both approaches were undertaken in the same software platform to provide consistency in the methods of calculation.

Methods

Overview

Existing studies provided only minimal assistance for decision makers concerned with the reimbursement of alternative chemotherapeutic agents [7,8], because no simultaneous comparison of the three drugs had been made. A new decision analytic model was therefore developed to address this issue to provide information for the relevant decision-maker (NICE). A full technical report is available relating to this work [9]; this article focuses on the methodological issues related to different approaches for synthesizing clinical data and the key results for the overall patient population. The objective is to highlight the importance of ensuring that appropriate techniques are applied to the parameterization of decision models to characterize uncertainty based on all relevant evidence. The model estimates costs from a UK National Health Service (NHS) perspective and health outcomes in terms of life-years gained (LYG) and quality-adjusted life-years (QALYs) for the full range of relevant treatment strategies. An overview of the basic structure of the model is provided in Figure 1. The uncertainty around the utility weights and estimated resource use or costs was characterized using appropriate probability distributions.

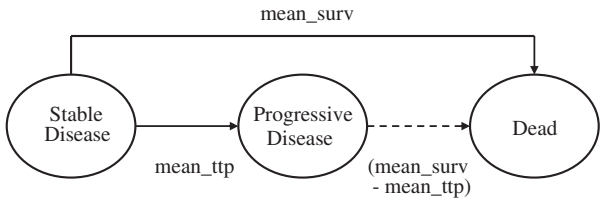

Figure 1 Structure of the economic model. Key: mean_surv = mean (overall) survival time; mean_ttp = mean time to progression.

The model calculates overall survival as the sum of two distinct periods: the progression-free period, and the time from progression to death, calculated as the difference between overall survival and progression-free survival (PFS). To calculate QALYs, overall survival is quality-adjusted using separate utility weights for the two periods of time during which the average patient is stable (i.e., progression-free) or in progression. The costs included relate only to the treatment period and comprised the costs of study drugs, premedication, monitoring, drug administration and the cost of managing adverse events.

Progression-Free and Overall Survival

A systematic review was conducted to identify randomized controlled trials (RCTs) and systematic reviews comparing the clinical effectiveness of licensed chemotherapies (PLDH, topotecan and paclitaxel) with any other second-line treatment, including best supportive care. A total of nine RCTs were identified [10–18]. Four of these studies were excluded from the model because the comparator groups (all unlicensed treatments, either in terms of indication, dosage, and route of administration or length of the chemotherapy cycle) provided no evidence on the relationship between the licensed treatments examined in this study [13–16]. The unlicensed comparator in each of these four trials was uniquely represented, meaning the separate pairwise comparisons could not be linked to provide indirect evidence about the relative treatment effects of licensed comparators. The remaining five trials assessed comparators that were used within their licensed indications. Of the five included studies, three included patients from the overall patient population [10–12], and two included only participants with platinum sensitive disease (relapse greater than 6 months after first-line therapy) [17,18]. In the TAR for NICE we considered all five trials and examined cost-effectiveness in different subgroups. Nevertheless, for the purpose of this article we focus on the three studies that examined the overall patient population, to demonstrate the methods for evidence synthesis. Table 1 shows the comparisons made in these three trials.

Table 1 Comparisons made in the included trials

Trial	PLDH	Topotecan	Paclitaxel
A* [10]	✓	✓	
B [11]		✓	✓
C [12]	✓		✓

*Update provided to NICE of previously published trial [34].
PLDH, pegylated liposomal doxorubicin hydrochloride.

No trial compared all the relevant treatment comparators simultaneously and there are three different pair-wise comparisons of the relevant comparators in the three included studies in the overall patient population. The first step in reviewing the available evidence was to examine the characteristics of each trial and the data reported. All trials were conducted in comparable patient populations with similar inclusion and exclusion criteria. The main distinction between the trials was the length of follow-up. Trial A had a median follow-up of approximately 3 years; trial B had a median follow-up of approximately 4 years; trial C did not report median follow-up, but the maximum follow-up was approximately 3 years, so we may infer that the median follow-up was less than 3 years. Survival data were presented as median weeks overall and PFS and in the form of hazard ratios between treatments. The hazard ratio represents the most accurate of these measures for comparing survival between treatments, because it is specifically designed to allow for censoring and time to an event [19]. Furthermore, the use of the (log) hazard ratio and its variance allows studies to be pooled using conventional meta-analytic approaches. Trial C [12] provided data on overall survival, but not PFS.

Table 2 presents the hazard ratios for overall survival and PFS extracted from the trials. It is clear that the three trials provide inconsistent information on the relative effectiveness of the three comparators. The results from trial A would suggest that topotecan is superior to paclitaxel, and the results from trial B suggest that PLDH is superior to topotecan. This would lead one to infer that PLDH would appear superior to paclitaxel in trial C, but this is not the case. Trials A and C were smaller than trial B, and this incongruous result could be put down to random chance. In this sense one may expect some inconsistency in most networks of evidence, although perhaps not as evident as in this example. Another factor to consider is the differing lengths of follow-up and what impact we may expect this to have on the hazard ratios. As each clinical trial estimated the hazard ratio using a Cox proportional hazards model [20], some assumption about the independence of the hazard ratio with respect to time has already been made. We similarly assumed that the hazard ratio was independent of length of follow-up, and thus the two formal approaches are based on the assumption that the hazard ratios estimated in the included trials are exchangeable.

Methods for Evidence Synthesis

A simple, informal approach to this network of evidence might be to make inferences based only on the direct comparisons made within each trial. For example, one might present three separate "pair-wise" analyses (PLDH vs. topotecan, topotecan vs. paclitaxel and PLDH vs. paclitaxel) based on the results of each individual trial. Nevertheless, clearly this approach does not fulfill the objective of providing a simultaneous direct comparison of all the relevant alternatives, and cannot produce a cardinal ranking of the alternatives or provide information about the associated decision uncertainty. Hence, for the purposes of decision-making, this approach does not provide an appropriate analytic framework.

An alternative, more formal approach, that is commonly applied, is to compare the relevant treatments on the basis of a common comparator [21]. In this example, because there is no single comparator which is common to each of the studies, only two of the three trials could be included at any one time. For example, if topotecan were selected as a common baseline, one could perform a cost-effectiveness analysis of all three treatments on the basis of the hazard ratios reported in trials A and B, to the exclusion of the information provided by trial C. Although this approach has certain advantages in comparison to the informal approach (i.e., it can provide a cardinal ranking and associated decision uncertainty), it also has a number of important limitations. Most importantly, by considering only two of the three trials, this approach makes selective use of the relevant evidence-base and, in doing so, ignores the information provided in the trial which has been omitted. Depending on the results of the omitted study, its exclusion could have important conse-

Table 2 Hazard ratios extracted from included trials

Trial	Overall survival Hazard ratio* (95% CI)	Progression-free survival Hazard ratio* (95% CI)
A: Topotecan vs. Paclitaxel [11]	0.914 (0.681–1.226)	0.811 (0.603–1.092)
B: Topotecan vs. PLDH [10]	1.216 (1–1.478)	1.118 (0.928–1.347)
C: Paclitaxel vs. PLDH [12]	0.931 (0.702–1.234)	n/a

*Hazard ratio less than one favors topotecan in trials A and B, and paclitaxel in trial C.
CI, confidence interval; n/a, not available; PLDH, pegylated liposomal doxorubicin hydrochloride.

quences for both the incremental cost-effectiveness ratio (ICER) and the associated decision uncertainty. Consequently, in those situations where there is no common comparator across all relevant trials, this approach does not provide an appropriate framework for decision making which is able to take account of all relevant evidence.

It is possible to incorporate all this evidence simultaneously in the form of an MTC model [2,3]. Such a model provides an explicit analytical framework to identify the most cost-effective treatment strategy given the combined weight of evidence from all the relevant clinical trials. There are several examples in the literature of statistical models for combining mixed comparison evidence to provide a consistent set of treatment effect estimates, relative to a common baseline [22]. Using a similar approach, a model was developed to estimate a set of hazard ratios relative to a common baseline, using the Bayesian inference software program WinBUGS [23]. One of the advantages of a Bayesian approach is the ability to make explicit probability statements about hypotheses, for example, the probability that a particular strategy is the most cost-effective alternative.

To correctly incorporate data from every trial, a Bayesian MTC model, assuming fixed treatment effects, was used to combine the (log) hazard ratios [2]. In brief, the technique extends the assumptions made in simple meta-analyses to include the principle of transitivity: if the true differences between three pair-wise comparisons of treatments *X*, *Y*, and *Z* are θ_{XY}, θ_{XZ}, and θ_{YZ}, then we expect:

$$\theta_{XZ} = \theta_{XY} + \theta_{YZ} \quad (1)$$

An additional assumption is that treatment effects can be expressed on an appropriate scale, such as log hazard ratio. The analyst must decide whether the patient populations and other trial characteristics are homogenous enough to justify synthesizing the relative treatment effects across the trials. The variances around the reported hazard ratios were used to incorporate the uncertainty around the estimated treatment effects. The analysis assumes that the (log) hazard ratios, observed in the clinical trials, are normally distributed about a true underlying effect size, θ, according to the precision (= 1/variance), τ^2, also observed in the trials. The underlying treatment effects are given independent vague priors, N (0, 0.001). The term "vague" is used to denote that prior information in the form of expert opinion or prior data is not included in the analysis, and hence these parameters are assigned very diffuse distributions. In other examples where prior information is available, this can be translated into informative priors for the relevant model parameters. The main premise of the analysis is that had paclitaxel been included as a comparator in trial A, or PLDH included as a comparator in trial B, the observed relative treatment effect of paclitaxel compared with PLDH would have been the same as that observed in trial C.

$$\text{Log}(HR_{top\_pac}) \sim N(\theta_{top\_pac}, \tau^2_{top\_pac}); \quad \theta_{top\_pac} \sim N(0,0.001) \quad (2)$$

$$\text{Log}(HR_{top\_PLDH}) \sim N(\theta_{top\_PLDH}, \tau^2_{top\_PLDH}); \quad \theta_{top\_PLDH} \sim N(0,0.001) \quad (3)$$

$$\text{Log}(HR_{pac\_PLDH}) \sim N(\theta_{pac\_PLDH}, \tau^2_{pac\_PLDH}); \quad \theta_{pac\_PLDH} = \theta_{top\_PLDH} - \theta_{top\_pac} \quad (4)$$

Where top_pac = topotecan vs. paclitaxel; top_PLDH = topotecan vs. PLDH; pac_PLDH = paclitaxel vs. PLDH. The full WinBUGS code for the evidence synthesis is available elsewhere [9].

Cost-Effectiveness Analysis

The advantages of the MTC model in comparison to the other approaches are examined in detail by undertaking separate cost-effectiveness analyses using the different methods. The hazard ratios from each approach are used to calculate expected costs and QALYs, using the same probabilistic decision-analytic model. The following sections describe in more detail how the hazard ratios were used in calculating mean quality-adjusted survival.

Absolute PFS and overall survival were calculated for a specified baseline regimen to apply the estimated hazard ratios for the other two regimes. An active treatment was chosen to represent the baseline regimen because no trial provided a comparison with supportive care (no chemotherapy). Topotecan was selected as the baseline comparator for the formal approaches, because trial data for topotecan were presented over the longest period of follow-up (approximately 4 years), and in the greatest detail. For the informal approach of three separate pair-wise comparisons, paclitaxel was used as the baseline in the trial that did not include topotecan. None of the trials provided an estimate of the absolute hazard of progression or death for an individual treatment, and so an exponential approximation was used to calculate the absolute hazards from the reported median PFS and overall survival. The baseline absolute hazard (h) and its variance was calculated according to the following formulae:

$$h = -\text{LN}(0.5)/t \quad (5)$$

$$\text{Var}\ (h) = h^2/r \quad (6)$$

Where t = median weeks survival; r = number of events.

Using this approach, the baseline absolute hazard (h) can then be converted into a mean survival time, for PFS and overall survival, by simply taking the inverse of the hazard ($1/h$). This conversion from median to mean survival is necessary, because the decision about whether an intervention is cost-effective is

made on the basis of the expected costs and effects at the population level [24]. Because survival data typically follow a skewed (i.e., not symmetrical) distribution, the median does not provide a good estimate of the mean. The estimated hazard ratios were then applied to the baseline absolute hazard to calculate the absolute hazard of progression and death for paclitaxel and PLDH. These were also converted into mean PFS and overall survival by taking the inverse of the absolute hazard.

Table 3 shows the WinBUGS code used to undertake the two different formal approaches to calculating the relative treatment effects for overall survival. Similar models were used to estimate PFS and the rate of adverse events. When basing the analysis on direct comparisons against topotecan, trial C is excluded. The MTC model allows the information from trial C to be incorporated, based on the assumptions outlined above. Each model was run for 20,000 iterations, from which the data from the first 10,000 were discarded to allow the model to "burn-in" or converge. Ten thousand iterations were used to ensure full coverage of the distribution of incremental net benefits in the probabilistic sensitivity analysis.

The output from the meta-analyses undertaken in WinBUGS was imported directly into Microsoft Excel 2000. This output consisted of 10,000 draws from the posterior distributions for PFS and overall survival (and adverse events) for each drug. These data incorporate the uncertainty around expected survival. The survival estimates were then combined with data on resource use and cost to obtain the mean estimates for the outcomes of interest and their associated uncertainty. Given that mean costs and QALYs gained are estimated with uncertainty, the outputs from the simulations were then used to generate cost-effectiveness acceptability curves (CEACs) for the alternative analyses [25]. These show the probability that each strategy is the most cost-effective given alternative maximum values that the health service may be willing to pay for an additional QALY. Previous studies have estimated the threshold for cost-effectiveness to be in the range of £20,000 to £40,000 per QALY in the UK NHS [26].

Results

Using the model structure developed for the assessment report [9], a cost-effectiveness analysis based solely on trial A would conclude that PLDH dominates topotecan, because it is estimated to be the more effective and cheaper alternative. A cost-effectiveness analysis based solely on trial B would produce an ICER of £33,532 per QALY gained with topotecan compared with paclitaxel. A cost-effectiveness analysis based solely on trial C would not be able to estimate QALYs because of the lack of data on PFS; one may wish to infer that paclitaxel would dominate PLDH, on the

Table 3 WinBUGS code for estimating treatments effects based on a common comparator, and using an MTC model

Common comparator	MTC model
model {	**model {**
# priors for basic parameters	# priors for basic parameters
dab ~ dnorm(0.001) #LHR a vs. b	**dab ~ dnorm(0, 0.001)** #LHR a vs. b
dac ~ dnorm(0.001) # LHR a vs. c	**dac ~ dnorm(0, 0.001)** # LHR a vs. c
la ~ dnorm(0.001) # Log hazard rate for a	**la ~ dnorm(0, 0.001)** # Log hazard rate for a
	# define functional parameters
	dbc <- dac - dab
# define absolute hazards on log scale	# define absolute hazards on log scale
ltopo ~ dnorm(la,ptopo)	**ltopo ~ dnorm(la,ptopo)**
lb <- la - dab	**lb <- la - dab**
lc <- la - dac	**lc <- la - dac**
# define absolute hazards natural scale	# define absolute hazards natural scale
log(a) <- la	**log(a) <- la**
log(b) <- lb	**log(b) <- lb**
log(c) <- lc	**log(c) <- lc**
# convert to mean survival	# convert to mean survival
OSa <- 1/a	**OSa <- 1/a**
OSb <- 1/b	**OSb <- 1/b**
OSc <- 1/c	**OSc <- 1/c**
# likelihood	# likelihood
yab ~ dnorm(dab,pab)	**yab ~ dnorm(dab,pab)**
yac ~ dnorm(dac,pac)	**yac ~ dnorm(dac,pac)**
}	**ybc ~ dnorm(dbc,pbc)**
	}
# data list (log hazard ratios, precision, baseline hazard and precision)	# data list (log hazard ratios, precision, baseline hazard and precision)
list(yab = –0.0899,yac = 0.1956, pab = 44.4520,pac = 100.6718, ltopo = –4.4558, ptopo = 114.9304)	**list(yab = –0.0899,yac = 0.1956,ybc = –0.0715, pab = 44.4520, pac = 100.6718,pbc = 48.2933, ltopo = –4.4558, ptopo = 114.9304)**

MTC, mixed treatment comparison. Text in bold is the actual Winbugs code. Text not in bold is annotation of the Winbugs code.

Table 4 Results from the WinBUGS model and decision-analytic model based on using a common comparator, and using an MTC model

	Common comparator			MTC model		
WinBUGS output	Log hazard ratio for overall survival (95% credible interval)			Log hazard ratio for overall survival (95% credible interval)		
Topotecan vs. paclitaxel	–0.092 (–0.390 to 0.206)			0.060 (–0.162 to 0.287)		
Topotecan vs. PLDH	0.196 (0.001 to 0.389)			0.129 (–0.046 to 0.304)		
Paclitaxel vs. PLDH	*Implied 0.288 (–0.0641 to 0.646)*			0.069 (–0.153 to 0.294)		
Cost-effectiveness	Topotecan	Paclitaxel	PLDH	Topotecan	Paclitaxel	PLDH
Mean PFS (weeks)	24.50	20.13	27.49	24.50	20.13	27.49
Mean OS (weeks)	86.03	79.70	104.79	86.03	92.06	98.08
Quality-adjusted survival (weeks)	34.21	30.86	40.91	34.21	34.63	38.86
Total cost (£)	11,394	6,354	7,714	11,394	6,354	7,714
ICER	D	—	7,033	D	—	16,714

D, dominated; ICER, incremental cost-effectiveness ratio in terms of cost per quality-adjusted life-year; MTC, mixed treatment comparison; OS, overall survival.

basis of survival, but this would be both inconsistent and inferior to the analyses of trials A and B based on cost per QALY [27]. This informal approach clearly cannot inform a decision between all relevant comparators because it does not reconcile the information provided by the three trials in a coherent analytical framework.

Table 4 presents the log hazard ratios for overall survival and their associated 95% credible intervals from the two formal analyses. It also displays the results from the decision-analytic model based on each set of hazard ratios. Despite the introduction of potentially inconsistent evidence from trial C, the 95% credible intervals from the MTC model are marginally smaller compared with the model that excludes trial C. The data from trial C reduce the amount by which PLDH is estimated to be superior to paclitaxel, although the direction of effect does not change. What does change is the direction of effect for topotecan vs. paclitaxel, which is reversed in the analysis using the MTC model to incorporate information from all three trials. Although this results in topotecan being dominated by paclitaxel in the decision-analytic model, qualitatively the adoption decision is unaffected; if we assume that society is willing to pay more than £17,000 for an additional QALY PLDH is found to be the optimal treatment strategy in both analyses. Figure 2 shows the CEACs for the two analyses. The reduction in the amount by which PLDH is estimated to be superior to paclitaxel increases the uncertainty in the adoption decision for cost-effectiveness thresholds in the range of £20,000 to £40,000 per QALY.

The indirect hazard ratio for paclitaxel compared with PLDH from the approach that excludes trial C is 1.33 ($e^{0.288}$). If we compare this to the direct comparison of these drugs in trial C, shown in Table 1, we see that this is outside the range of the 95% confidence interval reported in the trial. The hazard ratios from the MTC model all lie within the 95% confidence intervals of the direct comparisons reported in the clinical trials.

Discussion

If no attempt is made to synthesize data from multiple clinical trials, a set of individual trial-based cost-effectiveness analyses will be insufficient to inform a decision that must recommend one or more optimal strategies among all relevant alternatives. In this example, the inconsistency between the trial estimates in isolation meant that the series of pair-wise comparisons provided conflicting evidence. Nevertheless, this problem occurs even without conflicting evidence. Consider two pair-wise comparisons based on trials A (PLDH vs. topotecan) and B (topotecan vs. paclitaxel). PLDH dominates topotecan in trial A, and so we may wish to compare it to paclitaxel in an incremental analysis. Nevertheless, the costs and effects of paclitaxel have been estimated against a different topotecan baseline compared with PLDH, and so calculating the ICER from the two separate trial-based analyses would be incorrect. Added to that is the inability of this approach to inform decision makers adequately about uncertainty. The limitations of the "naive" approach are evident even in this very simple example. Clearly, as the number of relevant trials and comparisons increases, so too will the difficulty of reconciling the information provided by a list of inconsistent, disparate pair-wise comparisons.

When faced with a network of evidence that does not feature a common comparator among all the trials, more traditional methods of meta-analysis could not make use of all the available data [28–30]. The choice of common comparator, and therefore the choice of which studies to exclude, would affect the results of the analysis. In this example we chose topotecan as the common baseline; however, the results would have been completely different had we chosen paclitaxel as the common comparator. The exclusion of available data can lead to greater uncertainty in estimated treatment effects, which will in turn affect the decision uncertainty. If decision makers are also responsible for issuing recommendations about

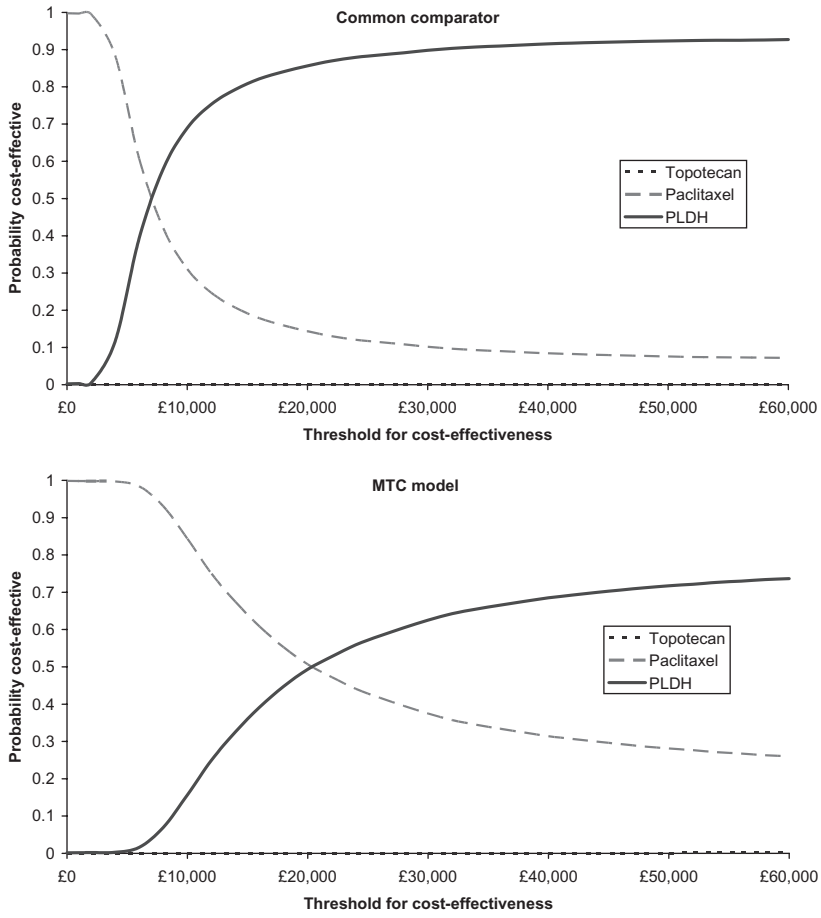

Figure 2 Cost-effectiveness acceptability curves generated from decision-analytic model using relative hazards calculated based on a common comparator (topotecan), and using a mixed treatment comparison (MTC) model. PLDH, pegylated liposomal doxorubicin hydrochloride.

further research, the decision to do so is likely to be a function of the decision uncertainty, which can be formally represented using value of information analysis [31–33]. If the value of collecting more information on any model parameter exceeds the costs of collecting that information, a gain can be made from instigating the necessary research. The failure to incorporate data that are already available may lead to erroneous recommendations. Because of the inconsistent nature of the available evidence in this example, the indirect hazard ratio estimated for paclitaxel compared with PLDH, when using topotecan as a common comparator, is actually outside of the 95% confidence interval reported from the direct head-to-head comparison in trial C. This demonstrates, to some extent, the strength of the assumption to exclude available data. In other words, this approach assumes that trial C contributes no information at all to the paclitaxel vs. PLDH comparison.

An MTC model provides an analytic framework to incorporate evidence in situations where there exists both direct head-to-head evidence and indirect evidence relative to a common comparator. Clearly when indirect evidence is used to estimate treatment effects it is not possible to rule out the introduction of bias, and the results should be interpreted accordingly, as they should for any meta-analysis. The validity of the result is dependent on the assumption that the relative treatment effects would be equal were they observed in any of the included trials. In this example, the three trials provide inconsistent evidence as to the relative effectiveness of the three comparators. Although this may be the product of random chance, particularly where trials are small, it is important to consider whether there is a systematic explanation for the inconsistency. In this example the length of follow-up differs between the trials. If the hazard ratio for overall survival varies with the length of follow-up, then it would be inappropriate to synthesize the hazard ratios from all three trials on the basis that they were directly exchangeable. This also applies to the approach where two trials were combined based on a common comparator. The Bayesian nature of the MTC model would allow the incorporation of prior information about the expected change in hazard ratio over time; however, none was available in this case. The MTC approach is, however, based on only a few additional assumptions over standard meta-analysis. It is a method to explicitly incorporate indirect evidence and quantify its uncertainty. The choice of which trials to synthesize must be

made in both formal approaches demonstrated here, with consideration of the trial characteristics and how each can contribute to the decision problem.

In this example, the use of an MTC model marginally reduced the uncertainty around the estimated treatment effects by incorporating all the available evidence. Nevertheless, it also reduced the amount by which the optimal treatment strategy was estimated to be superior to the next best alternative. This did not affect the adoption decision based on current evidence if we assume that society is willing to pay more than £17,000 per additional QALY, but it did increase the uncertainty in the adoption decision for the range of values of cost-effectiveness threshold commonly used in the UK. This ability of additional information to change the point estimates of relative treatment effects, at the same time as reducing the uncertainty around them, means that incorporating all available evidence could affect the uncertainty around the adoption decision in either direction. Thus, the effect of omitting some of the available evidence cannot be informally predicted.

If executed with the same analytical rigor as standard meta-analyses, MTC models provide a robust method for formally synthesizing both direct and indirect evidence to calculate the expected value and associated uncertainty around treatment effects and other parameters for input into decision-analytic models. This can only further improve on the assistance decision-analytic models provide to decision makers who wish to rank competing alternatives in terms of cost-effectiveness and assess the associated decision uncertainty.

Source of financial support: This project was funded by the Health Technology Assessment Programme (project number 04/03/01) and commissioned on behalf of NICE. The views expressed are those of the authors and do not necessarily reflect those of the Department of Health.

References

1 Hjelmgren J, Berggren F, Andersson F. Health economic guidelines—similarities, differences and some implications. Value Health 2001;4:225–50.
2 Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15:2733–49.
3 Ades AE, Cliffe S. Markov Chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making 2002;22:359–71.
4 Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). Lancet 2002; 360:711–15.
5 Forbes C, Shirran L, Bagnall A-M, et al. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer. Health Technol Assess 2001;5:1–110.
6 Forbes C, Wilby J, Richardson G, et al. A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer. Health Technol Assess 2002;6:1–119.
7 Capri S, Cattaneo G. Cost-minimization analysis of pegylated liposomal doxorubicin versus topotecan for the treatment of ovarian cancer in Italy. Clin Ther 2003;25:1826–45.
8 Smith DH, Adams JR, Johnston SRD, et al. A comparative economic analysis of pegylated liposomal doxorubicin versus topotecan in ovarian cancer in the USA and the UK. Ann Oncol 2002;13:1590–7.
9 Main C, Ginnelly L, Griffin S, et al. Topotecan, Pegylated Liposomal Doxorubicin Hydrochloride and Paclitaxel for Second-Line or Subsequent Treatment of Advanced Ovarian Cancer. Available from: http://www.nice.org.uk/pdf/Assessmentreport_Ovariancancer.pdf. [Accessed October 13, 2005].
10 Schering-Plough Ltd. Caelyx (Pegylated Liposomal Doxorubicin Hydrocholoride) in the Treatment of Recurrent Ovarian Cancer in the United Kingdom. A Submission to the National Institute for Clinical Excellence. Welwyn Garden City, Herts: Schering-Plough Ltd., 2004.
11 ten Bokkel Huinink W, Lane SR, Ross GA. Long-term survival in a phase III, randomised study of topotecan versus paclitaxel in advanced epithelial ovarian carcinoma. Ann Oncol 2004;15:100–3.
12 Johnson, Johnson Pharmaceutical Research, Development. L.L.C. Clinical Study Abbreviated Report. A phase III, randomized, open-label, comparative study of caelyx versus paclitaxel HCI in patients with epithelial ovarian carcinoma following failure of first-line, platinum-based chemotherapy, Protocol 30–57. Saunderton, Bucks: Johnson & Johnson Pharmaceutical Research and Development, 2004.
13 Gore M, Oza A, Rustin G, et al. A randomised trial of oral versus intravenous topotecan in patients with relapsed epithelial ovarian cancer. [See comment]. Eur J Cancer 2002;38:57–63.
14 Piccart MJ, Green JA, Lacave AJ, et al. Oxaliplatin or paclitaxel in patients with platinum-pretreated advanced ovarian cancer: a randomized phase II study of the European Organization for Research and Treatment of Cancer Gynecology Group. J Clin Oncol 2000;18:1193–202.
15 Omura GA, Brady MF, Look KY, et al. Phase III trial of paclitaxel at two dose levels, the higher dose accompanied by filgrastim at two dose levels in platinum-pretreated epithelial ovarian cancer: an intergroup study. J Clin Oncol 2003;21:2843–8.
16 Rosenberg P, Andersson H, Boman K, et al. Randomized trial of single agent paclitaxel given weekly versus every three weeks and with peroral versus intravenous steroid premedication to patients with ovarian cancer previously treated with platinum. Acta Oncol 2002;41:418–24.
17 The ICON and AGO Collaborators. Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with

relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. Lancet 2003;361:2099–106.
18 Cantu MG, Buda A, Parma G, et al. Randomized controlled trial of single-agent paclitaxel versus cyclophosphamide, doxorubicin, and cisplatin in patients with recurrent ovarian cancer who responded to first-line platinum-based regimens. J Clin Oncol 2002;20:1232–7.
19 Hougaard P. Fundamentals of survival data. Biometrics 1999;55:13–22.
20 Cox DR. Regression models and life tables (with discussion). J R Stat Soc [Ser B] 1972;34:187–220.
21 Palmer S, Sculpher M, Philips Z, et al. Management of non-ST-elevation acute coronary syndromes: how cost-effective are glycoprotein IIb/IIIA antagonists in the UK National Health Service? Int J Cardiol 2005; 100:229–40.
22 Bridle C, Palmer S, Bagnall AM, et al. A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder. Health Technol Assess 2004;8:1–200.
23 Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5, (Version ii). Cambridge: MRC Biostatistics Unit, 1996.
24 Meltzer D. Addressing uncertainty in medical cost–effectiveness analysis: implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost–effectiveness analysis to set priorities for medical research. J Health Econ 2001;20:109–29.
25 Fenwick E, Claxton K, Sculpher M. Representing uncertainty. the role of cost-effectiveness acceptability curves. Health Econ 2001;10:779–89.
26 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. BMJ 2001;323:1300–3.
27 NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Clinical Excellence, 2004.
28 Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds., The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994.
29 Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making 1998;18:37–43.
30 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719–48.
31 Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ 1996;5:513–24.
32 Claxton K, Neuman PJ, Araki SS, Weinstein MC. The value of information: an application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care 2001;17:38–55.
33 Claxton K, Eggington S, Ginnelly L, et al. A Pilot Study of Value of Information Analysis to Support Research Recommendations for the National Institute for Clinical Excellence. 2004. Available from: http://www.nice.org.uk. [Accessed October 5, 2005].
34 Gordon AN, Fleagle JT, Guthrie D, et al. Recurrent epithelial ovarian carcinoma: a randomized phase III study of pegylated liposomal doxorubicin versus topotecan. J Clin Oncol 2001;19:3312–12.

Volume 9 • Number 4 • 2006
VALUE IN HEALTH

Probabilistic Analysis and Computationally Expensive Models: Necessary and Required?

Susan Griffin, MSc, Karl Claxton, PhD, Neil Hawkins, PhD, Mark Sculpher, PhD

Center for Health Economics, University of York, York, UK

ABSTRACT

Objective: To assess the importance of considering decision uncertainty, the appropriateness of probabilistic sensitivity analysis (PSA), and the use of patient-level simulation (PLS) in appraisals for the National Institute for Health and Clinical Excellence (NICE).

Methods: Decision-makers require estimates of decision uncertainty alongside expected net benefits (NB) of interventions. This requirement may be difficult in computationally expensive models, for example, those employing PLS. NICE appraisals published up until January 2005 were reviewed to identify those where the assessment group utilized a PLS model structure to estimate NB. After identifying PLS models, all appraisals published in the same year were reviewed.

Results: Among models using PLS, one out of six conducted PSA, compared with 16 out of 24 cohort models. Justification for omitting PSA was absent in most cases. Reasons for choosing PLS included treatment switching, sampling patient characteristics and dependence on patient history. Alternative modeling approaches exist to handle these, including semi-Markov models and emulators that eliminate the need for two-level simulation. Stochastic treatment switching and sampling baseline characteristics do not inform adoption decisions. Modeling patient history does not necessitate PLS, and can depend on the software used. PLS addresses nonlinear relationships between patient variability and model outputs, but other options exist. Increased computing power, emulators or closed-form approximations can facilitate PSA in computationally expensive models.

Conclusions: In developing models analysts should consider the dual requirement of estimating expected NB and characterizing decision uncertainty. It is possible to develop models that meet these requirements within the constraints set by decision-makers.

Keywords: cost-effectiveness analysis, decision uncertainty, decision-analytic modeling, patient-level simulation, probabilistic analysis.

Introduction

Economic evaluation of health-care technologies is increasingly recommended internationally for informing the allocation of health-care resources [1]. For almost every set of technologies considered, it will be necessary to combine information on costs and effects from several sources, and modeling techniques will be employed. When building any decision model, it is important to consider how to handle uncertainty and variability, because these affect the value and interpretation of the model output. Decision-makers require unbiased estimates of the costs and effects of alternative interventions for identifiable patient groups, and the ability of a model structure to provide these can depend on how uncertainty, variability, and heterogeneity are handled in the structure. There is also a need to provide an assessment of whether current evidence is sufficient to support the decision to adopt a technology. By formally estimating decision uncertainty, the value of obtaining additional information on model parameters can be assessed [2–4].

In this article we explore methods available to address uncertainty, variability, and heterogeneity within decision models. We explore how the current methods for conducting probabilistic sensitivity analysis (PSA) to characterize decision uncertainty can conflict with the use of computationally expensive model structures. The premise is that the purpose of a decision model is to provide unbiased estimates of expected cost and effects, and of decision uncertainty, in a timely fashion and within resource constraints as determined by the decision-maker that commissions the model. In this article, the focus is on one particular decision-maker, the National Institute for Health and Clinical Excellence (NICE) in the UK. This focus has a number of advantages: 1) NICE is an agency with a history of using decision-analytic cost-effectiveness models as a basis for deciding whether to support the use of particular health-care technologies in the UK National Health Service (NHS) [5]; 2) by focusing on a single decision-maker, the case studies identified in the review will have been developed with consistent

Address correspondence to: Susan Griffin, Alcuin A Block, University of York, Heslington, York YO10 5DD, UK. E-mail: scg3@york.ac.uk

10.1111/j.1524-4733.2006.00107.x

  1098-3015/06/244 244–252

time and resource constraints; 3) NICE has specified a reference case that defines those methods considered most appropriate for informing decisions about the adoption of new technologies [6], which makes the case studies identified in the review comparable in this respect; and 4) the potential conflict between PSA and computationally expensive model structures is pertinent, because the reference case now calls for the use of PSA as the appropriate way by which the combined implications of uncertainty in all model parameters be reflected [7].

Although the review is specific to NICE, issues relating to the provision of unbiased estimates of costs and effects and quantifying decision uncertainty, while reflecting the complexity of the disease process and treatment effect under time and resource constraints, are more general and relevant to decision models developed in different contexts.

Review Methods

The review allows an exploration of the potential conflict between the use of computationally expensive model structures and the implementation of PSA in models submitted to NICE. For the purposes of this review, computational expense refers to the limited resources available with which to produce model results, given the constraints of the decision-making process. Models may be computationally expensive for a number of reasons. For our review, the use of patient-level simulation (PLS) was selected as an indicator of computational expense because its use should be readily apparent in models submitted to NICE and, for a given model structure, analysis using PLS is more computationally expensive than a cohort analysis.

All technology appraisals detailed on the NICE Web site that were published up until January 2005 were reviewed to identify those where the independent technology assessment group had utilized a model structure involving PLS to provide an estimate of cost-effectiveness. After identifying models that employed PLS, appraisals published in the same year were identified and reviewed. The benefit of comparing models that employed PLS with models published in the same period is the consistency of constraints facing model developers. Given the computing expense of PSA, it might be expected that it would be more common among models employing a cohort framework. We identified common stated reasons for choosing a PLS rather than a cohort framework, and assessed their implications for model structure with respect to uncertainty, variability, and heterogeneity. For each case study, we ascertained whether a PSA had been undertaken. In cases where a PSA had not been undertaken, we explored the availability of alternative modeling techniques, such as less computationally expensive modeling structures or emulators. In those cases where PSA was performed, we discuss the techniques used.

Uncertainty, Variability, and Heterogeneity

In this section, we distinguish uncertainty, variability, and heterogeneity, explore the implications of each for model structure and interpretation of results, and briefly review the methods available for characterizing decision uncertainty.

Decision Uncertainty

Decision uncertainty can be regarded as epistemic uncertainty [8] which relates to model parameters that have a definite value, but which cannot be known with certainty. For example, the risk of future individual developing cervical cancer has a definite value, but one which we can only estimate with uncertainty. This uncertainty can be characterized with a distribution and can be reduced with further investigation. Epistemic uncertainty is not confined to model parameters, and may exist in the determination of model structure. This discussion concentrates on parameter uncertainty, although it can be generalized to other sources of uncertainty [9].

The decision to adopt a particular technology should be based on expected net benefit (NB) so that, when comparing mutually exclusive treatment strategies for a particular disease area, the optimal strategy is simply the one with the highest expected NB [10]. Nevertheless, uncertainty is important for two reasons: 1) in nonlinear models, or multilinear models with correlated parameters, unbiased estimates of expected NB require a characterization of uncertainty; and 2) decisions based on expected NB are only appropriate if there is also some consideration of whether current evidence is sufficient for allocating health-care resources, based on an assessment of the consequences of decision uncertainty [11]. If the decision uncertainty, and/or the consequences of adopting a suboptimal treatment strategy are large, the decision-maker may require further evidence on which to base the adoption decision [4].

Variability

Expected costs and effects are not only uncertain but also vary across individuals with identical observed characteristics. This variability can be regarded as aleatory uncertainty [8] which arises as a result of stochastic variation. It cannot be reduced through measurement, but can be characterized with empirical distributions. For example, the rate at which an individual's cervical cancer develops will vary between patients. We can describe the distribution of the rate of cancer progression by counting the number of patients who progress at different rates. Nevertheless, further investigation would not reduce variation in the rate of

progression. Another example is where, given a probability of an event occurring, such as death, the realization of that event can be imagined as being governed by a lottery. So we may know with certainty that the probability of death is, for example, 5%, but we do not know which 5% of people will die.

Variability in itself is not relevant to an adoption decision based on expected NB. Nevertheless, it may be necessary to explicitly represent variability in model structures to obtain an unbiased estimate of expected NB if, within a patient population which is homogeneous in observed baseline characteristics, there is a nonlinear relationship between a characteristic that varies between patients and NB.

For example, suppose the outcome of interest, cost (C), is a nonlinear function of some patient characteristic (x) with mean, μ, that varies between patients (i) according to a normal distribution with variance σ:

$$C_i = kx_i^2 \quad (1)$$

$$x_i \sim N(\mu, \sigma) \quad (2)$$

The expected value of x across all patients, $E[x_i]$, cannot be used to derive the expected value of C; an estimate of $E[x_i^2]$ is required because $E[x_i]^2 \neq E[x_i^2]$. In this instance, an analysis which failed to account for the variability in x across patients would provide biased estimates of expected costs. The use of PLS accounts for variability in all included parameters, regardless of whether there exist any nonlinear relationships between these parameters and model output. Nevertheless, there are a number of other methods by which we could also address this issue.

For example, by repeatedly sampling from μ and σ we can estimate $E[x_i^2]$ as an input to the model. Alternatively it may be possible to derive a linear approximation to the model. That is, we find a mathematical function of μ and σ that gives us $E[x_i^2]$, that is, determine function G such that:

$$E[x_i^2] = G(\mu, \sigma^2) \quad (3)$$

The requirement to account for variability in model structure under these circumstances does not negate the need to estimate decision uncertainty. Failure to account for uncertainty will also lead to biased estimates of cost and effect in a nonlinear model. Consequently, under these circumstances, both variability and uncertainty must be characterized in order properly to inform an adoption decision. In model structures which are linear with respect to variability but nonlinear with respect to uncertainty, unbiased estimates of NB require the characterization of uncertainty but not variability. Where models are linear or multilinear (with independent parameters), it is not necessary to represent variability or uncertainty to obtain unbiased estimates of expected NB. Nevertheless, it will still be necessary to represent uncertainty in a model structure to address the question of whether current evidence provides sufficient basis for the adoption decision.

Heterogeneity

Heterogeneity can be regarded as variation as a result of observed characteristics on which it is possible to condition model parameters and therefore expected NB. Such heterogeneity must be observable at the time at which the treatment decision is taken. For example, the risk of developing cervical cancer may depend on family history. In principle, this can be observed and subsequent decisions, such as the decision of whether to screen, based on this observation. This contrasts with variation in the rate of disease progression which is unobservable at the time at which the decision to screen is made. Thus, one could not decide to screen only those patients whose cancer would develop at a fast rate. When estimating the cost-effectiveness of an intervention for a heterogeneous population, one can condition on the observed characteristics and separate the overall group into homogenous subpopulations within which patients have identical observed characteristics. The model can then be run separately for each homogenous group to generate estimates of cost-effectiveness conditional on each set of observed characteristics. Adoption decisions can then made for each of these mutually exclusive and identifiable patient groups [12].

Probabilistic Sensitivity Analysis

In this section, we consider the role of PSA in dealing with uncertainty in decision models. One- and multiway sensitivity analyses cannot reflect the combined uncertainty in all model parameters, and so are inappropriate for informing decision uncertainty. PSA, when conducted properly, provides a more rigorous approach by requiring that all input parameters in a model be specified as full probability distributions, rather than as point estimates, to indicate the uncertainty of the estimates [13]. PSA can be used accurately to estimate expected NB in a nonlinear model, and also to reveal the effect of the combined uncertainty in all model parameters.

Decisions based on expectation are only appropriate if the consequences of the uncertainty surrounding the decision are also considered. This informs the necessary question about whether current evidence is sufficient or whether further research is needed. Formal methods are available to estimate this value of information [2] and these are now recommended, although not required, by NICE. Nevertheless, for the purposes of this review it is sufficient to note that an appropriate characterization of decision uncertainty is a prerequisite for any assessment of the consequences of decision uncertainty, whether or not this achieved using formal

methods. Clearly, decisions based on point estimates without any consideration of uncertainty will lead to the adoption of technologies with inadequate and poor quality evidence [3,10]. Therefore, adoption decisions cannot be separated from an assessment of whether the evidence is sufficient to support such decisions. For example, as well as making the adoption decision, NICE also makes recommendations for future research, specifies a review date for guidance, in part based on when new evidence is expected to be available, and has issued guidance conditional on further evidence being collected to inform future decisions.

Implementing PSA

PSA is commonly conducted using Monte Carlo simulation. The model is run repeatedly, and each run uses new random draws from distributions describing the uncertainty surrounding the value of each of the model parameters. This propagates parameter uncertainty through the model, which is then reflected in the results, and can be used to describe the likelihood that a treatment decision is optimal. This is distinct from the use of Monte Carlo simulation in PLS where the model parameters are fixed and a random number is used to determine the path of each individual patient. This propagates variability, and sometimes heterogeneity, into the model results. Enough simulations must be run to ensure that this variation does not affect identification of the optimal treatment decision. Thus, executing PSA within PLS requires a two-level simulation where each set of probabilistic inputs is held constant and the required number of patients is simulated through the model [14]. This can make PSA an order of 1,000 or 10,000 more computationally expensive in a PLS structure as compared with a cohort structure, and it is sometimes for this reason that PSA is omitted from models employing PLS.

Alternative methods to conduct PSA exist in the form of analytical model solutions and emulators. A closed form solution of expected NB, and possibly the associated uncertainty, may be tractable. That is, the analyst could define a closed-form approximation or simplification that gives the expectation of nonlinear functions using the model parameters. One such example would be the use of a Taylor series expansion [15]. Emulators take the form of nonparametric statistical models of the outputs of a model, such that those outputs can be recalculated with minimal time and computational expense when varying the model inputs according to the associated uncertainty [14]. Nevertheless, there are some limitations associated with the use of emulators; for example, in the number of uncertain parameters that may be included. It is not yet known whether they provide a directly exchangeable alternative to Monte Carlo simulation. Therefore, the implementation of PSA through Monte Carlo simulation can be viewed as current practice.

Results of the Review

The search results are shown in Figure 1. Although not required by NICE guidance during this period, PSA was performed in some technology appraisals. One out of six (17%) models that used PLS, and 16 out of 24 (67%) cohort models conducted PSA. None of the eight cohort models that omitted PSA cited computational expense as the reason, whereas two of the PLS models did. Nevertheless, justification for omission of PSA was not present in the majority of cases.

Case Studies

The review identified six assessment reports submitted to NICE where the estimates of expected costs and outcomes were based on a model structure using PLS. Details of the included appraisals are given in Table 1.

We identified four reasons for choosing a PLS structure over a cohort framework. These were treatment switching, sampling from patient characteristics, dependence on patient history, including previous events and time-in-state, and uncertainty and variability. The following review section provides a more detailed description of these reasons in the context of their use in the case studies, and assesses their implications for model structure.

Treatment Switching

In some disease areas, patients will be treated with a sequence of interventions. These may involve different drugs or different dosages of the same drugs. The decision regarding whether to move a patient to the next treatment in a sequence may be based on patient characteristics or patient history and therefore subject to variability. Nevertheless, for a given set of eligibility criteria for treatment switching, there will not be uncertainty around whether a patient proceeds to the next treatment. In other words, patients may vary in their characteristics and history, and as a result there will be variation in the number of patients switching treatment, but for a given set of characteristics and history there will not be random variation in the number of patients switching.

Case study 1 assessed the cost-effectiveness of imatinib for gastrointestinal stromal tumors (GIST) [16]. Current guidelines at the time of the assessment recommended an initial dose of 400 mg daily, with the option of proceeding to a higher dose in the event of a poor response or disease progression, and withdrawal of treatment in the absence of benefit after 8 weeks. Nevertheless, because of a paucity of data, the best starting dose of imatinib and best treatment pattern were highly uncertain.

The model had four health states: progressive disease, treatment with 400 mg imatinib, treatment with 600 mg imatinib and death. Patients in the imatinib

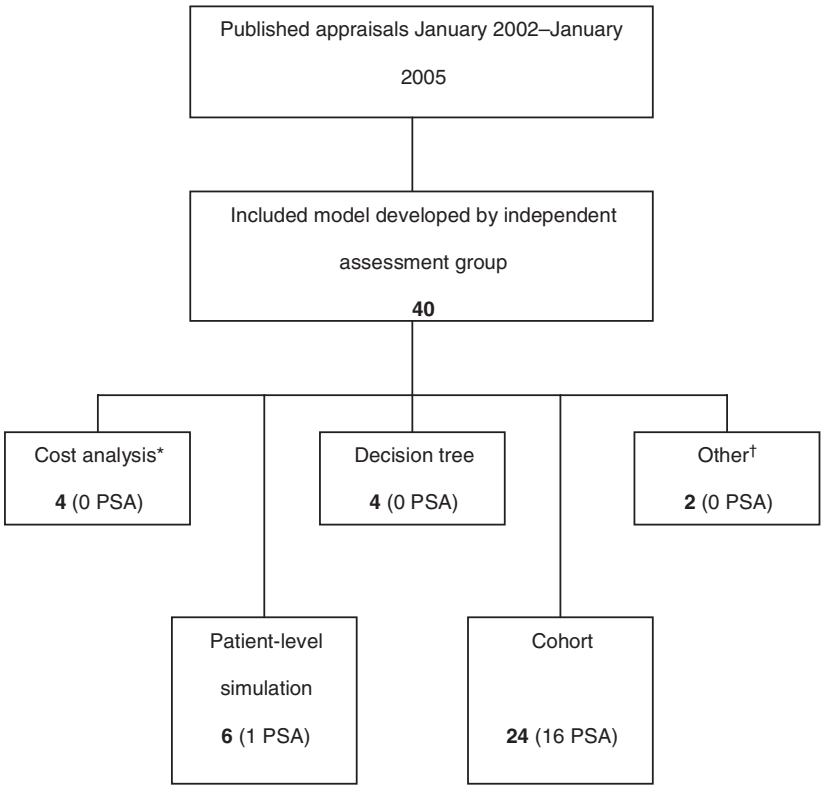

Figure 1 Search results. *Two cost-minimization analyses, two considered the cost results alongside trial results; †One meta-model of industry submission, one model type unclear. PSA, probabilistic sensitivity analysis.

treatment group began with 400 mg daily: patients whose disease progressed could move to 600 mg daily, or move to the progressive disease state. Patients who failed treatment with 600 mg imatinib daily moved to the progressive disease state, from which the only transitions were to remain in state or die. Patients could die at any stage in the model. Patients in the control group (i.e., no imatinib) began the model in the progressive disease state, and could remain in state or die. The cycle length was 4 weeks, and the time horizon was 10 years.

For those patients who failed to respond to 400 mg imatinib, a random number was generated to determine whether they would be moved to 600 mg, or straight to the progressive disease state. The probability of receiving 600 mg was based on the number of patients who had responded after crossing over from 400 mg to 600 mg imatinib in a clinical trial [17].

In reality, the decision to move from a dose of 400 mg to 600 mg would not be based on random chance. Although there may be uncertainty about the number of patients who would respond to 600 mg after progressing on 400 mg imatinib, this could not be identified before the treatment decision being taken. The treatment strategy in practice would involve moving all eligible patients onto 600 mg. When response is

Table 1 Examples of the use of patient-level simulation in NICE assessment reports

Case study	Title	Gave justification for choice of PLS?	Estimate decision uncertainty?
1	Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumors—a systematic review and economic evaluation [18]	Yes	No
2	The clinical effectiveness and cost effectiveness of prevention and treatment of osteoporosis [20]	Yes	Yes
3	The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation [22]	Yes	No
4	The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults [23]	Yes	No
5	The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy [24]	Yes	No
6	Coronary artery stents: rapid systematic review & economic evaluation [29]	No	No

NICE, National Institute for Clinical Excellence; PLS, patient level simulation.

assessed after 4 weeks, those patients not responding, as observed in the clinical trial, would then move to the progressive state. This could be compared with a strategy of not moving patients to a higher dose of the drug after failure on 400 mg. This alternative approach was taken by the developers of the GIST model in a subsequent evaluation. This approach enables decisions to be made about the best treatment pattern, rather than including current variability in treatment patterns in the model results. The situation where variability in patient characteristics may introduce uncertainty into the number of patients switching treatment is discussed in the later section on patient histories.

Sampling Baseline Patient Characteristics

Many characteristics of interest will vary between patients with the same disease, and some will be observable at the time at which the treatment decision is made. In other words, they are observable, able to be measured with precision, at baseline. If these characteristics determine the likelihood of future events, parameters in the model will depend on that observation. It is important to separate the heterogeneous population into homogenous subgroups, and to model these groups separately; otherwise the results of the model relate to the average patient, that is, the mean of the distributions describing the variation of the characteristics of interest. In practice, the treatment decision can and should be made conditional on the observed characteristic of each patient, not the expected characteristic.

Case study 2 considers treatments for osteoporosis for primary and secondary prevention of fractures [18]. The model is an update of a previously constructed model by the same assessment group [19]. This, in effect, relaxes the time constraint normally imposed on the development of such models for NICE. The structure of the model is described as being similar to a Markov model with a cycle length of 1 year, the difference being that patients are entered into the model individually, and their history is tracked. The states in the secondary prevention model are hip fracture, wrist fracture, vertebral fracture, proximal humerus fracture, breast cancer, coronary heart disease, and death. The model also tracks the residential status of each patient to assign costs. The authors state the belief that reflecting the increased risk of recurrent fractures after an initial fracture, and tracking the residential status of patients in the model, would be difficult in a cohort model.

The secondary prevention model considers the cost-effectiveness of treatment strategies for women presenting at baseline with hip, vertebral, wrist or proximal humerus fractures. Thus, the PLS is employed, in part, to track the presenting fracture site for each woman, because the model assumes a different baseline risk of subsequent fractures for each initial fracture site. Because of the difference in baseline risk of future events for each initial fracture site, and the fact that this characteristic is known when the treatment decision is taken, these could have been assessed as separate subgroups; it is important to be able to make separate treatment recommendations for each group. This would considerably reduce the number of states required to represent that portion of the model in a cohort framework. This alternative represents a different characterization of the decision problem which justifies a simplification of the model structure. This conditioning on baseline characteristics is distinct from conditioning on events that occur throughout the model process, for example, the location of new fracture sites after treatment has been initiated. In this example, the PLS model structure was also used to record patient history in the model, and this issue is the focus of the next section on patient histories.

Dependence on Patient Histories

Observable variation within groups homogenous in baseline characteristics may arise as a result of subsequent events that occur within the model structure. To condition model parameters on this observed variation, it is necessary to record these events in some way. The method with which to do this will depend on the choice of model structure.

Case studies 3 and 4 examined treatments for rheumatoid arthritis [20,21]. There is a low likelihood of long-term use for any one disease-modifying antirheumatic drug (DMARD), because they are not always effective, lose effectiveness over time, or cause adverse effects. Case study 5 assessed the cost-effectiveness of "newer" antiepileptic drugs (AEDs) in children [22]. Lack of effect on seizure rate and intolerable side effects means that many patients with epilepsy are treated with a sequence of drugs. The models compared fixed treatment sequences. The discontinuation rate of each treatment was modeled as a Weibull distribution with a shape parameter not equal to one (i.e., the hazard rate was not constant). In other words, the probability of discontinuing each treatment was dependent on the time spent on that treatment. Also, the availability of future treatment options was affected by toxic reactions to previous drugs. The PLS structure allows the analyst to record the realized time spent receiving each drug for each individual patient.

A separate assessment examined the use of newer AEDs in adults [23]. The decision problem and the events to be reflected in the model were very similar to the model of AEDs in children. As with case studies 3, 4, and 5, time to treatment discontinuation was a function of time spent on the drug, but this was facilitated in a cohort model by employing a semi-Markov framework. This semi-Markov model was built in the statistical programming language R [24], which can manipulate n-dimensional arrays and track the time

spent in each state. This alternative model structure enabled PSA to be undertaken to provide an estimate of decision uncertainty, without sacrificing the time-dependent structure of the model.

Probabilistic Sensitivity Analysis with PLS

As can be seen in Table 1, case study 2 formally assessed decision uncertainty [18]. This was made feasible through the use of an emulator employing Gaussian processes [14]. The full model was run 80 times using different values for the inputs to estimate a non-parametric relationship between the input parameters and the outputs of the model. This "model of a model" could then be analyzed relatively quickly to produce an estimate of decision uncertainty. The use of 80 runs of the model, as compared with perhaps 10,000 runs for PSA, considerably reduces the computing power and time required to estimate decision uncertainty within a PLS model structure.

Discussion

If it is accepted that adoption decisions should be made with consideration of the associated decision uncertainty, then we may say that models submitted to decision-makers have a dual requirement to estimate expected NB and characterize decision uncertainty. The use of PLS is often justified with reference to the first of these. In other words, the claim is made that it would not be possible to estimate NB accurately using a cohort framework for that particular decision problem. In our review, we have identified four reasons for choosing PLS, and showed that none of these necessarily preclude the use of a cohort framework. We have also identified an example where a formal estimate of decision uncertainty was obtained alongside a computationally expensive PLS model structure, showing that the two are not mutually exclusive.

This review indicates that the most common justifications for choosing a PLS are the need to incorporate time and history dependence in transition probabilities. In a Markov model, these would be handled using tunnel states, and if the number of states required is very large, the Markov framework may become unwieldy and inefficient. An alternative was illustrated whereby time- and state-dependent transitions were represented in a cohort framework using semi-Markov processes [25,26]. To employ this method to track elements of patient history within a cohort framework, it may be necessary to build the model in appropriate software. The use of alternative software can provide flexibility to design an alternative model structure, and may also reduce analysis time. This is no panacea, because there will be limits to the gain in analysis time available with alternative software, and improved hardware. For example, in the updated rheumatoid arthritis case study [21], the use of Borland Delphi [27] instead of TreeAge DATA 3.5 [28] sped up the analysis. Nevertheless, in this particular case the gain in analysis time was not enough to facilitate PSA. Dissemination and training are also required to allow further use of alternative software, and this may represent an additional constraint.

Time- and state-dependent transitions are easily handled within PLS. If the use of Monte Carlo simulation to conduct PSA is too computationally expensive, the requirement to characterize decision uncertainty could still be met, as evidenced by the use of an emulator in one case study [18]. Nevertheless, the use of emulators is still in development, and there are currently some limitations, for example, in the number of uncertain input parameters that can be included. More research is necessary to validate such models before it is known whether they are exchangeable with conducting PSA by means of Monte Carlo simulation.

We have shown that where there is a nonlinear relationship between a characteristic that varies between patients and the model output, it is necessary to account for this variability. This is distinct from the issue of baseline patient characteristics that confer a different baseline risk of subsequent events (heterogeneity), but refers to variability within homogenous patient subgroups. Importantly, this does not counter the need to address decision uncertainty because any nonlinear model which requires an assessment of variability will also need to assess uncertainty to estimate expected NB. A PLS can address the issue of variability, but this can also be addressed within a cohort framework by employing a two-level simulation. A third way to address this issue would be to find a closed-form approximation to the model which simplifies the analysis greatly. The alternatives identified here should allow the same level of detail in characterizing the decision problem as is possible with a PLS framework, but with lower computational expense. Nevertheless, they do require specialized knowledge to be able to correctly model that decision problem.

So how generalizable are the results of the review detailed here? An increasing number of models in the literature have been developed for—or to influence—a specific decision-making body at a particular point in time. Such models are typically developed with similar constraints as the models for NICE described here. Some models are developed over a longer-time period, and perhaps with more generous funding. Usually such analyses are not concerned with the decision problem of a particular organization, but have a wider set of aims and objectives. Nevertheless, even models developed in this way still face binding resource and time constraints, not least from the research funder. Furthermore, the purpose of all decision models is ultimately to inform real decisions at some point in time. Therefore, all analysts will simultaneously have to tackle the tasks of quantifying decision uncertainty

while reflecting the complexity of the disease process and effect of the intervention under time and resource constraints. Hence, the general issues highlighted by the NICE case studies are relevant to decision models developed in different contexts.

In conclusion, if the dual requirement of models to estimate NB and to characterize decision uncertainty is accepted, then the failure to fulfill the latter requirement will limit its value for decision-making. The claim that a model has been structured to provide a more appropriate estimate of expected NB but is too complex for PSA given the analysts' constraints would leave the decision-maker without a key element of information. There are often alternative modeling approaches to handle particular characteristics of a decision problem that can be used to reduce the complexity of the model, or to facilitate the conduct of PSA within a complex model structure, and both of these do not reduce the ability of the analyst to estimate NB correctly. This makes it possible to produce probabilistic models to estimate expected NB and characterize decision uncertainty within the constraints of the decision-making process.

Source of financial support: The work was funded by the UK Medical Research Council as part of a Health Services Research Collaboration program entitled "Methodological issues relating to decision analysis for resource allocation in health care." Mark Sculpher also receives funding via a Career Award in Public Health funded by the UK NHS Research and Development Programme. An earlier version of this article was presented at the UK Health Economists' Study Group in Oxford in January 2005. The authors are grateful for colleagues' comments and suggestions. All views and any errors are the sole responsibility of the authors.

References

1 Hjelmgren J, Berggren F, Andersson F. Health economic guidelines—similarities, differences and some implications. Value Health 2001;4:225–50.
2 Claxton K, Neuman PJ, Araki SS, Weinstein MC. The value of information: an application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care 2001;17:38–55.
3 Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ 1996;5:513–24.
4 Claxton K, Eggington S, Ginnelly L, et al. A Pilot Study of Value of Information Analysis to Support Research Recommendations for the National Institute for Clinical Excellence. 2004. Available from: http://www.nice.org.uk [Accessed September 1, 2005].
5 NICE. Guide to the Technology Appraisal Process. London: National Institute for Clinical Excellence, 2004.
6 NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Clinical Excellence, 2004.
7 Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005; 14:339–47.
8 Bedford T, Cooke R. Probabilistic Risk Analysis: Foundations and Methods. Cambridge: Cambridge University Press, 2001.
9 Ginnelly L. Characterising Structural Uncertainty in Decision Analytic Models: Review and Application of Currently Available Methods. Oxford: Health Economists' Study Group, January 2005.
10 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999;18:342–64.
11 Sculpher M, Claxton K. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty—when is there sufficient evidence? Value Health 2005;8:433–46.
12 Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 2004;82:661–87.
13 Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Med Decis Making 2002; 22:290–308.
14 Stevenson M, Oakley J, Chilcott J. Gaussian process modeling in conjunction with individual patient simulation modeling. a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. Med Decis Making 2004; 24:89–100.
15 Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making 2004;24:207–27.
16 Wilson J, Connock M, Song F, et al. Imatinib for the treatment of patients with unresectable and/or metastatic gastro-intestinal stromal tumours—a systematic review and economic evaluation. 2003. Available from: http://www.nice.org.uk [Accessed September 1, 2005].
17 Demetri GD, von Mehren M, Blanke CD, et al. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 2002;347:472–80.
18 Stevenson M, Jones M, de Nigris E, et al. The clinical effectiveness and cost effectiveness of prevention and treatment of osteoporosis. 2004. Available from: http://www.nice.org.uk [Accessed September 1, 2005].
19 Kanis J, Brazier J, Stevenson M, et al. Treatment of established osteoporosis: a systematic review and cost-utility analysis. Health Technol Assess 2002; 6(29):1–146.
20 Jobanputra P, Barton P, Bryan S, Burls A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. Health Technol Assess 2002;6(21):1–110.
21 Clark W, Jobanputra P, Barton P, Burls A. The clinical and cost-effectiveness of anakinra for the treat-

ment of rheumatoid arthritis in adults. 2004. Available from: http://www.nice.org.uk [Accessed September 1, 2005].
22 Bryan S, Connock M, Cummins C, et al. The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. 2004. Available from: http://www.nice.org.uk [Accessed September 1, 2005].
23 Wilby J, Kainth A, Hawkins N, et al. Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation. Health Technol Assess 2005;9.
24 Development Core Team. A Language and Environment for Statistical Computing. Vienna: R Development Core Team, 2004.
25 Hawkins N, Sculpher M, Epstein D. Cost-effectiveness analysis of treatments for chronic disease: using R to incorporate time dependency of treatment response. Med Decis Making 2005;25:511–20.
26 Hawkins N, Epstein D, Drummond M, et al. Assessing the cost-effectiveness of new pharmaceuticals in epilepsy in adults: the results a probabilistic decision model. Med Decis Making 2005;25:493–511.
27 Borland Software Corporation. Borland Delphi. Scotts Valley, CA: Borland Software Corporation, 2004.
28 Treeage Software, Inc. TreeAge DATA 3.5. Williamstown, MN: Treeage Software, Inc., 2004.
29 Hill R, Bagust A, Bakhai A et al. Coronary artery stents: a rapid systematic review and economic evaluation. Health Technol Assess 2004;8:iii–iv; 1–242.

Med Decis Making OnlineFirst, published on December 29, 2009 as doi:10.1177/0272989X09344746

Exploring the Research Decision Space: The Expected Value of Information for Sequential Research Designs

Susan Griffin, MSc, Nicky J. Welton, PhD, Karl Claxton, PhD

Purpose. *To investigate the expected value of partial perfect information (EVPPI) and the research decisions it can address.* ***Methods.*** *Expected value of information (EVI) analysis assesses the expected gain in net benefit from further research. Where the expected value of perfect information (EVPI) exceeds the costs of additional research, EVPPI can be used to identify parameters that contribute most to the EVPI and parameters with no EVPPI that may be disregarded as targets for further research. Recently, it was noted that parameters with low EVPPI for a one-off research design may be associated with high EVPPI when considered as part of a sequential design. This article examines the characteristics and role of conditional and sequential EVPPI in EVI analysis.* ***Results.*** *The calculation of EVPPI is demonstrated for single parameters, groups of parameters, and conditional and sequential EVPPI. Conditional EVPPI is the value of perfect information about one parameter, conditional on having obtained perfect information about another. Sequential EVPPI is the value of perfect information for a sequential research design to investigate first one parameter, then another. Conditional EVPPI differs from the individual EVPPI for a single parameter. Sequential EVPPI includes elements from the joint EVPPI for the parameters and the EVPPI for the first parameter in sequence. Sequential designs allow abandonment of research on the second parameter on the basis of additional information obtained on the first.* ***Conclusions.*** *The research decision space addressed by EVI analyses can be widened by incorporating sequential EVPPI to assess sequential research designs.* ***Key words:*** *value of information; uncertainty; decision making.* ***(Med Decis Making XXXX;XX:xx–xx)***

In cost-effectiveness analysis, the expected net benefits (NBs)[1] of a set of mutually exclusive health technologies can be compared, based on some threshold value for the cost per unit of health outcome, in order to identify the intervention that would represent the best use of available resources. The NBs include the health care resources utilized by patients receiving the health technologies and the health outcomes experienced by those patients. The NB of each alternative will be estimated with uncertainty. It has been argued that, in the absence of sunk costs[2] or adverse impact on the prospects of further research,[3] the decision of which technology to adopt can be based on expected NB and that information about decision uncertainty should inform questions about whether further research is required.[4] To this end, expected value of information (EVI) analysis can be used to estimate whether there is an expected gain in NB to be made by obtaining further information to inform the adoption decision.[5–8] Throughout this article, the gain in NB through further research refers to the improvement in NBs associated with the choice of health technology. This article will examine in more detail the way in which EVI analysis can be used to evaluate a range of research decisions.

The expected value of perfect information (EVPI) describes the value of further research that would eliminate all of the parameter uncertainty in the decision problem and provides an upper bound on the

Received 19 December 2007 from the Centre for Health Economics, University of York, York, United Kingdom (SG, KC); and the Department of Community Based Medicine, University of Bristol, Bristol, United Kingdom (NW). This work has previously been presented at the Health Economists' Study Group meeting in Brunel in September 2007 and at the Society for Medical Decision Making annual meeting in Pittsburgh in October 2007. Financial support for this study was provided by a grant from the Medical Research Council. The funding agreement ensured the authors' independence in designing the study, writing, and publishing the report. Revision accepted for publication 18 June 2009.

Address correspondence and reprint requests to Susan Griffin, MSc, University of York, Alcuin A. Block, York, United Kingdom YO10 5DD; e-mail: scg3@york.ac.uk.

DOI: 10.1177/0272989X09344746

value of additional information. The expected value of partial perfect information (EVPPI) describes the value of obtaining additional information on a subset of parameters, that is, eliminating uncertainty in only some aspects of the decision problem. It has been suggested that EVPPI can be used to identify those parameters that contribute most to the EVPI for the whole decision problem and that might be worthy of further research. Additionally it has been suggested that those parameters or groups of parameters associated with low or no EVPPI may be disregarded as potential targets for further research.[7,9–12] It has been emphasized that the sum of the EVPPI for each parameter in the model need not sum to the EVPI for the decision problem as a whole and that combinations of parameters may be associated with positive EVPPI were they to be investigated jointly even if individually they all exhibit zero EVPPI.[13,14]

Most existing EVI analyses[15–17] in the field of health care have focused on questions about one-off research designs that can be used to address the following research questions:

1) Do not conduct further research?
2) Conduct research to inform a single parameter in isolation?
3) Conduct research to inform a group of parameters simultaneously?
4) Conduct research to inform all the parameters simultaneously, that is, the whole decision problem?

We take each in turn to illustrate how EVI analysis can inform these research decisions. The research decision space is then expanded to include sequential designs and to show how EVI analysis can be used to inform a more complete set of research decisions commonly faced.

ONE-OFF RESEARCH DESIGNS

Is There Value in Any Further Research?

The value of a decision based on current information and the value of acquiring additional information are based on the decision maker's objective function. Given an objective to maximize health gains subject to an exogenous budget constraint, the value of information can be expressed in terms of net health benefits. These are calculated by converting costs onto the same scale as health using the incremental cost-effectiveness ratio (the ratio of the additional costs to additional health outcomes) of the health care program(s) that would be displaced by adopting a new intervention. With current information, a decision maker should select the intervention (j) that maximizes expected net benefit. The net benefit of each intervention can be calculated as a known function of a set of parameters. Given a decision problem with 2 uncertain parameters, θ_1 and θ_2, the value of a decision based on current information can be expressed as the following:

$$\max_j E_{\theta_1,\theta_2} NB(j,\theta_1,\theta_2). \quad (1)$$

With perfect information, the decision maker could select the intervention that maximizes the net benefit for particular values of θ_1 and θ_2: $\max_j NB(j,\theta_1,\theta_2)$. However, the true values of $\theta_1$and θ_2 are unknown, so the expected value of a decision taken with perfect information is found by averaging the maximum net benefit over the joint distribution of θ_1 and θ_2:

$$E_{\theta_1,\theta_2} \max_j NB(j,\theta_1,\theta_2). \quad (2)$$

The value of a decision made with perfect information (equation 2) can then be compared to the value of the same decision based on current information (equation 1) in order to determine the expected value of perfect information (EVPI):

$$EVPI = E_{\theta_1,\theta_2} \max_j NB(j,\theta_1,\theta_2) - \max_j E_{\theta_1,\theta_2} NB(j,\theta_1,\theta_2). \quad (3)$$

Additional information only has value when it would lead to the adoption of a different intervention than the one that would be selected on the basis of current information. The information gained from further research is nonrival and can be used to inform decisions for current and future patients in the same population. Therefore, the population EVPI can be expressed as the per patient EVPI multiplied by the future population of patients expected to benefit from the information. The future population of patients expected to benefit will be a function of the duration of research and the expected time horizon for the current decision problem.[18] Assuming an incidence of I in each period t, an expected time horizon of T, and discount rate r, the discounted population expected to benefit from research that reports at time τ can be expressed as the following:

$$P = \sum_{t=\tau}^{T} \frac{I_t}{(1+r)^t}. \quad (4)$$

The expected duration of research is likely to differ depending on the parameter being investigated

and the research design (e.g., whether it is a randomized trial or an observational cohort study). So we can express the population to benefit from research on parameter θ_1, that is, P_{θ_1}, in terms of the time required for research on that parameter to report, τ_{θ_1}.

The EVPI represents the upper bound for the value that could be gained from further research. Where the population EVPI is positive, it must be compared to the potential costs of further research (C) in order to establish whether there may be a positive potential payoff (Π). The costs of research include not only investigative and organizational costs but also any opportunity cost to patients included in the research. For example, if the additional research were to include randomization between alternative treatments, then a proportion of patients would not receive the intervention with the maximum expected net benefit on the basis of current evidence. Also those included in the research will not be able to benefit from the results; that is, the population that can benefit is used up in the research itself. For there to be any potential value to further research, the maximum achievable value must at least exceed the minimum expected cost of research:

$$\Pi_{\theta_1,\theta_2} = EVPI.P_{\theta_1,\theta_2} - \min(C_{\theta 1}, C_{\theta 2}). \quad (5)$$

Equation 5 represents a necessary condition. Where this first necessary condition is met, we may then go on to calculate the value of obtaining further information for a range of more specific research questions. If the expected payoff is negative, then investing in further research would not represent a good use of available resources, as any value gained from the additional information would be outweighed by the cost of obtaining that information.

Is There Value in Further Research to Inform a Single Parameter in Isolation?

A study could be designed to gather information about one specific parameter. With perfect information about θ_1, the decision maker could select treatment j that maximizes the expected net benefit over the remaining uncertain parameter θ_2. As the true values of θ_1 are unknown, this must then be averaged over the distribution of θ_1:

$$E_{\theta_1} \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2). \quad (6)$$

By comparing the value of a decision made with perfect information about θ_1 to the value of a decision made with current information (equation 1), we can quantify the expected value of further research to inform a single parameter in isolation:

$$EVPPI_{\theta_1} = E_{\theta_1} \max_j E_{\theta 2|\theta 1} NB(j,\theta_1,\theta_2) - \max_j E_{\theta_1,\theta_2} NB(j,\theta_1,\theta_2). \quad (7)$$

The EVPPI about each parameter must be multiplied by the population of patients who stand to benefit from the additional information,P_{θ_1}, and compared to the potential costs of research to determine whether there may be a positive potential payoff from that research:

$$\Pi_{\theta_1} = EVPPI_{\theta_1}.P_{\theta_1} - C_{\theta_1}. \quad (8)$$

Is There Value in Further Research to Inform Groups of Parameters Simultaneously?

Further research could inform more than one parameter so that the results could reduce the uncertainty in a number of inputs simultaneously. Such research could take the form of a single study gathering information on a set of parameters or a set of concurrent studies each investigating a different parameter. In the case of a 2-parameter model, obtaining perfect information on both θ_1 and θ_2 would eliminate all of the decision uncertainty, and so in this 2-parameter case, the joint EVPPI for θ_1 and θ_2 is equivalent to the EVPI ($EVPPI_{\theta_1,\theta_2} = EVPI$). In order to establish whether there may be a positive potential payoff from research to update both parameters simultaneously, the joint EVPPI should be multiplied by the population expected to benefit and compared to the costs of research:

$$\Pi_{\theta_1,\theta_2} = EVPPI_{\theta_1,\theta_2}.P_{\theta_1,\theta_2} - C_{\theta_1,\theta_2}. \quad (9)$$

This also answers question 4: ''Is there value in further research to inform the whole decision problem?'' But more generally, that question can be informed by evaluating the following:

$$\Pi_{\theta 1,\ldots,\theta n} = EVPPI_{\theta 1,\ldots,\theta n}.P_{\theta 1,\ldots,\theta n} - C_{\theta 1,\ldots,\theta n} \quad \text{for the total number } (N = 1,\ldots,n) \text{ of uncertain parameters.} \quad (10)$$

The value of information calculations considered above concerns the value of perfect information so they can only provide a necessary condition for deciding to conduct further research. In practice, the type of research that can be conducted will provide

only imperfect sample information. The EVPI calculations described above can be extended to estimate the expected value of sample information (EVSI) based on the possible sample results and posteriors predicted from sampling the prior distributions of the parameters and their likelihood.[19] The EVPI provides an upper bound to the EVSI, but where the EVSI exceeds the expected costs of research, a sufficient condition for conducting further research is met. All of the equations so far described can be extended to calculate EVSI instead of EVPI. However, their implementation poses additional computation burden, as predicted posteriors for alternative sample sizes need to be established based on a range of possible sample results.[8,19,20]

Summary

EVI analyses are based on an assessment of the value of a decision based on current evidence (equation 1) and the value of that same decision based on additional evidence about all of the input parameters (equation 2), individual input parameters (equation 6), or groups of parameters (in this example, equivalent to equation 2). These elements of EVI analysis can be combined to answer a variety of decisions regarding one-off research designs, that is, whether there would be value in conducting any single additional study or set of concurrent additional studies. If the potential payoff from an additional study is expected to be positive, then investing in further research may represent a valuable use of current resources. The single-study or simultaneous study design that provides the largest potential payoff may be regarded as the priority for research funding. However, these types of one-off research designs are not the only research decisions that are available to policy makers.

SEQUENTIAL RESEARCH DESIGNS

Decisions about one-off research designs cover only a portion of the total research decision space. The research decision space can be expanded to consider the value of a series of studies, each investigating a different parameter or group of parameters in turn. This poses a new question not previously addressed fully in the literature:

5) Is there value in conducting a sequence of research to inform a set of parameters?

Brand and Small and Chao and others[21,22] have discussed the value of additional information in terms of updating decisions in a sequential or iterative expected value decision-making framework but did not address directly a means to establish the value of sequential data collection. Recently, it has been noted that additional information about one parameter will modify the EVPPI associated with remaining parameters.[23] This would indicate that the value of a sequential research design to investigate first θ_1 then θ_2 cannot be evaluated simply on the basis of $EVPPI_{\theta 1}$ and $EVPPI_{\theta 2}$. What is required is an estimate of the value of obtaining additional information on θ_1, followed by an estimate of the value of obtaining additional information on θ_2, conditional on having already obtained additional information on θ_1.

Conditional EVPPI

If research is conducted to obtain additional information on θ_2, this in combination with the information already collected for θ_1 provides the decision maker with perfect information on both parameters. Thus, the expected net benefits of a decision taken with perfect information on θ_2 and θ_1 are simply equation 2, regardless of whether this information was obtained sequentially or simultaneously. Again, we must compare this to the expected net benefits of the decision made with current information. But now, current information includes the information already collected on θ_1. Therefore, the decision based on current information is based on perfect knowledge of θ_1, and its value is given by equation 5. Thus, the conditional EVPPI for θ_2 given perfect information about θ_1 can be expressed as the following:

$$EVPPI_{\theta_2|\theta_1} = E_{\theta_1,\theta_2} \max_j NB(j,\theta_1,\theta_2) - E_{\theta_1} \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2). \quad (11)$$

This value can be multiplied by the population expected to benefit from the additional information and compared to the cost of research to determine the payoff from additional research on a second parameter conditional on having already conducted further research on the first:

$$\Pi_{\theta_2|\theta_1} = EVPPI_{\theta_2|\theta_1}.P_{\theta_2} - C_{\theta_2}. \quad (12)$$

This, however, would not answer the question regarding the value of a sequential research design

because it does not incorporate the value or cost of the information obtained on the first parameter in sequence.

Sequential EVPPI

The total value of a sequential design can be obtained by combining the value of the information gained on the first parameter in the sequence with the conditional EVPPI for the second parameter in sequence. The value of a sequence evaluating first θ_1 then θ_2 would be the sum of $EVPPI_{\theta_1}$ and $EVPPI_{\theta_2|\theta_1}$ only if the decision to conduct research about θ_2 did not depend on the resolution of θ_1. However, such an a priori research decision would negate all the potential advantages of a sequential research design. In evaluating sequential research, it is important to allow for a change in research decision conditional on the new information as it is collected. So for particular resolutions of θ_1, there may be no expected value from obtaining additional information on θ_2 ($EVPPI_{\theta_2|\theta_1} \leq C_{\theta_2}$), and further research should not be conducted. In these circumstances, the sequential research design allows the decision maker to avoid additional research on θ_2 when it is not sufficiently valuable, that is, issue a 'stop' decision. For particular resolutions of θ_1 where $EVPPI_{\theta_2|\theta_1} > C_{\theta_2}$, the decision may be to 'go' and conduct research on to θ_2. The choice set for the decision maker now includes the intervention (j) and the decision about whether to proceed with research on to the next parameter in sequence (r). Maximizing the value of the decision with perfect information within the brackets in equation 13 determines whether a 'stop' or 'go' decision is made to research θ_2 for each particular value of θ_1:

$$\begin{aligned} seqEVPPI_{\theta_1,\theta_2} = E_{\theta_1} \max\Big\{ & E_{\theta_2|\theta_1} \max_j NB(j,\theta_1,\theta_2), \\ & \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2) \Big\} \\ & - \max_j E_{\theta_1,\theta_2} NB(j,\theta_1,\theta_2). \end{aligned} \tag{13}$$

In a sequential design, the cost of research into the first parameter in the sequence is always incurred. However, research on the second parameter in the sequence will only proceed and costs be incurred if the expected payoff is positive. Therefore, the expected maximum payoff from a sequential design that investigates θ_1 first then θ_2 ($\Pi_{seq\theta_1,\theta_2}$) can be expressed as follows:

$$\begin{aligned} \Pi_{seq\theta_1,\theta_2} = & E_{\theta_1} \max\Big\{ \Big(E_{\theta_2|\theta_1} \max_j NB(j,\theta_1,\theta_2) - \\ & \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2) \Big).P_{\theta_2+\theta_1} - C_{\theta_2}, 0 \Big\} \\ & + \Big(E_{\theta_1} \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2) \\ & - \max_j E_{\theta_1,\theta_2} NB(j,\theta_1,\theta_2) \Big).P_{\theta_1} - C_{\theta_1} \\ = & E_{\theta_1} \max\Big\{ \Big(E_{\theta_2|\theta_1} \max_j NB(j,\theta_1,\theta_2) - \\ & \max_j E_{\theta_2|\theta_1} NB(j,\theta_1,\theta_2) \Big).P_{\theta_2+\theta_1} - C_{\theta_2}, 0 \Big\} \\ & + \Pi_{\theta 1}. \end{aligned} \tag{14}$$

Equation 14 provides an answer to question 5: ''Is there value in further research to inform a set of parameters sequentially?'' It describes the expected maximum payoff from a sequential design before any of the parameters have been investigated further. Sequential EVPPI is composed of the same elements (equations 1, 2, and 6) that are used to answer questions 1 to 4, alongside some assessment of the costs and duration of research. Therefore, the simulations required to estimate equations 1, 2, and 6 can also be used to estimate each of the elements of equations 13 and 14, supposing that the Monte Carlo simulation is employed. Alternative means to estimate value of information could well reduce the sampling time required.[24]

COMPARISONS BETWEEN ALTERNATIVE RESEARCH DESIGNS

For the 2-parameter decision problem used to illustrate the EVI calculations, there are only 4 possible mutually exclusive research decisions (no research, θ_1 alone, θ_2 alone, or θ_1 and θ_2) when considering one-off designs. The research decision space widens to 6 alternatives when considering the possible sequential designs (θ_1 then θ_2, or θ_2 then θ_1). It would be feasible to enumerate the potential payoffs for all of these possible one-off and sequential research designs. The decision maker could then select a design that would maximize the potential payoff from further research, including consideration of the option for no further research. However, as the number of parameters (n) increases, the number of potential research designs and the research decision space increases dramatically. For a given number of parameters, n, the number of one-off research designs is equal to 2^n; the number of

sequential research designs of any length greater than 2 and including studies that investigate only a single parameter is given by the number of permutations $\sum_{p=2}^{n} \frac{n!}{(n-p)!}$. However, a sequence could include single studies investigating groups (combinations) of parameters. The number of unique combinations that can be formed from a set of parameters n is given by $\sum_{d=2}^{n-1} \frac{n!}{d!(n-d)!}$, where d is the number of parameters investigated jointly, and for each unique combination, a new set of sequences is possible of number $\sum_{p=2}^{n-d} \frac{(n-d+1)!}{(n-d-p+1)!}$. It would not be correct to determine the number of theoretically possible sequences based on the number of parameters in a decision model, as this is arguably not relevant to research design in practice. In reality, some parameters in a decision model will not be amenable to further investigation, and others will be naturally grouped. In other words, a research design to investigate a particular parameter of interest may automatically, or at near zero marginal cost, collect additional information on a number of related parameters, and so we may think of groups of parameters in terms of research sets. The true n is not the number of parameters in the decision model but the number of research sets. Yokota and Thompson[17] emphasize that any value of information analysis requires a definition of the set of available actions and information collection strategies, and so the need to define the research sets is not limited to consideration of sequential research designs.

Furthermore, if, when evaluating the benefit of a sequence beginning $\theta_1\theta_2$, the payoff from proceeding to the second parameter in sequence is never positive for any value of θ_1, there is no need to investigate any sequence beginning with that ordered pair, no matter how long. Optimizing over the whole research decision space may be computationally impractical, but investigating whether any sequential design of length $2(\frac{n!}{(n-2)!})$ is worth pursing may be a valuable addition to the current value of information calculations.

Comparing per Patient Value of Information

It is less easy to discuss the value of information in isolation of costs for a sequential design than it is when considering a one-off design, but it may be of worth to consider the bounds. The expected value of information from a sequential design will be at least as great as the value of information about the first parameter in the sequence. Furthermore, the value of information of the sequential design cannot exceed the value of information for both parameters in combination. When there is no additional value from investigating the second parameter in sequence, the sequential design collapses back to the EVPPI for the first parameter in sequence. Where there is always additional value from investigating the second parameter in sequence, the value of information gathered from a sequential design will be equal to the value obtained from a one-off research design investigating both parameters simultaneously.

Comparing Population Benefit

We may assume the duration of a study investigating a number of parameters simultaneously is determined by the parameter for which the data takes longest to collect. We may also assume that collecting data on parameters concurrently will take less time than investigating those parameters sequentially, so $\tau_{\theta_1,\theta_2} = \max(\tau_{\theta_1}, \tau_{\theta_2}) < \tau_{\theta_1} + \tau_{\theta_2}$. The information on the first parameter in a sequential design will be available in the same time as would be taken to investigate that parameter individually. However, conditional on that information, the research into the second parameter may or may not proceed. Thus, for some resolutions of θ_1, research will proceed on θ_2, the results of which would be available at time $t = \tau_{\theta_1} + \tau_{\theta_2}$. For other resolutions of θ_1, no further research would be conducted. Patients from time τ_{θ_1} benefit from the additional information on θ_1 (P_{θ_1}), and for some resolutions, a smaller population of patients, $P_{\theta_1+\theta_2}$, benefits from additional information on θ_2. In contrast, a design that evaluated θ_1 and θ_2 simultaneously would provide information on both θ_1 and θ_2 at time $t = \tau_{\theta_1,\theta_2}$. The population of patients expected to benefit from additional information on θ_1 is smaller, but the population expected to benefit from additional information on θ_2 is larger. Potentially, if τ_{θ_1} is short and τ_{θ_2} is long, the population benefit from a sequential research design could exceed the population benefit from a one-off research design investigating both parameters simultaneously.

Comparing Costs

The total cost of investigating 2 parameters sequentially will in general exceed the cost of investigating them simultaneously. This may in part be

due to economies realized by running a single trial to investigate both parameters (e.g., each study may have some element of fixed cost). However, even in the absence of any economies from a joint design, the opportunity costs of the sequential design that does investigate both parameters will generally exceed the opportunity costs of the joint design because the length of time taken to gain the additional information on both parameters will tend to be greater for a sequential design, $C_{\theta_1} + C_{\theta_2} > C_{\theta_1,\theta_2}$. This suggests that where the payoff from research into the second parameter in sequence is always positive (for any resolution of the first parameter), a sequential design will be dominated by a one-off research design that investigates both parameters simultaneously.

However, the value of a sequential design derives from the ability to avoid the cost of research on the second parameter in sequence on the basis of additional information collected on the first parameter in sequence. This suggests that sequential designs may be most valuable when the costs of investigating the second parameter in the sequence are significant and the benefits of investigating the second parameter in the sequence are sensitive to realizations of the first parameter. In these circumstances, it will be more likely that research about the second will be unnecessary.

If the first parameter is relatively cheap to investigate in terms of cost and research time, then the issue of whether to proceed to the second parameter could be addressed with little additional opportunity cost compared to a concurrent research design and potentially great savings. A parallel can be drawn with Chao and others'[22] value of flexibility, in which the timing of the 'second-stage' decision based on updated information decreases with the timing of the information (length of the 'first stage'). For example, a lengthy randomized trial to establish the rate of long-term adverse events from a medical intervention could be circumvented if it was first established in a small observational study that the health-related quality of life impacts of such events were too small for the true rate to impact on the decision. In contrast, if the small observational study indicated a large impact on health-related quality of life for those same adverse events, then further support would be provided for investing in the randomized trial.

DISCUSSION

Most previous EVI analyses have focused on questions about one-off research designs. When assessing the individual EVPPI for each parameter in the model, it may be tempting to prioritize further studies in order of the size of the potential payoffs for each design. In this article, we have discussed how to calculate the payoff from alternative sequential research designs that account for learning from the information gathered on each parameter in sequence and allow a choice of whether to proceed with research on the next parameter in sequence. Widening the research decision space to incorporate sequential designs is important because the sequence with the largest expected potential payoff may imply a different order of research to that indicated by naïvely ordering the parameters in terms of the EVPPI. This additional information to inform the order of research will however come at the cost of increased analysis time required to perform the value of information analyses; thus, the use of sequential EVPI in practice will be linked to how time consuming it is to calculate for each given decision problem.

Sequential EVPPI can be used to identify the parameter or group of parameters that should be investigated first because it accounts for the impact of any additional information on the need for further research. However, it does not determine the order of research beyond the first study. Once the results of the first study or set of studies are available, the previous EVI calculations become obsolete and a new set incorporating the additional information is required. In this example, once additional information has been collected for θ_1, the question of whether to proceed with research on θ_2 will be informed by a new calculation of Π_{θ_2}.

This article focuses on perfect information to describe alternative research decisions. Of course, the value of perfect information represents the maximum value of further research, and in practice, actual research will only provide imperfect sample information. Even so, estimating the maximum payoff from further research is computationally feasible and can provide a useful indication of which research designs are likely to represent a worthwhile use of available resources. If the expected cost in a research proposal exceeds the EVPI, then it will also exceed the EVSI and it may be disregarded as a valuable use of scarce resources. Currently, optimizing over all possible research design space on the basis of the EVSI is generally computationally expensive and may be prohibitive in many common circumstances. The computational expense of EVSI calculations would increase if sequential research designs were also evaluated. Therefore, using this

type of EVPI analysis to prioritize the research decision space within a particular decision problem is of value and can be used to focus expensive EVSI calculations on those research designs which are likely to be most valuable.

REFERENCES

1. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18(Suppl):S68–80.

2. Eckermann S, Willan A. Expected value of information and decision making in HTA. Health Econ. 2006;26(2):195–209.

3. Griffin S, Claxton K, Palmer S, Sculpher, M. Dangerous omissions: the consequences of ignoring decision uncertainty. Health Econ. Forthcoming.

4. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18:342–64.

5. Raiffa H, Schlaifer RO. Applied Statistical Decision Theory. Boston: Division of Research, Harvard Business School; 1961.

6. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. Med Decis Making. 1998;18:95–109.

7. Reichard EG, Evans JS. Assessing the value of hydrogeologic information for risk-based remedial action decisions. Water Resour Res. 1989;25(7):1451–60.

8. Dakins M. The value of the value of information. Human and Ecological Risk Assessment. 1999;5(2):281–9.

9. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. Med Decis Making. 2007;27:448–70.

10. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment Programme. Health Technol Assess. 2004;8:1–103, iii.

11. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis: an application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care. 2001;17: 38–55.

12. Oakley J. Decision-Theoretic Sensitivity Analysis using Value of Information. Los Alamos, NM: Los Alamos National Library; 2005.

13. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.

14. Thompson KM, Graham JD. Going beyond a single number: using probabilistic risk assessment to improve risk management. Human and Ecological Risk Assessment. 1996;2(4):1008–34.

15. Claxton K, Eggington S, Ginnelly L, et al. A Pilot Study of Value of Information Analysis to Support Research Recommendations for the National Institute for Health and Clinical Excellence. York, UK: Centre for Health Economics, University of York; 2005. Research paper 4.

16. Yokota F, Thompson KM. Value of information literature analysis: a review of applications in health risk management. Med Decis Making. 2004;24:287–98.

17. Yokota F, Thompson KM. Value of information analysis in environmental health risk management decisions: past, present, and future. Risk Anal. 2004;24(3):635–50.

18. Philips Z, Claxton K, Palmer S. The half-life of truth: appropriate time horizons for research decisions? Med Decis Making. 2008;28:287–99.

19. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making. 2004;24:207–27.

20. Conti S, Claxton K. Dimensions of Design Space: A Decision-Theoretic Approach to Optimal Research Design. York, UK: Centre for Health Economics, University of York; 2008.

21. Brand KP, Small MJ. Updating uncertainty in an integrated risk assessment: conceptual framework and methods. Risk Anal. 1995;15(6):719–31.

22. Chao H, Peck SC, Wan YS. Managing uncertainty: the tropospheric ozone challenge. Risk Anal. 1994;14(4):465–75.

23. Welton NJ, Ades AE, Caldwell DM, Peters TJ. Research prioritization based on expected value of partial perfect information: a case study on interventions to increase uptake of breast cancer screening. J R Stat Soc Ser A Stat Soc. 2008;171:807–41.

24. Cox DC, Baybutt P. Methods for uncertainty analysis: a comparative survey. Risk Anal. 1981;1(4):251–8.

HEALTH ECONOMICS
Health Econ. (2010)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hec.1586

DANGEROUS OMISSIONS: THE CONSEQUENCES OF IGNORING DECISION UNCERTAINTY

SUSAN C. GRIFFIN*, KARL P. CLAXTON, STEPHEN J. PALMER and MARK J. SCULPHER
Centre for Health Economics, University of York, UK

SUMMARY

Institutions with the responsibility for making adoption (reimbursement) decisions in health care often lack the remit to demand or commission further research: adoption decisions are their only policy instrument. The decision to adopt a technology also influences the prospects of acquiring further evidence because the incentives to conduct research are reduced and the ethical basis of further clinical trials maybe undermined. In these circumstances the decision maker must consider whether the benefits of immediate access to a technology exceeds the value of the evidence which maybe forgone for future patients. We outline how these expected opportunity losses can be established from the perspective of a societal decision maker with and without the remit to commission research, and demonstrate how these considerations change the appropriate decision rules in cost-effectiveness analysis. Importantly, we identify those circumstances in which the approval of a technology that is expected to be cost-effective should be withheld, i.e. when an 'only in research' recommendation should be made. We demonstrate that a sufficient condition for immediate adoption of a technology can provide incentives for manufacturers to reduce the price or provide additional supporting evidence. However, decisions based solely on expected net benefit provide no such incentives, may undermine the evidence base for future clinical practice and reduce expected net health benefits for the patient population. Copyright © 2010 John Wiley & Sons, Ltd.

Received 3 November 2008; Revised 30 July 2009; Accepted 13 November 2009

KEY WORDS: value of information; decision analysis

1. INTRODUCTION

Appropriate social decisions about new health-care technologies require a number of questions to be addressed: (i) is the new technology expected to be cost-effective based on existing evidence; (ii) is additional evidence required to support its use; (iii) if so, what type of evidence will be most valuable and (iv) can the evidence needed be provided once the technology is approved for use? If not, or if it is less likely or more costly to provide, then the decision maker must consider whether the benefits of immediate access to a technology exceed the value of the evidence which maybe forgone for future patients.

It has been argued that in the absence of sunk cost or other irreversibilities the decision to adopt a technology can be based on expected cost-effectiveness. The question of whether further evidence is required can be regarded as a conceptually separate although simultaneous decision based on the same principles and analysis (Claxton, 1999). However, there are good reasons to believe that adoption or reimbursement of a technology may damage the prospects of further research being conducted. We will show that in these circumstances, adoption or reimbursement becomes the only policy instrument available and decisions can no longer be based on expected cost-effectiveness. Instead there must be

*Correspondence to: Centre for Health Economics, Alcuin A Block, University of York, Heslington, York, YO10 5DD, UK.
E-mail: scg3@york.ac.uk

some formal assessment of the opportunity loss of immediate adoption (the value of research which may be forgone) and failure to do so is a dangerous omission and may lead to the adoption and diffusion of technologies for which current evidence is insufficient, and undermine the evidence base for future clinical practice. We outline how these expected opportunity losses can be established from the perspective of a societal decision maker with and without the remit to commission research and demonstrate how these considerations change the appropriate decision rules in cost-effectiveness analysis.

2. IS THE TECHNOLOGY EXPECTED TO BE COST-EFFECTIVE?

If the objective of a health-care system is to maximise health gains from available resources then the decision to adopt or reimburse an alternative (j) should be based on the costs (C_j), health outcomes (Q_j) and the cost-effectiveness threshold (λ). The cost-effectiveness of an alternative j can be expressed in terms of net health benefit ($\mathrm{NB}_j = Q_j - C_j/\lambda$) (Phelps and Mushlin, 1991; Stinnett and Mullahy, 1998). Expected net health benefit ($E(\mathrm{NB}_j)$) will be uncertain and a decision must be made before it is known how the uncertain parameters (θ) which determine $E(\mathrm{NB}_j)$ will resolve. With current information, and in the absence of irreversibility (Palmer and Smith, 2000) or any costs associated with reversing a decision (Eckermann and Willan, 2008b), the decision maker can choose the intervention that generates the maximum expected net benefit:

$$\max_j \; E_\theta \mathrm{NB}(j, \theta) \tag{1}$$

Net health benefit per patient can be expressed for the population of current and future patients (I_t) over the effective life time of the technology (T) discounted at rate r:

$$\text{Population } E(\mathrm{NB}_j) = E_\theta \mathrm{NB}(j, \theta) \cdot \sum_{t=1}^{T} \frac{I_t}{(1+r)^t}; \tag{2}$$

These expected population net health benefits are illustrated in Figure 1[1] for a new technology (j^*) and current clinical practice (j^0). The expected incremental cost-effectiveness ratio (ICER) of the new technology is £25 000 per QALY so is expected to be cost-effective when $\lambda \geq$ £25 000 and $E(\mathrm{NB}_{j^*}) \geq E(\mathrm{NB}_{j^0})$. The difference in population net health benefit between j^* and j^0 represents the expected benefit of immediate adoption based on existing evidence.

Therefore, a necessary and sufficient condition for adopting the new technology is $\lambda \geq \mathrm{ICER}$ or $E(\mathrm{NB}_{j^*}) \geq E(\mathrm{NB}_{j^0})$. This does not mean, however, that uncertainty is irrelevant or that decisions can be based on limited and poor quality evidence, because the question of whether additional evidence is required to support the adoption or reimbursement decision must also be addressed (Claxton, 1999).

3. IS MORE EVIDENCE REQUIRED?

If the uncertainty in costs and effects could be immediately resolved, i.e. with perfect information, the decision maker could select the alternative that maximises the net benefit for each possible value of θ ($\max_j \mathrm{NB}(j, \theta)$). However, the true values of θ are unknown, so the expected value of a decision taken with perfect information is found by averaging the maximum net benefit over the joint distribution of θ.

$$E(\mathrm{NB}^{**}) = E_\theta \max_j \mathrm{NB}(j, \theta) \tag{3}$$

[1]The figures used throughout are intended to be illustrative. All the figures are based on a published analysis of paclitaxel and PLDH for the treatment of ovarian cancer. (Main *et al.*, 2006).

DOI: 10.1002/hec

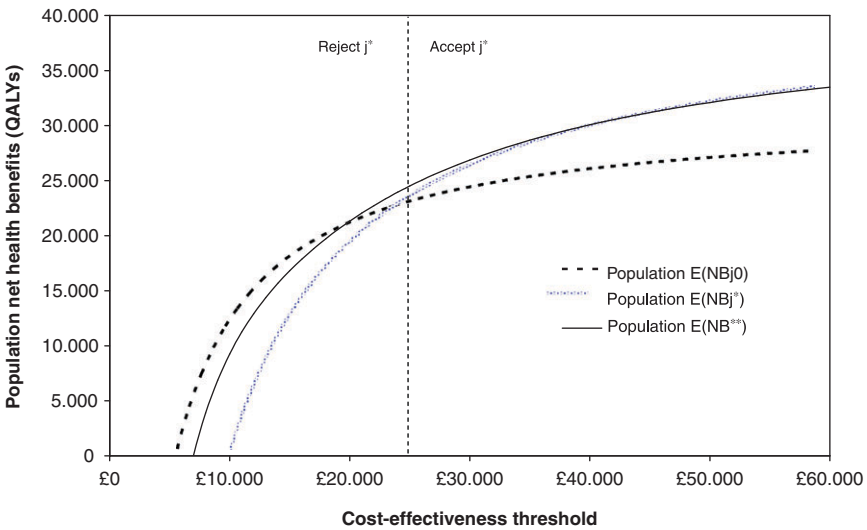

Figure 1. Decisions based on expected cost-effectiveness.

This is the maximum expected net health benefit that can be achieved and can also be expressed for the population of current and future patients: (Phillips *et al.*, 2008; Eckermann and Willan, 2008a)

$$\text{Population } E(\text{NB}^{**}) = E_\theta \max_j \text{NB}(j,\theta) \cdot \sum_{t=1}^{T} \frac{I_t}{(1+r)^t} \tag{4}$$

This is also illustrated in Figure 1. Resolving uncertainty by conducting further research improves expected net health benefits ($E(\text{NB}^{**}) \geq \max_j E(\text{NB}_j)$) because better decisions can be made which generate more health outcome from the available resources. Therefore, evidence can be valuable in the same way as access to a new cost-effective technology. The maximum social value of conducting further research is the difference between the population values of $E(\text{NB}^{**})$ and $\max_j E(\text{NB}_j)$, which is the expected value of perfect information (EVPI). This provides an upper bound on the value of evidence since actual research will not resolve all the uncertainty and will not be immediately available. Nevertheless it does provide a necessary condition: research will only be potentially worthwhile if the EVPI exceeds the expected costs of research (Cr), expressed in health terms. It should be apparent that there can be circumstances in which the value of additional evidence exceeds the value of access to a cost-effective technology, e.g. when £25 000 $\leq \lambda \leq$ £30 000 in Figure 1. In particular, when $\lambda =$ ICER and $E(\text{NB}_{j^*}) = E(\text{NB}_{j^0})$ there is no expected gain in health benefits from adopting the new technology but the EVPI is positive and reaches a maximum.

Adopting new technologies based on expected cost-effectiveness may be appropriate if the decision of whether further evidence is needed to support its use is made at the same time and on the same basis. However, in the current policy environment those institutions with responsibility for making adoption and reimbursement decisions do not always formally address the question of whether further evidence is required. Even when they do, they often do not have the remit to prioritise and commission research or have the powers to require that additional research is conducted by the manufacturers of a technology.

This may not be of concern for reimbursement decisions if the prospects of acquiring information through research are unaffected by the decision to adopt a technology: decisions can continue to be based on expected cost-effectiveness. However, this is unlikely to be the case for a number of reasons: (i) the adoption of a technology removes incentives on the manufacturer to conduct further research; (ii) the early diffusion of a technology means that future clinical trials are less likely to be supported or

DOI: 10.1002/hec

regarded as ethical by the clinical community, even when pubic funds are made available for such research; and (iii) patients are unlikely to enroll in clinical trials once they have unrestricted access to the new technology.

It has been demonstrated that irreversible aspects of adoption decisions or any costs associated with subsequently reversing decisions (Eckermann and Willan, 2008b; Palmer and Smith, 2000) ought to be taken into account in both adoption and research decisions. Below we demonstrate that evidence forgone due to early adoption may also be an important form of irreversibility that ought to be accounted for. We demonstrate those circumstances when it may be better to deny the approval of a technology, which is expected to be cost-effective, even when the costs of reversal are zero, if the value of evidence forgone exceeds the benefits of early access to the technology.

4. CAN THE EVIDENCE BE PROVIDED?

The sufficient condition for adopting a new technology needs to be extended to account for the value of information that might be forgone. This requires some assessment of the probability that research will be conducted (α) and when it might report (τ), at which point the decision to adopt or reject the technology can be revised. Initially we consider the situation in which the new technology is expected to be cost-effective compared with current practice ($E(\text{NB}_{j^*}) \geq E(\text{NB}_{j^0})$)[2] and where the reimbursement authority does not have the remit to commission research.[3]

4.1. Expected net benefits of rejection

If approval of j^* is withheld, patients receive current practice, j^0, and the associated lower net benefits of $E(\text{NB}_{j^0})$. If research is not conducted and this decision is not revised over the life time of the technology ($t = 1,\ldots,T$), these expected net benefits accrue to the population of current and future patients. However, if research is conducted and reports at time $\tau < T$, the decision can be revised and the maximum that subsequent patients are expected to receive is net benefits of $E(\text{NB}^{**})$. If the decision maker does not control research commissioning then some assessment of the probability that research will be conducted if the technology is rejected is required (α_R). Therefore, the expected net benefit of rejecting j^* (B_R) can be expressed as:

$$B_R = E(\text{NB}_{j^0}) \cdot P_{t<\tau} + (1 - \alpha_R) \cdot E(\text{NB}_{j^0}) \cdot P_{t>\tau} + \alpha_R \cdot E(\text{NB}^{**}) \cdot P_{t>\tau}, \tag{5}$$

where the population before any research reports at $t = \tau$ is:

$$P_{t<\tau} = \sum_{t=1}^{\tau} \frac{I_t}{(1+r)^t},$$

and the population following research reports at $t = \tau$ is:

$$P_{t>\tau} = \sum_{t=\tau}^{T} \frac{I_t}{(1+r)^t}$$

[2]If $\lambda >$ ICER the new technology is not expected to be cost effective and should be rejected, in which case we assume the prospects for further research are unaffected. However, it is possible that some types of evidence may be more difficult to acquire (e.g. long-term adverse events) if the technology is rejected. This is discussed in more detail in Section 5.

[3]The decision maker need only consider the expected benefits of any research, whether commissioned by other public bodies or conducted by manufacturers within or outside their jurisdiction. However, if they were responsible for both adoption and research decisions they would need to decide whether to commission research taking account of the costs of conducting research as well as the value of the evidence it will produce. This is discussed in more detail in Section 6.

DOI: 10.1002/hec

By rejecting a new technology in favour of current practice, the prospects for future research are unaffected[4]: the manufacturer of the new technology retains an incentive to conduct research to obtain approval and no change in clinical practice occurs to damage existing trials or remove the ethical basis for future research.

4.2. Expected net benefits of adoption

If the new technology is approved on the basis of current evidence, patients begin to receive j^* immediately with higher net expected benefits of $E(\mathrm{NB}_{j^*})$. If research is not conducted and this decision is not revised, these expected net benefits accrue to the population of current and future patients over the life time of the technology ($P_{t<\tau}+P_{t>\tau}$). However, if research is conducted and reports at time $\tau<T$, the decision can be revised and again the maximum subsequent patients receive would also be $E(\mathrm{NB}^{**})$. Therefore some assessment of the probability that research will be conducted if the technology is adopted is required (α_A). The expected net benefit of adopting j^* (B_A) can be expressed as:

$$B_A = E(\mathrm{NB}_{j^*}) \cdot P_{t<\tau}+(1-\alpha_A) \cdot E(\mathrm{NB}_{j^*}) \cdot P_{t>\tau}+\alpha_A \cdot E(\mathrm{NB}^{**}) \cdot P_{t>\tau} \quad (6)$$

A sufficient condition to adopt a technology should take account of both the expected benefits of access to the technology and the opportunity loss of the value of the evidence forgone. This can be expressed as the difference between (6) and (5):

$$B_A - B_R = [E(\mathrm{NB}_{j^*}) - E(\mathrm{NB}_{j^0})](P_{t<\tau}+(1-\alpha_R)P_{t>\tau}) - (\alpha_R-\alpha_A)[E(\mathrm{NB}^{**}) - E(\mathrm{NB}_{j^*})]P_{t>\tau} \quad (7)$$

If adoption reduces the prospects of further research being conducted ($\alpha_A<\alpha_R$), or increases the time taken for research to report, there will be an opportunity loss of the value of information forgone ($E(\mathrm{NB}^{**}) - E(\mathrm{NB}_{j^*})$) for future patients ($P_{t>\tau}$) which may be greater than the expected benefits of immediate access to the new technology ($E(\mathrm{NB}_{j^*}) - E(\mathrm{NB}_{j^0})$) for the population of patients before any research which might be conducted reports ($P_{t<\tau}+(1-\alpha_R)P_{t>\tau}$). It should be clear that the standard sufficient condition for adopting a technology based on expected cost effectiveness is a special case, i.e. if the decision to adopt the technology does not affect the prospects of further research ($\alpha_A=\alpha_R$), then the new technology should be adopted when $\lambda \geq \mathrm{ICER}$ and $E(\mathrm{NB}_{j^*}) \geq E(\mathrm{NB}_{j^0})$.[5]

If adoption reduces the prospects for further research then observing $\lambda \geq \mathrm{ICER}$ is no longer sufficient. This is illustrated in Figure 2 where $\alpha_A=0$, $\alpha_R=1$ and $\tau=2$ based on the same example used in Figure 1. The expected benefits of adopting the technology are the same ($B_A = \mathrm{Population}\, E(\mathrm{NB}_{j^*})$ in Figure 1) since $\alpha_A=0$ in (6) above. But the expected benefit of rejecting the technology is greater than the Population $E(\mathrm{NB}_{j^0})$ because (5) also includes the higher expected net benefit of $E(\mathrm{NB}^{**})$ available in two years time. The benefit of rejecting the technology has increased so the sufficient condition is no longer that $\lambda \geq \mathrm{ICER}$ but $\lambda \geq$ £39 000. Indeed, adopting this technology when $\lambda = \mathrm{ICER}$ would impose a maximum expected opportunity loss of 1879 QALYs or £47 m. Therefore, the expected benefits of immediate access must not simply be greater than zero but sufficiently great to offset the value of any information that maybe forgone.

4.3. Sufficient conditions for approval

Some assessment of α_A, α_R and τ is required to establish a sufficient condition for approval. Combinations of α and τ can be found for which $B_A = B_R$ for particular values of λ. The bold curve in Figure 3 represents those combinations of α_R and τ where $B_A = B_R$ when $\lambda =$ £30 000 and $\alpha_A = 0$. This represents the boundary beyond which the sufficient condition for approval is met, i.e. to the north-east the new technology meets the sufficient condition and should be approved on the basis of current

[4]See footnote 1.

[5]If it is assumed that perfect information will become immediately available following rejection then the technology should always be rejected. This is what would be implicitly assumed by simply comparing the EVPI to the incremental net benefit without accounting for the timing and probability of research.

DOI: 10.1002/hec

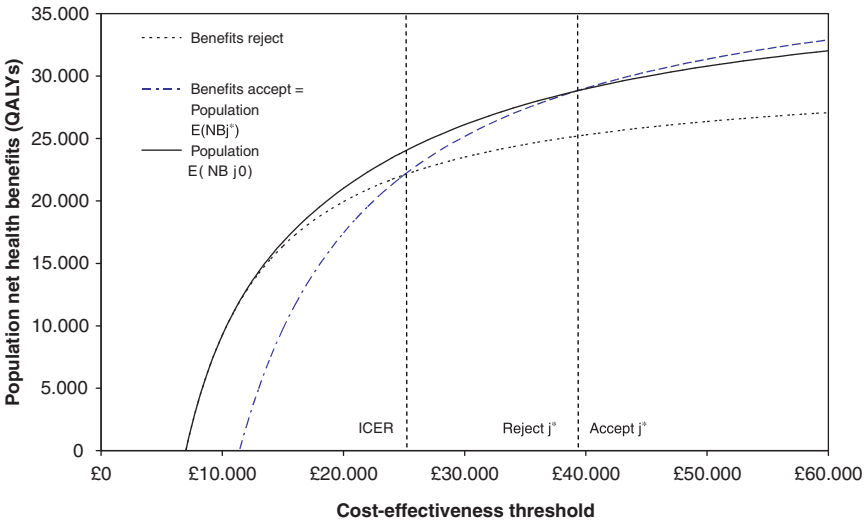

Figure 2. Expected benefits of adoption and rejection.

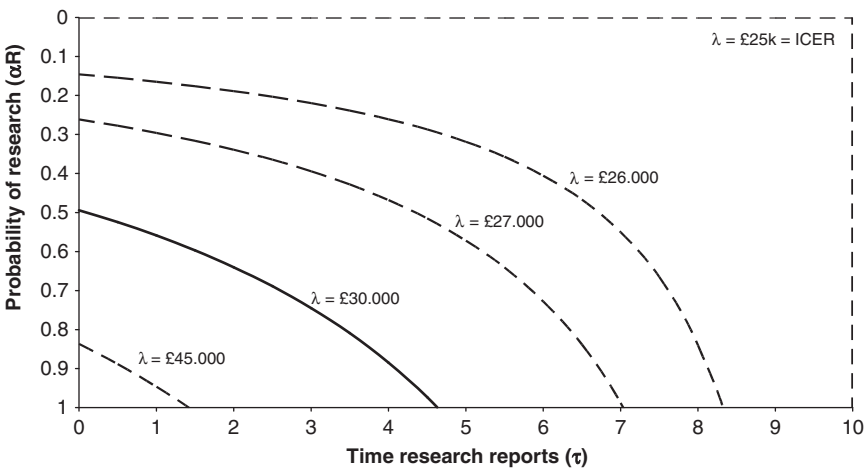

Figure 3. Approval boundary ($\alpha_A = 0$).

evidence. As α_R approaches 1 and τ approaches 0 the technology is increasingly likely to be rejected on the basis of current evidence.

For example, if future research is not expected to report for some time (e.g. $\tau = 4.5$) then the new technology should be approved even though it may displace research that is almost certain to be conducted. Similarly, if research is less likely to be conduced (e.g. $\alpha_R = 0.4$) then unless it is expected to report quickly (e.g. $\tau = 1$) the technology should also be approved. However, if there is research which is likely to be conduced (e.g. $\alpha_R = 0.8$) and is likely to report soon (e.g. $\tau = 2$) then it may be better to withhold approval even though the technology is expected to be cost-effective. Boundaries are also illustrated for a range of cost-effectiveness thresholds. When $\lambda = \text{ICER} = £25\,000$ there is no region of acceptance based on current evidence, i.e. standard decision rules are not a sufficient condition for

DOI: 10.1002/hec

approval. As λ is increased the boundary shifts to the south-west and it is more likely that the technology will be accepted on the basis of current evidence.

4.4. Price, evidence and approval

If manufacturers believe that decision makers approve technologies solely on the basis of expected cost-effectiveness, they have no incentive to provide evidence other than that required for licensing. They also have an incentive to price their technology such that its ICER is just below the cost-effectiveness threshold. By doing so they can minimise research and development costs and fully capture the surplus to the health-care system. (Claxton, 2007) In contrast, decision rules that consider the opportunity loss of adoption do provide incentives for manufacturers to provide more evidence to support the technology or to reduce price. Indeed manufacturers face a trade-off between price reductions or investing in evaluative research.

The adoption decision can be altered by reducing uncertainty surrounding cost-effectiveness, even if this leaves the ICER unchanged. For example, if a trial was expected to report in two years with a probability of 0.8 then the technology would not meet the sufficient condition for approval: this point is to the south-west of the original boundary for approval (for $\lambda = £30\,000$) in Figures 3 and 4. However, approval could have been gained if the manufacturer had provided additional evidence to support the technology. If the ICER and therefore B_A remains unchanged the cost-effectiveness of the new technology would be less uncertain so the EVPI per patient ($E(\mathrm{NB}^{**}) - E(\mathrm{NB}_{j*})$) will be reduced and B_R would be lower. If the EVPI per patient is reduced from 0.032 to 0.020 QALYs, the boundary for approval shifts to the south-west (from the dashed line in Figure 4 to the solid line) to the point at which $B_A = B_R$ when $\alpha = 0.8$ and $\tau = 2$. The new technology now meets the sufficient condition and the manufacturer would have gained market access.

Instead of investing in additional research, the manufacturer could reduce the price of the new technology, reducing the ICER and increasing the expected benefits of adoption. For example, if the total cost is reduced by 3% the technology appears more cost-effective. The ICER falls to £23 753 and $E(\mathrm{NB}_{j*}) - E(\mathrm{NB}_{j^0})$ rises from 0.0321 to 0.040 QALYs so B_A increases. In this case decision uncertainty and EVPI per patient are also reduced from 0.032 to 0.025. The boundary for approval in Figure 4 again shifts to the south-west. The same technology with the same evidence base but with a lower price meets the sufficient condition and should be approved on the basis of current evidence.[6] In summary, both price and the level of decision uncertainty can be altered to increase the benefits of immediate adoption ($B_A - B_R$), which means that the sufficient condition for adoption is met at lower combinations of α_R and τ.

Decision rules that consider the value of evidence forgone would reduce the number of new technologies approved compared with decisions based solely on expected net benefit. The decision to reject technologies with ICERs higher than the threshold would be unaffected,[7] but some technologies with ICERs below the threshold would also be rejected. Technologies more likely to be rejected on these grounds would be those priced such that the ICER is close to the threshold and those were there is more uncertainty surrounding cost-effectiveness and where the EVPI is highly relative to the expected additional net benefits. Such decisions would provide clear incentives to manufacturers to either invest in sufficient evidence to support the technology or price it in such a way that it appears so cost-effective that decision uncertainty and EVPI are reduced to acceptable levels. Other things being equal, those that do invest, thereby reducing the uncertainty surrounding the cost-effectiveness of their product at launch, will be able to gain approval at higher prices than those that do not, proving rewards for socially valuable research.

[6]Reducing price will also reduce the EVPI if uncertainty is associated with whether j^* is sufficiently more effective to justify the additional cost. However, if the effectiveness of j^* compared with j^0 is also uncertain then the effect of price reductions will be limited and will not continually reduce EVPI, i.e. there will still be positive EVPI even if the new technology was free.
[7]See footnote 1.

DOI: 10.1002/hec

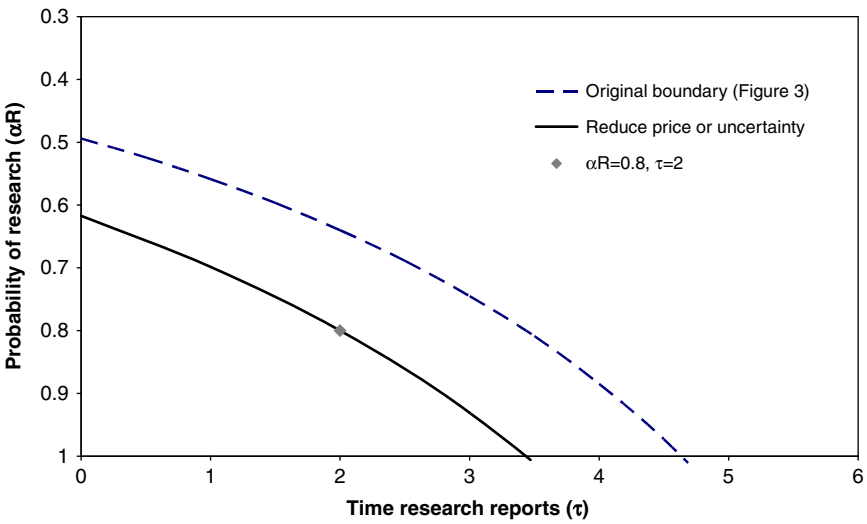

Figure 4. Price, evidence and approval ($\lambda = £30\,000$, and $\alpha_A = 0$).

5. WHAT TYPE OF RESEARCH?

The sufficient conditions for adoption outlined above only consider the value of information for the decision problem as a whole. However, there are different types of evidence that maybe valuable requiring different types of research. The expression for B_A and B_R can be extended to consider the value of information associated with particular groups of parameters that contribute to net benefit (Ades *et al.*, 2004). This may be useful in two ways: (i) the value of information associated with particular parameters can indicate to manufacturers and those responsible for commissioning publicly funded research, not only if further evidence maybe required, but what type of evidence is needed and, by implication, the appropriate research design; (ii) it may be that some types of research are more likely to be displaced by an adoption decision than others. Therefore, setting aside the question of whether the private or public sector should bear the costs of research, the opportunity loss of adoption will depend, not only, on the overall EVPI for the decision but whether the EVPI is associated with parameters which require particular research designs.

For example, if there are two uncertain parameters, $\theta_1, \theta_2 \cup \theta$, which determine expected net health benefit in (1), there are three, mutually exclusive research possibilities[8]: (i) research on θ_1 only; (ii) research on θ_2 only or (iii) as previously, research informing all the uncertainties (θ_1 and θ_2 together). The value of perfect parameter information (EVPPI) per patient can be calculated for each of these possibilities. For example the value of research on θ_1 is described by:

$$\text{EVPPI}_{\theta_1} = E_{\theta_1} \max_j E_{\theta_2} \text{NB}(j, \theta_1, \theta_2) - \max_j E_{\theta} \text{NB}(j, \theta), \tag{8}$$

In this case, since there are only two parameters $\text{EVPPI}_{\theta_1,\theta_2} = E(\text{NB}^{**}) - E(\text{NB}_{j*}) = \text{EVPI}$.

An assessment of the probability that each of these three types of research will be conducted is required if the technology is rejected or approved:

$$\alpha_R = \alpha_R^{\theta_1} + \alpha_R^{\theta_2} + \alpha_R^{\theta_1,\theta_2}, \text{ and } \alpha_A = \alpha_A^{\theta_1} + \alpha_A^{\theta_2} + \alpha_A^{\theta_1,\theta_2} \tag{9}$$

[8]Sequences of research could also be considered, e.g. research on θ_1 followed by θ_2 or θ_2 followed by θ_1. The EVPI for such sequences can be established (Griffin *et al.*, 2009). However, for simplicity but without loss of generality sequential designs are not explicitly considered here.

DOI: 10.1002/hec

If research is conducted then some assessment of time at which these different types of study may report will also be required (τ_{θ_1}, τ_{θ_2}, τ_{θ_2}). Now a sufficient condition will depend on the value of the different types of research which may be forgone and the time at which they may report. The second term in the expression for $B_A - B_R$ in (7), which represents the expected value of information forgone by adopting the technology, must be expanded to include each of these elements:

$$(\alpha_R^{\theta_1} - \alpha_A^{\theta_1})\text{EVPPI}_{\theta_1} \cdot P_{t>\tau\theta_1} + (\alpha_R^{\theta_2} - \alpha_A^{\theta_2})\text{EVPPI}_{\theta_2} \cdot P_{t>\tau\theta_2} + (\alpha_R^{\theta_1,\theta_2} - \alpha_A^{\theta_1,\theta_2})\text{EVPI} \cdot P_{t>\tau\theta_1,\theta_2} \quad (10)$$

For example, if θ_1 is an estimate of relative effect of j^* then a randomised clinical trial maybe required to provide more precise and unbiased estimates. Immediate approval may mean that this type of experimental research would be regarded as unethical or many not be able to recruit patients even if public funds were available so $\alpha_R^{\theta_1} > \alpha_A^{\theta_1} = 0$. Alternatively, θ_2 may be an estimate of quality of life associated with clinical events and could be informed by an observational study which may be undertaken whether or not approval is granted, i.e. $\alpha_R^{\theta_2} = \alpha_A^{\theta_2}$. In these circumstances, even if the overall EVPI is high, the technology maybe still be approved if most value is associated additional evidence about θ_2. Even if $\text{EVPPI}_{\theta_1} > \text{EVPPI}_{\theta_2}$, a trial may take some considerable time to report but results from an observational study could quickly be available ($\tau_{\theta_1} > \tau_{\theta_2}$) and the technology may still be approved. It is also possible that some types of evidence maybe more likely to be acquired (or made more quickly available) once the technology is approved, e.g. evidence of longer-term adverse events may become more quickly available with wider use of the technology. In these circumstances there will be a trade-off between the value of those types of evidence gained and those forgone following adoption. In principle (10) need not necessarily be positive since $\alpha_R^{\theta i} < \alpha_A^{\theta i}$ is possible.

6. NECESSARY AND SUFFICIENT CONDITIONS

The characterisation of the sufficient conditions for approval described above has been simplified to more easily demonstrate the essential characteristics of the decision problem. However, they suffer from a number of limitations that, in principle, can be over come by extending the analysis in a number of ways.

6.1. Assessment of α and τ

Some assessment of the probability and timing of research if the technology is approved or rejected is required. However, any assessment of α_R, α_A and τ will be uncertain, particularly for decision makers without the remit to prioritise and commission research. This uncertainty can be characterised by assigning prior distributions to α_R, α_A and τ to represent decision makers' beliefs. The expectation of B_R and B_A taken over these distributions can then be established. The uncertainty associated with the probability of research will not change the expectation of B_R or B_A because both are multi-linear in α_R and α_A. However, the uncertainty associated with its timing will influence expectation because both B_R and B_A are non-linear in τ due to the discounting term in (5) and (6).

Also, α and τ have been assumed to be independent and τ to be independent of whether the technology is adopted or rejected. However, adoption may delay as well as displace research. Of course, more complex characterisations of the prospects and timing of research are possible. For example, by specifying the probability of research reporting in each discreet period ($t = 1, \ldots, T$), or by using a continuous distribution of time to event. In both cases the expectation of B_R or B_A would be influenced by uncertainty due to the non-linearity introduced by discounting.

6.2. A necessary condition

The sufficient conditions described above are based on the possibility of forgoing research which would have resolved all uncertainty and provided perfect information, so the value evidence forgone (second

DOI: 10.1002/hec

term in (7)) is the upper bound on these opportunity losses. Therefore, $B_A - B_R \geq 0$ can only be a sufficient but not necessary condition for immediate approval based on current evidence. A necessary condition would require some assessment of the Expected Value of the (imperfect) Sample Information (EVSI) that may be forgone (Ades *et al.*, 2004).

This would require some assessment of the expected sample size and design of future research as well as α_R, α_A and τ. There will be combinations of α_R, α_A and τ where $B_A - B_R < 0$ based on perfect information (the sufficient condition for approval is not met), but would be positive based on sample information (a necessary condition would be met), i.e. if perfect information was forgone then approval should be withheld but since only imperfect sample information will be forgone the technology should be approved on the basis of current evidence. Although the calculation of EVSI over the range of expected sample sizes and research designs would be computationally expensive in most practical applications[9] it is possible to establish a threshold value of EVSI below which the technology should be approved on the basis of current evidence for particular value of α_R, α_A and τ.

For example, if $\alpha_R = 1$, $\alpha_A = 0$ and $\tau = 2$ the sufficient condition for approval is not met when $\lambda = £30\,000$ (see Figure 3). The technology should only be adopted (pass the necessary condition) if the EVSI per patient from this research was expected to be less than 0.001 QALYs or £286. Since the EVSI approaches the EVPI as sample size becomes large and the EVPI per patient is 0.033 QALYs or £990, it seems unlikely that well-designed research with adequate sample size would provide an EVSI lower than £286. Therefore, the technology fails both the sufficient and necessary conditions (the value expected to be forgone is too high) and approval should be withheld until the research reports. Alternatively if research is less likely and will take longer to report ($\alpha_R = 0.8$, $\alpha_A = 0$ and $\tau = 3$) approval should be granted if the EVSI per patient is expected to be less than 0.0285 QALYs or £855. Since this is very close to the EVPI, even well-conducted research is likely to generate a lower EVSI unless the sample size is expected to be very large. Therefore, although the technology fails the sufficient condition (see Figure 3) it may pass the necessary condition and be approved for immediate use based on existing evidence.

6.3. Prioritising and commissioning research

The sufficient and necessary conditions outlined above are relevant to decision making authorities who do not have the remit or budget to prioritise and commission research. In these circumstances the authority need only consider the expected benefits of any research (whether commissioned by public bodies or conducted by manufacturers within or outside the authorities jurisdiction (Eckermann and Willan, 2009)) based on an assessment of the probability the research will be undertaken and when it is likely to report. However, if the authority was responsible for both adoption and research decisions they would be able to decide whether to commission research or not, assess the time at which it will report, but also take account of the costs of conducting research as well as the value of the evidence it will produce.

Therefore, the sufficient condition for adoption needs to be amended so that the costs as well as the benefits of research are taken into account. For example, if research will not be possible following adoption due to ethical and recruitment considerations ($\alpha_A = 0$) the benefits of approval from (6) simplify to:

$$B_A = E(\mathrm{NB}_{j^*}) \cdot P_{t<\tau} + E(\mathrm{NB}_{j^*}) \cdot P_{t>\tau} \tag{11}$$

The benefits of rejecting the technology, however, now depend on whether the expected benefits of conducting research ($E(\mathrm{NB}^{**}) \cdot P_{t>\tau}$) exceed the costs (Cr) expressed in terms of

[9]A parametric approach to EVSI is computationally less demanding, see Raiffa and Schlaiffer 1967 and has been suggested for the analysis of health technologies, e.g. Claxton 1999 and Willan and Pinto 2005. However, the computational simplicity requires a comparison of only two alternatives, an assumption that incremental net benefit is normally distributed and that any sample will provide information on incremental net benefit directly rather than on the parameters of a model commonly used to estimate it.

DOI: 10.1002/hec

health.[10] If they do, then additional research maybe worthwhile and could be commissioned, so $\alpha_R = 1$ and (5) becomes:

$$B_R = E(\mathrm{NB}_{j^0}) \cdot P_{t<\tau} + E(\mathrm{NB}^{**}) \cdot P_{t>\tau} - Cr \tag{12}$$

However, if the costs exceed the expected benefits, research would not be worthwhile even if it was possible to conduct ($\alpha_R = 0$) and the technology should be adopted, i.e. B_R becomes $E(\mathrm{NB}_{j^0}) \cdot P_{t<\tau} + E \times (\mathrm{NB}_{j^0}) \cdot P_{t>\tau}$ and $B_A \geq B_R$ because $E(\mathrm{NB}_{j^*}) \geq E(\mathrm{NB}_{j^0})$.

As previously the expected benefits of research in (12) above is the upper bound because commissioned research will not resolve all uncertainty and provide perfect information. Therefore, the difference between (11) and (12) still provides only a sufficient condition to approve the technology when the decision maker is responsible for both adoption and research decisions. As discussed above these conditions can be extended to include the expected benefits of different types of evidence and the costs of associated research as well as the EVSI net of the cost of research, i.e. the Expected Net Benefits of Sampling (ENBS) when considering necessary conditions.

7. DISCUSSION

Decisions to adopt or reimburse a new technology based on expected cost-effectiveness can only be justified (in the absence of other sources of irreversibility or costs of reversal) if its approval and widespread use has no effect on the prospects of acquiring further evidence that maybe needed. However, there are good reasons to believe that the approval or reimbursement of a technology will damage the prospects of further research being conducted, particularly when any decision provides widespread and mandatory access to the technology, as in the case of the National Institute for Health and Clinical Excellence (NICE) in the UK. In these circumstances the decision to adopt a technology cannot be separated from the question of whether the evidence to support such a decision is sufficient. The decision maker must consider whether the benefits of immediate access to a technology exceed the value of the evidence that maybe forgone.

Some informal assessment of these issues may already be implicit in adoption and reimbursement decisions. For example, NICE makes suggestions for further research when issuing guidance on the use of technologies and these recommendations are passed to the NHS research and development programme for consideration. However, it is not clear how the research recommendations are formulated and they do not have the legal standing to require either the manufacturers to conduct the research or for it to be prioritised for public funding. NICE also claims to consider the uncertainty surrounding cost-effectiveness in formulating guidance (NICE, 2004). However, it is not made explicit how an assessment of decision uncertainty is 'taken into account' particularly when assessments of value of information are often not available. NICE, however, does have the remit to make 'only in research' recommendations and has used this in formulating guidance on a number of occasions (Chalkidou *et al.*, 2007). The question of when an 'only in research' recommendation should be made has been subject to recent debate (Chalmers, 2007; NICE citizens council, 2007).

The circumstances in which the approval of a technology that is expected to be cost-effective should be withheld, i.e. when an 'only in research' recommendation should be made, can be established based on an explicit and transparent assessment of the value of evidence which may be forgone and the expected benefits of immediate approval. Such assessments are entirely consistent with estimates of expected costs-effectiveness, which are already required in many jurisdictions.

Establishing clear decision rules based on an explicit and transparent assessment may have a number of advantages: First, withholding the approval of a technology which is expected to be cost-effective

[10]Cr will include the fixed and variable resource costs of the research and any other opportunity costs to those patients enrolled the study not already captured in (12) and (13), all appropriately discounted and expressed in health terms using λ.

DOI: 10.1002/hec

based on current evidence may be difficult to sustain (given the commercial interests of manufacturers and often the views of the clinical and patient communities) without an explicit and accountable demonstration of the value of evidence forgone for future patients from immediate adoption. Second, formal assessment would ensure consistency in decision-making within and between technology appraisals since estimates of the value of evidence are based on the same analysis used to provide estimates of the cost-effectiveness of the technologies. And finally, it is only with explicit decision rules and transparent assessment that accountable and predictable decisions can be made. Only then will manufactures face the type of incentives which will encourage the provision of sufficient evidence (of the right type) to be offered earlier in support of a technology.

The estimates of the opportunity loss of adoption have only considered the value of evidence forgone for the current decision problem, which only includes existing technologies and the patient population with a particular indication. This may underestimate opportunity losses in two ways. First, information about existing technologies will be relevant to the evaluation of new future technologies when they become available (the existing technology maybe an important comparator and may also contribute to the network of evidence required to estimate relative effect of a new technology). Second, evidence about a technology for one indication may well be relevant when considering a future extension of the licensed indication (e.g. patient group or the sequence and combination of therapy). Forgoing evidence now will clearly impact on these future decisions. Increasingly adoption and reimbursement decision are being made earlier in the life cycle of health technologies and closer to licence. At this point, often the only evidence to support the use are licensing trials and the body of evidence from both publicly funded and sponsored research has not matured. Therefore, earlier adoption is likely to incur greater opportunity losses and failure to account for this may have a significant impact on the future evidence base for clinical practice.

The possibility of making reimbursement or approval decisions conditional on evidence being provided (i.e. coverage with evidence development), (Hutton *et al.*, 2007) or negotiating some form of risk sharing agreement (An OFT market study, 2007; Cooksey, 2006) will need to consider when additional evidence is needed, the type of evidence required and whether this can or will be gathered once the technology is approved for use. For example, the type of observational registry data that are often envisaged will be unable to provide more precise estimates of relative treatment effect because a comparable control group will not be available. Therefore, setting aside the question of whether manufacturers or the public sector should bear the costs of research, the opportunity loss of adoption will depend not only on the overall value of information for the decision, but also on whether the value is associated types of evidence that will require particular research designs.

Failure to take account of the value of evidence that maybe forgone particularly by early adoption of technologies is a dangerous omission and may lead to the approval and diffusion of technologies for which current evidence is insufficient. If reimbursement authorities are not able to commission or demand that research be conducted then it may be better to deny the approval of a technology which is expected to be cost-effective and issue an 'only in research recommendation'. If not, adoption and reimbursement decisions may undermine the evidence base for future clinical practice and reduce expected net health benefits for the population of current and future patients.

ACKNOWLEDGEMENTS

This work was funded by a programme grant from the UK Medical Research Council as part of the Health Services Research Collaboration. None of the authors have any conflicts of interest to report.

An earlier version of this paper was presented at the Health Economists Study Group meeting held in York, July 2006. We would like to acknowledge helpful comments from two anonymous referees and the helpful discussion and comments from colleagues. We would particularly thank Anirban Basu for insightful and valuable comments on an earlier version of this paper.

DOI: 10.1002/hec

REFERENCES

Ades AE, Lu G, Claxton K. 2004. Expected value of sample information calculations in medical decision modeling. *Medical Decision Making* **24**: 207–227.

An OFT market study. 2007. *The Pharmaceutical Price Regulation Scheme*. Office of Fair Trading: London.

Chalkidou K, Hoy A, Littlejohns P. 2007. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *Journal of the Royal Society of Medicine* **100**: 453–460.

Chalmers I. 2007. Addressing uncertainties about the effects of treatments offered to NHS patients: whose responsibility? *Journal of the Royal Society of Medicine* **100**: 440–441.

Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* **18**: 342–364.

Claxton K. 2007. OFT, VBP: QED? *Health Economics* **16**: 545–558.

Cooksey D. 2006. *A Review of UK Health Research Funding*. HM Treasury: London.

Eckermann S, Willan A. 2008a. Time and EVSI wait for no patient. *Value Health* **11**: 522–526.

Eckermann S, Willan AR. 2008b. The option value of delay in health technology assessment. *Medical Decision Making* **28**: 300–305.

Eckermann S, Willan AR. 2009. Globally optimal trial design for local decision making. *Health Economics* **18**: 203–216.

Griffin S, Welton N, Claxton K. 2009. Exploring the research decision space: the expected value of information for sequential research designs. *Medical Decision Making*, DOI: 10.1177/0272989X09344746.

Hutton J, Trueman P, Henshall C. 2007. Coverage with evidence development: an examination of conceptual and policy issues. *International Journal of Technology Assessment in Health Care* **23**: 425–435.

Main C, Ginnelly L, Griffin S, Norman G, Barbieri M, Mather L, Stark D, Palmer S, Riemsma R. 2006. Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer. *Health Technology Assessment* **10**: i–148.

NICE. 2004. *Guide to the Technology Appraisal Process*. National Institute for Clinical Excellence: London.

NICE Citizens Council. 2007. *Only in Research*. National Institute for Health and Clinical Excellence: London.

Palmer S, Smith PC. 2000. Incorporating option values into the economic evaluation of health care technologies. *Journal of Health Economics* **19**: 755–766.

Phelps CE, Mushlin AI. 1991. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care* **7**: 12–21.

Phillips Z, Claxton K, Palmer S. 2008. The half-life of truth: what are appropriate time horizons for research decisions? *Medical Decision Making* **28**: 287–299.

Raiffa H, Schlaiffer R. 1967. *Applied Statistical Decision Theory*. Wiley Interscience: New York.

Stinnett AA, Mullahy J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* **18**(suppl): S68–S80.

Willan AR, Pinto EM. 2005. The expected value of information and optimal clinical trial design. *Statistics in Medicine* **24**: 1791–1806.

Glossary

ENBS	Expected net benefit of sampling information
EVPI	Expected value of perfect information
EVSI	Expected value of sample information
HRQL	Health related quality of life
IV	Instrumental variable
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
RCT	Randomised controlled trial

References

1. Griffin SC, Barber JA, Manca A, Sculpher MJ, Thompson SG, Buxton MJ, et al. Cost effectiveness of clinically appropriate decisions on alternative treatments for angina pectoris: prospective observational study. *British Medical Journal* 2007;334(7594):624.
2. Griffin SC, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using Bayesian methods: an applied case study in ovarian cancer. *Value in Health* 2006;9(2):123-131.
3. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic analysis and computationally expensive models: necessary and required? *Value in Health* 2006;9(4):244-252.
4. Griffin S, Welton N, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs *Medical Decision Making* 2010;30(2):155-162.
5. Griffin S, Claxton K, Palmer S, Sculpher M. Dangerous Omissions: The Consequences of Ignoring Decision Uncertainty. *Health Economics* 2010;*Published online in Wiley Interscience*.
6. Williams A. Cost-benefit analysis: bastard science? And/or insidious poison in the body politick? *Journal of Public Economics* 1972;1:199-225.
7. Williams A. The cost-benefit approach. *British Medical Bulletin* 1974;30(3):252-256.
8. Department of Health. The NHS Constitution: the NHS belongs to us all. London: Department of Health, 2010.
9. Wanless D. Securing good health for the whole population. London: HM Treasury, 2004.
10. Culyer AJ. The normative economics of health care finance and provision. *Oxford Review of Economic Policy* 1989;5:34-58.
11. Claxton K, Sculpher M, Culyer AJ. Mark versus Luke? Appropriate methods for the evaluation of public health interventions. *CHE research paper* 2007;31.
12. Brouwer WBF, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics* 2000;19:439-459.
13. Smith RD, Richardson J. Can we estimate the 'scoial' value of a QALY? Four core issues to resolve. *Health Policy* 2005;74:77-84.
14. Sugden R, Williams A. *The Principles of Practical Cost-Benefit Analysis*. Oxford: Oxford University Press, 1979.
15. Epstein D, Chalabi Z, Claxton K, Sculpher M. Efficiency, equity and budgetary policies: informing decisions using mathematical programming. *Medical Decision Making* 2007;27:128-137.
16. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996;15(5):641.
17. Birch S, Gafni A. Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be? *Journal of Health Economics* 1992;11:279-296.
18. Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. *Canadian Medical Association Journal* 1993;148(6):913-917.
19. Johannesson M, Weinstein M. On the decision rules of cost-effectiveness analysis. *Journal of Health Economics* 1993;12:459-467.
20. Appleby J, Devlin N, Parkin D, Buxton MJ, Chalkidou K. Searching for cost effectiveness thresholds in the NHS. *Health Policy* 2009;91:239-245.
21. McCabe C, Claxton K, Culyer AJ. The NICE Cost-effectiveness Threshold. What it is and What that Means. *Pharmacoeconomics* 2008;26(9):733-744.
22. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *International Journal of Technology Assessment in Health Care* 1989;5:559-575.

23. Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986;5:1-30.
24. Torrance GW, Thomas WH, Sackett DL. A Utility Maximization Model for Evaluation of Health Care Programs. *Health Services Research* 1972;7(2):118-133.
25. Raiffa H, Schlaifer R. *Probability and Statistics for Business Decisions*. New York: McGraw Hill, 1959.
26. Yokota F, Thompson KM. Value of information literature analysis: a review of applications in health risk management. *Medical Decision Making* 2004;24:287-298.
27. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 1998;18:95-109.
28. Claxton K, Posnett J. An economic approach to clinical trial design and research priority setting. *Health Economics* 1996;5:513-524.
29. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 1999;18:342-64.
30. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. *Journal of Health Economics* 2000;19:755-766.
31. Eckermann S, Willan A. Expected value of information and decision making in HTA. *Health Economics* 2006;16(2):195-209.
32. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines - similarities, differences and some implications. *Value in Health* 2001;4:225-250.
33. Tarn TYH, Smith MD. Pharacoeconomic guidlines around the world. *ISPOR connections* 2004;Aug 15 2004.
34. Rubin DB. Estimating the causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974;66(5):688-701.
35. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
36. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 1957;54:297-312.
37. Fisher RA. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain* 1925;33:505-513.
38. Cronbach LJ. *Designing evaluations of educational and social programes*. San Francisco: Jossey-Bass, 1982.
39. Sackett DL. Inference and decision at the bedside. *Journal of Clinical Epidemiology* 1989;42(4):309-316.
40. Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technology Assessment* 2005;9(38):i-168.
41. Black N. Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal* 1996;312:1215.
42. Griffin S, Rice N, Sculpher M. Economic evaluation of public health interventions. In: Killoran A, Kelly M, editors. *Evidence-based public health: effectiveness and efficiency*. Oxford: Oxford University Press, 2009.
43. Sekhon JS. Opiates for the Matches: Matching Methods for Causal Inference. *Annual Review of Political Science* 2009;12:487-508.
44. Rosenbaum P, Rubin DB. The central role of propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
45. Newhouse JP, McClennan M. Econometrics in outcomes research: the use of instrumental variables. *Annual Review of Public Health* 1998;19:17-34.

46. Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;47(1):153-161.
47. Timmins N. NHS goes to the PROMS. *BMJ* 2008;336(7659):1464.
48. Brazier J, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics* 2010;11(2):215-225.
49. Petrillo J, Cairns J. Converting condition-specific measures into preference-based outcomes for use in economic evaluation. *Expert Reviews of Pharmacoeconomics & Outcomes Research* 20080;8(5):453-461.
50. Mortimer D, Segal L. Comparing the Incomparable? A Systematic Review of Competing Techniques for Converting Descriptive Measures of Health Status into QALY-Weights. *Medical Decision Making* 2008;28(1):66-89.
51. Chuang L-H, Kind P. Converting the SF-12 into the EQ-5D: An Empirical Comparison of Methodologies. *Pharmacoeconomics* 2009;27(6):491-505.
52. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons, 1987.
53. Schafer JL. *Analysis of Incomplete Multivariate Data.* New York: Chapman & Hall, 1997.
54. Schafer JL. Multiple Imputation: A Primer. *Statistical Methods in Medical Research* 1999;8:3-15.
55. Hemingway H, Crook AM, Dawson JR, Edelman J, Edmondson S, Feder G, et al. Rating the appropriateness of coronary angiography, coronary angioplasty and coronary artery bypass grafting: the ACRE study. *Journal of Public Health Medicine* 1999;21(4):421-429.
56. Kind P. The EuroQoL instrument: an index of health-related quality of life. In: Spilker B, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2nd ed. Philadelphia: Lippincott-Raven, 1996.
57. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999;18(6):681-694.
58. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *International Journal of Technology Assessment in Health Care* 1986;2(1):53-63.
59. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics* 2004;13(5):461-475.
60. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology & Community Health* 2002;56:119-127.
61. Rutten F. Economic evaluation and health care decision-making. *Health Policy* 1996;36(3):215-229.
62. Victoria CG, Habicht J-P, Bryce J. Evidence-Based Public Health: Moving Beyond Randomized Trials. *American Journal of Public Health* 2004;94:400-405.
63. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics* 2006;15(7):677-687.
64. Dobrow MJ, Goel V, Upshur REG. Evidence-based health policy: context and utilisation. *Social Science & Medicine* 2004;58:207-217.
65. Mulrow CD. Systematic Reviews: Rationale for systematic reviews. *British Medical Journal* 1994;309:597.
66. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technology Assessment* 1998;2(19).

67. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002;21(4):589-624.
68. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research*. London: John Wiley, 2000.
69. Glass GV. Primary, secondary and meta-analysis of research. *Educational Research* 1976;5:3-8.
70. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007;7(3).
71. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 1997;50(6):683-691.
72. Hasselblad V. Meta-analysis of Multitreatment Studies. *Medical Decision Making* 1998;18:37-43.
73. Ades AE, Cliffe S. Markov Chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Medical Decision Making* 2002;22:359-371.
74. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996;15:2733-2749.
75. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted. *Statistics in Medicine* 2002;21(11):1559-1573.
76. Rothstein HR, Sutton AJ, Borenstein M. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester: John Wiley & Sons, Ltd, 2006.
77. NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Clinical Excellence, 2008.
78. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2009.
79. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004;23(20):3105-3124.
80. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *Journal of Clinical Epidemiology* 2008;61(5):455-463.
81. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008;17(3):279-301.
82. Ades AE, Sculpher MJ, Sutton AJ, Abrams KR, Cooper N, Welton N, et al. Bayesian Methods for Evidence Synthesis in Cost-Effectiveness Analysis *Pharmacoeconomics* 2006;24(1):1-19.
83. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Resaerch* 2001;10(4):277-303.
84. Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics in Medicine* 2000;19(24):3359-3376.
85. Briggs A, Sculpher MJ, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006.
86. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *British Medical Journal* 1996;312:71-72.
87. Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2006;169(1):5-35.

88. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health* 2003;6(1):9-17.
89. Main C, Ginnelly L, Griffin S, Norman G, Barbieri M, Mather L, et al. A systematic review of the clinical effectiveness and cost effectiveness of topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer (up-date). *Health Technology Assessment* 2006;10(9):i-292.
90. Philips Z, Ginnelly L, Sculpher MJ, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004;8(36):i-158.
91. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 2006;15(12):1295-1310.
92. Akehurst R, Anderson PD, Brazier J, Brennan A, Briggs A, Buxton MJ, et al. Decision Analytic Modelling in the Economic Evaluation of Health Technologies: A Consensus Statement. *Pharmacoeconomics* 2000;17(5):443-444.
93. Box GEP, Tiao GC. *Bayesian inference in statistical analysis*. Reading: Addision-Wesley, 1973.
94. Cooper K, Brailsford SC, Davies R. Choice of modelling techniques for evaluating health care interventions. *Journal of the Operational Research Society* 2006;58:168-176.
95. Gass SI. Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research* 1983;31(4):603-631.
96. Snee RD. Validation of Regression Models: Methods and Examples. *Technometrics* 1977;19(4):415-428.
97. Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. *Health Economics* 2010;19(1):43-55.
98. Sculpher M, Fenwick E, Claxton K. Assessing Quality in Decision Analytic Cost-Effectiveness Models: A Suggested Framework and Example of Application *Pharmacoeconomics* 2000;17(5):461-477.
99. Quade ES. Pitfalls in modeling. In: Majone G, Quade ES, editors. *Pitfalls of Analysis*. Chichester: John Wiley & Sons, 1980:31-42.
100. Karnon J, Brennan A, Akehurst R. A Critique and Impact Analysis of Decision Modelling Assumptions. *Medical Decision Making* 2007;27(4):491-499.
101. Claxton K. Exploring Uncertainty in Cost-effectiveness Analysis. *Pharmacoeconomics* 2008;26(9):781-798.
102. O'Hagan A, Buck CE, Daneshkhah A, Eiser JE, Garthwaite PH, Jenkinson DJ, et al. *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: Wiley, 2006.
103. Howard R. Decision Analysis: Practice and Promise. *Management Science* 1988;34(6):679-695.
104. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision-analytic models: a review and application of methods. *Value in Health* 2009;12(739-49).
105. Eschenbach TG. Spiderplots versus Tornado Diagrams for Sensitivity Analysis. *Interfaces* 1992;22(6):40-46.
106. Hamby DM. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment* 1994;32(2):135-154.
107. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* 1997;6:327-340.

108. Critchfield GC. Probabilistic Analysis of Decision Trees Using Monte Carlo Simulation. *Medical Decision Making* 1986;6(2):85-92.
109. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Medical Decision Making* 1985;5(2):157-177.
110. Iman RL. Latin Hypercube Sampling. In: Melnick EL, Everitt BS, editors. *Encyclopedia of Quantitative Risk Analysis and Assessment*: John Wiley & Sons, 2008.
111. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics* 2001;10:779-789.
112. Draper D. Assessment and Propogation of Model Uncertainty. *Journal of the Royal Statistical Society: Series B* 1995;57(1):45-97.
113. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian Model Averaging: A Tutorial. *Statistical Science* 1999;14(4):382-417.
114. Claxton K, Sculpher MJ, McCabe C, Briggs A, Akehurst R, Buxton MJ, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* 2005;14(4):339-347.
115. Koerkamp BG, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink M. Uncertainty and Patient Heterogeneity in Medical Decision Models. *Medical Decision Making* 2010;30:194-205.
116. Drummond M. Evaluation of health technology: Economic issues for health policy and policy issues for economic appraisal. *Social Science & Medicine* 1994;38(2):1593-1600.
117. Iman RL, Helton JC. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis* 1988;8(1):71-90.
118. Downing DJ, Gardner RH, Hoffman FO. An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models. *Technometrics* 1985;27(2):151-163.
119. O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. *Health Economics* 2007;16(10):1009-1023.
120. Pratt J, Raiffa H, Schlaifer R. Statistical decision theory. Cambridge M: MIT Press, 1995.
121. Hoffman FO, Hammonds JS. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis* 1994;14:707-712.
122. Chilcott J, Brennan A, Booth A, Karnon J. The role of modelling in prioritising and planning clinical trials. *Health Technology Assessment* 2003;7(23):i-138.
123. Claxton K, Eggington S, Ginnelly L, Griffin S, McCabe C, Philips Z, et al. A pilot study of Value of Information analysis to support research recommendations for the National Institute for Clinical Excellence. York: Centre for Health Economics. Research Paper 4., 2005:Chapter 4.
124. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS health technology assessment programme. *Health Technology Assessment* 2003:Chapter 6.
125. Eckermann S, Karnon J, Willan AR. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics* 2010;28(9):699-709.
126. Welton NJ, Ades AE, Caldwell DM, Peters TJ. Research Prioritisation Based on Expected Value of Partial Perfect Information: a Case Study on Interventions to Increase Uptake of Breast Cancer Screening. *Journal of the Royal Statistical Society: Series A* 2008;171:807-841.

127. Fleurence RL, Torgerson DJ. Setting priorities for research. *Health Policy* 2004;69(1):1-10.
128. Philips Z, Claxton K, Palmer S. The half-life of truth: appropriate time horizons for research decisions? *Medical Decision Making* 2008;28:287-299.
129. Sculpher M, Claxton K. Establishing the Cost-Effectiveness of New Pharmaceuticals uncer Conditions of Uncertainty - When Is There Sufficient Evidence? *Value Health* 2005;8(4):433-446.
130. Ades AE, Lu G, Claxton K. Expected value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making* 2004;24:207-227.
131. Brennan A, Kharroubi SA, O'Hagan A, Chilcott J. Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms. *Medical Decision Making* 2007;27(4):448-470.
132. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *The European Journal of Health Economics* 2008;9(3):251-259.
133. Dakins M. The Value of the Value of Information. *Human and Ecological Risk Assessment* 1999;5(2):281-289.
134. Raiffa H, Schlaifer RO. *Applied Statistical Decision Theory*: Division of Research, Harvard Business School, 1961.
135. Thompson KM, Graham JD. Going beyond a single number: Using probabilistic risk assessment to improve risk management. *Human and Ecological Risk Assessment* 1996;2(4).
136. Conti S, Claxton K. Dimensions of design space: a decision-theoretic approach to optimal research design. York: Centre for Health Economics, 2008.
137. Willan AR, Pinto EM. The expected value of information and optimal clinical trial design. *Statistics in Medicine* 2005;24:1791-1806.
138. Brennan A, Kharroubi SA. Expected value of sample information for Weibull survival data. *Health Economics* 2007;16(11):1205-1225.
139. Cooksey D. A review of UK health research funding. London: HM Treasury, 2006.
140. Willan A, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Economics* 2010;19(5):549-561.
141. Fenwick E, Claxton K, Sculpher M. The Value of Implementation and the Value of Information: Combined and Uneven Development. *Medical Decision Making* 2008;28(1):21-32.
142. Banta HD, Thacker SB. The case for reassessment of health care technology: once is not enough. *Journal of the American Medical Association* 1990;264:235-240.
143. Sculpher MJ, Drummond MF, Buxton MJ. The iterative use of economic evaluation as part of the process of health technology assessment. *Journal of Health Services Research and Policy* 1997;2:26-30.
144. Towse A, Garrison LP. Can't Get No Satisfaction? Will Pay for Performance Help?: Toward an Economic Framework for Understanding Performance-Based Risk-Sharing Agreements for Innovative Medical Products. *Pharmacoecon* 2010;28(2):93-102.
145. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *Journal of the Royal Society of Medicine* 2007;100:453-460.
146. Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *International Journal of Technology Assessment in Health Care* 2007;23:425-435.

147. McCabe CJ, Stafinski T, Edlin R, Menon D. Access with Evidence Development Schemes: A Framework for Description and Evaluation *Pharmacoeconomics* 2010;28(2):143-152.
148. Drummond M, Sculpher M, Torrence G, O'Brien B, Stoddart G. *Methods for the economic evaluation of health care programmes*. Third ed. Oxford: Oxford University Press, 2005.
149. Eckermann S, Willan AR. The option value of delay in health technology assessment. *Medical Decision Making* 2008;28:300-305.
150. Chalkidou K, Lord J, Fischer A, Littlejohns P. Evidence-Based Decision Making: When Should We Wait For More Information? *Health Affairs* 2008;27(6):1642-1653.
151. Eckermann S, Willan A. Time and EVSI wait for no patient. *Value in Health* 2008;11:522-526.
152. Eckermann S, Willan AR. Globally optimal trial design for local decision making. *Health Economics* 2009;18(2):203-216.
153. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technology Assessment* 2004;8(49).
154. Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: A taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy* 2010;96(3):179-190.
155. McCabe C, Chilcott J, Claxton K, Tappenden P, Cooper C, Roberts J, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified *BMJ* 2010;340:1786.
156. Raftery J. Multiple sclerosis risk sharing scheme: a costly failure. *BMJ* 2010;340:1672.