THE UNIVERSITY of York

Defining and Characterising Structural Uncertainty in Decision Analytic Models

CHE Research Paper 9

Defining and Characterising Structural Uncertainty in Decision Analytic Models

Laura Bojke
Karl Claxton
Stephen Palmer
Mark Sculpher

March 2006

Background

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The new CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Disclaimer

Papers published in the CHE Research Paper (RP) series are intended as a contribution to current research. Work and ideas reported in RPs may not always represent the final position and as such may sometimes need to be treated as work in progress. The material and views expressed in RPs are solely those of the authors and should not be interpreted as representing the collective views of CHE research staff or their research funders.

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/inst/che/pubs. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321458 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/inst/che

Abstract

An inappropriate structure for a decision analytic model can potentially invalidate estimates of cost-effectiveness and estimates of the value of further research. However, there are often a number of alternative and credible structural assumptions which can be made. Although it is common practice to acknowledge potential limitations in model structure, there is a lack of clarity about methods to characterize the uncertainty surrounding alternative structural assumptions and their contribution to decision uncertainty.

A review of decision models commissioned by the NHS Health Technology Programme was undertaken to identify the types of model uncertainties described in the literature. A second review was undertaken to identify approaches to characterise these uncertainties.

The assessment of structural uncertainty has received little attention in the health economics literature. A common method to characterise structural uncertainty is to compute results for each alternative model specification, and to present alternative results as scenario analyses. It is then left to decision maker to assess the credibility of the alternative structures in interpreting the range of results.

The review of methods to explicitly characterise structural uncertainty identified two methods: 1) model averaging, where alternative models, with different specifications, are built, and their results averaged, using explicit prior distributions often based on expert opinion and 2) Model selection on the basis of prediction performance or goodness of fit. For a number of reasons these methods are neither appropriate nor desirable methods to characterize structural uncertainty in decision analytic models.

When faced with a choice between multiple models, another method can be employed which allows structural uncertainty to be explicitly considered and does not ignore potentially relevant model structures. Uncertainty can be directly characterised (or parameterised) in the model itself. This method is analogous to model averaging on individual or sets of model inputs, but also allows the value of information associated with structural uncertainties to be resolved.

1. Introduction

Decision-analytic models are now established practice in formal decision-making processes.[1-3] One of the requirements for decision-making is that uncertainty regarding the adoption decision must be appropriately characterised and quantified. Uncertainty in decision analytic models presents in many forms.[4] However, the issue of parameter uncertainty is the best researched with other types of uncertainty receiving less attention in the Health Technology Assessment (HTA) literature, of particularly structural uncertainty which remains the most illusive class of uncertainty to define.

An inappropriate structure for a decision analytic model can potentially invalidate estimates of cost-effectiveness and estimates of the value of further research. Although it is common practice to acknowledge potential limitations in model structure, there is a lack of clarity about methods to characterize the uncertainty surrounding alternative structural assumptions and their contribution to decision uncertainty. This paper reviews the ways in which structural uncertainty has been described in the Health Technology Assessment (HTA) literature and identifies methods that have been employed to characterise these types of structural uncertainties. The results of these searches are presented and the potential relevance of these methods for the HTA decision-making process is assessed.

2. Current perspectives for decision-making in the UK healthcare system

Evidence-based medicine has been at the forefront of healthcare decision-making since the thalidomide scandal in the 1960s.[5] There is stringent regulation on the licensing of pharmaceuticals, with an emphasis on evidence of efficacy and safety. Decisions about the reimbursement of pharmaceuticals are increasingly informed by formal assessment of cost-effectiveness.[6]

The increased emphasis on formal methods of evidence-based medicine[5] has lead to the development of the guidelines industry. Guidelines were originally intended as 'clinical policy' documents, however, they now act as much wider decision-making tools for the individual practitioner/patient consultation. Guidelines, such as those produced by the National Institute for Health and Clinical Excellence (NICE), can also used to allocate scarce resources and act as an efficiency tool.[7]

Spiralling costs and formal audit of clinical services has lead to an increased emphasis on explicit consideration of the budget constraints apparant in the modern UK National Health Servive (NHS). As part of this, NICE was established in 1999 to provide guidance on new and existing interventions. Since 2001 the guidance issued by NICE has been made compulsory.[5, 8]

Within the context of informing NICE decisions, the NHS Health Technology Assessment Programme (HTA) commissions reports from independent assessment teams. These assessment teams come from a small number of academic units and all have recognised capacity and expertise in secondary research.[9] Because of its role in informing public policy there is a requirement for NICE assessment teams to employ robust methodology in the evaluation of technologies.[10]

3. Requirements for the decision-making process

In order to inform a decision-making process, an evaluation must address two questions; 1) is a technology cost-effective based on current evidence? and 2) would further research represent good value for money?[10]

To address the first of these two questions, the analytical framework needs to have particular characteristics, specifically:

- Comparison of the new technology with all relevant comparisons. For many technologies this should include a 'do nothing' strategy.

- A consistent perspective on costs and benefits. Many have argued for a societal perspective; however for many decision-makers, such as NICE, a third party or payer perspective is adopted e.g. the NHS.
- A clear objective function must be specified. For NICE the objective is to maximise health (represented by quality-adjusted life-years (QALYs)) given the budget constraint.
- All relevant evidence must be incorporated.

The second question requires that any uncertainty regarding an adoption decision must be explicitly characterised. This then provides a route to quantifying the cost of making a wrong decision and hence the value of acquiring additional information

In many circumstances trial data alone is of limited value in meeting these requirements.[11] Phase III trials usually compare to placebo or 'do-nothing' whereas cost-effectiveness can only be established by comparing the new intervention to all forms of current practice. Phase III trials are also usually quite short and, therefore, cannot be used to measure costs and effects over an adequate time period without any form of extrapolation. It is also unlikely that there is only one trial available that is relevant to address a given decision problem. Where there are multiple sources of data, methods to synthesise this data must be employed in order to incorporate all relevant evidence.

4. Role of decision analytic modelling in decision-making

Decision analysis has previously been described as a method to "guide the decision-maker to compare relevant alternatives and select the most appropriate solution for the decision situation".[12] In the context of healthcare decision-making, decision analysis provides a statistical or mathematical process that brings together multiple sources of evidence on a range of parameters in order to quantify the costs and outcomes of all possible competing interventions.

Decision analysis provides a systematic and explicit approach to decision-making under conditions of uncertainty.[13] It is intended to assist conventional decision-making rather than replace it. Models can also help to avoid subjective decisions, either about the data within the model or concerning any model uncertainties.[14] The majority of decisions about the cost effectiveness of an intervention are based on uncertain information. A decision model can explicitly represent this uncertainty and quantify it through the use of probabilistic sensitivity analysis (PSA).[11]

PSA can be conducted for as many uncertain parameters as the model may contain, concurrently.[15,16] Uncertainty about a parameter is represented by a probability distribution. The choice of distribution is informed by the type and shape of the data observed, for example cost data are usually represented by gamma distributions and probability data by beta distributions.[17, 18] Markov Chain Monte Carlo (MCMC) simulation methods[15,16] are then used to simulate the expected costs and outcomes of the various interventions, by sampling from the distributions that feed into the estimation. In this way, PSA enables uncertainty surrounding all input parameters to be propagated through the decision model in order that uncertainty surrounding the decision itself can be quantified. Since probabilistic analysis allows the quantification of the probability of making the wrong decision, these results can also be used to estimate the expected costs of the uncertainty surrounding a decision based on current evidence. These costs can thus be interpreted as the expected value of perfect information (EVPI), since having perfect information can eliminate the possibility, and hence costs, of making the wrong decision.[19]

Because of the limitations of trial based analysis and the requirements of the decision-making process, decision analytic models are increasingly being used to guide policy decisions regarding the optimum allocation of health care expenditures.[10] The process of decision analytic modelling is now seen as central to the process of HTA in general, and it plays a key role in the NICE technology appraisal process.[10] In the period 2004-2005, 39 out of 61 NICE appraisals considered an independent decision analytic model (in addition to manufacturer models) as evidence to inform the decision-making process.

5. Issues of uncertainty in decision analytic models

Although analysts seek to develop models and incorporate data that most accurately inform the costs and outcomes associated with a particular disease and intervention, some degree of uncertainty is present in the majority of models; uncertainty about the true parameter values, the type of model used and the model results. More formally the dimensions of uncertainty have been categorised as parameter, heterogeneity, methodological and structural.[4]

Within health technology assessment, analysts have tended to focus almost entirely on quantifying and assessing the impact of parameter uncertainty. As a consequence, methods for dealing with parameter uncertainty are well-developed and are becoming an integral part of current best practice.[20] Issues of heterogeneity have similarly been discussed at length in the HTA literature and it is common practice to employ methods to adjust for patient characteristics, geographical location or to compute cost-effectiveness results for particular sub-groups of patients or locations.[21] Many issues of methodological uncertainty have also been resolved by guidelines encouraging the harmonisation of economic evaluation techniques.[22] Structural uncertainty, however, has received relatively little attention, although many guidelines for good practice in decision modelling recognize the need to explore structural assumptions [23-25] and the evidence supporting the chosen model structure.

6. Defining structural uncertainty

Structural uncertainty remains the most problematic class of uncertainty to define. Indeed, it is often simply used to classify those types of uncertainties that do not easily fit into the categories of parameter, methodological or heterogeneity.

In order to examine how structural uncertainty has been described and resolved in the HTA literature, a review of decision models commissioned by the NHS HTA programme from 1997 to 2005 was undertaken (www.hta.org)

Of the 241 HTA reports that were published in the period 1997-2005 (August), 90 (37%) include some form of decision analytic model. Of these 90 reports, only 14 (15.5%) suggest or discuss issues of structural uncertainty. The type of structural uncertainty and methods employed in each of these reports are shown in Table 1 below.

The discussion of structural uncertainty has been somewhat limited in the models reviewed here. Despite these limitations it is evident that models produced more recently seem more likely to have formally considered issues related to structural uncertainty. Indeed, 12 (85%) of the 14 reports were published between 2004 and 2005. This could reflect the fact that the issue of structural uncertainty is gaining more exposure in HTA generally.

The structural uncertainties discussed in the 12 reports were: the length of treatment effect, time horizon of the model, inclusion of specific events or comparators and statistical issues relating to modelling survival. Thus, on the basis of this review we can classify structural uncertainty of the following general types: (i) inclusion/exclusion of potentially relevant comparators; (ii) inclusion/exclusion of potentially relevant events; (iii) statistical models to estimate specific parameters and (iv) clinical uncertainty or lack of clinical evidence. Each of these types is described in more detail below.

Table 1: Structural uncertainty described in HTA decision models

HTA Report	Type of structural uncertainty	Methods employed to characterise uncertainty
Cuzick, 1999[26]	Alternative assumptions regarding parameters relating to HPV test	Alternative model scenarios were presented
Berry, 2002[27]	Alternative assumptions regarding correlation between sensitivity and specificity and time frame of the model	
Clegg, 2003[28]	Describes testing for the robustness of the model structure, but does not undertake any analysis	
Garside, 2004[29]	Different life times of the model	Alternative model scenarios were presented
Kaltenhaler, 2004[30]	Alternative assumptions regarding background risk and LFT results	
Claxton, 2004[31]	Alternative assumptions regarding effectiveness of screening	Scenarios not formally presented
Jones, 2004[32]	Alternative assumptions about the length of treatment and inclusion/exclusion of events	Alternative model scenarios were presented
Main, 2004[33]	Different life times of the model and length of treatment	
Green, 2005[34]	Alternative assumptions about QOL and long term costs	
McCormack, 2005[35]	Alternative assumptions about method of surgical repair and additional effect of complications	
Stevenson, 2005[36]	Including/excluding fractures not clinically identified	
Wilson, 2005[37]	Alternative assumptions regarding correlation between survival curves	
Robinson, 2005[38]	Inclusion/exclusion of additional strategies	
Tillin, 2005[39]	Different life times of the model, replacement and failure rates.	

6.1 Inclusion/exclusion of relevant comparators

Guidelines on good modelling practice advocate the use of a broad range of feasible mutually exclusive strategies.[40] The selection of comparators should be informed by current evidence or opinion and, if relevant, should include a 'do nothing' strategy.[40]

In reality the choice of comparators is often governed by the scope of the model and the analyst may not necessarily choose this scope. The scope is often a direct consequence of the question(s) specified by the decision-maker. If the question is "to compare intervention A and intervention B, for the treatment of disease X", then this implies a much narrower scope of comparators than the question "compare all interventions available for the treatment of disease X".

The construction of a model should always bear in mind its intended audience or customer,[40] however when reviewing all available evidence to inform a decision, a decision-maker (such as NICE) may be faced with 2 alternative models with different sets of comparators. In some situations the most cost-effective option may be different according to which model is believed to be accurate. Even if the excluded comparators are not cost-effective, excluding them will change EVPI estimates.

The model by Robinson et al[38] explored the inclusion of additional potentially relevant strategies as alternate scenarios. Although inclusion of additional strategies did not affect the base case results, when combined with other sensitivity analysis, the inclusion of additional strategies did change the adoption decision. Unfortunately in this model EVPI was not calculated, although one can suppose that these alternate scenarios would have produced quite different EVPI estimates.

6.2 Inclusion/exclusion of relevant events – related/unrelated events

All decision models are simplifications[41] of an actual disease and healthcare consumption process. Simplification is a necessity in order for the modelling process to inform the decision-maker in a timely manner. However, the process of simplification will inevitably require certain assumptions to be made (i.e. the extent to which potential events can be ignored since they are unlikely to differ between interventions). These assumptions should be supported by evidence and choices between alternative assumptions should be justified and made transparent.[40]

A key part of the development of any model is the decision about which events, stages or health states should be included or excluded from the model. It is also important to recognise that the information (current evidence or clinical opinion) used as the basis for these decisions is not static, and new information relevant to the model development may arise at a later stage. In certain instances a decision may have been made to exclude a particular event on the basis that it was deemed to be unrelated to the interventions under consideration at the time the model was being developed. However, it may be that subsequent evidence becomes available that contradicts this assumption. It would then be important to determine whether the inclusion of this event would alter the results from the model.

In the model by Jones, et al[32] non vascular deaths were included and excluded from the sub-group models for stroke, myocardial infarction (MI), Transient Ischemic Attack (TIA) and Peripheral Artery Disease (PAD). Alternative assumptions about the length of treatment were also explored in scenario analysis. Including or excluding non vascular deaths had a profound effect on the cost-effectiveness results particularly for the TIA and stroke models assuming a lifetime treatment duration. In both subgroups ASA-MR-dipyridamole was the most cost effective strategy in the model excluding treatment effect on vascular deaths and dominated in the model including the treatment effect on vascular deaths. EVPI was not calculated in the models. However, like the previous example, because of the sensitivity of the adoption decision to the inclusion and exclusion of events, we can suppose that these alternate models would have produced quite different EVPI estimates.

6.3 Statistical models used to estimate specific parameters

Decision models are using increasingly sophisticated statistical techniques to derive estimates of parameters, particularly when estimates are available from more than one source. As an example, meta-analysis techniques[42] are used as part of many systematic review processes to combine evidence from multiple sources and to explore the impact of study heterogeneity. Two general types of meta-analysis model exist, the fixed or random effects model. The fixed effect model assumes that the included studies are all estimating a common effect size, that is we assume that there is no between-study heterogeneity. If there is reason to believe that the studies may differ, in terms of the underlying effect size that they are estimating, a random effects model may be more appropriate.[42] A random effects model, therefore, accounts for both within- and between-study variation.[42] In many circumstances fixed and random effects models will give the same result and we are uncertain about which is the most appropriate model to use to estimate treatment effect.

Another issue of statistical uncertainty is the estimation of correlation between parameters in the model. Where parameters in the model are interdependent, attempts should be made to correlate these within the model, for example sensitivity and specificity trade offs in diagnostic test accuracy.[43] In the model by Berry et al[27] scenarios were presented to explore just this issue. Pairs of sensitivity and specificity values were predicted from the clinical trial data and used in the model as opposed to the uncorrelated distributions. The cost-effectiveness results were not sensitive to the alternate structural assumptions. Neither decision uncertainty nor EVPI was calculated in this model, instead probabilistic analysis results were used to calculate confidence intervals for results.

In Wilson et al[37] a weakness in the industry model was identified; time to failure (TTF) and survival curves were independently calculated and that no attempt had been made to correlate the two curves. Scenarios were therefore presented allowing the TTF and survival curves to be correlated. Cost-effectiveness results were similar, with imatinib regarded as cost-effective after approx 5-years in both models. A further model scenario was presented which incorporated alternative estimates of survival as well as allowing TTF and survival curves to be correlated. In this scenario, imatinib is not regarded as cost-effective until 10-years and even then its cost-effectiveness is in the balance with an

incremental cost-effectiveness ratio of £29,789.[37] As with the model by Berry et al[27] neither decision uncertainty nor EVPI was calculated.

6.4 Clinical uncertainty or lack of clinical evidence

In many situations a decision model may be commissioned on the basis of a lack of clinical evidence (in particular RCT evidence) to inform a decision. In these circumstances some of the relationships in the model may not have data to inform them and expert opinion may be sought to enable the model to be developed. However, expert opinion, in the absence of sufficient clinical data to inform this, may be contradictory, that is 2 different clinical experts may present 2 different expert opinions. The HTA models by Tillin et al,[39] McCormack et al[35] and Garside et al[29] all explored issues of clinical uncertainty (length of treatment effect, strategy following recurrence, effect of complications) through the use of scenario analysis.

Garside, et al[29] concluded that the model results were relatively insensitive to alternative clinical assumptions. Likewise Tillin et al[39] concluded that alternative assumptions regarding relapse and replacement did not change the cost-effectiveness of the alternative strategies. Adopting a 5-year as opposed to 25-year time horizon did change the cost-effectiveness results considerably. Decision uncertainty and EVPI were not calculated for these two models and it is therefore unclear how sensitive these would be to alternative structural assumptions.

McCormack et al[35] found that changing the assumption about which type of surgical repair to use following relapse reduces the cost-effectiveness of the totally extrapentoneal (TEP) strategy slightly as does changing the assumptions about the effect of serious complications. Decision uncertainty is increased slightly for the alternative assumption regarding relapse but actually decreases when serious complications for operative mortality are included. EVPI was not calculated for these alternative scenarios, although one can suppose that given the limited effect on decision uncertainty, the value of further research to resolve these uncertainties would be minimal.

7. Methods used to characterise structural uncertainties in HTA models

Of the 14, two of the reports did not include any application of methods to characterise specific structural uncertainties.[28,31] Of the 12 reports in which structural uncertainties were analytically evaluated, methods were limited to running alternative scenarios representing the different assumptions or model structures. The differences between results from alternative scenarios and the base case models were then discussed. No attempt was made to combine estimates from the alternative models.

This method, emerged from the statistics literature.[44] In the absence of a ‘best’ model the analyst is required to compute results for each alternative model specification, and to present alternative results as sensitivity analyses.[45] The alternative models and their results are then presented to the decision-maker. Analysts may choose to present one particular model as the most likely or conservative estimate, which is usually stated as the base-case analysis. Alternatively the analyst may leave the decision-maker to choose which model he/she thinks is most credible. The limitations of each model will be made available to the decision-maker. The decision-maker will use this information to generate an implicit weight for each of the models. The model with the highest weight will be regarded as the ‘true’ model and all other models discarded. By discarding models, this method fails to include all relevant evidence, a requirement of the decision-making process.

Although this method can be useful, in as much as it illustrates the potential impact of structural uncertainties, there are a number of potential problems. The weights applied to each model are subject to the decision-makers’ interpretation of the limitations as presented. This process of interpretation is, on the whole, internalised, and as such is difficult to replicate given an alternative set of decision-makers. Related to this is the issue of multiple decision-makers contributing to a single decision, as is the case with the appraisal committee that make recommendations for NICE in the UK. Each member of the committee may apply different weights to each of the alternative models. There may be circumstances in which there is no agreement about which model has the most weight and thus represents the ‘best’ model.

Most importantly, by removing the uncertainty associated with choosing between multiple alternative models from the actual modelling process, presenting scenarios can offer little help in truly quantifying structural uncertainty. It cannot, therefore, inform decision-makers about the decision to undertake further research to resolve this uncertainty.[46] An assessment of alternative model structures must therefore quantify the uncertainty in a meaningful way.[46]

8. Review of methods to characterise structural uncertainty

Given the types of structural uncertainties that have been discussed in the HTA literature and the limitations of the scenario analyses that have been used to characterise these, methods were sought to explore these uncertainties in a more quantifiable and explicit manner. The characteristics and principles of the methods identified is discussed along with their performance in terms of meeting the requirements of the decision-making process.

8.1 Search strategies

A systematic search was conducted to look for papers relating to the identification, assessment or quantification of structural uncertainty in quantitative models, not just restricted to decision analytic models. As very little on structural uncertainty has been published in the health economics/HTA literature, the searches were not restricted to medical or economics databases. Because of the enormity of the available literature, as a result of widening the database search, searches were restricted to identifying key words in the title. In addition, citations from relevant papers were also obtained, this method is known as 'pearl growing'.[47] No date restrictions were placed on the searches. Only English language papers were included.

The search strategy used and databases searched are shown below:

Search strategy:

Science Citation Index 1980-2004:

model*)

and (robustness or uncertain* or forecast* or accura* or fit for purpose or predict* or inference)

and (structur* or develop* or
construct* or specification or selection)
[title]

Databases searched:

NHS Economic Evaluation Database (public version)
NHS Economic Evaluation Database (admin version)
EconLit
OHE Health Economic Evaluations Database
ScienceDirect

8.2 Papers included in the review

The systematic searches identified 40 potentially relevant papers. Although the searches were limited to key words contained within the title of a paper, a number of review papers were identified. The bibliographies of these papers were checked and potentially relevant articles were obtained. In total, 68 full papers were obtained and screened for inclusion in the review. Of these, 41 papers looked at methods to characterise and quantify structural uncertainty and thus were included in the review. A list of included and excluded studies and a brief summary of papers is available in Appendix 1.

8.3 Available methods to characterise structural uncertainty

As anticipated, aside from scenario analysis[45] very little has been carried out in the health economics literature to address the issue of structural uncertainty. The problem of structural uncertainty is often

regarded as secondary to parameter uncertainty, in which methods have developed significantly. General discussion of the causes of structural uncertainty is also absent from the health economics literature. Outside of HTA, in areas such as mathematics, statistics, operational research and environmental modelling, authors have recognised the potential errors arising from model uncertainty and the possibility that model results will be biased when analysts ignore such issues.[48-50] It is in these other disciplines that nearly all of the methods to deal with structural uncertainty have been developed (Appendix 1).[44] Two methods to address the issue of structural uncertainty are apparent, these are:

8.3.1 Weighting the plausibility of the structural assumptions using priors

Model averaging[51] is a method which has not been used in the health technology literature. This requires the analyst to build alternative models, with different assumptions, and then average across these models weighting each by the plausibility (prior)[52] of their assumptions. Weights are commonly based on expert opinion, or for statistical models more formal methods such as an analytic hierarchy process.[53] Although the combination of these types of models may not necessarily improve predictive performance in the true sense,[54] from a decision-making perspective it does help to deal with the issue of structural uncertainty more explicitly then simply presenting alternative scenarios. Decision-makers are not themselves faced with multiple models to which they have to attach subjective weights. In addition, all relevant evidence is considered.

Bayesian methods for model averaging appear to be well developed.[55] The problem of averaging across models can be viewed in a Bayesian sense as one in which a decision-maker needs to make the best possible use of information on a model structure he/she has available.[56] The most widely used method of Bayesian model averaging (BMA) works on the premise that given alternative ways of modelling an intervention effect or other parameter ($\mathrm{M}\kappa$), with Δ as the quantity of interest (such as net benefit), the posterior distribution of Δ given the data (*D*) is:

$$\Pr(\Delta|D) = \sum_{K=1}^{K} pr(\Delta|\mathrm{M}\kappa, D)\, pr(\mathrm{M}\kappa|D)$$

57

Thus, Δ is an average of the posterior distributions for each of the models considered, weighted by their posterior model probability.[57] In some circumstances, the number of models available can be unfeasibly large. Madigan and Raftery[58] proposed the Occam's window method to reduce the number of possible models to average across. If a particular model performs significantly worse (in terms of prediction performance) than the model which makes the best predictions, it should be discarded. This results in the exclusion of more complex models which are less well supported by the data than simpler models. However, methods that exclude models on the basis of poor performance do not meet the requirement to include all relevant evidence.

BMA methods have been applied frequently in the operational research and forecasting literature[59] and, more recently, when faced with alternative models to fit survival data[50,60,61,111,62] and environmental prediction.[63,64] The increased use of the technique (and indeed use of Bayesian analysis in general) is primarily due to the increased computer power now available to undertake MCMC.[65] When applying BMA techniques to HTA decision models, there is an issue of determining the posterior distribution of Δ given *D,* when *D* may not be available. It is often the case that a model is required in the absence of any real information on *D*.

Non Bayesian methods for model combination have also been proposed. Shannon and Banks[66] discuss the combination of classification trees using maximum likelihood estimation (MLE). In this method a central tree is defined and a distance parameter, representing the amount of rearrangement between alternative models. Using this distance parameter the MLE of the central classification tree is estimated. This MLE represents the 'best' tree structure.[66] As with BMA techniques, estimating the MLE for many of the structural uncertainty problems described in the HTA literature would not be possible.

As an alternative to BMA and MLE, more general techniques of model averaging have been described in the BMA literature.[48] Previous work looking at the technique has focussed on taking the

mean of results of all possible models, weighted by the likelihood that particular model specifications are correct.[48] These weights can be subjective assessments or based on criteria such as minimum variance.[67] Model averaging can easily be applied to HTA models in this basic form, without the use of Bayesian methods to update priors or measures of goodness of fit. Model averaging can simply take the weighted average of the possible scenarios at each iteration of the model. Average EVPI can also be calculated as an average of scenarios at each iteration of the model.

Model averaging, although using all available evidence on alternative model structures, does not fully incorporate the additional uncertainty introduced when choosing between alternative model structures. The structural uncertainty is essentially ignored when models are averaged across.

8.3.2 Model selection

An alternative to model averaging, often employed in statistical sciences, is to choose the 'best' model on the basis of some measure of prediction performance, fit or criteria.[68-76] Applications of model selection have employed both Bayesian and frequentist methods and criteria.

The model selection versus model averaging debate is still ongoing in the statistical, mathematical and economics literature. Gutierrez-Pena[77] argues that it may not always be possible (or desirable) to average across models and focuses on finding the model which maximises utility (representing the consequences of a particular action from the space of decisions given a particular state of the world) given a particular data set.[77] Bunn[78] regard the combination of models as model failure because a single comprehensive model cannot be determined or agreed. However, even in areas outside of HTA where statistical models can be ranked according to some criterion, it may not be advantageous to determine the 'best' model. By choosing the 'best' model, useful evidence may be discarded[79] violating he requirements of the decision-making process. Uncertainty relating to the choice of 'best' model is also ignored.

In terms of their suitability for characterising structural uncertainty in HTA decision models, many of the methods to assess model performance[49] such as Residual Mean Squared Error (MSE), Finite-Prediction-Error (FPE),[49] minimum variance criteria[80] and subjective probabilities[78, 81] are not directly applicable. These involve the quantification of a statistic that describes goodness of fit, prediction performance, or probability of error.[54] In HTA decision modelling, where there are many competing objectives, it is not possible to identify one particular parameter whose performance must be maximised by a fitted model. In those circumstances where the structural uncertainty relates to the choice of statistical model, selection can still be difficult when there is only a small amount of data.[81]

Kashyap[82] and Zellner[83] describe an alternative Bayesian approach to model selection in which the relative likelihood of each model being correct is determined by a set of posterior probabilities derived from the alternative models. This technique, as with the other performance measures, requires the data to act as 'judge, juror and executioner' that is the data determines the structure of the model, populates and validates the model.

8.4 Suitability of methods to characterise structural uncertainty to decision analytic models

As discussed in sections 8.3.1 and 8.3.2, model selection and model averaging are neither appropriate nor desirable methods to characterize structural uncertainty in decision analytic models. When faced with a choice between multiple models, another method can be employed which allows structural uncertainty to be explicitly considered and does not ignore potentially relevant model structures. Uncertainty can be directly characterised (or parameterised) in the model itself. Although this method itself has not been discussed in the literature to date, it is analogous to model averaging on individual or sets of model inputs,[48] but it has additional benefits.

To parameterise structural uncertainty, uncertain parameters are added to the model to represent the choice between multiple model scenarios. At each iteration of the model, a value from the uncertain parameter will be drawn and then linked via an 'if' function to one of the multiple model scenarios. Therefore, an alternative model scenario is chosen at each iteration of the model and structural uncertainties are represented in much the same way as parameter uncertainties.

These uncertain parameters can be specified using a number of different distributions, such as beta or uniform. The distribution chosen will depend on the number of scenarios and the weights attached to each scenario. Often little is known about the distribution of the uncertain parameters, therefore efforts should be made to assign the most appropriate distribution. Expert opinion can be sought for this.

Like model averaging, alternative structural assumptions can be given equal or unequal weights depending on what prior knowledge is known about the likelihood of the scenarios representing the 'true' scenario. However, unlike model averaging where the objective is to simply synthesise all evidence on the structure of a decision model to assess if a treatment is cost-effective, by parameterising the uncertainty directly in the model, estimates of the value of further research on the uncertain parameters can also be made. Simulation output from the 'parameterised model' can provide the decision-maker with the maximum monetary value of conducting further research to eliminate these structural uncertainties. In the HTA models reviewed above that have estimated parameter uncertainty through the use of probabilistic sensitivity analysis, parameterising structural uncertainty simply requires specifications of an additional distributions, or sets of distributions to represent the choice between alternative structural assumptions.

9. Discussion

Decision analytic models represent an explicit way to synthesise evidence currently available on the outcomes and costs of alternative health care interventions and are, therefore, a powerful tool for decision-making. The results derived from a decision analytic model will depend on how the model structure has been defined and the data used to populate the model. Recent developments in the analysis of uncertainty in HTA models have typically focused on parameter uncertainty.[4] Such analyses are, of course, based on the premise that the model has been correctly specified.

To date, presenting alternative model assumptions as scenario analysis is the only method that has been used to address structural uncertainties in the HTA literature. Whilst scenarios can be useful, in as much as they illustrate that structural uncertainty is a potential issue, a decision-maker is still faced with the choice of multiple models. Any uncertainty he/she then has about choosing the 'true' model is not explicitly incorporated into the modelling process.

Model selection is a method commonly applied in mathematics and statistics. However this method is not directly applicable to HTA decision-making. This is primarily because of the focus on model selection through prediction performance or goodness of fit. This is a criteria not commonly used to assess decision analytic models. Decision models are not intended to predict the data and so measures of goodness of fit are often not available. However, for alternative models used to estimate specific parameters such as survival, there may be actual data available to generate such measures. Similar tests are performed when choosing between parametric distributions such as weibull and exponential.[84]

A more formal method to account for structural uncertainty has been proposed, model averaging. This prevents the problem of a decision-maker being faced with multiple models and incorporates all relevant evidence, but does not reflect the uncertainty introduced when there are multiple ways to model a given decision problem.

Parameterising structural uncertainty directly in a decision model is analogous to methods for characterising parameter uncertainty, and as such structural uncertainties can be characterised by propagating uncertain distributions in the model using Monte Carlo simulation methods.[15] In addition to providing an unbiased estimate of decision uncertainty, the output of these simulations can also be used to calculate partial EVPI. Decision-makers can utilise this information when deciding if to commission further research. No other method of characterising structural uncertain provides any information on the value of reducing structural uncertainties that are apparent for a particular decision problem. It remains to be seen how feasible it is to apply this method to decision models.

References

1. National Institute for Clinical Excellence. *Guide to the technology appraisal process*. London: National Institute for Clinical Excellence; 2004.
2. Canadian Coordinating Office for Health Technology Assessment (CCOHTA). *Guidelines for Authors of CCOHTA Health Technology Assessment Reports*. Ottawa; 2003.
3. Scottish Medicines Consortium. *Guidance to Manufacturers for completion of New Product Assessment Form (NPAF)*. Glasgow: NHS Scotland; 2005.
4. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000;17:479-500.
5. Stevens A, Milne R. Health technology assessment in England and Wales. *Int J Technol Assess Health Care* 2004;20:11-24.
6. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines - similarities, differences and some implications. *Value in Health* 2001;4:225-250.
7. Giacomini MK, Cook DJ, Streiner DL, Anand SS. Using practice guidelines to allocate medical technologies. An ethics framework. *Int J Technol Assess Health Care* 2000;16:987-1002.
8. Rutten F. Health technology assessment and policy from the economic perspective. *Int J Technol Assess Health Care* 2004;20:67-70.
9. Woods K. Health technology assessment for the NHS in England and Wales. *Int J Technol Assess Health Care* 2002;18:161-5.
10. National Institute for Clinical Excellence. *A guide for manufacturers and sponsors: contributing to a technology appraisal*. London: National Institute for Clinical Excellence; 2004.
11. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002;360:711-5.
12. Nobre FF, Trotta LT, Gomes LF. Multi-criteria decision making--an approach to setting priorities in health care. *Stat Med* 1999;18:3345-54.
13. Weinstein MC, Fineberg HV, Elstein AS, Frazier HS, Neuhauser D, Neutra RR, et al. *Clinical Decision Analysis*. Philadelphia: W. B. Saunders, 1980.
14. Crane VS, Gilliland M, Tuthill EL, Bruno C. The use of a decision analysis model in multidisciplinary decision making. *Hosp Pharm* 1991;26:309-13, 319-22.
15. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probalistic sensitivity analysis using Monte Carlo simulation. *Medical Decision Making* 1985;5:157-17.
16. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. *Comput Biomed Res* 1986;19:254-65.
17. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999;8:257-262.
18. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions: *Health Technology Assessment*; Vol 3: No 2; 1999.
19. 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.
20. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* 2005;14:339-347.
21. 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. Vol 8(49).* 2004.
22. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 2nd edition ed. New York: Oxford University Press, 1997.
23. Hay J, Jackson J. Panel 2: methodological issues in conducting pharmacoeconomic evaluations - modeling studies. *Value in Health* 1999;2:78-81.
24. ISPOR task force. *Principles of good practice for decision analytic modeling in health care evaluation. Draft document,*; 2001.
25. Ramsey SD. Evaluating evidence from a decision analysis. *Journal of the American Board of Family Practice.* 1999;12:395-402.
26. Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, et al. A systematic review of the role of human papillomavirus testing within a cervical screening programme. London: *Health Technology Assessment,* Vol 3, No 14; 1997.
27. Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, et al. The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review: *Health Technology Assessment*, Vol 6, No 7; 2002.

28. Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A. The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation. *Health Technology Assessment*, Vol 6, No 12. 2002.
29. Garside R, Stein K, Wyatt K, Round A, Price A. The effectiveness and cost-effectiveness of microwave and thermal balloon endometrial ablation for heavy menstrual bleeding: a systematic review and economic modelling. *Health Technology Assessment*, Vol 8, No 3. 2004.
30. Kaltenthaler K, Bravo Vergel Y, Chilcott J, Thomas S, Blakeborough T, Walters SJ, et al. A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography. *Health Technology Assessment*, Vol 8, No 10. 2004.
31. 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,* Vol 8, No 31; 2004.
32. Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, et al. Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation. *Health Technology Assessment; 8(38)*; 2004.
33. Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher S, et al. A rapid and systematic review of the clinical effectiveness and costeffectiveness of clopidogrel used in combination with aspirin compared to aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes (ACS). London: *Health Technology Assessment*, Vol 8, No 40; 2004.
34. Green C, Dinnes J, Takeda A, Shepherd J, Hartwell D, Cave C, et al. Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris) for the treatment of severe sepsis in adults:a systematic review and economic evaluation. *Health Technology Assessment*, Vol 9, No 11. 2005.
35. McCormack K, Wake B, Perez J, Fraser C, Cook J, McIntosh E, et al. Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation. *Health Technology Assessment*, Vol 9, No 14. 2005.
36. Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J. A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis. *Health Technology Assessment*, Vol 9, No 22. 2005.
37. Wilson J, Connock M, Song F, Yao G, Fry-Smith A, Raftery J, et al. Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation. *Health Technology Assessment*, Vol 9, No 25. 2005.
38. Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, et al. Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. *Health Technology Assessment;* Vol 9: number 27; 2005.
39. Tillin T, Chambers M, Feldman R. Outcomes of electrically stimulated gracilis neosphincter surgery. *Health Technology Assessment*, Vol 9, No 28. 2005.
40. Philips Z, Ginnelly L, Sculpher M, 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*; Vol 8: No 36; 2004.
41. Russell LB. Modelling for cost-effectiveness analysis. *Stat Med* 1999;18:3235-3244.
42. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: John Wiley & Sons, 2000.
43. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001;20:2865-2884.
44. McKay MD MJ, Upton SC,, Wendelberger JR, Morrison JD. *Model uncertainty in stochastic simulation models.Contributions to the development of a statistical framework for joint mobility modelling*. Los Alamos: Los Alamos National Labratory; 1997.
45. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics.* 2000;17:501-13.
46. Snowling SD, Kramer JR. Evaluating modelling uncertainty for model selection. *Ecol. Model.* 2001;138:17-30.
47. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating Partial Expected Value Of Perfect Information in Cost-Effectiveness Models via a Two Level Monte-Carlo Algorithm. In: *66th Health Economics Study Group*; 2005; Oxford. 2005.
48. Chatfield C. Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society* 1995;158:419-466.

49. Engle RF, Brown SJ. Model Selection for Forecasting. *Appl Math Comput* 1986;20:313-327.
50. Augustin N, Sauerbrei W, Schumacher M. *The practical utility of incorporating model selection uncertainty into prognostic models for survival data*. Glasgow; 2004.
51. Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society* 1995;57:45-97.
52. Scott M. *Uncertain models and modelling uncertainty.* Presentation at Department of Environmantal Modelling. University of Nottingham, 2004.
53. Duong QP. Model selection and ranking - an AHP approach to forecasts combination. *Math Comput Model* 1988;11:282-285.
54. Bates JM, Granger CWJ. The combination of forecasts. *Operational Research Quarterly* 1969;20:451-468.
55. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995;82.
56. Clemen RT. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* 1989;5:559-583.
57. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Statistical Science* 1999;14:382-417.
58. Madigan D, Raftery AE. Model selection and accounting for model uncertainty in graphical models using Occams window. *Journal of the American Statistics Association* 1994;89:1535-1546.
59. Chatfield C. Model uncertainty and forecast accuracy. *Journal of Forecasting* 1996;15:495-508.
60. Volinsky CT, Madigan D, Raftery AE, Kronmal RA. Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. *Journal of the Royal Statistical Society, Series C* 1997;46:433-448.
61. Raftery AE, Madigan D, Volinsky CT. Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian Statistics* 1994;5:323-349.
62. Viallefont V, Raftery AE, Richardson S. Variable selection and Bayesian model averaging in case-sontrol studies. *Statistics in Medicine* 2001;20:3215-3230.
63. Conrad Lamon III E, Clyde MA. Accounting for model uncertainty in prediction of chlorophyll a in Lake Okeechobee. *Journal of Agricultural, Biological, and Environmental Statistics* 2000: 297-322.
64. Raftery AE, Balabdaoui F, Gneiting T, Polakowski M. *Using Bayesian model averaging to calibrate forecast ensembles: Technical report 440*. Washington: Department of Statistics, University of Washington; 2003.
65. Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the royal Statistical Society. Series B* 1995;57:473-484.
66. Shannon WD, Banks D. Combining classification trees using MLE. *Stat Med* 1999;18:727-40.
67. Dickinson JP. Some statistical results on the combination of forecasts. *Operational Research Quarterly* 1973;24:253-260.
68. Wagenmakers EJ. How many parameters does it take to fit an elephant? Book review of "Model selection and multimodel inference: A practical information-theoretic approach". *J Math Psychol* 2003;47:580-586.
69. Laud PW, Ibrahim JG. Predictive model selection. *Journal of the Royal Statistical Society* 1995;Series B:247-262.
70. O'Hagan A. Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society* 1995;Series B:99-138.
71. Berger JO, Pericchi LR. The intrinsic Bayes factor for model selection and prediction. *J Am Stat Assoc* 1996;91:109-122.
72. Kittler J, Messer K, Sadeghi M. Model selection by predictive validation. *Pattern Anal Appl* 2002;5:245-260.
73. San Martini A, Spezzaferri F. A predictive model selection criterion. *Journal of the Royal Statistical Society* 1984;Series B:296-303.
74. Carlin BP, Kass RE, Lerch FJ, Huguenard BR. Predicting working memory failure - a subjective Bayesian- approach to model selection. *J Am Stat Assoc* 1992;87:319-327.
75. George EI. Bayesian model selection. In: *Encyclopedia of statistical sciences update 3*. New York: Wiley, 1999.
76. Shen XT, Huang HC, Ye J. Inference after model selection. *J. Am. Stat. Assoc.* 2004;99:751-762.
77. Gutierrez-Pena E, Walker SG. A Bayesian predictive approach to model selection. *J Stat Plan Infer* 2001;93:259-276.
78. Bunn DW. Combining forecasts. *European Journal of Operational Research* 1988;33:223-229.
79. Bunn DW. A Bayesian approach to the linear combination of forecasts. *Operational Research Quarterly* 1975;26:325-329.

80. Dickinson JP. Some comments on the combination of forecasts. *Operational Research Quarterly* 1975;26:205-210.
81. Bunn DW, Kappos E. Synthesis or selection of forecasting models. *European Journal of Operational Research* 1982;9:173-180.
82. Kashyap RL. A Bayesian comparison of different classes of dynamic models using empirical data. *IEEE Trans* 1977;AC-22:715-727.
83. Zellner A. *An introduction to Bayesian Inference in Econometrics*. New York, 1971.
84. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technolgy Assessment. Vol 3(10)*; 1999.
85. Baldelli DH, Pena RSS. Uncertainty modeling in aerospace flexible structures. *J Guid Control Dyn* 1999;22:611-614.
86. Besag JE, Green P, Higdon D, Mengerson K. Bayesian computation and stocastic systems. *Statistical Science* 1995;10:3-66.
87. Blockley D. Model Uncertainty in Structural Reliability - Comment. *Struct Saf* 1983;1:233-235.
88. Bracke MB, Spruijt BM, Metz JH, Schouten WG. Decision support system for overall welfare assessment in pregnant sows A model structure and weighting procedure. *J Anim Sci* 2002;80:1819-34.
89. Buntine W. Learning classification trees. *Statistical Computing* 1992;2:63-73.
90. Dawid AP, Guttman I. Conjugate Bayesian-inference for structural models. *Communications in Statistics Part a-Theory and Methods* 1981;10:739-748.
91. Dilks DW, Canale RP, Meier PG. Development of Bayesian Monte-Carlo techniques for water-quality model uncertainty. *Ecol Mode.* 1992;62:149-162.
92. Ditlevsen O. Model uncertainty in structural reliability. *Struct Saf* 1982;1:73-86.
93. Freedman DA, Navidi W, Peters SC. In the impact of variable selection in fitting regression equations. In: Dijkstra TK, editor. *On model uncertainty and its statistical implications*. Berlin: Springer, 1988.
94. Geisser ES. A Bayes approach for combining correlated estimates. *Journal of the American Statistical Association. Series A* 1965;60:602-607.
95. Gewke J. Simulation methods for model criticism and robustness analysis. *Bayesian statistics 6* 1998:000-000.
96. Granger CWJ, Hashem Pesaran M. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 2000;19:537-560.
97. Groenewald PCN, Vandermerwe AJ. Model selection - using normal priors and predictive sample re- use. *South Afr Stat J* 1988;22:131-151.
98. Hasselman T. Quantification of uncertainty in structural dynamic models. *J Aerosp Eng* 2001;14:158-165.
99. Lange T. A multi-criterial decision approach to structural modeling with uncertainty conditions. *Systems Analysis Modelling Simulation* 1989;6:147-154.
100. McDowell RM, Jaworska JS. Bayesian analysis and inference from QSAR predictive model results. *SAR QSAR Environ Res* 2002;13:111-25.
101. McKay MD MJ, Upton SC,. Evaluating prediction uncertainty in simulation models. *submitted to Computer Physics Communications* 2004.
102. Morris PA. Decision analysis expert use. *Management Science* 1974;20:1233-1241.
103. Morris PA. Combining expert judgements: A Bayesian approach. *Management Science* 1977;23:679-693.
104. Ossen A, Ruger SM. Weight space analysis and forecast uncertainty. *Journal of Forecasting* 1998;17:471-480.
105. Roberts HV. Probabalistic prediction. *J Am Stat Assoc* 1965;60:50-62.
106. Sauerbrei W, Augustin N, Hollander N. Incorporating model selection uncertainty into prognostic models for survival data. *Controlled Clin Trials* 2003;24:P105.
107. Sayed AH, Nascimento VH. Design criteria for uncertain models with structured and unstructured uncertainties. In: *Robustness in Identification and Control*, 1999. p. 159-173.
108. Schoemaker PJH. When and how to use scenario planning: A heuristic approach with illustrations. *Journal of Forecasting* 1991;10:549-564.
109. Tuljapurkar S. Stochastic population forecasts and their uses. *Int J Forecast* 1992;8:385-91.
110. van Noortwijk JM, Cooke RM, Kok M. A Bayesian failure model based on isotropic deterioration. *European Journal of Operational Research* 1995;82:270-282.
111. Wang D, Zhang W, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Statistics in Medicine* 2004;23:3451-3467.
112. West M, Harrison PJ. *Bayesian forecasting and dynamic linear models*. New York, 1989.

113. Wilson CA. Feeder cattle forecasting models - an econometric study of development and performance. *Am J Agr Econ* 1993;75:1326-1332.
114. Zhao XD, Xie JX, Leung J. The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research* 2002;142:321-344.

Appendix 1: Papers included in review of methods to characterise structural uncertainty

Author (s), date	Title	Discipline	Focus	Relevant
Augustin, et al. 2004 [50]	The practical utility of incorporating model selection uncertainty into prognostic models for survival data.	Statistics	Looks at methods to account for model selection uncertainty in survival analysis	✓
Baldelli, 1999[85]	Uncertainty modelling in aerospace flexible structures	Aerospace modelling	Representing parametric uncertainties in the modes of a flexible structure	X
Bates, et al, 1969 [54]	The combination of forecasts	Operational research	Discusses issues relating to the combination of forecasts	✓
Berger, et al, 1996 [71]	The intrinsic Bayes factor for model selection and prediction.	Statistics	Bayes factor for multiple model comparison and prediction	✓
Besag, et al, 1995 [86]	Bayesian computation and stochastic systems	Statistics	Use of Monte Carlo Markov Chain methods	X
Blockley, 1983 [87]	Model uncertainty in structural reliability - comment	Engineering	Interval probability theory for structural reliability calculations	X
Bracke, 2002[88]	Decision support system for overall welfare assessment in pregnant sow: A model structure and weighting procedure	Agricultural modelling	Simulation model to assess welfare of sows	X
Bunn, 1975 [79]	A Bayesian approach to the linear combination of forecasts	Operational research	Discusses Bayesian approaches to combining forecasts	✓
Bunn, 1988 [78]	Combining forecasts	Operational research	Discusses the rationale for combining forecasts	✓
Bunn, et al, 1982 [81]	Synthesis or selection of forecasting models	Operational research	Discusses methods to select the most accurate model, based on prediction performance indicators	✓
Buntine, 1992 [89]	Learning classification trees	Statistics	Building decision tress and how to update information using Bayesian methods	X
Carlin, et al, 1995 [65]	Bayesian model choice via Markov chain Monte Carlo methods	Statistics	Framework for Bayesian model choice	✓
Carlin, et al, 1992 [74]	Predicting working memory failure - a subjective Bayesian- approach to model selection	Statistics	Bayesian approach to hypothesis testing	✓
Chatfield C, 1995 [48]	Model uncertainty, data mining and statistical inference (with discussion).	Operational research	Discusses type of model uncertainty	✓
Chatfield, 1996[59]	Model uncertainty and forecast accuracy	Operational research	General overview of 'model uncertainty' Proposes use of Bayesian model averaging techniques.	✓
Clemen, 1989 [56]	Combining,forecasts: a review and annotated bibliography. International	Operational research	Review of methods to combine forecasts	✓
Conrad Lamon, et al, 2000 [63]	Accounting for model uncertainty in prediction of chlorophyll a in Lake Okeechobee	Environmental modelling	Application of Bayesian model averaging techniques to prediction	✓

Dawid, et al, 1981 [90]	Conjugate Bayesian-inference for structural models	Statistics	Use of Bayesian inference	X
Dickinson JP, 1973 [67]	Some statistical results on the combination of forecasts	Operational research	Looks at implications of combining forecasts	✓
Dickinson, 1975 [80]	Some comments on the combination of forecasts	Operational research	Looks at implications of combining forecasts using a minimum variance criterion	✓
Dilks, et al, 1992 [91]	Development of Bayesian Monte-Carlo techniques for water- quality model uncertainty	Environmental modelling	Representing parameter uncertainty	X
Ditlevsen, 1982 [92]	Model Uncertainty in Structural Reliability	Engineering	Dealing with model uncertainty in structural reliability analysis	X
Draper, 1995 [51]	Assessment and propagation of model uncertainty	Statistics	Bayesian approaches to model uncertainty	✓
Duong, 1988 [53]	Model selection and ranking - an AHP approach to forecasts combination	Mathematics	Model selection in statistical inference problems	✓
Engle, et al, 1986 [49]	Model selection for forecasting	Mathematics	Discusses various criteria to measure the performance of a model	✓
Freedman, 1988[93]	In the impact of variable selection in fitting regression equations	Econometrics	Selection of parameters for regression models	X
Geisser, 1965 [94]	A Bayes approach for combining correlated estimates	Statistics	Bayesian random effects meta-analysis	X
George, 1999 [75]	Bayesian model selection	Statistics	Reviews Bayesian model selection and discussed BMA in the context of decision theory	✓
Gewke, 1998 [95]	Simulation methods for model criticism and robustness analysis	Statistics	Use of Bayes factor and uncertainty about prior distributions	X
Granger, et al, 2000 [96]	Economic and statistical measures of forecast accuracy	Operational research	Looks at applications of forecasting to game theory	X
Green PJ, 1995 [55]	Reversible jump Markov chain Monte Carlo computation and Bayesian model determination	Statistics	Use of Bayesian methods in model choice and averaging	✓
Groenewald, et al, 1988 [97]	Model selection - using normal priors and predictive sample re- use	Statistics	Applying model selection criteria	✓
Gutierrez-Pena, et al, 2001 [77]	A Bayesian predictive approach to model selection	Mathematics	Discusses predictive model selection criterion	✓
Hasselman, 2001 [98]	Quantification of uncertainty in structural dynamic models	Engineering	Application of methods to evaluate the accuracy of structural dynamic models	X
Hoeting JA, et al, 1999 [57]	Bayesian model averaging: a tutorial	Statistics	Discusses Bayesian model averaging in detail	✓
Kashyap, et al 1977 [82]	A Bayesian comparison of different classes of dynamic models using empirical data	Computer simulation	Bayesian methods of comparing different types of dynamical structures	✓
Kittler, et al, 2002 [72]	Model selection by predictive validation	Computer simulation	Focuses on model selection through model validation	✓
Lange, 1989 [99]	A multi-criteria decision approach to structural modelling with uncertainty conditions	Statistics	Problems and solutions to uncertainty conditions in large scale systems models	X
Laud, et al, 1995 [69]	Predictive model selection	Statistics	Bayesian methods for model selection	✓
Madigan, et al, 1994 [58]	Model selection and accounting for model uncertainty in graphical models using Occams window	Statistics	Choosing what model to include in a Bayesian model averaging framework	✓

McCabe, et al, 2000 [45]	Testing the validity of cost-effectiveness models	Health Technology Assessment	Guidelines of good practice in decision models	✓
McDowell, et al, 2002 [100]	Bayesian analysis and inference from QSAR predictive model results	Environmental modelling	Assessing accuracy of QSAR models using sensitivity and specificities and using sequential models to improve accuracy	X
McKay, et al, 2004 [101]	Evaluating prediction uncertainty in simulation models	Statistics	General background on types of model uncertainty, goes onto discuss parameter uncertainty in detail	X
McKay, et al 1997 [44]	Model uncertainty in stochastic simulation models	Statistics	Review and application of methods to handle all types of model uncertainty	✓
Morris, 1974 [102]	Decision analysis expert use	Management science	Discusses use of expert opinion in decision-making	X
Morris, 1977 [103]	Combining expert judgements: A Bayesian approach	Management science	Discusses issue of combining evidence from multiple sources of expert opinion	X
O'Hagan A, 1995 [70]	Fractional Bayes factors for model comparison (with discussion)	Statistics	Discusses and applies Bayesian methods for model comparison, in particular frictional Bayes factor	✓
Ossen, et al, 1998[104]	Weight space analysis and forecast uncertainty	Operational research	Methods for deriving weights to improve forecast uncertainty in neural networks	X
Raftery,et al, 2003 [64]	Using Bayesian model averaging to calibrate forecast ensembles	Statistics	Discusses the use of Bayesian model averaging	✓
Raftery, et al, 1994 [105]	Accounting for model uncertainty in survival analysis improves predictive performance	Statistics	Discussion and application of BMA in survival analysis	✓
Roberts, 1965[105]	Probabilistic prediction	Statistics	Application of predictive distributions	X
San Martini, et al, 1984 [73]	Predictive model selection criterion	Statistics	Applying model selection criteria	✓
Sauerbrei, et al, 2003 [106]	Incorporating model selection uncertainty into prognostic models for survival data	Statistics	Discussion and application of BMA in survival analysis	✓
Sayed, et al, 1999 [107]	Design criteria for uncertain models with structured and unstructured uncertainties	Engineering	Discussion and application of weighted game-type cost criterion for evaluating model uncertainties	X
Schoemaker, 1991 [108]	When and how to use scenario planning: A heuristic approach with illustrations	Operational research	Looks at use of scenario analysis in decision-making	X
Shannon, et al, 1999 [66]	Combining classification trees using MLE	Statistics	Maximum Likelihood Estimation applied to choice of classification tree	✓
Shen, et al, 2004 [76]	Inference after model selection	Statistics	Discussion and application of optimal approximation methods for model selection.	✓
Snowling, et al, 2001 [46]	Evaluating modelling uncertainty for model selection	Environmental modelling	General discussion of structural (model) uncertainty	✓
Tuljapurkar, 1992 [109]	Stochastic population forecasts and their uses	Operational research	Discussion and application of stochastic forecasting methods	X

Van Noortwijk, 1995 [110]	A Bayesian failure model based on isotropic deterioration	Operational research	Accounting for uncertainty in predicting the failure of hydraulic structures using Bayesian methods	X
Viallefont V, et al, 2001 [62]	Variable selection and Bayesian model averaging in case-control studies	Statistics	Application of BMA to take account of uncertainty in choosing covariates for a case-control study	✓
Volinsky, et al, 1997 [60]	Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke	Statistics	Looks at the use of Bayesian model averaging applied to clinical data	✓
Wagenmakers, 2003 [68]	Review of "Model selection and multimodel inference: A practical information-theoretic approach"	Mathematics	Review of a book looking at model selection	✓
Wang, et al, 2004 [111]	Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression	Statistics	Looks at BMA for model selection in logistic regression compared to stepwise procedure	✓
West M, et al, 1989 [112]	Bayesian forecasting and dynamic linear models	Statistics	Theory and application of forecasting and dynamic models	X
Wilson, 1993 [113]	Feeder cattle forecasting models - an econometric study of development and performance	Agricultural modelling	Measuring forecasting ability for 2 models predicting future price of feeder cattle	X
Zellner, 1971 [83]	An introduction to Bayesian Inference in Econometrics	Econometrics	Use of Bayesian methods in econometrics	✓
Zhao, et al, 2002 [114]	The impact of forecasting model selection on the value of information sharing in a supply chain	Operational research	Information requirement of a supply chain management process	X