1. Introduction

This report adopts the view of the ISPOR Task Force on Good Research Practices in Modeling Studies that the purpose of a model is to inform medical decisions and health-care resource allocation. Modelers employ quantitative methods to gain qualitative insight. The tools of formal analysis are employed to structure the clinical, epidemiological, and economic evidence base in service of better clinical practice decisions and public health priorities. Viewed from this policy perspective, the value of a model-based analysis lies not in its ability to generate a precise point estimate for a specific outcome; rather, the value lies in the clarity with which the model reveals to stakeholders the logical and conditional relationship between inputs (both structural assumptions and source data) and a recommended course of action. The systematic examination and responsible reporting of uncertainty – specifically as it pertains to the choices facing the decision maker – are hallmarks of good modeling practice.

The extent to which an uncertainty analysis can be considered 'fit for purpose' in part depends on the decision the modeling seeks to support. Uncertainty analysis can be seen to serve two main purposes: firstly, to assess confidence in a chosen course of action and, secondly, to ascertain the value of collecting addition information to better inform the decision.

Many models are designed to help decision makers to optimize a well-specified, quantitative objective (subject, perhaps, to one or more limiting constraints). The model generates point estimates of the objective function for each possible course of action; the “best” choice is the one that maximizes this objective. If the decision maker has to make a resource allocation decision now, has no role in commissioning or mandating further research and cannot delay the decision or review it in the future, then the role of uncertainty analysis is limited and that decision should be based only on expected values of effectiveness. Nevertheless, since the conditional relationship between inputs and outputs will typically differ from one course of action to another, decision makers may want to gauge their confidence in the appropriateness of this “best choice” by exploring its robustness in the face of changes in the model’s inputs. This may be particularly important in situations where the decision would be difficult to reverse or modify as new data emerge. Even when inputs are believed to be known with great precision, examination of the robustness of a chosen course of action may help decision makers to assess the extent to which it can be generalized to other settings or target populations.

Increasingly models are being developed to guide the decisions of particular bodies - for example, organizations responsible for deciding whether a health system reimburses a new pharmaceutical product. Such decision makers who have the authority to delay decisions or to review them in the future, based on research they commission or mandate, should be interested not just in expected cost-effectiveness, but also in a thorough analysis of uncertainty and the value of additional research. Such information, as well as assessments of factors such as the costs of reversing a decision shown to be suboptimal as further information emerges, and the cost of research and the likelihood of it being undertaken, can influence the array of decisions available. By helping decision
makers to determine the expected incremental improvement in the objective function, uncertainty analysis conveys not only qualitative information about the critical uncertainties surrounding a decision but also quantitative information about the decision maker’s priorities in allocating resources to further research.

Of course, many models are developed for general dissemination and without a specific decision maker in mind. Such models could be used to inform the decisions of a range of possible decision makers with varying decision-making responsibilities. Here there is a case for undertaking a full uncertainty analysis, thus allowing different types of decision maker to take from them what they require given the decisions with which they are charged.

Recommendations:

1.1 The systematic examination and responsible reporting of uncertainty are hallmarks of good modeling practice. All modeling studies should therefore include an assessment of uncertainty as it pertains to the decision problem being addressed.

1.2 The role of the decision maker should be considered when presenting uncertainty analyses. In particular, the description of analytic perspective should include an explicit statement regarding what is assumed about the power of the decision makers to delay or review decisions and to commission or mandate further research.
2. Background and terminology

It is important to be precise concerning the terminology which will be drawn on in this paper, but the meaning of which is sometimes confused in the literature. In particular, first-order uncertainty is distinguished from both parameter uncertainty and from heterogeneity. Furthermore, each of these concepts is argued to have an analogous form within a simple ‘regression-type’ model in the statistical literature and in which, as in regression analysis, the structural uncertainty associated with the model itself must also be considered. Table 1 summarizes the concepts used in this paper together with our preferred terminology for this paper, lists other terms that have been used, and provides an analogous link to statistical regression. Further discussion of the concepts is given below.

Table 1

<table>
<thead>
<tr>
<th>Preferred term</th>
<th>Concept</th>
<th>Other terms sometimes employed</th>
<th>Analogous concept in regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-order uncertainty</td>
<td>Random variability in outcomes between identical patients</td>
<td>Variability, Monte Carlo error, Unobserved or unexplained heterogeneity</td>
<td>Error term</td>
</tr>
<tr>
<td>Parameter uncertainty</td>
<td>The uncertainty in estimation of the parameter of interest</td>
<td>Second-order uncertainty</td>
<td>Standard error of the estimate</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>The variability between patients that can be attributed to characteristics of those patients</td>
<td>Variability, Observed or explained heterogeneity</td>
<td>The Beta coefficients (or the extent to which the dependent variable varies by patient characteristics)</td>
</tr>
<tr>
<td>Structural uncertainty</td>
<td>The assumptions inherent in the presentation of the decision modeling form</td>
<td>Model uncertainty</td>
<td>The form of the regression model (e.g. linear, log-linear etc.)</td>
</tr>
</tbody>
</table>

Firstly, it should be understood that the term 'parameter uncertainty' is not the same as the uncertainty around the realization of individual, patient-level, events or outcomes. This 'first-order uncertainty', relates to the fact that, within a given sample of patients, individuals facing the same probabilities and outcomes will experience the effects of a disease or an intervention differently, just as a fair coin might come up heads or tails on any given toss. For example, the first patient in a sample might respond to a treatment but the next may not; the first may experience an adverse
effect to the therapy but the second may not; the first may stay in hospital for two days and the
second for three days. Parameter uncertainty, sometimes called ‘second-order uncertainty’, relates
to the fact that the probabilities that govern outcomes are themselves uncertain, owing to the fact
that they are estimated quantities. For example, one hundred tosses of a fair coin will not always
lead to exactly fifty realizations of “heads” and fifty realizations of “tails.” Estimates of the coin’s
probability of landing “heads” based upon one hundred observations will be uncertain. The size of
sample informing that estimate, together with the variance in the data, contributes to determining
the level of parameter uncertainty. Although these two types of uncertainty are clearly linked,
uncertainty about actual outcomes for individual patients is different from uncertainty surrounding
population parameter values (e.g. the probability of response, the probability of an adverse event or
the mean length of hospital stay). The distinction is analogous to the difference between the
standard deviation of a sample as the estimate of how individual observations within a sample vary
and the standard error that summarizes the precision of an estimated quantity.

Further adding to the confusion between first-order uncertainty and second-order (parameter)
uncertainty is the fact that the term ‘variability’ is used to refer to the former (Briggs et al 2006;
Griffin et al 2006; Claxton 2008), but is also used to refer to ‘heterogeneity’ (Hunink et al 2001;
Groot Koerkamp et al 2010), the differences in parameter values across patients or patient
subgroups. Heterogeneity is the extent to which the between-patient variability in a particular
parameter can be explained by the patients’ characteristics, for example, age- and sex-specific
mortality rates.

An analogy can be made with a simple regression model of the form

$$Y = X\beta + \varepsilon,$$

where an outcome variable Y depends on covariates X with coefficients β and stochastic error term
ε. In this example, the vector of coefficients β, represents the parameters of the model and will be
estimated with uncertainty represented by the standard error of the coefficients from the fitted
regression. The extent to which the predicted values Y vary with the known covariates X represents
heterogeneity, and the stochastic error term ε represents the unexplained variability.

Just as a linear regression model imposes a structural relationship between independent and
dependent variables, so decision analytic models are characterized by a series of assumptions which
are reflected in the model’s structure but which are not formally expressed numerically. These
would include, for example, the types of adverse events included, the duration of a treatment effect,
the time dependency of probabilities, and the prognostic implications of surrogate endpoints or
clinical events. Although these structural assumptions are not typically formally quantified, it is
uncertain whether they are an accurate expression of reality. As such, and analogously with
statistical modeling, any representation of uncertainty in a decision model is conditional on the
structural assumptions of the model; therefore, in principle, structural uncertainty in the
construction of the model is a further level of uncertainty that should be reflected in the analysis of
model uncertainty.

Although the overall structure of a model is made up of the multiplicity of assumptions and analytic
decisions that are made by the analyst(s) in the construction of a model, it is useful to distinguish
two broad categories of model that reflect both the underlying structure and relate to the concepts
of uncertainty outlined above. Discrete-event simulation models are structured around events that occur at the patient level and therefore require the use of individual-level patient simulation. By contrast, state transition models are most often evaluated using a ‘cohort’ approach such that individual-level simulation is not required. In the former case, assessment of second-order uncertainty first requires the elimination of first-order uncertainty. In the latter case, second-order uncertainty can be addressed directly. Each of these approaches to structuring a decision model forms part of a separate paper in the Task Force series on Modeling Practices.

Recommendation:

2.1 Terminology to describe concepts relating to parameter estimation and representation of uncertainty varies within the medical decision modeling field and in comparison to related fields. Authors should be aware of this and seek to carefully define their use of terminology to avoid potential confusion.

3. Parameter estimation and uncertainty

All decision models will have parameters that need to be estimated. In populating models with parameter estimates, analysts should conform to the broad principles of evidence-based medicine. For example, analysts should: seek to incorporate all evidence, rather than cherry picking the best single source of evidence for that parameter; use best practice methods to avoid potential biases in parameter estimates that might arise (for example, when estimating treatment effectiveness from observational sources); and employ formal evidence synthesis techniques (meta analysis and network meta analysis) as appropriate. As part of this process of estimating input parameters the estimation of uncertainty should be seen as equally integral to decision analysis, recognizing that the steps that need to be taken to estimate a parameter link directly with those necessary to conduct uncertainty analysis. Standard statistical methods for estimation generate a point estimate together with some measure of precision such as standard errors or 95% confidence intervals. This is true whether these methods are implemented within a uni- or multi-variable framework, although the latter will also provide a measure of covariance between estimated parameters. Consequently, whether a modeler uses primary data sources to estimate input parameters for a decision model or derives this information from one secondary source or a pooling of several such sources, the process of estimation generates a point estimate (typically a mean), a measure of precision and, potentially, a measure of covariance. These different types of information from estimation should feed directly into the implementation of the uncertainty analysis for the decision model.

This is true whatever the technical specification of the uncertainty analysis. In the case of a deterministic one-way sensitivity analysis, it is necessary to specify a point estimate for the parameter together with a defensible range; these may be taken directly from the process of estimation, with the latter based, for example, on a 95% confidence interval. A two-way sensitivity analysis will be more useful if it can draw on any available information on the covariance between the two parameters of interest, or on the logical relationship between the parameters. For example, if clinical event rates with two treatments are related through a hazard ratio, then a two-way sensitivity analysis might be presented in terms of the event rate with one of the treatments and the hazard ratio.
Representation of uncertainty will depend upon the type of sensitivity analysis that is planned. For deterministic sensitivity analysis, then, an interval estimate is required. For probabilistic sensitivity analysis, the whole of the distribution is specified and that distribution will have parameters that will require specification (see below).

**Recommendation:**

3.1 All decision models will have parameters that need to be estimated. In populating models with parameter estimates, analysts should conform to the broad principles of evidence-based medicine. For example, analysts should: seek to identify and incorporate all relevant evidence, rather than cherry picking the best single source of evidence for that parameter; use best practice methods to avoid potential biases in parameter estimates that might arise (for example, when estimating treatment effectiveness from observational sources); and employ formal evidence syntheses techniques (meta analysis and network meta analysis) as appropriate.

**Consistency of approach between deterministic and probabilistic sensitivity analysis (SA)**

A general principle is that estimates of uncertainty for parameters that are (or could be) estimated from data should be consistent with standard statistical approaches. For example, where data exist on a particular parameter, it would be natural to assume that the mean is the point estimate and a standard 95% confidence interval could be used as the parameter’s range in a deterministic SA. The underling distributional assumption used to calculate the interval could form the basis of an equivalent distributional assumption for a probabilistic analysis. One exception would be where a formal Bayesian approach is taken involving the use of subjective prior information: in this case the standard distributional assumption relates to the data likelihood that is combined with the prior information to form the posterior density. In this case, consistency between deterministic SA and the fully Bayesian probabilistic approach would be retained if the interval estimate for deterministic SA were based on the 95% highest density region of the Bayesian posterior distribution.

A number of forms of sensitivity analysis do not require formal ascertainment of the parameter uncertainty. These include threshold analysis, where the value of a parameter (or several parameters in a multi-way analysis) that would be needed to change the decision from that based on expected values is identified. This form of analysis is closely linked to ‘even if’ type approaches which identify extreme values of parameters that still do not change the decision -- with the implication that the parameter in question is unlikely to influence the decision. Such analyses may be sufficient under a fortiori conditions such that there is little decision uncertainty given reasonable assumptions about parameter uncertainty. Another form of analysis that does not require estimation of uncertainty is identification of the quantitative relationships between input and output parameters of the model. For example, it might be determined that a 10% increase in the value of a particular parameter leads to a 20% increase in expected effectiveness and a 5% decrease in expected cost. This sort of analysis is unlikely to be sufficient in and of itself since a highly uncertain parameter with low sensitivity could easily have more impact on the outcome parameter than a more sensitive parameter that has been estimated more precisely. Therefore, completely arbitrary
analyses, such as the presentation of the effect on model outputs of varying each input parameter by +/- 50% is not recommended.

Recommendations:

3.2 Whether employing deterministic SA methods (point estimate and range) or probabilistic SA (parameterized distribution) the link to the underlying evidence base should be clear.

3.3 While completely arbitrary analyses, such as the presentation of the effect on model outputs of varying each input parameter by +/- 50%, can be used as a measure of sensitivity, such analyses should not be used to represent uncertainty.

Estimation and choice of distribution for probabilistic SA and interval estimation

It was argued above that whether estimating a distribution for a parameter to feed into a PSA or for an interval estimate in a deterministic SA, there should be a consistent relationship between the uncertainty estimate and the underlying evidence informing parameter estimation. Where the expected value of a parameter is of interest, then it is of note that the central limit theorem indicates that the sampling distribution of the arithmetic mean will follow a normal sampling distribution (with sufficient sample size), whatever the underlying distribution of the data. This suggests that if there is much information available to inform the estimate of a parameter then the normal distribution could be used in a probabilistic SA, or a standard confidence interval in a deterministic SA. Consistency between each form of uncertainty estimate is maintained since the arithmetic mean and standard error inform both the parameters of the distribution and the confidence interval calculation.

In contrast, if there is less information on which to base parameter estimates then other distributional forms than the normal distribution might be considered. The general principle remains that the underlying assumptions for specifying the distribution and/or defining the interval estimate for uncertainty analysis should follow standard statistical methods. For example, beta distributions are a natural match for binomial data; gamma or log normal distributions can be used for right skew parameters; log normality would be a standard assumption for relative risks or hazard ratios; and a logistic distributional assumption works well for odds ratios. These distributions can be used directly in a probabilistic SA or used to define the uncertainty interval for a deterministic SA.

Sometimes there is very little information on a parameter, either because there are very few studies informing the parameter estimation, or because there are no data at all and expert opinion must be relied upon. In these situations, it is imperative that the uncertainty related to such parameter estimates be fully explored. Analysts should adopt a conservative approach such that, in the absence of evidence on a given parameter, a very broad range of possible estimates is considered. A wide interval should be assumed in deterministic sensitivity analysis, or a very broad prior distribution should be assumed in a probabilistic analysis. On no account should parameters be excluded from a sensitivity analysis on the grounds that ‘there is not enough information from which to estimate uncertainty’. In choosing distributional forms for parameters in a probabilistic sensitivity analysis, favor should be given to continuous distributions that provide a realistic portrayal of uncertainty over the theoretical range of the parameter of interest. Hence careful consideration
should be given to whether distributions like the triangular should have any role in a PSA. Analysts should note that formal elicitation methods for eliciting probability distributions from experts have been developed in the Bayesian literature (O’Hagan et al, 2006).

Recommendations:

3.4 Analysts should give consideration to using commonly adopted standards from statistics for point estimate and interval estimation for input parameters, such as 95% confidence intervals, or distributions based on agreed statistical methods for a given estimation problem. Where departures from these standards are deemed necessary (or where no such standard exists for a given estimation problem), these should be justified.

3.5 Where there is very little information on a parameter, analysts should adopt a conservative approach such that the absence of evidence is reflected in a very broad range of possible estimates. On no account should parameters be excluded from a sensitivity analysis on the grounds that ‘there is not enough information from which to estimate uncertainty’.

3.6 In choosing distributional forms for parameters in a probabilistic sensitivity analysis, favor should be given to continuous distributions that provide a realistic portrayal of uncertainty over the theoretical range of the parameter of interest. Hence careful consideration should be given to whether distributions like the triangular should have any role in a PSA.

Multivariate estimation and correlation

Where parameters are estimated as part of a regression model to capture the effect of subject/patient characteristics (covariates) on the parameter estimates, then it is helpful to recognize that the dependent variable in the regression is a functional parameter of the coefficients of the model. Therefore, the uncertainty in the functional parameter can be defined in terms of the uncertainty (and correlation) in the basic parameters (coefficients). The covariance matrix defines these uncertainties, and the assumption of multivariate normality is the appropriate assumption for the linear predictor of the regression. The covariance matrix can be used together with the assumption of multivariate normality to specify the interval for the functional parameter or as the basis for the probabilistic SA (Briggs et al, 2006).

In probabilistic SA, model parameters are assigned probability distributions to reflect uncertainty. However, parameters are typically not all independent of one another. For example, suppose two uncertain parameters in a model are the probabilities of disease progression with and without treatment. Part of this uncertainty may derive from uncertainty regarding the natural history of the disease, and part of it may derive from uncertainty about the efficacy of treatment. The component of uncertainty related to the natural history of disease would affect the probabilities of progression with or without treatment, whereas the component of uncertainty related to the efficacy of treatment would affect the relationship (e.g., relative risk reduction) between the probabilities of progression with and without treatment. In this example, it would be wrong to regard the probabilities of progression with and without treatment as coming from independent distributions,
and to conduct the PSA accordingly. However, it might be reasonable to regard the natural history probability of progression and the risk reduction with treatment as independent. A recommended approach to specifying the parameter distributions in this type of situation is to define the parameters for probabilistic SA in a way that makes it plausible that they are independent. In this example, the baseline progression probability and the relative risk reduction would be assigned distributions in the probabilistic SA, with the on-treatment probability of progression derived as the product of the two. While this method of defining parameters in a way that induces mutual independence offers a practical and sufficient approach in many situations, more sophisticated methods that explicitly quantify joint distributions of correlated parameters may also be considered.

Recommendation:

3.6 Correlation among parameters should be considered. Jointly estimated parameters, such as those from a regression analysis, will have direct evidence on correlation which should be reflected in the analysis. Independently estimated parameters will have no such evidence, but this should not necessarily lead to an assumption of independence. Possible approaches are (1) to include a correlation coefficient as a parameter to the model where concern exists that an unknown correlation between parameters could be important, or (2) to reparameterize the model so that the uncertain parameters can be reasonably assumed to be independent.
4. Calibration methods and structural uncertainty

In the discussion of parameter fitting above the assumption is that input parameters are specified (together with appropriate correlations) within a given structure and the outcomes of interest for the decision model estimated. However, there is emerging interest in calibration methods that combine knowledge over model parameter inputs, model structure and model outputs (or calibration targets) that assist in ensuring the consistency of inputs and outputs. Common calibration targets include overall and disease-specific mortality, and event incidence rates.

Bayesian evidence synthesis (or multi-parameter evidence synthesis) is a form of calibration that has been developed in the health technology assessment space using Markov chain Monte Carlo (MCMC). This approach involves the specification of a model structure comprising input (or basic) parameters and functional parameters that are functions of multiple input parameters, for which an external data source exists (Ades & Cliffe, 2002). The MCMC estimates a joint set of posterior distributions for the model's input parameters, based on the likelihood of the functional parameter(s). Despite the increased viability of use by health services researchers, by way of the freely available MCMC software WinBUGS (Lunn et al 2009), most applied examples of Bayesian evidence synthesis have involved relatively simple model structures. For more complex models, standard calibration approaches can be applied to identify the best fitting set of input parameter values, or multiple sets of values, which can then form the basis for sensitivity analyses (Kim et al 2007; Karnon et al 2009). Steps in a calibration process include: identifying calibration targets; selecting individual and aggregated measures of goodness-of-fit (GOF); defining the parameter space; selecting a search strategy; defining convergence thresholds; and specifying a stopping rule (Vanni et al 2011).

The use of calibration methods to estimate parameters or adjust estimated values, emphasizes the important role the model structure plays in defining the relationship between model inputs and model outputs. Structural uncertainty is frequently ignored in the presentation of decision models, despite suggestions that it may have a much greater impact on results than parameter uncertainty (Brisson & Edmunds, 2006). Recent approaches to tackling this issue have sought to parameterize structural uncertainties into the model (Bojke et al, 2009; Chapman et al, in press). This is trivial for nested model structures (for example, a constant hazard function of an event of interest could be replaced by a more flexible hazard function) but is much more challenging for non-nested model structures which could require complete redesign/rebuilding of the model. While it may be feasible to internalize structural uncertainty by adding additional parameters to the model, any given research team will be limited in the extent to which they can fully incorporate this form of uncertainty. In such situations, analysts are encouraged to be as explicit as possible as to the structural assumptions made that might have a material impact on the overall findings and suggest alternative assumptions that future modeling exercises might employ.

Recommendation:

4.2 Where uncertainties in structural assumptions were identified in the process of conceptualizing and building a model, those assumptions should be tested in a sensitivity analysis. Consideration should be given to opportunities to parameterize these uncertainties for ease of testing. Where it is not possible to perform structural sensitivity analysis it is nevertheless important that analysts be aware of the potential for this form of uncertainty to
be at least as important as parameter uncertainty for the decision maker. (Linked to conceptual modeling recommendations)
5. Reporting Uncertainty Analyses

As has been emphasized earlier, uncertainty analyses can be either deterministic or probabilistic, and often it is appropriate to report aspects of both types within a single evaluation. For example, deterministic threshold analysis may be reported for key model parameters (e.g., for what parameter values the model outcome is less than a particular threshold or one alternative dominates another), while for the same study a full probabilistic sensitivity analysis might also be reported to convey the overall uncertainty in results. The guiding principle is that the method of reporting should be tailored to guide the decision that the analysis is meant to inform.

When additional assumptions or parameter values are introduced for purposes of uncertainty analyses, such as distributional parameters for probabilistic sensitivity analyses, or parameter ranges for deterministic sensitivity analyses, these values should be disclosed and justified. Technical appendices are often appropriate for this purpose. When model calibration is used to derive parameters, uncertainty around the calibrated values should also be reported, and this uncertainty should be reflected in either deterministic or probabilistic sensitivity analyses, or both. The remainder of this section comments on appropriate reporting of (a) deterministic sensitivity and threshold analyses, (b) probabilistic sensitivity analyses and associated value of information analyses, and (c) uncertainty surrounding calibrated models.

Reporting deterministic sensitivity analysis

Many different methods may be used to convey how results may depend on individual parameters, multiple parameters jointly, or model structure. One-way sensitivity analyses may be reported using a display known as a “tornado diagram” (Figure 5.1), because of its characteristic shape. The horizontal axis is the outcome of interest. Along the vertical dimension, different parameters are arrayed, and horizontal bars represent the range of the outcome measure associated with the specified range of each parameter. The point estimate of the outcome corresponding to base-case values of all parameters is indicated by a vertical line that cuts through all the horizontal bars. Most commonly, the longest bar, reflecting the parameter that generates the widest uncertainty in results, is placed at the top of the diagram, and the other bars are arrayed in descending order of length. A tornado diagram should be accompanied by a legend or table that indicates the ranges of values (upper and lower bounds) used for each parameter, with justification for those ranges.

A table may be used instead of a tornado diagram to convey the results of one-way sensitivity analyses, or the ranges of results of one-way sensitivity analysis may be described in the text of the report. For example, the text might state that “the outcome ranged from XXX to YYY when parameter Z was varied from A to B.”

Often, uncertainty in a parameter may be represented by several discrete values, instead of a continuous range. For example, outcomes from different clinical studies, utility surveys, or cost datasets may lead to different values. It is acceptable to report alternative outcomes under each of these discrete assumptions, instead of or in addition to tornado diagrams or uncertainty intervals.

Structural uncertainty, such as different qualitative assumptions about the relationships between biologic variables, or different functional forms of risk functions, to name only two examples, may be represented deterministically by reporting results under each discrete structural assumption or set
of structural assumptions. Quantitative sensitivity analyses may be embedded within structural
sensitivity analyses by reporting them separately under each possible structural assumption.

In presenting one-way sensitivity analyses in cost-effectiveness analyses, care should be taken to
avoid reporting negative Incremental Cost Effectiveness Ratios (ICERs), which are meaningless.
Instead, the range of ICERs should be limited to results corresponding to positive incremental health
consequences (QALYs) and costs -- quadrant I in the cost-effectiveness plane. Results for which
incremental costs are positive and incremental health consequences are negative should be
indicated qualitatively as “Dominated”, and results for which incremental costs are negative and
incremental health consequences are positive should be indicated as “Dominant”. ICERs
corresponding to negative incremental costs and health consequences – quadrant III – should be
distinguished from conventional ICERs (in quadrant I).

Threshold analyses are deterministic sensitivity analyses that answer the question: “For what
range(s) of parameter value(s) is a specified decision criterion met?” The decision criterion could be
an ICER less than some real or hypothetical threshold, or possibly dominance (cost savings with
positive net benefit). Threshold analyses may be applied to individual parameters one at a time, or
they may apply to combinations of parameters in two-way and multi-way sensitivity analyses.
Results of one-way threshold analyses are easily reported in text form, in statements such as, “The
ICER remains less than $100,000 per QALY as long as the value of X is greater than A”, or “Alternative
1 dominates Alternative 2 if the value of Z is less than B”.

Results of two-way and multi-way sensitivity analyses require graphical or tabular displays. An
example of a graphical display of a two-way threshold analysis is shown in Figure 5.2. The axes
represent possible values of the two parameters. The quadrant is partitioned into regions
corresponding to various cost-effectiveness results, the boundaries of which represent specified
ICER thresholds, or the thresholds of dominance of one comparator over the other. As in one-way
analyses, it is important to specify which alternative dominates the other in a region of the graph,
and which comparator is more effective and costly when an ICER threshold is indicated. Three-way
threshold analyses may superimposed on two-way threshold graphs by overlaying multiple
threshold curves on the same graph (Figure 5.3), but this often leads to visual overload and
confusion, and it only works if the third parameter can be represented as taking on discrete values.

Threshold analyses are especially useful, perhaps necessary, when reporting deterministic sensitivity
analyses in evaluations involving three or more comparators. In those situations, the relevant
question may be “Which alternative is cost-effective at a decision threshold of $XXX per QALY?” To
portray the answer to this question for two parameters, a threshold graph might partition the
quadrant to show which alternative is cost-effective at various ICER thresholds and for different
combinations of parameters (Figure 5.4). Results for different decision criteria (e.g., ICER
thresholds, or dominance) are best presented in separate panels of a graphical display.

When the base case result of an analysis strongly favors one alternative, then a threshold analysis
may be presented as a worst-case or “even if” analysis. For example, a statement of the following
type may be made: “Even if the risk reduction is as low as XXX, the ICER remains below $20,000 per
QALY”. Two-way or multi-way versions of “even if” statements might state: “Even if the relative risk
reduction with alternative A is as low as XXX and the cost of treatment is as high as $YYY, alternative
A dominates alternative B.”
Reporting Probabilistic Sensitivity Analyses

When a PSA is performed, the probability distributions around each parameter, or set of parameters, should be disclosed. This entails specifying the functional form (e.g., beta, normal, lognormal, chi-square) as well as the parameters used for each distribution. Sometimes it is feasible to incorporate this information into the table that is used to list model parameters. If not, a table that details the probability distributions and their parameters may be included in a technical appendix.

In addition to reporting the parameters of the probability distributions, justification for these parameters should be provided. This justification may come from empiric data from a clinical study, from which the mean and standard error (or full likelihood function) are used as the mean and standard deviation (or probability distribution) in the PSA. Realistically, published sources do not always report sampling distributions around all endpoints, so subjective assessments may be the only possibility. In those cases, these probability distributions should be identified as “subjective”. Since such subjective distributions represent the analyst’s (or an expert’s) personal judgment, and need not correspond to the perceptions of decision makers, alternative specifications and parameters should be provided so that users of the analysis can select the parameter distributions that most closely reflect their own subjective judgments. A rule of reason applies in this regard: parameters that exert little leverage on the overall model uncertainty can be left as subjective. A third approach to parameterizing the probability distributions, in addition to using sampling distributions from clinical studies or subjective judgments, is to perform a full Bayesian evidence synthesis. While a full Bayesian evidence synthesis is theoretically the best approach, it is impractical in most situations owing to incomplete data and computational challenges.

Perhaps the best measure of uncertainty surrounding a choice in cost-effectiveness analysis is the expected value of perfect information (EVPI), since this measure combines the probability of incorrect decision making with the consequential loss function. The higher the EVPI, the larger the ex-ante opportunity cost of an incorrect decision, i.e., the more costly is the uncertainty. Total EVPI is commonly reported in monetary terms, using Net Monetary Benefit. (An alternative is to measure EVPI in QALYs, using Net Health Benefit.) Since both Net Monetary Benefit and Net Health Benefit depend on the cost-utility threshold, EVPI should be reported in the text at a specified cost-per QALY value or values, or in graphical form as a function of cost-per QALY values (Figure 5.5).

Expected values of partial perfect information (EVPPI) for key parameters or sets of parameters should be reported in the same way as total EVPI. Expected value of sample information (EVSII) analyses should be reported similarly, but with the additional proviso that the parameters governing the assumed study from which sample information is obtained (e.g., sample size, individual-level or patient-level variation in outcomes) must also be specified.

When a PSA is performed without an accompanying value of information analysis, options for presenting results include cost-effectiveness acceptability curves (CEACs), and distributions of net monetary benefit or net health benefit. When more than two comparators are involved, CEACs for each comparator should be plotted on the same graph (Figure 5.6. shown with EVPI analysis overlaid).
Reporting uncertainty owing to calibrated parameters

When model calibration is used to estimate parameters that are not directly observable from data, uncertainty owing to the calibration process should be reported. As for other model parameters, such reporting may be either deterministic or probabilistic.

Deterministic reporting of calibration-related uncertainty entails repeating the calibration under different sets of assumptions, and reporting the range of results. Results include both the range of calibrated parameter sets themselves (as model inputs) and the resulting range of model outputs such as incremental cost-effectiveness ratios. Examples of different assumptions that enter into a calibration process include, but are not limited to: (1) objective function used for evaluating goodness of fit, (2) computational process or algorithm used to determine the best-fitting parameter set, and (3) choice of target data for calibrated endpoints. The results across this range of calibration scenarios may be reported as discrete point estimates, or as a range. Threshold analyses may be performed for calibrated parameters as for directly estimated parameters.

Probabilistic reporting of calibration-related uncertainty can be done in many ways. If a formal Bayesian approach to calibration is used, then the posterior distributions of the calibrated parameters should be reported. These posterior distributions should then be used as inputs to the overall PSA, along with the probability distributions of the directly estimated parameters. If a less formal approach to calibration is used, then a discrete distribution of parameter estimates can be generated based on (a) the various approaches taken to calibration, including goodness-of-fit measure, computational method, and target endpoint values, and possibly (b) the best N fitting parameter sets within each calibration method. The discrete distribution may assign equal probability to each of the resulting sets of parameter values, or if unequal probabilities can be justified, they may be used.

Recommendations:

5.1 Uncertainty analyses can be either deterministic or probabilistic, and often it is appropriate to report aspects of both types within a single evaluation. Tornado diagrams, threshold plots, or simple statements of threshold parameter values, are all appropriate ways of reporting results from deterministic sensitivity analyses.

5.2 When additional assumptions or parameter values are introduced for purposes of uncertainty analyses, such as distributional parameters for probabilistic sensitivity analyses, or parameter ranges for deterministic sensitivity analyses, these values should be disclosed and justified. Technical appendices are often appropriate for this purpose.

5.3 When model calibration is used to derive parameters, uncertainty around the calibrated values should also be reported, and this uncertainty should be reflected in either deterministic or probabilistic sensitivity analyses, or both.

5.4 When the purpose of a probabilistic sensitivity analysis is to guide decisions about acquisition of information to reduce uncertainty, results should be presented in terms of expected value of information.
For economic studies, when a PSA is performed without an accompanying expected value of information analysis, options for presenting results include cost-effectiveness acceptability curves (CEACs), and distributions of net monetary benefit or net health benefit. When more than two comparators are involved, CEACs for each comparator should be plotted on the same graph.
References

model: consistency of evidence and the accurate assessment of uncertainty. Med. Decsn Makng, 22,
359–371.

Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic


Brisson M & Edmunds WJ. Impact of Model, Methodological, and Parameter Uncertainty in the
Economic Analysis of Vaccination Programs. Medical Decision Making 2006;26:434–446.


Griffin S, Claxton K, Hawkins N, Sculpher MJ. Probabilistic analysis and computationally expensive

Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MGM: Uncertainty and


Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural
uncertainty in decision models. Medical Decision Making (in press).

Karnon J, Czoski Murray C, Smith KJ, Brand C, A hybrid cohort individual sampling natural history
model of age-related macular degeneration: assessing the cost-effectiveness of screening using
probabilistic calibration, Medical Decision Making 2009, 29: 304-316.


future directions". Statistics in Medicine 28: 3049–3067

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E.
and Sons, 2006.

Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, Legood R, Calibrating models in
Figure 5.1: Tornado diagram showing impact of uncertainty on the outcome of a decision model (in descending order)

Figure 5.2: Two-way threshold analysis
Figure 5.3: Three-way threshold analysis
Figure 5.4: Two-way threshold graph with 3 or more comparators

(a) Two-way SA for three treatments at $50,000 WTP for a QALY

(b) Two-way SA for three treatments at $100,000 WTP for a QALY
Figure 5.5: EVPI as a function of the cost-per-QALY threshold

Figure 5.6: CEACs with 3 or more comparators