

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6

Andrew H. Briggs, DPhil^{1,*}, Milton C. Weinstein, PhD², Elisabeth A.L. Fenwick, PhD¹, Jonathan Karnon, PhD³, Mark J. Sculpher, PhD⁴, A. David Paltiel, PhD⁵, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force

¹Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow, Glasgow, UK; ²Harvard School of Public Health, Boston, MA, USA; ³School of Population Health and Clinical Practice, University of Adelaide, SA, Australia; ⁴Centre for Health Economics, University of York, York, UK; ⁵Yale School of Medicine and Yale School of Management, New Haven, CT, USA

ABSTRACT

A model's purpose is to inform medical decisions and health care resource allocation. Modelers employ quantitative methods to structure the clinical, epidemiological, and economic evidence base and gain qualitative insight to assist decision makers in making better decisions. From a policy perspective, the value of a model-based analysis lies not simply in its ability to generate a precise point estimate for a specific outcome but also in the systematic examination and responsible reporting of uncertainty surrounding this outcome and the ultimate decision being addressed. Different concepts relating to uncertainty in decision modeling are explored. Stochastic (first-order) uncertainty is distinguished from both parameter (second-order) uncertainty and from heterogeneity, with structural uncertainty relating to the model itself forming another level of uncertainty to consider. The article argues that the estimation of point estimates and uncer-

tainty in parameters is part of a single process and explores the link between parameter uncertainty through to decision uncertainty and the relationship to value of information analysis. The article also makes extensive recommendations around the reporting of uncertainty, in terms of both deterministic sensitivity analysis techniques and probabilistic methods. Expected value of perfect information is argued to be the most appropriate presentational technique, alongside cost-effectiveness acceptability curves, for representing decision uncertainty from probabilistic analysis.

Keywords: best practices, heterogeneity, sensitivity analysis, uncertainty analysis, value of information.

Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Background to the Task Force

A new Good Research Practices in Modeling Task Force was approved by the International Society for Pharmacoeconomics and Outcomes Research Board of Directors in 2010, and the Society for Medical Decision Making was invited to join the effort. The Task Force coauthors and members are expert developers and experienced model users from academia, industry, and government, with representation from many countries. Several teleconferences and hosted information sessions during scientific meetings of the Societies culminated in an in-person meeting of the Task Force as a whole, held in Boston in March 2011. Draft recommendations were discussed and subsequently edited and circulated to the Task Force members in the form of a survey where each one was asked to agree or disagree with each recommendation, and if the latter, to provide the reasons. Each group received the results of the survey and endeavored to address all issues. The final drafts of the articles were available on the ISPOR and Society for Medical Decision Mak-

ing Web sites for general comment. A second group of experts was invited to formally review the articles. The comments received were addressed, and the final version of each article was prepared. (A copy of the original draft article, as well as the reviewer comments and author responses, is available at the ISPOR Web site: <http://www.ispor.org/workpaper/Model-Parameter-Estimation-and-Uncertainty-Analysis.asp>). A summary of these articles was presented at a plenary session at the ISPOR 16th Annual International Meeting in Baltimore, MD, in May 2011, and again at the 33rd Annual Meeting of the Society for Medical Decision Making in Chicago, IL, in October 2011. These articles are jointly published in the Societies' respective journals, *Value in Health* and *Medical Decision Making*. Other articles in this series [1–6] describe best practices for conceptualizing models, building and applying particular types of models, and transparency and validation. This article addresses best practices for parameter estimation and uncertainty analysis and is intended to apply to all types of models. Examples are cited throughout, without implying endorsement or preeminence of the articles referenced.

* Address correspondence to: Andrew H. Briggs, Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ, UK.

E-mail: andrew.briggs@glasgow.ac.uk.

1098-3015/\$36.00 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2012.04.014>

Introduction

This report adopts the Task Force's view that a model's purpose is to inform medical decisions and health care resource allocation. Modelers employ quantitative methods to structure the clinical, epidemiological, and economic evidence base and gain qualitative insight to assist decision makers in making better decisions. From a policy perspective, a model-based analysis's value lies not simply in its ability to generate a precise point estimate for a specific outcome but also in the systematic examination and responsible reporting of uncertainty surrounding this outcome and the ultimate decision being addressed. These are the hallmarks of good modeling practice.

The extent to which an uncertainty analysis can be considered "fit for purpose" in part depends on the decision(s) the modeling seeks to support. Uncertainty analysis can serve two main purposes: assess confidence in a chosen course of action and ascertain the value of collecting additional information to better inform the decision.

Many models are designed to help decision makers maximize a given outcome (e.g., cases identified in a screening model or quality-adjusted life-years in a cost-effectiveness model), subject, perhaps, to one or more limiting constraints (such as a fixed budget). The model generates point estimates of the outcome for each possible course of action; the "best" choice is the one that maximizes the outcome subject to the constraint. 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 the decision should be based only on expected values (although some commentators have argued that for nonlinear models, probabilistic sensitivity analysis (PSA) is required to generate appropriate expected values [7]). Nevertheless, decision makers may want to gauge confidence in the "best choice's" appropriateness by exploring its robustness to changes in the model's inputs.

Increasingly, models are developed to guide decisions of particular bodies (e.g., organizations responsible for deciding whether to reimburse a new pharmaceutical). Such decision makers who have the authority to delay decisions or review them later, based on research they commission or mandate, should be interested not just in expected cost-effectiveness but also in a thorough uncertainty analysis 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 undertaking it, can influence the array of decisions available. Thus, 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.

Many models are developed for general dissemination, without a specific decision maker in mind. Such models could inform a range of decision makers with varying responsibilities. Here, there is a case for undertaking a full uncertainty analysis, thus allowing different decision makers to take from the analysis what they require given the decisions with which they are charged.

Best practices

VI-1 *The systematic examination and responsible reporting of uncertainty are hallmarks of good modeling practice. All modeling studies should include an uncertainty assessment as it pertains to the decision problem being addressed.*

VI-2 *The decision-maker's role should be considered when presenting uncertainty analyses. The analytic perspective description should include an explicit statement regarding what is assumed about the decision-makers' power to delay or review decisions and to commission or mandate further research.*

Background and Terminology

It is important to be precise concerning the terminology used in this article, which is sometimes confused in the literature (reflecting the multidisciplinary nature of decision modeling in health care). In particular, stochastic (first-order) uncertainty is distinguished from both parameter (second-order) uncertainty and from heterogeneity. Furthermore, each concept is argued to have an analogous form within a "regression-type" model in statistics. As in regression analysis, the structural uncertainty associated with the model itself must also be considered. Table 1 summarizes the concepts used here and preferred terminology, lists other terms used, and provides the link to statistical regression.

The term "parameter uncertainty" is not the same as the uncertainty around the realization of individual events or outcomes. This "stochastic uncertainty" relates to the fact that individuals facing the same probabilities and outcomes will experience the effects of a disease or intervention differently, just as a fair coin might come up heads or tails on any given toss (e.g., the first patient in a sample might respond to a treatment but the next may not; the first may not experience an adverse effect but the second may; the first may stay in hospital for 2 days and the second for 3 days). Parameter uncertainty ("second-order uncertainty") relates to the fact that the probabilities that govern outcomes are themselves uncertain, because they are estimated quantities (e.g., 100 tosses of a fair coin will not always lead to 50 realizations of "heads" and fifty "tails"). Estimates of the probability of "heads" based on 100 observations are uncertain. The sample size informing that estimate and variance in the data contribute to determining the parameter uncertainty. Parameter uncertainty also arises

Table 1 – Uncertainty for decision modeling: Concepts and terminology.

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

from the existence of multiple, conflicting studies, problems with a source study's internal or external validity, generalizability from a study to a real-world setting, and lack of empirical data. Although these two types of uncertainty are clearly linked, uncertainty about actual outcomes for individuals differs from uncertainty surrounding population parameter values (e.g., the response probability, adverse event probability, or mean length of hospital stay). The distinction is analogous to the difference between standard deviation (estimate of how individual observations within a sample vary) and standard error (precision of an estimated quantity).

Parameter uncertainty may be represented via deterministic sensitivity analysis (DSA) or via PSA. In a DSA, parameter values are varied manually to test the sensitivity of the model's results to specific parameters or sets of parameters. In a PSA, (preferably) all parameters are varied simultaneously, with multiple sets of parameter values being sampled from *a priori* defined probability distributions. The outputs from a PSA may inform several different forms of analysis, including confidence intervals, cost-effectiveness planes (showing the distributions of costs and effects for each evaluated technology or service), cost-effectiveness acceptability curves (showing the probability of cost-effectiveness for each option), and value of information analyses. The latter involves the estimation of the expected value of perfect information (EVPI), which may be estimated for the model as a whole or for specific parameters or sets of parameters (expected value of partial perfect information).

Further adding to the confusion between first- and second-order (parameter) uncertainty is the use of "variability" to refer to the former but also to differences in parameter values across patients or patient subgroups. "Heterogeneity" refers to the extent to which between-patient variability can be explained by patients' characteristics (e.g., age- and sex-specific mortality). Its relevance lies in the identification of subgroups for whom separate cost-effectiveness analyses should be undertaken. Such analyses may inform alternative decisions regarding the service provision to each subgroup, or contribute to a weighted analysis of the aggregate group.

An analogy is a simple regression model of the form:

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

where an outcome variable Y depends on covariates X . The vector of coefficients β represents the model parameters and is estimated with uncertainty represented by the coefficients' standard error from the fitted regression. The extent to which the predicted values (\hat{Y}) vary with known covariates X represents heterogeneity and the stochastic error term ε represents unexplained variability, which we call stochastic uncertainty.

Just as a linear regression imposes a structural relationship between independent and dependent variables, so a decision-analytic model is characterized by assumptions reflected in its structure but not formally expressed numerically (e.g., types of adverse events included, duration of treatment effects, time dependency of probabilities, and prognostic implications of surrogate end points or clinical events). Although these structural assumptions are not typically formally quantified, it is uncertain whether they express reality accurately. As such, and analogously with statistical modeling, any representation of uncertainty in a decision model is conditional on its structural assumptions. Therefore, in principle, the structural characteristics are a further level of uncertainty to be considered.

Although a model's overall structure results from many assumptions and analytic decisions, it is useful to distinguish two broad categories of models that reflect both the underlying structure and relate to the uncertainty concepts outlined above. Patient-level stochastic simulations (e.g., discrete-event simulations [3] and state-transition microsimulations [2]) are structured around events occurring at the individual level and requiring simulation of numerous virtual patients. For these models, assess-

ment of parameter uncertainty requires elimination of stochastic uncertainty (sometimes called Monte Carlo error). In cohort models, parameter uncertainty can be addressed without concern for stochastic uncertainty.

Methodological uncertainty has been identified as a specific type [8,9] that can be as important as parameter uncertainty [10]. In common with others [9,11], we think a "reference case" should be applied. Nevertheless, disagreement about appropriate methodology may be a reason to undertake sensitivity analysis.

Best practices

VI-3 Terminology to describe concepts relating to parameter estimation and representation of uncertainty varies within the health-care 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.

Parameter Estimation and Uncertainty Analysis

All models have parameters that must be estimated. In doing so, analysts should conform to evidence-based medicine principles (e.g., seek to incorporate all evidence, rather than selectively picking a single source; use best-practice methods to avoid potential biases, as when estimating treatment effectiveness from observational sources; employ formal evidence synthesis techniques [12,16]). Uncertainty analysis is equally integral: the steps 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 multivariable framework, although the latter will also provide a measure of covariance between estimated parameters. Consequently, whether primary data sources are used to estimate input parameters or information derives from one or more secondary source, the estimation generates a point estimate, a measure of precision, and, potentially, one of covariance. These types of information from estimation should feed directly into the uncertainty analysis.

This is true whatever the uncertainty analysis's technical specification. For a one-way DSA, it is necessary to specify the parameter's point estimate and a defensible range; these may be taken directly from the estimation process, with the latter based, for example, on a 95% confidence interval. A two-way uncertainty analysis will be more useful if informed by the covariance between the two parameters of interest, or on the logical relationship between them (e.g., a two-way uncertainty analysis might be represented by the control intervention event rate and the hazard ratio with the new treatment).

Representation of uncertainty depends on the uncertainty analysis planned. For DSA, an interval estimate representing beliefs about the parameter's plausible range is required. For PSA, a distribution is specified via its parameters.

Best practices

VI-4 All decision models have parameters that need to be estimated. Populating models with parameter estimates should conform to evidence-based medicine principles.

Consistency of approach between deterministic and probabilistic uncertainty analysis

Uncertainty estimates for parameters estimated (or estimable) from data should be consistent with standard statistical approaches. The underlying distributional assumption used to calculate the 95% confidence interval can be the basis for an uncertainty analysis distributional assumption. One exception is when

taking a formal Bayesian approach involving subjective prior information, where the standard distributional assumption relates to the data likelihood combined with the prior information to form the posterior density. In this case, consistency between DSA and the fully Bayesian probabilistic approach would be retained if the DSA interval estimate were based on the Bayesian posterior's 95% highest density region [13].

Some uncertainty analyses do not require formal ascertainment of parameter uncertainty. These include threshold analysis, where the parameter's value needed to change the decision is identified. This is closely linked to "even if" approaches that identify extreme parameter values that still do not change the decision—with the implication that these parameters are unlikely to influence the decision. Such analyses may be sufficient if there is little decision uncertainty given reasonable assumptions about parameter uncertainty. Another form not requiring uncertainty estimation is identification of the quantitative relationships between inputs and outputs (e.g., it might be determined that a 10% increase in a particular parameter's value leads to a 20% increase in expected effectiveness and a 5% decrease in expected cost). This sort of analysis is unlikely to be sufficient since a parameter with low sensitivity, but highly uncertain, could easily have more impact on the model outputs than a more sensitive parameter but estimated more precisely. Therefore, completely arbitrary analyses, such as the effect on outputs of varying each input by $\pm 50\%$, are not recommended as a representation of uncertainty.

Best practices

VI-5 *Whether employing deterministic methods (point estimate and range) or probabilistic (parameterized distribution), the link to the underlying evidence base should be clear.*

VI-6 *While completely arbitrary analyses, like presentation of the effect on outputs of varying inputs by $\pm 50\%$, can be used as a measure of sensitivity; such analyses should not be used to represent uncertainty.*

Estimation and choice of distribution for PSA and interval estimation

If there is much information available to inform a parameter's estimate, then by the central limit theorem—the sampling distribution of the arithmetic mean will follow a normal distribution (with sufficient sample size), whatever the data's underlying distribution—the normal distribution can be used in a PSA and a standard confidence interval in a DSA. 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. Nevertheless, the reality of multiple data sources suggests that reliance on a single study is likely to underestimate uncertainty. This suggests a broader uncertainty analysis than based on study data alone, leaving open some subjective assessment for ensuring appropriate representation of uncertainty, even where single large studies are available for estimating parameters and their associated uncertainty.

The general principle remains that assumptions for specifying the distribution and/or defining the interval for uncertainty analysis should follow standard statistical methods (e.g., beta distributions are a natural match for binomial data; gamma or log normal for right skew parameters; log normal for relative risks or hazard ratios; logistic for odds ratios [14]). These distributions can be used directly in PSA or to define the interval (plausible range) for a DSA.

Sometimes there is very little information on a parameter, because either there are very few studies informing the estimation or there are no data and expert opinion must be relied upon. Here, it is imperative that uncertainty related to such estimates be fully explored. A conservative approach should be adopted with an appropriately broad range of possible estimates elicited from each

expert, reflected in how opinions are combined across experts and incorporated into the uncertainty analysis. On no account should parameters be excluded from an uncertainty analysis on the grounds that "there is not enough information to estimate uncertainty." Continuous distributions providing a realistic portrayal of uncertainty over the parameter's theoretical range should be favored in PSA. Hence, careful consideration should be given to whether convenient-to-fit, but implausible, distributions (e.g., uniform or triangular) should have any role in PSA. Formal methods for eliciting probability distributions from experts have been developed [15].

Best practices

VI-7 *Use commonly adopted statistical standards for point and interval estimation (e.g., 95% confidence intervals, or distributions based on agreed statistical methods for a given estimation problem). Where departures from these standards are deemed necessary (or no such standard exists), these should be justified.*

VI-8 *Where there is very little information on a parameter; adopts a conservative approach such that the absence of evidence is reflected in a very broad range of possible estimates. Never exclude parameters from uncertainty analysis on the grounds that there insufficient information to estimate uncertainty.*

VI-9 *Favor continuous distributions that portray uncertainty realistically over the theoretical range of the parameter. Careful consideration should be given to whether convenient-to-fit but implausible distributions (such as the Triangular) should have a role in PSA.*

Multivariate estimation and correlation

When regression is used to capture the effect of subject characteristics on parameter estimates, the dependent variable is a functional parameter of the regression coefficients. Therefore, uncertainty in the functional parameter can be defined in terms of uncertainty (and correlation) in the coefficients. The covariance matrix defines these uncertainties, and the assumption of multivariate normality is appropriate for the regression's linear predictor. These can be used to specify the interval for DSA or as the basis for PSA [14].

In PSA, parameters are typically not all independent of one another. For example, if two uncertain parameters are disease progression probabilities with and without treatment, part of the uncertainty may derive from doubts regarding the disease's natural history and part from imprecise measurement of treatment efficacy. The component related to natural history would affect the progression probabilities with or without treatment, whereas the component related to efficacy would affect the relationship (e.g., relative risk) between progression probabilities with and without treatment. It would be wrong to regard the progression probabilities as coming from independent distributions and conduct the PSA accordingly. It might be reasonable, however, to regard the natural history progression probability and the risk reduction with treatment as independent. Parameter distributions in this situation should be defined 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 PSA, with the on-treatment progression probability derived as their product. Where this application of relative risks can result in "out-of-range" parameters, consider switching to odds ratios. 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 [14].

Best practices

VI-10 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 to include a correlation coefficient as a parameter where concern exists that an unknown correlation could be important; or to re-parameterize the model so that that the uncertain parameters can be reasonably assumed to be independent.

Calibration Methods and Structural Uncertainty

There is emerging interest in calibration methods that combine knowledge over parameter inputs, structure, and outputs (or calibration targets) to assist in ensuring consistency of inputs and outputs. Common calibration targets include overall and disease-specific mortality and event incidence rates.

Bayesian (or multiparameter) evidence synthesis is a calibration approach developed in health technology assessment by using Markov chain Monte Carlo via freely available WinBUGS software [16]. This approach involves specification of a structure comprising inputs and parameters that are functions of multiple inputs, for which an external data source exists [17]. The Markov chain Monte Carlo estimates a joint set of posterior distributions for the input parameters, based on the functional parameters' likelihood. Most applied examples of this approach have involved relatively simple structures. For more complex models, standard calibration approaches can be applied to identify the best fitting set of inputs, or multiple sets of values, which can then form the basis for uncertainty analysis [18–21]. Steps in calibration include identifying calibration targets, selecting individual and aggregated measures of goodness of fit, defining the parameter space, selecting a search strategy, defining convergence thresholds, and specifying a stopping rule [22].

The use of calibration to estimate parameters or adjust estimated values emphasizes the important role model structure plays in defining the relationship between inputs and outputs. Structural uncertainty is frequently ignored though it may have a much greater impact on results than parameter uncertainty [10]. Recent approaches to this issue have sought to parameterize structural uncertainties into the model [23–25]. This is trivial for nested structures (e.g., a constant hazard function could be replaced by a more flexible function) but is much more challenging for nonnested structures, which could require complete redesign/rebuilding of the model. While it may be feasible to internalize structural uncertainty by adding parameters to the model, any given research team will be limited in the extent to which it can fully incorporate this form of uncertainty. In such situations, analysts are encouraged to be as explicit as possible regarding the structural assumptions that might have an impact on the findings and suggest alternative assumptions that future modeling exercises might employ.

Best practices

VI-11 Where uncertainties in structural assumptions were identified in the process of conceptualizing and building a model, those should be tested in uncertainty analysis. Consideration should be given to opportunities to parameterize these uncertainties for ease of testing. Where it is impossible to perform structural uncertainty analysis, it is important to be aware that this uncertainty may be at least as important as parameter uncertainty.

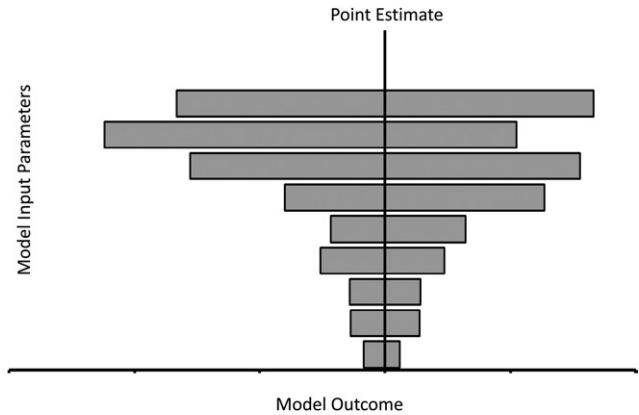


Fig. 1 – Tornado diagram showing the impact of uncertainty on the outcome of a decision model.

Reporting Uncertainty Analyses

Often, it is appropriate to report aspects of both DSA and PSA (e.g., may report deterministic threshold analysis for key parameters and PSA to convey overall uncertainty). The guiding principle is that the reporting method should be tailored to guide the decision at issue.

When additional assumptions or parameter values are introduced for purposes of uncertainty analysis (e.g., distribution parameters for PSA, parameter ranges for DSA), these should be disclosed and justified. Technical appendices are often appropriate for this purpose. When calibration is used to derive parameters, uncertainty around the calibrated values should be reported, and this uncertainty should be reflected in DSA, PSA, or both.

Reporting deterministic sensitivity analysis

Many methods may be used to convey how results depend on individual parameters, multiple parameters jointly, or model structure. One-way DSA may be reported by using a “tornado diagram” (Fig. 1). The horizontal axis is the outcome; along the vertical, parameters are arrayed, and horizontal bars represent the outcome range associated with the specified each parameter's range. The outcome point estimate corresponding to base-case values is indicated by a vertical line cutting through all horizontal bars. Commonly, the longest bar (reflecting the parameter generating the widest uncertainty) is placed at the top and the other bars are arrayed in descending order of length. A tornado diagram should be accompanied by a legend or table indicating the upper and lower bounds of values for each parameter, with their justification in terms of the evidence base. A table may be used instead of a tornado diagram or the results ranges provided in the text of the report (e.g., the text might state that “the outcome ranged from X to Y when parameter Z was varied from A to B”). It is important that the range A to B represent a defensible range for parameter Z, not an arbitrary one.

Often, uncertainty in a parameter may be represented by several discrete values, instead of a continuous range, sometimes called “scenario analyses” (e.g., evidence from clinical studies, utility surveys, or cost data sets may lead to different values). It is acceptable to report alternative outcomes under each of these discrete assumptions to complement other uncertainty analysis.

Structural uncertainty may be represented deterministically by reporting results under each set of structural assumptions. Quantitative uncertainty analysis may be embedded within structural uncertainty analysis by reporting them separately under each structural assumption.

In presenting one-way uncertainty analysis, reporting negative incremental cost-effectiveness ratios (ICERs) should be avoided as

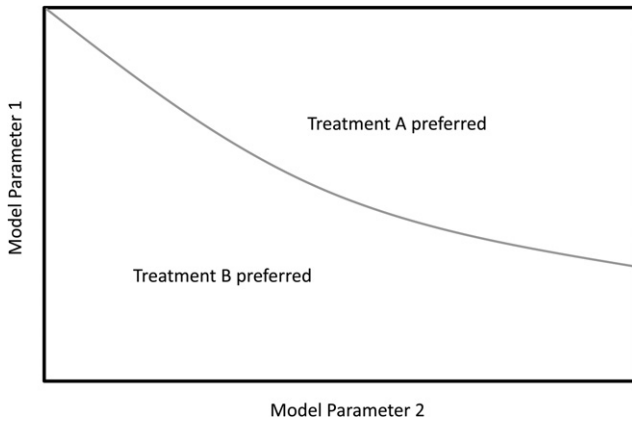


Fig 2 – Two-way threshold analysis.

they are meaningless [26,27]. Instead, the ICER range should be limited to results corresponding to positive incremental health consequences and costs—quadrant I in the cost-effectiveness plane. Results for which incremental costs are positive and health consequences negative should be indicated qualitatively as “dominated” and those with negative incremental costs and positive health consequences as “dominant.” ICERs corresponding to negative incremental costs and health consequences—quadrant III—should be distinguished from ICERs in quadrant I.

Results of one-way threshold analyses are easily reported in text (e.g., “The ICER remains less than Y as long as the value of X is greater than A,” or “Alternative 1 dominates 2 if the value of Z is less than B”). Results of two-way and multiway uncertainty analysis require graphical or tabular displays (Fig. 2). The axes represent possible values, and the quadrant is partitioned into regions corresponding to various ICERs, the boundaries representing specified ICER thresholds, or thresholds of dominance. As in one-way analyses, it is important to specify which alternative dominates and which comparator is more effective and costly when an ICER threshold is indicated. Three-way threshold analyses may be superimposed on two-way graphs by overlaying threshold curves (Fig. 3), but this often leads to visual overload and confusion and will work only if the third parameter can be represented as taking on discrete values.

Threshold analyses are especially useful, perhaps necessary, when reporting DSA involving three or more comparators. In those situations, the relevant question may be “Which alternative is cost-effective at a threshold of X?” To portray the answer

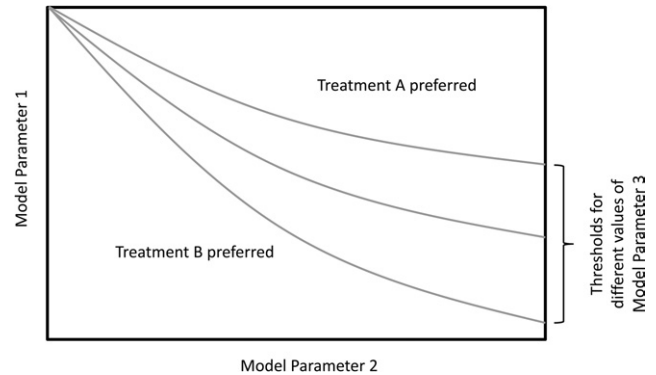


Fig 3 – Three-way threshold analysis.

for two parameters, partition the quadrant to show which alternative is cost-effective at various ICER thresholds and for different combinations of parameters (Figs. 4a and b). 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, a threshold analysis may be presented as a worst-case or “even if” analysis (e.g., “Even if the risk reduction is as low as X, the ICER remains below Y” or “Even if the relative risk reduction with alternative A is as low as X and the cost of treatment is as high as Y, alternative A dominates B.”) Threshold values can easily be combined with the tornado presentation by marking them on the horizontal bars.

Reporting probabilistic sensitivity analysis (PSA)

When reporting a PSA, the specific distribution (e.g., beta, normal, lognormal) as well as its parameters should be disclosed. Sometimes, it is feasible to incorporate this information into the table listing model parameters. If not, a table detailing the distributions may be included in a technical appendix. Justification for the distributions chosen should be provided. This may be directly from empiric data, a full Bayesian evidence synthesis, or subjective. As personal judgments need not correspond to the decision makers’ perceptions, alternative specifications and parameters should be provided so that users can select the distributions most closely reflecting their own judgments. A rule of reason applies in this regard: parameters that exert little leverage on the overall uncertainty can be left as subjective.

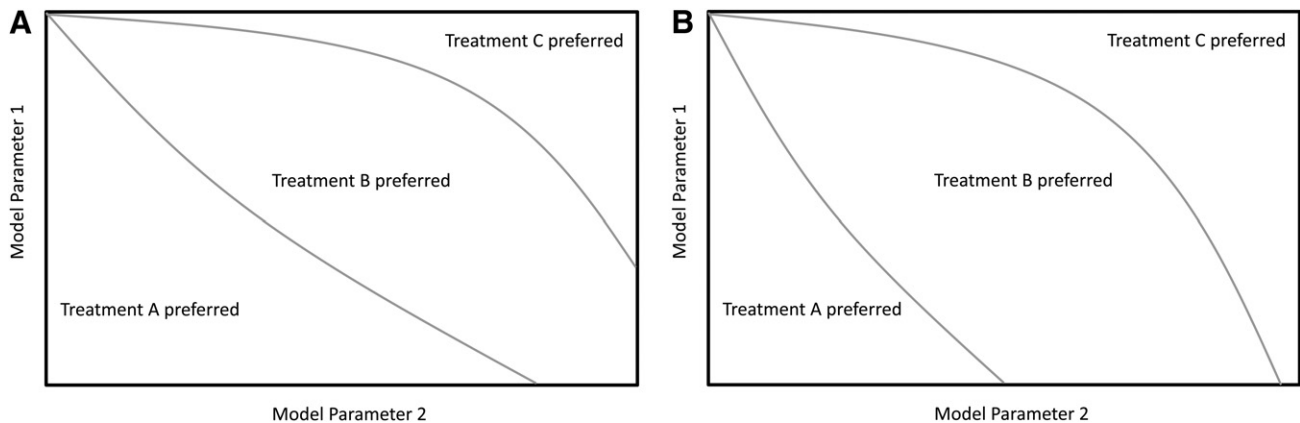


Fig. 4 – Two-way threshold graph with 3 or more comparators. (A) Two-way SA for 3 treatments at \$50,000 WTP for a QALY. (B) Two-way SA for 3 treatments at \$100,000 WTP for a QALY.

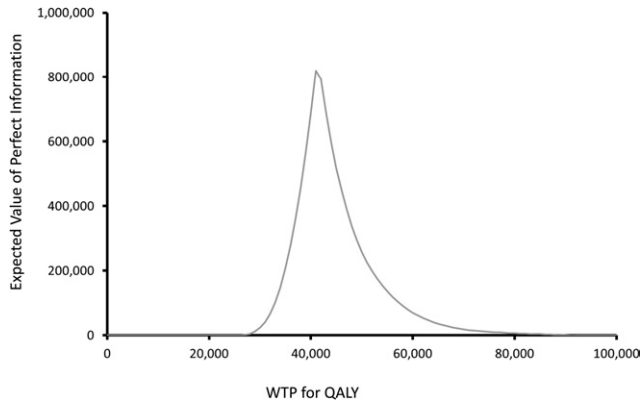


Fig. 5 – EVPI as a function of the cost-per-QALY threshold. EVPI, expected value of perfect information; QALY, quality adjusted life-year; WTP, willingness to pay.

Perhaps the best measure of uncertainty surrounding a particular decision in cost-effectiveness analysis is the EVPI, since this measure combines the probability of incorrect decision making with the consequential loss function [28]. The higher the EVPI, the larger the opportunity cost of an incorrect decision viewed at the point at which the uncertain decision is being made (i.e., the more costly is the uncertainty). Total EVPI is commonly reported in monetary terms, using net monetary benefit (an alternative is to express EVPI by using net health benefit.) Since both net monetary benefit and net health benefit depend on the ICER threshold, EVPI should be reported for specified ICER threshold(s), or in graphical form as a function of ICER thresholds (Fig. 5).

Expected value of partial perfect information can be estimated to identify key parameters and should be reported as for EVPI. Because of the likely correlation between individual parameters in expected value of partial perfect information, it may be preferable to report values for groups of parameters, which might also be the focus of future research effort [29,30]. Expected value of sample information analyses [31] should be reported similarly, but with

the additional proviso that the factors 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 and distributions of net monetary benefit or net health benefit. When more than two interventions are involved, cost-effectiveness acceptability curves for each should be plotted on the same graph (Fig. 6), with or without the inclusion of a cost-effectiveness acceptability frontier.

Reporting uncertainty owing to calibrated parameters

When model calibration is used to estimate parameters 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 involves presentation of the range of calibrated input parameter values across the convergent parameter sets and the resulting range of outputs (e.g., ICERs). The results may be reported as discrete point estimates or as a range.

Probabilistic calibration-related uncertainty can be reported in many ways. If a formal Bayesian approach to calibration is used, then the posterior distributions of the calibrated parameters should be reported. If a less formal approach to calibration is used, then a discrete joint distribution of parameter estimates can be generated on the basis of all convergent input parameter sets. The discrete distribution may assign equal probability to each resulting parameter value set, or probability weights may be applied that reflect the relative goodness of fit of the component parameter sets.

Analogous to the reporting of structural uncertainty, the results of separate calibration analyses using alternative methodological approaches should be reported under each discrete calibration process. Alternative approaches may include the objective function used for evaluating goodness of fit, the computational process or algorithm used to identify convergent parameter sets, and the importance weights attached to different calibration targets.

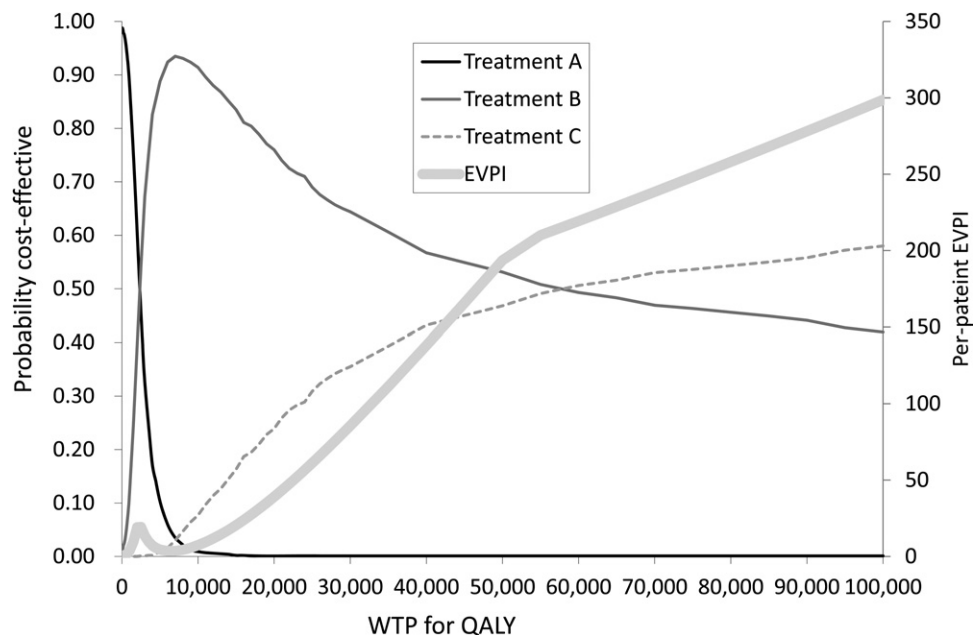


Fig. 6 – CEACs with three or more treatment strategies. CEACs, cost-effectiveness acceptability curves; EVPI, expected value of perfect information; QALY, quality adjusted life-year; WTP, willingness to pay.

Best practices

VI-12 It is appropriate to report both deterministic and probabilistic uncertainty analyses 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.

VI-13 When additional assumptions or parameter values are introduced for purposes of uncertainty analyses; these values should be disclosed and justified.

VI-14 When model calibration is used to derive parameters; uncertainty around the calibrated values should be reported and reflected in deterministic or probabilistic sensitivity analyses, or both.

VI-15 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.

VI-16 For economic studies, when a probabilistic SA is performed without an accompanying expected value of information analysis, options for presenting results include cost-effectiveness acceptability curves and distributions of net monetary benefit or net health benefit. When more than two comparators are involved, curves for each comparator should be plotted on the same graph.

Source of financial support: The authors have no other financial relationships to disclose.

REFERENCES

- [1] Caro JJ, Briggs AH, Siebert U, Kuntz K. Modeling good research practices—overview: a report of the ISPOR-SMDM modeling good research practices task force-1. *Value Health* 2012;15:796–803.
- [2] Roberts M, Russel L, Paltiel AD, et al. Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. *Value Health* 2012;15:6:804–811.
- [3] Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. *Value Health* 2012;15:812–820.
- [4] Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task force-4. *Value Health* 2012;15:821–827.
- [5] Pitman R, Fisman D, Zaric GS, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5. *Value Health* 2012;15:828–834.
- [6] Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. *Value Health* 2012;15:843–850.
- [7] Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;26:781–98.
- [8] Briggs A, Sculpher M, Buxton MJ. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ* 1994;3:95–104.
- [9] Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 2007.
- [10] Brisson M, Edmunds WJ. Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. *Med Decis Making* 2006;26:434–46.
- [11] National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Clinical Excellence, 2008. Available from: www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. [Accessed December 1, 2012].
- [12] Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313–24.
- [13] Everitt BS. *The Cambridge Dictionary of Statistics*. Cambridge: Cambridge University Press, 1998.
- [14] Briggs AH, Sculpher MJ, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006.
- [15] O'Hagan A, Buck CE, Daneshkhan A, et al. *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: John Wiley and Sons, 2006.
- [16] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat Med* 2009;28:3049–67.
- [17] Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making* 2002;22:359–71.
- [18] 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. *Med Decis Making* 2009;29:304–16.
- [19] Karnon J, Vanni T. Calibrating models in economic evaluation: a comparison of alternative measures of goodness-of-fit, parameter search strategies, and convergence criteria. *Pharmacoeconomics* 2011;29:51–62.
- [20] Kim JJ, Kuntz KM, Stout NK, et al. Multiparameter calibration of a natural history model of cervical cancer. *Am J Epidemiol* 2007;166:137–50.
- [21] Taylor DCA, Pawar V, Kruzikas DT, et al. Incorporating calibrated model parameters into sensitivity analyses: deterministic and probabilistic approaches. *Pharmacoeconomics* 2012;30:119–26.
- [22] Vanni T, Karnon J, Madan J, et al. Calibrating models in economic evaluation: a seven step approach. *Pharmacoeconomics* 2011;29:35–49.
- [23] Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health* 2009;12:739–49.
- [24] Chapman CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc Ser A Stat Soc* 2009;172(2):383–404.
- [25] Jackson CH, Bojke L, Thompson SG, et al. A framework for addressing structural uncertainty in decision models. *Med Decis Making* 2011;31:662–74.
- [26] Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ* 1998;7:723–40.
- [27] Stinnett AA, Mullahy J. The negative side of cost-effectiveness analysis [letter]. *JAMA* 1997;277:1931–2.
- [28] 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.
- [29] Brennan A, Kharroubi S, O'Hagan A, et al. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med Decis Making* 2007;27:448–70.
- [30] Groot Koerkamp B, Hunink MG, Stijnen T, Weinstein MC. Identifying key parameters in cost-effectiveness analysis using value of information: a comparison of methods. *Health Econ* 2006;4:383–92.
- [31] Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making* 2004;24:207–27.