# A systematic review of research guidelines in Decision-Analytic Modelling

Penaloza, Maria; Barton, Pelham; Jowett, Sue; Sutton, Andrew

*DOI:*
10.1016/j.jval.2014.12.014



*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Penaloza, M, Barton, P, Jowett, S & Sutton, A 2015, 'A systematic review of research guidelines in Decision-Analytic Modelling', *Value in Health*, vol. 18, no. 4, pp. 512-29. https://doi.org/10.1016/j.jval.2014.12.014

# A systematic review of research guidelines in Decision Analytic Modelling

Maria Cristina Peñaloza Ramos, MA [1], Pelham Barton, PhD [1,*], Sue Jowett, PhD [1], Andrew John Sutton, PhD [1]

**[1] Health Economics Unit, University of Birmingham, UK**

Key words: decision-analytic modelling, modelling, guidelines, good practice, methods

*Corresponding author

Health Economics Unit, Public Health Building, University of Birmingham, Edgbaston, Birmingham B15 2TT. United Kingdom, email: P.M.Barton@bham.ac.uk Tel: + 44 (0)121 414 3170, Fax: + 44 (0)121 414 8969

Running title: A systematic review of guidelines in DAM

**Acknowledgements:** We would like to thank the Health Economics Study Group Conference in Warwick, 26-28 June 2013 and Glasgow 24th – 26th June 2014 for the valuable comments and feedback on an early version of this paper, and for acting as a test of face and content validity. Thanks to Dr. Randoph Bruno and Dr. Lazaros Andronis for proof reading and comments to earlier versions of this manuscript.

**Contributors:** MCP undertook the review and analyses, wrote the first draft. All authors contributed to refined manuscript and approved the final version.

**Competing interests:** None declared. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Health Economics Unit or the University of Birmingham. No funding was received for this manuscript.

**Abstract**

**Introduction: D**ecision-analytic modelling (DAM) has been increasingly used to aid decision making in healthcare. The growing use of modelling in economic evaluations has led to increased scrutiny of the methods used.

**Aim:** To perform a systematic review to identify and critically assess good practice guidelines, with particular emphasis on contemporary developments.

**Methods:** A systematic review of English language articles was undertaken to identify papers presenting guidance for good practice in DAM in the evaluation of healthcare. The inclusion criteria were: papers providing guidance or criteria against which to assess good practice in DAM and studies providing criteria or elements for good practice in some areas of DAM. The review covered the period January 1990 to March 2014 and included the following electronic bibliographic databases: Cochrane Library; Cochrane Methodology Register and HTA; NHS EED; MEDLINE and PubMed (Embase). Additional studies were identified by searching references.

**Results:** 33 papers were included in this review. A practical five-dimension framework was developed which describe the key elements of good research practice that should be considered and reported to increase the credibility of the results obtained from DAM in the evaluation of healthcare.

**Discussion:** This study is the first to critically review all available guidelines and statements of good practice in DAM since 2006. The development of good practice guidelines is an ongoing process and important efforts have been made to identify what is good practice and to keep these guidelines up-to-date.

**Key points for decision makers**

- This study is the first, since 2006, to critically review available guidelines in DAM
- A dramatic evolution of good practice guidelines for DAM was found accompany with important efforts to keep these guidelines up-to date, functional and helpful
- This study proposes a practical framework to serve as a reference point towards the thorough consultation of good practice guidelines.
- This study did not review guidelines in light of current practice, which may be seen as a limitation; further research is required to assess the adherence of current practice to guidelines

## 1. Introduction

DAM in health care has been widely used to synthesize clinical and economic evidence and to inform resource allocation decisions for the purpose of allowing scarce health care resources to be allocated more efficiently (1). In simple terms, in DAM, a model is structured to represent clinical pathways to examine whether an intervention, compared for example to current practice, is cost effective (2). Building a model requires consideration of important elements including the complexity of the clinical area; the available evidence related to the disease; as well as other issues such as the scope or boundaries of the model; the appropriate time horizon; the perspective of the analysis; the availability of data and a formal synthesis of evidence within the model (2, 3). The increasing use of DAM in the economic evaluation of health care interventions and health technology assessments requires the use of sound analytic methods and consideration of the requirements of good practice.

The aim of this study was to perform a review to identify and critically assess good practice guidelines, highlighting areas where these have failed to provide recommendations, with emphasis being given to more recent developments. In this study we define DAM as a method that "uses mathematical relationships to define a series of possible consequences that would flow from a set of alternative options being evaluated" (4) (p.6).

## 2. Methods

A systematic review of articles written in English was undertaken with the aim of identifying published guidelines on DAM in healthcare. The following types of studies were included: guidelines for DAM or Health Technology Assessments (HTA) and other published papers on good practice in DAM. On the basis of an assessment of their title and abstract (if available), papers were deemed potentially relevant for inclusion if they: 1) provided general guidance in DAM for health care or health technology assessment (HTA); or 2) provided general criteria against which to assess good practice in DAM (for example a checklist).

This review excluded guidelines on: 1) trials or economic evaluations alongside clinical trials; 2) other non-DAM studies including statistical or econometric models; and 3) conference abstracts or other non-DAM papers.

### 2.1. Search strategy

An initial exploratory approach was undertaken using search terms used in a previous review (5) and this helped inform the final search terms used in this review (see Supplemental Material, Appendix 1, Tables 1 and 2). Relevant literature was also obtained by checking the references of the included articles.

The following bibliographic databases were searched: The Cochrane Library, Cochrane Methodology Register (CMR), Cochrane Health Technology Assessments, NHS Economic Evaluation Database, Embase, and MEDLINE. To avoid duplication, the PROSPERO database of prospectively registered systematic reviews in health and social care was searched for any existing or ongoing reviews that addressed similar topics, and none were identified. This review covered the period from January 1990 to March 2014. This is a period that reflects the development of guidelines for DAM in healthcare and the consolidation of good practice guidelines.

### 2.2. Selection of papers for review

Titles and abstracts (if available) were screened against the inclusion criteria to identify potentially relevant papers. In total, 33 studies, corresponding to general guidance or elements of good practice in DAM were included in this review. A flow chart showing the study selection process is shown in Figure 1. The methodological quality of the papers included in this study was not comprehensively assessed using formal checklists due to the diversity of the literature included and the nature of the review.

#### 2.2.1. Data extraction

All studies were manually searched and data were extracted by the first author from each paper using a data extraction form. The data extraction form was developed to retrieve and organise information from each paper based on its main topic, model structure, model uncertainty, model transparency, and validation. The data extraction form was developed through a process in which the content of the papers informed the “areas” that the data were extracted under. This approach was utilised to ensure that the review did not miss any information related to the model building process. The data was extracted as free text and in the form of a ‘yes/no’ response.

## 3. Results

The DAM guidelines identified in this study have responded to the need to: reflect on how good practice in the field has been defined; the need to keep pace with the rapid progress in the way that economic evaluation methodology has progressed since the 1980s; and as a means to ensure that guidelines for good practice remain current, effective, and helpful. More comprehensive guidelines, for example Philips (5) or the set of the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making (ISPOR-SMDM) guidelines have been developed as part of bigger projects i.e., a Health Technology Assessment project involving experts from prestigious academic institutions or as part of a 'task force' respectively (see section 3.1).

Recommendations and statements of good practice have been proposed following the application of different methods, for example: Philips' synthesized good practice guidance and accompanying checklist resulted after taking each theme and subtheme identified in a systematic review of guidelines followed by technical discussions among the research team of its relevance in relation to the development of general guidelines (5). Guidelines produced by ISPOR-SMDM resulted from a 'task force' consisting of expert developers and experienced users of models from academia, industry, and government, with representation from many countries. A decision was made to divide the DAM topic into six components and working groups respectively; three of these groups covered aspects relevant to all models such as the conceptualization of a model, the estimation of model parameters and handling of uncertainty, and the validation of models and issues of transparency. The other three components considered specific techniques: state-transition modelling, discrete event simulation, and dynamic transmission models. The working groups produced draft reports for each section, and in contrast to Philips there was no systematic attempt to review the literature. The first draft of recommendations represented the opinions of the experts in the Task Force and these were posted on the ISPOR and SMDM Web sites for comment by the general membership of the societies. A second group of experts—again, with broad representation of modellers and users of models—was invited to formally review the articles. Their comments were addressed and after receiving any additional comments and considering any further revisions, the final version of each article was prepared and released to the public (see section 3.1).

Of the 33 papers included in this review, 15 studies provided general guidelines for good practice or criteria in the form of a checklist. Eighteen papers were focused on particular elements of good practice, for example, model structure or uncertainty, or model transparency and validation.

### 3.1. Elements of good practice for DAM

Fifteen studies provided general guidelines for good practice; eight out of the 15 guidelines were released before 2012; (5-13) with the remainder making up the ISPOR-SMDM (14-20) set of guidelines. Tables 1a to 1c present a breakdown of the elements of good practice by the main themes of the guidance, i.e., model structure, identifying and synthesizing evidence, and model validity. These studies provided a source of complete information on the various stages that need to be covered in DAM. Some of the studies constituted a list of topics that need to be checked, or questions that modellers need to answer prior to constructing a model. Most commonly, guidelines have been presented as a series of good practice statements, starting with Weinstein (13), then Philips (5, 9) and more recently ISPOR-SMDM (14-20). DAM guidelines provide a set of principles that might lead, for example, to an appropriate model structure or else indicators of areas that require consideration in decision modelling (9).

To inform model construction and increase model credibility and validity, these guidelines provide a set of principles, checklists, or have stated the agreement of a common application (8, 10-12, 15-17, 19, 20). For example, guidelines have stated that model construction is likely to be influenced by the adoption of simplifying assumptions reflecting issues such as data availability, and that the design of a model should not be driven by the data at hand. Under these circumstances the identification of the explicit characteristics of the disease area that affect model selection, for example, the unit of representation, are considered important (11, 12, 17, 19, 20). Other aspects in model construction that arise from the application of models to specific groups of patients or specific settings, include the scope of the model, the model perspective, choice of model type, choice of utility structure (e.g. quality adjusted utility scale) and the interventions to be included in the model (10-12, 17, 19, 20). These guidelines identify the characteristics of individuals as a key element aiding the process of model selection, i.e., whether a model needs to represent individuals or groups or interactions between individuals (19). Furthermore, guidelines recommend that ‘the appropriate model type is determined by purpose, level of detail

and complexity' (p.809); (19) and the use of 'explicit processes' involving expert consultation, influence diagrams or similar should be used to convert the conceptualization of the problem into an appropriate model structure (19).

ISPOR-SMDM (14-20) recognised the difficulty for all models in achieving all the recommended best practice for model validation, i.e., face validity, internal validity, cross validity, external validity and predictive validity. Instead of establishing a minimum quality standard, guidelines recommend the adoption of optimal practices that all models should aim for (16). Among these, model transparency was identified as a key area of optimal practice that should be achieved by all models and is reflected by providing clear information on how the model was built, i.e., describing its structure, parameter values, and assumptions (16).

ISPOR-SMDM (14-20) reiterated statements of good practice emphasizing on its appropriate conduct and furthermore establishing grounds for usage, for example, the use of time horizons sufficiently large to capture all health effects and costs relevant to the decision problem in cohort simulations; (14-20) or insisting on the value of model simplicity as long as a model's face validity is not compromised (19).

### 3.2. Model structure

Good practice for selecting a model or the use of alternative model structures was discussed in ISPOR-SMDM (15, 17, 19) and in four out of 18 individual papers included in this review (21-24). Model structure should be considered in the initial stages in the process of model building (Table 2). Guidelines have suggested that prior to model building, researchers should identify: the problem and objective of the project; the analytical perspective of the model; the scope; the rationale for selecting the particular structure; the target population; the strategies and comparators; and then give justification for choosing the model type, the time horizon and the disease states (19, 21, 23, 24). These initial steps are important and will have important implications for the model structure, data requirements, and the reporting of the final results obtained from the model.

Guidelines for conceptualizing a model's structure have evolved from statements of general principles, for example by stating that the structure of a model should be consistent with the theory of the health condition and the available evidence (13), to more systematic processes describing how to select a model from competing

alternatives (21-24). ISPOR-SMDM (15, 17, 19) described the development and construction of a model as a process that starts with *model conceptualization* (19) which consists of a two-step process: problem conceptualization and model conceptualization. Problem conceptualization in this context is transforming knowledge of the healthcare process into a representation of the decision problem. Model conceptualization is the representation of the components of the problem using a particular decision-analytic method (Table 2). The nature of the problem and the project objectives are decisive in selecting the structure of a model. Furthermore, ISPOR-SMDM (15, 17, 19) suggested that the early specification of the decision problem and project objectives will improve model-building and the structure of the model (data requirements, analytic strategy and reporting) (19).

The importance of the choice of model structure stems from the fact that alternative model structures can impact on model results and thereby affect decision making (19, 21, 23). The appropriate model type should be determined according to its purpose, level of detail required, and complexity (19). As previously demonstrated, guidelines aid the selection of an appropriate modelling approach by providing an overview of competing approaches and highlighting examples of where each alternative technique should be employed (19, 21, 23). The most common issues affecting a model's selection are (15, 17, 19): 1) the unit of representation, does a model represent individuals or groups? The unit of representation affects the level of detail required for the variables that predict outcomes (19); 2) if the decision problem requires the modelling of the effect of an intervention on disease spread or use of limited resources, in other words, if interactions among individuals need to be represented then models designed for patient interactions are necessary (19); and 3) the time horizon is dictated by the problem scope. For example, decision trees are considered appropriate for models with very short time horizons, while longer horizons require the use of models such as State-Transition (for example a Markov) or Discrete Event Simulation DES (19).

Among the most difficult stages in the conceptualization of a model is the selection of the appropriate level of model complexity, as very simple models may lose face validity if they do not incorporate all the aspects that experts feel are required; whereas complex models may be difficult to build, debug, analyse, understand and communicate (19). Guidelines have generally supported the choice of simpler models as 'model simplicity is desirable for transparency, ease of analysis, validation and description' (19), while at the same time it is recognised that under certain circumstances, more complex models may be needed. Consensus-based guidelines

stating common grounds for the application of more complex model structures have been developed, i.e., State-transition models, DES and Dynamic Transmission models (17, 18, 20).

### 3.3. Model uncertainty and synthesis of information

ISPOR-SMDM (14) and an additional eleven individual papers (25-35) provided methodological guidelines for the analysis of model uncertainty (methodological, structural, parameter, heterogeneity and stochastic), and the use of sensitivity analysis. Step by step guidelines and checklists have been developed (Table 3) to aid researchers in accounting for uncertainty or to identify how uncertainty was incorporated in a model or to address special model circumstances, for example where the evidence is insufficient to give a clear representation of the uncertainty through parameter distributions (14, 26). The view presented by some of the studies included in this review is that many published models still fail to account correctly for the major sources of uncertainty, in particular structural uncertainty, indicating that a gap may still exist between techniques, guidelines, and what is done in practice (26, 31).

Assumptions adopted in decision models determine its final structure and can consider: the choice of relevant comparators and health states, or available clinical evidence that determines the type of adverse events, duration of treatment effects, time dependency of probabilities and prognostic implications of surrogate end points or the clinical events included (14). Structural uncertainties arise when these *structural* assumptions are not formally quantified and it is uncertain whether they accurately reflect reality (14). Current methods for addressing structural uncertainty include scenario analysis (presenting the results under different model structures); model averaging (presenting results of different models using different assumptions and an average across these models); parameterization of structural uncertainty; and in the absence of data or presence of weak data, expert elicitation to translate expert beliefs into probability distributions (30). Model structure plays an important role in defining the relationship between inputs and outputs to the point that it has been recognised that structural uncertainty may be at least as important, in terms of its impact on results, as parameter uncertainty (14). ISPOR-SMDM (14) highlighted the emerging interest in calibration methods as an aid to ensure consistency of inputs and outputs in a model. Calibration is used when data are available to match model outputs rather than model inputs: it is then necessary to determine parameter values which give model results that match the data (14).

Many techniques have been developed and have evolved which aim to capture the various sources of DAM uncertainty. However, there still remain some areas where more research is needed, such as: accounting for uncertainty surrounding quality of evidence for particular structural aspects; generalizability from one setting to another; and the way multiple sources of evidence should be combined (heterogeneity of parameter values from different sources) (26). ISPOR-SMDM (14) proposed the parameterization of structural uncertainties into a model as an approach to deal with issues around the quality of evidence, however this approach seems to become complex if a complete redesign/rebuilding of the model is required (nested structures). (14). Under these circumstances guidelines have stated that '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 and analysts are encourage to be explicit about the structural assumptions that might impact their findings and suggest alternative assumptions for future modelling exercises (14).

### 3.4. Model transparency and validation

Four papers discussed methods to assess the consistency or validity of models and model transparency, (Table 4) (16, 36-38). Model transparency reflects the extent to which a model's structure, equations, parameter values and assumptions can be reviewed, and a model is considered transparent if any interested reader with the necessary expertise who wants to evaluate the model is able to reproduce it (16). Model validation has been recommended to enhance the credibility of models and as an indicator of reliability in practice guidelines (9, 16, 36-38). Model transparency does not equal the accuracy of a model in making relevant predictions; a transparent model may yield the wrong answer, and vice versa, while a model may be correct and lack transparency. Thus, transparency and validation are both necessary for good practice in modelling (16).

Validation involves a set of methods for judging the accuracy of models when making predictions. More recent guidelines have used the terms '*model consistency'* or *'model validation'* to refer to five types of model validity: face validity (evaluation of model structure, data sources, assumptions and results), internal validity (the practical model should behave as the theoretical model predicts), cross validity (comparison of results with other models), external validity (comparing model results and real-world results) and predictive validity (comparing model results with prospective observed events) (9, 16).

Principles and methods to enable researchers to assess model validity have been discussed and in some cases demonstrated (16, 37, 38). However, results of a study (38) established that health economic models based on limited follow-up data from one source may not be generalizable either to longer follow-up periods or other contexts. Furthermore, in addition to the standard considerations of uncertainty about parameter estimates, it is also important to assess the implications of model uncertainty on results, in other words, to undertake independent model validation (16).

Best practice recommends that face validity (due to its subjective nature) should be judged by people who have expertise in the problem area, but who are impartial and preferably blinded to the results (16). Internal validity verifies that mathematical calculations are performed correctly and are consistent with the specification of the model. Methods to assess internal validity will depend on the model's complexity, but two main stages of internal validity involve the verification of individual equations and their accurate implementation. It should be noted that internal validity does not evaluate the accuracy of model's predictions (16). Cross validity involves examining different models and comparing their results to then identify and analyse the causes of differences and similarities in these results. External validation compares the results of a model with actual data however the difficulty in identifying 'alternative data' has been noted. Best practice to undertake external validation recommends following a formal process to compare a model's results to actual event data. Guidelines provide awareness of the important limitation that external validation can only address the parts covered by data sources (16). Predictive validity remains a highly desirable type of independent model validation due to its potential ability to demonstrate the accuracy of the results obtained from the DAM. However its results are potentially limited if there are changes in the design of the study or other factors outside the control of the study design change during the development of the study (16).

Even though the latest guidelines (16) have provided more detailed guidance on how best to ensure model transparency and undertake validity checks, which reflect the value of concise reporting of a model and advocate the quantification of uncertainties arising from differences in assumptions, (16) some quandaries seem to prevail. For example, in order to examine external validity, modellers are advised to use actual event data. However, that same data in many instances would already have been used to parameterise the model – as guidelines suggest that the most representative data sources should be used in developing a model.

**Discussion**

This review has critically compared statements of good practice in contemporary guidelines and identified areas where further work may be needed. This review found: 1) good practice guidelines have been developed and agreed; adherence to these guidelines is considered as best practice in DAM; 2) guidelines should be seen as tools that if followed will lead to the results obtained being more credible; 3) common grounds in the application of guidelines; and 4) some aspects of the guidelines related to DAM require further development, for example, the choice of model structure, assessment of structural uncertainty and achieving predictive validity.

Common grounds have been identified for the application of guidelines in aspects such as the specification of a model's structure, the inclusion of incident cases over the time horizon of an evaluation, the use of time horizons, parsimonious model structure, and subgroup analysis in DAM.

Most decision problems can be conceptualized using one of the available model types, whilst the choice of model structure is unlimited. There is general acceptance of the special circumstances under which complex modelling needs to be taken into consideration, while at the same time, overly complex models should be avoided if a simpler model can accurately reflect all aspects of the decision problem. More research should be undertaken of case studies comparing the economic efficiency of simple versus complex models, the use of hybrid models which are considered to be very flexible and accurate with no restriction on how time is handled (15), and the trade-off between model complexities versus model transparency. This should be done in light of the advances in computing that make complex calculations feasible and economically efficient, opening the way for the more generalised use of individual-based simulations (15).

Whether model structure should be informed by data availability or not remains another conflicting aspect in DAM. Current guidelines have argued the case for building a model first and then looking for the data to populate it, as this strategy will result in more appropriate and relevant model structures (19). However an apparent drawback of this approach which has already been argued by detractors is data availability. Alternatively, finding the data to populate the model might be possible perhaps by adopting more assumptions based on expert opinion (15, 19). Independent of the assumptions adopted, the model parameters should reflect

the uncertainty due to the gaps in the available data, which in an ideal world would trigger the need for value of information analyses to show the value of this required data (14).

Structural uncertainty remains an area of controversy; an inappropriate structure can invalidate the conclusions drawn from CE analyses, while choices made when structuring a model can significantly affect its results and the inferences from it. Until recently, even the definition of structural uncertainty was a matter of dispute; (27, 30) however contemporary guidelines have clarified this concept by using an analogy with linear regression, and it is now recommended as good practice to factor in structural uncertainties into a model (14).

Another area where issues have been raised has been with model validity. Guidelines have recognised that 'not all models will be able to achieve all these best practices' (14) while the 'inability to do so does not necessarily imply a model is not useful' (14). However recent guidelines seem to have provided a scope for analysts to use their own discretion to solve some issues, provided that the use of 'optimal practices', as described by methods and recommended practice is demonstrated (14). Some aspects of model generalizability demand further research because it relies on the availability of follow-up data ideally from the same source, and follow-up data from other sources may not be generalizable to longer follow-up periods or to new contexts (38).

There are some areas where there is a contradiction between the guidelines; however we believe that as with model validity, these issues can be solved at the discretion of analysts. A good example is when guidelines indicate the use of all feasible and practical comparators (5, 19). The same guidelines indicate that the choice of comparators is governed by the scope of the model, which is a direct consequence of the research question. In other words, even though a broad range of feasible strategies may be available, the choice of comparators is expected to answer to the decision problem. However, the inclusion or exclusion of potentially relevant comparators should be assessed as part of the structural uncertainty of the model (27).

This review has found that while guidelines have been developed and are available to aid researchers to inform the results of their studies and most importantly, to increase the credibility of their results, these guidelines lack practicality due to the extensive amount of information available and its complexity. Current standards of reporting could be improved if a single, comprehensive, user friendly and practical instrument is made available to direct researchers towards the key elements of good research practice in DAM which should be assessed and

reported to increase the credibility of their results. We aim to contribute towards this end by proposing a practical five-dimension framework to assess adherence to guidelines in DAM.

The framework we propose incorporates and reflects much of the evidence from this review, i.e. it has synthesized all contemporary guidelines in a checklist instrument. To ensure its consistency, we adopted the most up to date and agreed guideline statement when components in each dimension were superseded or contradictory. The framework uses the following five-dimension checklist: i) problem concept; ii) model concept; iii) synthesis of evidence; iv) analysis of uncertainty; and v) model transparency and validation (see table 5). This framework does not attempt to replace the guidelines provided by ISPOR-SMDM 2012 or any other contemporary guidelines; instead it attempts to serve as a reference point for the thorough consultation of good practice guidelines.

*Strengths and limitations*

This study constitutes a comprehensive review of more than a decade of developments in DAM, including the most contemporaneous guidelines. While this study has discussed all available general guidelines in a single document, the breadth of this field determined that this review focus on aspects that are considered general to all models (model structure, model conceptualization, model parameters, model uncertainty and model transparency and validation). The exclusion criteria adopted (abstracts, posters, conference papers and non-English language studies) may be considered as a limitation of this review, however these were required to guarantee consistency in the analysis; furthermore, a negligible number of non-English language studies were identified pertaining to applied studies. Some databases such as HEED, Psychinfo and Cinhal were not included in this review mainly because we took the view that the same references would be identified in Medline or their focus was applied research. This review does not address the choice of data and its processing to yield suitable inputs for the model; we took the view that this is a topic has been extensively developed in other fields such as epidemiology or statistics. Finally, as stated in the previous section, this review has excluded applied studies that are important for identifying which elements of guidelines pose greater challenges for analysts or correspond to deviances from guidelines in current practice. This undoubtedly triggers the need for future research on the adherence of current practice to guidelines and its impact on results of decision-modelling emphasizing for example, on issues around the reporting of uncertainty analysis or the assessment of structural uncertainty or around areas of increasing interest such as the practical use and feasibility of generic models.

## 4. Conclusions

The framework to judge the adequacy of decision-modelling has changed dramatically since it was first envisioned; ISPOR-SMDM constitutes the most contemporaneous, up-to date and agreed set of good practice guidelines.

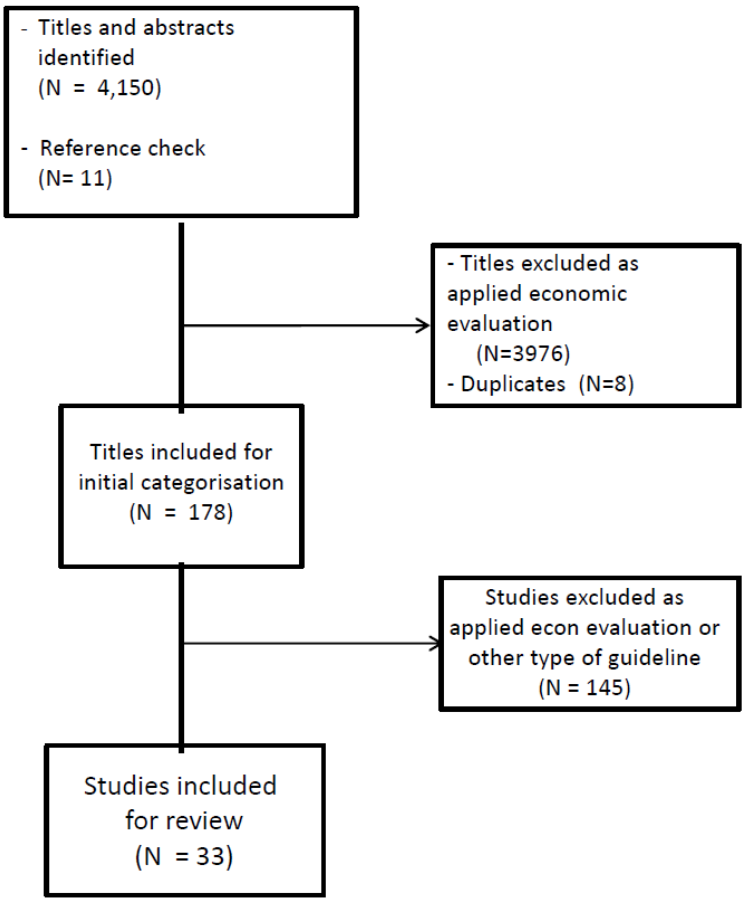


Figure. 1 – Flow chart

Table 1 General Guidelines

| Paper ID | | Model Structure | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Author(s) / year | Topic | Statement of Decision problem/objective | Scope / analytic perspective | Rationale for structure | Strategies and comparators | Model type | Time horizon | Disease states / pathways / time to events | Cycle length | Parsimony |
| Sonnenberg et al, 1994 [11] | Framework to judge adequacy | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Sculpher M et al, 2000 [10] | Framework for validity and quality | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| Soto J, 2002 [12] | Checklist for decision-analytic | ✓ | ✓ | ✓ | ✓ | | ✓ | | | |
| Weinstein MC et al, 2003 [13] | Good modelling practice | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ |
| Philips Z et al, 2004 [5] | General guidelines | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Philips Z et al, 2006 [9] | Framework for quality assessment | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| HTA, Canada, 2006 [6] | General guidelines in Canada | ✓ | ✓ | | ✓ | ✓ | ✓ | | | |
| Karnon J et al, 2007 [8] | Modelling issues | | | | | | ✓ | | | ✓ |
| Earnshaw J et al, 2008 [7] | Guidelines for economic evaluation | | | ✓ | | | ✓ | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ISPOR-SMDM, 2012 [14-20] | Good research practices | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

Table 1 General Guidelines (continuation)

| Paper ID | | Identifying and synthesizing evidence | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Author(s) / year | Topic | Baseline data | Bias in parameter estimates | Costs | Data identification | Data incorporation | Data modelling | Heterogeneity | Methodological | Parameter | Parameter estimates | Stochastic | Structural | Treatment effects | Utilities |
| Sonnenberg et al, 1994 [11] | Framework | | | | ✓ | | ✓ | ✓ | | ✓ | ✓ | | | | ✓ |
| Sculpher M et al, 2000 [10] | Framework | | | | ✓ | ✓ | | | | | | | | | |
| Soto J, 2002 [12] | Checklist | | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | | ✓ |
| Weinstein MC et al, 2003 [13] | Good modelling practice | ✓ | | | | ✓ | ✓ | ✓ | | | | | | | |
| Philips Z et al, 2004 [5] | General guidelines | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Philips Z et al, 2006 [9] | Framework | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| HTA, Canada, 2006 [6] | General guidelines | | | ✓ | ✓ | | | ✓ | | ✓ | ✓ | | | | ✓ |
| Karnon J et al, 2007 [8] | Modelling issues | | | | | | | ✓ | | | | | | | |
| Earnshaw J et al, 2008 [7] | Guidelines | ✓ | ✓ | | ✓ | | | | | ✓ | | | ✓ | ✓ | |
| ISPOR-SMDM, 2012 [14-20] | Good research practices | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

Table 1 General Guidelines (continuation)

| Paper ID | | Model validity | | | |
|---|---|---|---|---|---|
| Author(s) / year | Topic | Face/Internal/technical validity, verification or consistency | Cross validity | External validation | Predictive validation |
| Sonnenberg et al, 1994 [11] | Framework to judge adequacy | ✓ | | | |
| Sculpher M et al, 2000 [10] | Framework for validity and quality | ✓ | ✓ | | |
| Soto J, 2002 [12] | Checklist for decision-analytic modelling | | | ✓ | |
| Weinstein MC et al, 2003 [13] | Methodology regarded as good modelling practice | ✓ | ✓ | ✓ | ✓ |
| Philips Z et al, 2004 [5] | General guidelines | ✓ | | ✓ | ✓ |
| Philips Z et al, 2006 [9] | Framework for quality assessment | ✓ | ✓ | ✓ | |
| HTA, Canada, 2006 [6] | General guidelines in Canada | | | | |
| Karnon J et al, 2007 [8] | Modelling issues | | | | |
| Earnshaw J et al, 2008 [7] | Guidelines for economic evaluation | ✓ | | | |
| ISPOR-SMDM, 2012 [14-20] | Good research practices | ✓ | ✓ | ✓ | ✓ |

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

Table 2 Model Structure

| Author(s) and year | Area of guidelines | Criteria for selecting a modelling approach | Rationale for structure | Model-based evaluation | Parsimony | Key recommendations |
|---|---|---|---|---|---|---|
| Roberts (ISPOR_12) [19] | Choice of model structure | Justified in line with policy context and aiming to inform resource allocation | Whether a model represents individuals or groups or interactions between individuals | Expert consultation and conceptualization in two stages: problem conceptualization and model conceptualization | Model simplicity however preserving face validity | Early specification of the decision problem, modelling objectives and valuing outcomes will improve model efficiency (expert consultation, influence diagrams, concept mapping) |
| Siebert (ISPOR_12) [20] | Structures and model complexity | Whether decision-problem requires time-dependent parameters or time to an event or repeated events | Markov models can handle memory by creating states that include history; but then model complexity | STM are comprehensive and powerful tools to guide decision in health care | Decision tree has limited ability to reflect time; then STM seems the simplest option | Markov model if decision problem has a manageable number of health states; if not, use a individual-level state-transition model (STM) |
| Karnon (ISPOR_12) [17] | Structures and model complexity | DES justified on model flexibility | Constraint resources; patient's interactions; time dependencies | Value of DES to inform health care decisions; flexible and able to represent complex behaviour and interactions between individuals | Easy representation of complex systems | A good choice if individuals are subject to multiple or competing risks |
| Bentley TG et al, 2010 [22] | Structures and model complexity | Subsequent event risk dependent on prior event history; simulation of event or disease risks over time; improving validity | Recurrent events and time dependency | The ability of incorporate past history is restricted to the number of model health states | Trade-off between model bias and model complexity | Failing to incorporate prior event history in Markov models would overestimate the impact of an intervention; incorporate dependency by adding states that track event history; make subsequent event risks dependent on this history |
| Brennan A et al, 2006 [23] | Choice of model structure | Needs to be justified | Interactions between individuals or an individual level model; choice | Comparison of health technologies and synthesising evidence on costs and benefits | Simplest model that addresses objectives and structure of disease and treatment | Responsibility of developers to select the most appropriate modelling approach; taxonomy grid is a guidance |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | from taxonomy of model structures | | | |
| Barton P et al, 2004 [21] | Choice of model structure | Needs to be justified | Interactions versus not interactions between individuals | Two distinct and independent aspects: mean estimate of cost-effectiveness and exploration of uncertainty in the model inputs | Simplicity (relates to the size of the model) is seen as an advantage | Check dependence or independence among individuals; model simplicity is an advantage; model validation; challenge the need for a complex model |
| Karnon J, 2003 [24] | Structures and model complexity | Assess relative advantages of alternatives according to areas of treatment | Model flexibility and analytic input (complexity of model building) | Choice depend on flexibility vs. time availability; there may be circumstances where DES provides a more accurate representation of the data | A simpler model was the optimal technique as compared to a complex DES model | Results of different models (Markov or DES) may produce likely results; model flexibility (DES) may be outweighed by greater time required to evaluate its results |

Note: DES= Discrete Event Simulation; STM= State Transition Model

Table 3 Model Uncertainty & Synthesizing Evidence

| Author(s) / year | Area of guideline | General principles | Way of reporting | Methodological issues | Methods /recommendations |
|---|---|---|---|---|---|
| Briggs (ISPOR_12) [14] | Point estimate(s) & parameter uncertainty | Responsible reporting; use of terminology; justify its omission; decision maker's role; preferable parameterize uncertainty from structural assumptions if possible | Use tornado diagrams, threshold plots, or statements of threshold parameter to report DSA; describe assumption(s); report uncertainty around calibrated parameter(s); report EVPI if needed | Methodological, structural, patient heterogeneity; parameter uncertainty and stochastic uncertainty | For structural uncertainty, calibration approaches; for parameter uncertainty DSA or PSA; for point and interval estimates use CI or distributions; reflect absence of evidence |
| Bilcke J et al, 2011 [26] | Uncertainty: Step-by-step guide and checklist | Formulate decision-problem; specify sources of uncertainty; obtain information and evidence; report results; apportion uncertainty to sources | Report choices of normative approach(es); present sources of uncertainty; use distributions; assess the most influential sources of uncertainty; report results of PSA | Methodological, structural and parameter uncertainty | State if there are more than one approach that can be used; use distributions; assess the most influential sources of uncertainty; global sensitivity analysis; PSA |
| Jain R et al, 2011 [31] | Sensitivity Analysis | Report all sources of uncertainty; Strengths and limitations of SA should be acknowledge (interactions and correlations between parameters) | If long term analysis is needed, conduct CEA under various time horizons; use for instance, tornado diagrams, or threshold analysis to present results | Methodological, structural, parameter | Model averaging and parameterization for structural uncertainty; methodological uncertainty can be addressed by providing results for a 'reference case'; DSA or PSA. |
| Koerkamp BG et al, 2010 [32] | Uncertainty and patient heterogeneity | Consider range of assumptions for the natural course of a disease; provide model for every set of assumptions instead of using the single best model; trade-off between the realism of a model and time availability | Use tornado diagrams or threshold plots; describe assumption(s); report uncertainty; if the purpose of the PSA is the acquisition of information to reduce uncertainty, report EVPI | Parameter uncertainty, patient heterogeneity, stochastic uncertainty (first-order uncertainty) | PSA joint uncertainty; parameterization model structure uncertainty; first-order Monte Carlo analysis for stochastic; DSA for parameter uncertainty; EVPI if needed |

| | | | | | |
|---|---|---|---|---|---|
| Jackson CH et al, 2011 [30] | Structural uncertainty | Various sources: statistical models, evidence used, states or clinical events represented, or treatment strategies considered | Should be acknowledge, assess and reported | Structural uncertainty | Reference case model; assign distributions; use PSA; for non-parametised uncertainties use global model; if lack of data, elicit distributions |
| Strong M et al, 2010 [35] | How complex a model should be | Uncertainty in model structure is complex; it involves making judgements about model's ability to accurately represent a decision problem | Most commonly by PSA, however it will only quantify uncertainty about the costs and consequences; problem when a model lacks accuracy | Uncertainty about the model input values and model structure | To properly represent uncertainty about the costs and outcomes, structural uncertainty must be presented; structural uncertainty measured with model averaging or the discrepancy approach |
| Bojke L et al, 2009 [27] | Structural uncertainty | Impossible to accurately predict mean costs and outcomes; sources are treatment effects and type of model | Importance of differentiating parameter and structural uncertainty: if uncertainty can be parameterised, then there is parameter uncertainty | Parameter, methodological and structural (little attention given to structural uncertainty) | Model selection (not plausible); model averaging (difficulty determining posterior distributions); parameterising (directly representing uncertainty by adding other 'uncertain' parameters) |
| Briggs AH et al, 2003 [28] | Probabilistic probabilities over multiple branches | If there is a need to specify a distribution over multiple branches at a chance node | Have the Dirichlet distribution specified over multiple branches at a chance node | Inconsistencies performing sensitivity analysis if a node has more than two branches and the sum of the branching probabilities is different from 1 | Use Dirichlet distribution, a multivariate equivalent of the beta distribution |
| Kuntz KM et al, 2002 [33] | Patient heterogeneity | Cohorts are defined based on population characteristics; sometimes other characteristics may be overlooked (disease incidence | Heterogeneity bias may be evaluated as a function of 3 parameters: annual probability of developing the disease; RR of disease with vs. without the | The assumption that each health state contain a homogenous population group does not always hold: for instance, in presence of risk factors affecting | Adjust by introducing an heterogeneity factor; probability of transitioning to disease dependent on heterogeneity factor; transition probabilities |

| | | | | | |
|---|---|---|---|---|---|
| | | or progression), causing heterogeneity | factor; iii) baseline prevalence of the factor | the chance of developing disease | averages to that of the model without adjustment |
| Briggs AH et al, 1999 [29] | Uncertainty | Study designs included were modelling-type based approaches | The majority included some form of sensitivity analysis (one-way sensitivity analysis) | Mainly one-way SA; 5% attempted statistical analysis; 17% failed to provide any attempt to quantify uncertainty in their results | Reference case (comparability of results); potential for ICER to vary; avoid selective comparison; uncertainty; interval estimates; SA; probabilistic nature of reported range; descriptive statistics; estimate CI; present CEAC |
| Andronis L et al, 2009 [25] | Sensitivity analysis | DSA requires variables and sources to be justified; for PSA distributions should be placed around all parameters (excluded parameters should be justified) | Repeated analysis should be run using different models and methods where uncertainties exist | Methodological and structural uncertainty | Univariate, multivariate, PSA and DSA; distributions in line with logical bounds; if correlation is expected, use joint distributions (do not assume independence) |
| Sendi P et al, 2002 [34] | Uncertainty & opportunity costs | Univariate and multivariate SA to assess robustness; however SA does not inform joint uncertainty | Alternative approaches as a result of the intractability of the ICER: Net Health Benefit (NHB) and CEAC | ICER difficulty apparent if a distribution extends over more than one quadrant | NHB, however a problem if lambda is unknown; CEAC, however same problem with lambda; uncertainty can be accounted for using Bayesian methods |

Note: DSA= Deterministic Sensitivity analysis; PSA= Probabilistic Sensitivity Analysis; EVPI= Expected Value of Perfect Information; CI= Confidence Intervals; SA= Sensitivity Analysis; CEA= Cost-Effectiveness Analysis; RR= Risk Ratio; ICER= Incremental Cost-Effectiveness Ratio; CEAC= Cost-Effectiveness Acceptability Curve; NHB= Net Health Benefit

Table 4 Model Transparency and Validation

| Author(s) / year | Area of guideline(s) | Methodology | Rationale for model transparency and validation | Best practice | Recommendations |
|---|---|---|---|---|---|
| Eddy (ISPOR_12) [16] | Transparency and validation of models | Recommendations on optimal practice | Made available model's non-technical and technical documentation, written in sufficient detail to enable the reader to evaluate a model | Face validity of a model's structure, evidence, problem formulation and results; transparency and validation | Models are instruments to help decision makers answer complex questions; model confidence and credibility is demonstrated by clarity in model structure, equations, parameters, and assumptions, and by subjecting models to tests of validity |
| Karnon J, 2011 [37] | Model validation | Empirical comparison | Identification of input parameter(s) that produce output that best predict observed data | Probabilistic calibration of models produced improvements in model's accuracy, and reduced uncertainty | Widespread of model calibration (probabilistic calibration); a process of validation against more theoretically grounded approaches is valuable (Bayesian updating approach) |
| Goldhaber-Fiebert JD, 2010 [36] | External model validation | Literature review | Comparing model to independent data not used in the model | Heterogeneity in how results of model evaluation are reported | Evaluation via comparison(s) to independent studies; structured reporting format: empirical study description, baseline characteristics, study protocol, study outcomes, model outcomes and model consistency |
| Kim LG, 2010 [38] | Model validation | Use of internal, prospective and external validation | Indication of reliability of assumptions adopted | A model should be generated which fits all available data | Model based on limited data may not be generalizable; uncertainty from model assumptions as important as parameter uncertainty; new model should be generated which fits all available data; model validation should assess: key events, rate of accrual of events and absolute and incremental costs and effects |

Table 5 Framework to assess adherence to good practice guidelines in Decision-Analytic modelling (DAM)

| DIMENSION 1: PROBLEM CONCEPT | | | |
|---|---|---|---|
| Components of good practice | Questions for review | Yes , No, or NA | Attributes |
| Decision problem | Is there a written statement of the decision problem and scope of the study? | | A clear statement of the decision problem and scope would determine the interventions and health outcomes to be measured |
| | Are the objective(s) of the study and model structure consistent with the stated decision problem and scope? | | They are expected to be consistent |
| Analytical perspective | Has the perspective of the model been stated? | | Most common perspectives are: patient, health system (insurer) and society |
| Target population | Has the target population been identified? | | Target population should be defined in terms of features relevant to the decision (geography, patient characteristics, including comorbid conditions, disease prevalence and stage) |
| Health outcomes | Are the outcomes of the model stated and consistent with the perspective, scope and overall objective(s) of the model? | | Health outcomes may be events, cases of disease, deaths, life-years gained, quality-adjusted life-years, disability-adjusted life-years or other measures important to stakeholders and should be directly relevant to the question being asked |
| | Has any adverse effect of the intervention(s) been captured? | | Interventions may cause negative health consequences that need to be modelled and discussed as part of the study's results. The impact of assumptions regarding adverse effects of interventions should be assessed as part of the structural uncertainty analysis |
| Comparators | Is there a clear definition of the alternative interventions under evaluation? | | Usually the choice of comparators is governed by the scope of the model. Impact of assumptions adopted when deciding upon comparators should be assessed as part of the structural uncertainty analysis |
| | Is there a discussion around feasible options or justification for the exclusion of feasible options? | | The choice of comparators affects results and should be determined by the decision problem, not by data availability. All feasible and practical strategies as determined by the scope of the model should be considered. Constraining the range of strategies should be justified |
| Time horizon | Is the time horizon of the model justified and sufficient to reflect all important differences between options? | | Time horizon of the model should be long enough to capture relevant differences in outcomes across strategies (lifetime). Time horizon is dictated by the problem scope |

Note: NA= Not Apply

| DIMENSION 2: MODEL CONCEPT | | | |
|---|---|---|---|
| Components of good practice | Questions for review | Yes , No, or NA | Attributes |
| Choice of model type | Has the unit of representation been given? | | Usually stated in terms of groups or individuals. If groups are being modelled most frequently decision trees, Markov processes or infectious disease models are the correct choice; if individuals are being modelled then the choice is between DES, dynamic transmission models or agent-based models |
| | Is there a need to model the interaction between individuals in this model? Has this been discussed? | | If interactions between individuals is required (when the disease or treatment includes interactions between individuals) then DES, dynamic-transmission, or agent-based models may be the correct choice |
| | Does the decision problem require a short time horizon? | | For simple models or problems (short time horizon, few outcomes) a decision tree may be appropriate; time horizon should be large enough to capture all health effects and costs directed related to the decision problem |
| | Is it necessary to model time in discrete cycles? | | Continuously for Individual STM or in discrete cycles for Markov STM; if the assumption that transition probabilities do not depend on history is not required, then individual state-transition models are an alternative; If disease or treatment process need to be represented as health states, state transition models are appropriate (Markov type) |
| | Is there a need to model competition for resources or the development of waiting lists or queues? | | If the problem requires the ability of a model to incorporate interactions between individuals and other model parts for example to answer questions on resource allocation i.e., organ allocation for transplantation, distribution of antiretroviral medications in resource-poor environments, then a DES may be appropriate |
| | Has a type of model been chosen and discussed? | | It is expected that studies report on the reasons for choosing a type of model |
| Model structure | Has the starting cohort been defined by demographic and clinical characteristics affecting the transition probabilities or state values? | | If results may vary by subgroups (age, sex, risk factors) is advisable to report results for different cohorts |
| | Has health states and transitions reflecting the biological/theoretical understanding of the disease or condition been modelled? | | States should adequately capture the type of intervention (prevention, screening, diagnostics, and treatment) as well as the intervention's benefits and harms. States need to be homogeneous with respect to both observed and unobserved characteristics that affect transition probabilities |

Note: NA= Not Apply

| DIMENSION 3: SYNTHESIS OF EVIDENCE | | | |
|---|---|---|---|
| Components of good practice | Questions for review | Yes , No, or NA | Attributes |
| Data sources | Has transition probabilities and intervention effects been derived from representative data sources for the decision problem? | | Most common sources of data include population-based epidemiological studies, control arms of trials or literature |
| | Has (all) methods and assumptions used to derive transition probabilities and intervention effects been described/justified? | | Attention should be given to the use of transition probabilities and rates; conversion of transition probabilities from one time unit to another should be done through rates and never presented as percentages |
| | Has parameters relating to the effectiveness of interventions derived from observational studies been controlled for confounding? | | If results of meta-analyses were used as data sources then consider how potential confounders are addressed; consider the likelihood of increased heterogeneity resulting from residual confounding and from other biases across studies. Efficacy derived from RCT may have to be adjusted for compliance to reflect real-world effectiveness. Effectiveness derived from observational studies must be adjusted for confounding (e.g., using multivariate regression techniques or propensity scoring). Adjustment for time-varying confounding (confounders that simultaneously act as intermediate steps in the pathway between intervention and outcome) require special methods such as marginal structural analysis or g-estimation. When results from observational studies are used in the model, causal graphs can be used to explicitly state causal assumptions |
| | Has the quality of the data been assessed appropriately? | | Sources of data and data limitations are expected to be discussed |
| | Has expert opinion been used, are the methods described and justified? | | An expectation that strengths and limitations of assumptions adopted should be included |
| Utilities | Are the utilities incorporated into the model appropriate? | | methods used to obtain utility weights and methodology used to transform health estate estimates into quality of life scores |
| | Is the source for the utility weights referenced? | | Sources of data and data limitations are expected to be discussed |
| Cycle length and half cycle correction | Has the choice of cycle length been justified? | | It should be based on the clinical problem and remaining life expectancy |
| | Has the use of a half cycle correction been stated? | | Any assumption adopted is expected to be disclosed |
| Resources/ costs | Are the costs incorporated into the model justified and sources described? | | Sources of data and data limitations are expected to be discussed |

| | | | |
|---|---|---|---|
| | Has discount rates been reported and justified given the target decision-maker? | | |
| Patient heterogeneity | Has patient heterogeneity been considered? | | For example, in a cohort model states need to be homogeneous to observed or unobserved characteristics affecting transition probabilities to observed or unobserved characteristics affecting transition probabilities |
| Parameter precision | Has mean values and distributions around the mean and the source and rationale for the supporting evidence been clearly described for *each parameter* included in the model? | | Sources of data and data limitations are expected to be discussed |

Note: NA= Not Apply

| DIMENSION 4: ANALYSIS OF MODEL UNCERTAINTY | | | |
|---|---|---|---|
| Components of good practice | Questions for review | Yes , No, or NA | Attributes |
| Uncertainty | Has analyses of uncertainty pertaining to the decision problem been included and reported? If not, has the reasons been explained for its omission? | | Analysis of uncertainty is expected to be include as part of the DAM |
| Parameter estimation & uncertainty | Has one-way DSA or two-way sensitivity analysis been performed? | | Tornado diagrams, threshold plots or simple statements of threshold parameter values, are all appropriate. Uncertainty of parameters may be represented by several discrete values, instead of a continuous range, called 'scenario analyses'. It is a good practice to include the specification of parameter's point estimate and a 95% CI range. |
| | Has a Probabilistic Sensitivity Analysis (PSA) been included? | | The specific distribution (e.g. Beta, normal, lognormal) as well as its parameters should be disclosed. When PSA is performed without an accompanying EVPI, options for presenting results include CEAC 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. |
| | Has correlation among parameters been assessed? | | Lack of evidence on correlation among parameters should not lead to an assumption of independence among parameters |
| | If model calibration was used to derive parameters, has the uncertainty around calibrated values been tested using DSA or PSA? | | Calibration is commonly used to estimate parameters or adjust estimated values such as overall and disease specific mortality and event incidence rates |
| Structural uncertainty | Has a discussion about the inclusion/exclusion of assumptions affecting the structure of the model been included? | | For example: i) health states and the strategies adopted following the recurrence of events; ii) length of treatment effects; iii) types of adverse effects included; iv) duration of treatment effects; v) time dependency of probabilities (in a time dependent utility, the cost of delaying treatment as a function of the time a patient has remained in an untreated acute pathological state); vi) prognostic implications of surrogate end points or vii) clinical events. Although these structural assumptions are not typically quantified, it is uncertain whether they express reality accurately and for that reason they should be assessed as part of structural uncertainty analysis |
| Other reporting of uncertainty analyses | Has the EVPI being measured /discussed? | | If the purpose of a PSA is to guide decisions about acquisition of information to reduce uncertainty in the results, EVPI should be presented in terms of expected value of information. EVPI is commonly reported in monetary terms using net monetary benefit or net health benefits; EVPI should be reported for specified ICER thresholds |

Note: NA= Not Apply

| DIMENSION 5: MODEL TRANSPARENCY AND VALIDATION | | | |
|---|---|---|---|
| Components of good practice | Questions for review | Yes , No, or NA | Attributes |
| Transparency | Has a graphical description of the model been provided? | | |
| | Has all sources of funding and their role been identified? | | |
| | Has all methods used been customised to specific application(s) and settings? | | |
| | Has the report used nontechnical language and clear figures and tables to enhance the understanding of the model? | | |
| | Has limitations and strengths been acknowledged/discussed? | | |
| | Is there any reference as to whether technical documentation would be made available at request? | | |
| Validation | Is there any evidence of model's face validity? | | Can occur in several ways: the group that develop the model can appeal to members of the modelling group, people in the same organisation who did not build the model, or external consultants. Any reader can perform his/her own evaluation. Peer review (previous to publication) |
| | Has internal validity been assessed? | | Verification or technical validity; models should be subject to rigorous verification and the methods used should be described and results made available on request |
| | Has cross-validation been assessed? | | or external consistency (involves examining different models that address the same problem and comparing their results) its meaningfulness depends on the degree to which methods and data are independent. Modellers should search for modelling analyses of the same or similar problems and discuss insights gained from similarities and differences in results |
| | Has external validity been assessed? | | This compares the model's results with actual event data; a formal process needs to be developed including identifying suitable sources of data; results of external validation should be made available |
| | Has the model's predictive validity been assessed? | | If feasible given the decision problem and future's sources availability |

Note: NA= Not Apply

References


1. Drummond, M.F., et al., Methods for the Economic Evaluation of Health Care Programmes. Oxford Medical PublicationsResults of a model should be intuitive otherwise, counterintuitive results should be explained; all available data should be incorporated in the model; results of the model should be compared with results from previous models, and any differences explained. Vol. Third. 2005: Oxford University Press.
2. Petrou, S. and G. Alastair, Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ, 2011. **342**.
3. Jowett, S., Using Decision Analytical Modelling Techniques in Health Economics: An Application to Screening for and Treatment of Atrial Fibrillation. PhD Thesis. 2007, University of Birmingham.
4. Briggs, A.H., K. Claxton, and M.J. Sculpher, Decision Modelling for Health Economic Evaluation. Handbooks in Health Economic Evaluation. 2006: Oxford, University Press.
5. Philips, Z., et al., Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health technology assessment (Winchester, England), 2004. **8**(36): p. Sep.
6. Canadian Coordinating Office for Health Technology, A., Guidelines for the economic evaluation of health technologies: Canada (3rd Edition) (Special). 2006: Canadian Coordinating Office for Health Technology Assessment (CCOHTA).
7. Earnshaw, J. and G. Lewis, NICE guide to the methods of technology appraisal: Pharmaceutical industry perspective. PharmacoEconomics, 2008. **26**(9): p. 2008.
8. Karnon, J., A. Brennan, and R. Akehurst, A critique and impact analysis of decision modeling assumptions. Medical Decision Making, 2007. **27**(4): p. July/August.
9. Philips, Z., et al., Good practice guidelines for decision-analytic modelling in health technology assessment: A review and consolidation of quality assessment. PharmacoEconomics, 2006. **24**(4): p. 2006.
10. Sculpher, M., E. Fenwick, and K. Claxton, Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. [Review] [44 refs]. PharmacoEconomics, 2000. **17**(5): p. 461-477.
11. Sonnenberg, F.A., et al., Toward a peer review process for medical decision analysis models. Medical care, 1994. **32**(7 Suppl): p. Jul.
12. Soto, J., Health economic evaluations using decision analytic modeling: Principles and practices - Utilization of a checklist to their development and appraisal. International Journal of Technology Assessment in Health Care, 2002. **18**(1): p. 2002.
13. Weinstein, M.C., et al., Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR task force on good research practices - Modeling studies. Value in Health, 2003. **6**(1): p. January/February.
14. Briggs, A.H., et al., Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value in Health, 2012. **15**(6): p. 835-842.
15. Caro, J.J., et al., Modeling Good Research Practices - Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value in Health, 2012. **15**(6): p. 796-803.
16. Eddy, D.M., et al., Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Value in Health, 2012. **15**(6): p. 843-850.
17. Karnon, J., et al., Modeling using Discrete Event Simulation: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. Value in Health, 2012. **15**(6): p. 821-827.
18. Pitman, R., et al., Dynamic Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. Value in Health, 2012. **15**(6): p. 828-834.
19. Roberts, M., et al., Conceptualizing a Model: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Value in Health, 2012. **15**(6): p. 804-811.
20. Siebert, U., et al., State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. Value in Health, 2012. **15**(6): p. 812-820.
21. Barton, P., S. Bryan, and S. Robinson, Modelling in the economic evaluation of health care: Selecting the appropriate approach. Journal of Health Services Research and Policy, 2004. **9**(2): p. April.
22. Bentley, T.G., K.M. Kuntz, and J.S. Ringel, Bias associated with failing to incorporate dependence on event history in Markov models. Medical decision making : an international journal of the Society for Medical Decision Making, 2010. **30**(6): p. 2010-2Dec.
23. Brennan, A., S.E. Chick, and R. Davies, A taxonomy of model structures for economic evaluation of health technologies. Health Economics, 2006. **15**(12): p. December.

24. Karnon, J., Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. Health Economics, 2003. **12**(10): p. 01.
25. Andronis, L., P. Barton, and S. Bryan, Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health technology assessment (Winchester, England), 2009. **13**(29): p. Jun.
26. Bilcke, J., et al., Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Medical Decision Making, 2011. **31**(4): p. 675-692.
27. Bojke, L., et al., Characterizing structural uncertainty in decision analytic models: A review and application of methods. Value in Health, 2009. **12**(5): p. July/August.
28. Briggs, A.H., A.E. Ades, and M.J. Price, Probabilistic sensitivity analysis for decision trees with multiple branches: Use of the Dirichlet distribution in a Bayesian framework. Medical Decision Making, 2003. **23**(4): p. July/August.
29. Briggs, A.H. and A.M. Gray, Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technology Assessment, 1999. **3**(2): p. 1999.
30. Jackson, C.H., et al., A framework for addressing structural uncertainty in decision models. Medical Decision Making, 2011. **31**(4): p. 662-674.
31. Jain, R., M. Grabner, and E. Onukwugha, Sensitivity analysis in cost-effectiveness studies: From guidelines to practice. PharmacoEconomics, 2011. **29**(4): p. 2011.
32. Koerkamp, B.G., et al., The combined analysis of uncertainty and patient heterogeneity in medical decision models. Medical Decision Making, 2011. **31**(4): p. 650-661.
33. Kuntz, K.M. and S.J. Goldie, Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. Medical decision making : an international journal of the Society for Medical Decision Making, 2002. **22**(3): p. 2002-2Jun.
34. Sendi, P., A. Gafni, and S. Birch, Opportunity costs and uncertainty in the economic evaluation of health care interventions. Health Economics, 2002. **11**(1): p. 2002.
35. Strong, M., J.E. Oakley, and J. Chilcott, Managing structural uncertainty in health economic decision models: a discrepancy approach. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2012. **61**(1): p. 25-45.
36. Goldhaber-Fiebert, J.D., N.K. Stout, and S.J. Goldie, Empirically evaluating decision-analytic models. Value in Health, 2010. **13**(5): p. July-August.
37. Karnon, J. and T. Vanni, Calibrating models in economic evaluation: A comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. PharmacoEconomics, 2011. **29**(1): p. 2011.
38. Kim, L.G. and S.G. Thompson, Uncertainty and validation of health economic decision models. Health Economics, 2010. **19**(1): p. January.

**SUPPLEMENTAL MATERIAL (for possible WEB publication)**

**Appendix**

**Table 1: Search results from MEDLINE AND EMBASE**

| Steps | Search criteria | Number of hits |
|---|---|---|
| 1 | (checklist$ or check list$ or standards or standardi?ation or peer review or rules or critiquin or criteria or good or bad or correct$ or bias or fundamentals recommend$ or best or strength$ or weakness$ or quality or qualities or validity or guideline$ or validation or checkpoint$ or critically appraise or problems or limitations or rating scale$ or framework$ or protocol$ or audit or principles or methodolog$ or validate or validation or evaluating or properties or guidance or integrity or evaluation or pros or cons).m_titl. | 2261664 |
| 2 | limit 1 to abstracts | 1470011 |
| 3 | limit 2 to English language | 1266438 |
| 4 | limit 3 to yr="1990 -Current" | 1052800 |
| 5 | limit 4 to humans | 684348 |
| 6 | (economic model$ or Markov model$ or mathematical model$ or cost model$ or decision model$ or pharmacoeconomic$ model$ or decision tree$ or decision data or decision analytic$ or decision analysis or economic evaluation? or economic analysis).m_titl. | 30075 |
| 7 | limit 6 to abstracts | 24245 |
| 8 | limit 7 to English language | 21989 |
| 9 | limit 8 to yr="1990 -Current" | 19501 |
| 10 | limit 9 to humans | 12259 |
| 11 | 5 and 10 | 3930 |
| **12** | **remove duplicates from 11** | **2486** |

**Table 2: Search results from Cochrane Library**

| Steps | Search criteria | Number of hits |
|---|---|---|
| 1 | economic model* or economic analysis* or economic evaluation* or decision analytic* or decision analysis* or economic study* or economic submission* from 1990 to 2013, in Methods Studies, Technology Assessments and Economic Evaluations (Word variations have been searched) | 14677 |
| 2 | guideline* | 13983 |
| 3 | #1 and #2 | **1664** |