Guidance for the Conduct and Reporting of Decision and Simulation Models in the Context of Health Technology Assessment

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

This information does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services determination or policy.

Contract No.

Prepared by:

Investigators:

AHRQ Publication No.
<MonthYear>
This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted, for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested Citation: pending

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

William Lawrence, M.D., M.S.
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx-x). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.
Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Richard Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Yen-Ping Chiang, Ph.D.
Acting Deputy Director, Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

William Lawrence, M.D., M.S.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality
Contents
Introduction ................................................................................................................................... 5
Scope of the guidance .................................................................................................................. 5
Goals of modeling in EPC reports ............................................................................................... 6
Decision and Simulation Modeling Steps ................................................................................... 7
When is Decision and Simulation Modeling Worth the Effort? ................................................. 7
Guidance Development Process ............................................................................................... 8
Terminology and definitions ..................................................................................................... 8
Principles for best practice ........................................................................................................ 8
Recommendations for decision and simulation modeling in systematic reviews ................... 11
  Conceptualization and Structure .......................................................................................... 11
  Data ......................................................................................................................................... 13
  Consistency: Anticipating and correcting errors, model verification .................................... 17
  Consistency: Face validity, conceptual model validation ....................................................... 17
  Consistency: Operational model validation, confronting models with data ......................... 17
  Consistency: Predictive model validation ............................................................................. 18
  Reporting and interpreting results ......................................................................................... 18
References ................................................................................................................................... 20

Tables
Table 1. Definition of terms ........................................................................................................... 9
Table 2. Principles for best practice in decision and simulation modeling ................................. 10
Table 3. Conceptual definitions for the structure, data, consistency, reporting framework ....... 10
Introduction

Despite rigorous systematic reviews of efficacy and effectiveness of health care interventions, patients, providers and policymakers may remain in doubt about what they should do because of uncertainty, tradeoffs, and values. First, residual uncertainty may remain for meaningful patient-relevant outcomes from surrogate outcome measures or limited time followup or for subgroups from inadequate sample or inclusion/exclusion criteria. Second, tradeoffs occur, e.g., the U.S. Preventive Services Task Force (USPSTF) analysis of mammography for women in their 40s suggests a statistically significant reduction in breast cancer death but also potential harms, namely radiation exposure, overdiagnosis and overtreatment.\(^1\) Thus, optimal decisionmaking for individuals and populations may depend on their values (or preferences) for the outcomes. Lastly, just as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group considers resource utilization in guideline development, the USPSTF has examined modeling to estimate resource consumption for its recommendations on cancer screening. This document was designed to extend current Evidence-based Practice Center (EPC) guidance on developing methodological guidance for decision and simulation modeling.

We discuss principles and good practice recommendations for decision and simulation modeling. We believe this work to apply generally, but for concreteness, we place emphasis on models that could accompany systematic reviews produced by the EPC program. Such modeling exercises may be used to structure investigators’ thinking, and facilitate the communication of assumptions and results; synthesize data from disparate sources; inform decisionmaking; make predictions; or infer the impact of manipulations.

The goals of the guidance are to encourage the use of good modeling and reporting practices, while not being too prescriptive about how to develop specific models. We deemphasize issues specific to economic modeling, because economic assessments are not a priority of the EPC program.

Scope of the guidance

A model is a construct that represents salient aspects of reality in a simplified way. Models are physical (e.g., a scaled-down airplane wing tested in a wind tunnel) or theoretical constructs (e.g., a mathematical description of the flow of air around an airplane wing). Models that could be prepared in conjunction with systematic literature reviews are exclusively theoretical in nature; for this reason we do not discuss physical models further. The starting point for most theoretical models is a conceptual model, a simplified natural language or pictorial representation of reality. The analytic frameworks that are used to guide the conduct of systematic reviews prepared by the EPC program are conceptual models that often function as schematics of an underlying decision problem.\(^2\)\(^-\)\(^5\) For example, the analytic frameworks used in reviews of diagnostic tests, which often resemble decision trees.\(^6\) Although conceptual modeling is a prerequisite for the development of mathematical (quantitative) models, our current guidance focuses exclusively on the latter. Readers interested in the use of conceptual models in systematic reviews can consult relevant chapters of the methods guide.

Mathematical models are a large and diverse group of models that use variables, together with mathematical symbols that represent relationships between variables. The most common quantitative models encountered in systematic reviews are multivariable regression models (for
primary data analysis and meta-analysis). These models and other related techniques (e.g., neural networks) that aim to describe how a response (dependent variable) changes conditional on covariates (independent variables) are types of behavioral models. They describe how the response varies over values of the covariates, without necessarily referring to assumptions about the underlying mechanisms.\(^a\) The literature addressing these models is vast (e.g., in statistics or computer science) and not covered in this guidance. Instead, we address structural models, which attempt to capture mechanistic relationships among their components. Structural models include declarative (e.g., Markov models), functional (e.g., compartmental models), and spatial models (e.g., geographic information systems data models). In applied work, elements of these structural model subtypes are commonly combined (multi-models).

### Goals of modeling in EPC reports

This document does not provide detailed guidance to help investigators decide whether decision or simulation modeling should be undertaken. Issues related to the appropriateness of modeling in EPC reports are addressed by existing guidance and are not covered in this document.\(^7;8\) However, we briefly consider the potential goals of modeling when performed in conjunction with systematic reviews.\(^9-18\)

- **To inform decisionmaking under uncertainty:** The decisions that can be informed by modeling, even in the relatively narrow context of systematic reviews, are extremely diverse.\(^9;10\) They include decisions about patient-level care, the licensing of drugs or devices, healthcare policy decisions for populations, and decisions about the need to conduct additional research.

- **To structure investigator’s thinking, and facilitate the communication of data, assumptions and results:** modeling can help investigators organize their knowledge about the topic area, formalize the research question, and communicate assumptions and results to peers (e.g., topic or methodological experts) and stakeholders (e.g., patients, decisionmakers).\(^10\)

- **To synthesize data from disparate sources:** evidence on a specific research question may be available from multiple sources and a single study may contribute to the estimation of more than one model parameter (or functional combinations of parameters). Modeling provides the mathematical tools for synthesizing all evidence and facilitating assessments of consistency between sources. For example, models can be used to combine information from clinical trials of the impact of a treatment on intermediate outcomes can be combined with information from long-term cohort studies that assess of the association of the intermediate outcome with a clinical outcome of interest.

- **To make predictions:** predictions can refer to conditions similar to those already observed (sometimes referred to as “interpolations”; e.g., prediction of outcomes in a new study, similar to an existing one); or the future (forecasts), other populations, or other outcomes (sometimes referred to as “extrapolations”; e.g., predicting outcomes at longer-term followup times based on results of short-term clinical trials). They can also pertain to the prioritization and planning of future research.\(^19\) These predictions may be useful in themselves, even without reference to the anticipated effects of interventions.

---

\(^a\) In some cases behavioral models (e.g., regression) are used to estimate the parameters of structural models.
• To support causal explanations and infer the impact of interventions: Related to the above is prediction of the effects of possible or even hypothetical manipulations.20-22 When used this way, at least implicitly, models are claimed to encode structural causal mechanisms or emulate such mechanisms sufficiently well.

Conveying assumptions, synthesizing evidence, and informing decisions are probably the primary goals of decision and simulation models that would be developed in conjunction with systematic reviews.

**Decision and Simulation Modeling Steps**

Because decision and simulation models are used to achieve multiple goals and address diverse research questions, the model development and evaluation process is bound to differ across specific applications. Nonetheless, the key steps to develop quantitative models that are within the scope of this guidance can be identified23-29:

1. **Definition of the research question:** specify an answerable research question for the relevant stakeholder.
2. **Model conceptualization:** determine which components of a disease or process need to be represented in the model to address the research question, and describe their relationships.
3. **Data collection and processing:** identify data sources and process data to inform the model.
4. **Model implementation and mathematical manipulation:** ‘solve’ the model using mathematical or numerical analysis methods, or simulation techniques.
5. **Model evaluation:** detect model shortcomings by examining the model, and by comparing its output with data and other similar models.
6. **Reporting and interpretation of results:** present the model findings in a way that addresses the research question.

Model development is an iterative and dynamic process.27;30 Multiple iterations are typically needed between the phases outlined above because, at each step, the need for changes at earlier phases may become apparent. For example, the availability of some data (possibly preliminary or incomplete) often provides an incentive for modeling and simulation; as the model is conceptualized additional data needs may arise that require further data collection. Similarly, deficiencies that are detected at the evaluation phase may require restructuring of the model, supplemental data collection, or other modifications of the modeling strategy.

**When is Decision and Simulation Modeling Worth the Effort?**

Developing decision and simulation models, especially models that can be used to inform complex decisions or understand complex disease processes, is a demanding task. Similarly, choosing between alternative modeling approaches can be difficult because the correct choice is not always obvious early in the modeling process. Also, the same research question may be amenable to multiple modeling approaches, each with distinct strengths and disadvantages.

Although this document, and the cited references, can provide guidance on decision and simulation modeling methods, it is harder to define with certainty the circumstances under which
modeling is worth the investment of time and resources beyond those required for a systematic evidence review. In general, modeling is most useful when the research question is complex, data sources are multiple (and possibly conflicting), outcomes involve trade-offs, and choices are value-laden. The details of the research question, the availability of resources (e.g., analyst time and experience with the related methods), and the potential impact of modeling on future research, clinical practice and policy decisions, should also be considered when deciding about whether modeling efforts are likely to be beneficial.

**Guidance Development Process**

Development of this guidance document is the culmination of a multistep process of summarizing existing recommendations and soliciting stakeholder input. We first updated and expanded two systematic reviews of recommendations for the conduct and reporting of decision and simulation models, as described in detail in our companion paper. This was done with input from a multidisciplinary team of clinical, policymaking, and decision analysis experts. The results of our systematic review were discussed in-person with a panel of 28 stakeholders including patient representatives, providers of care, purchasers of care, payers, policy makers (including research funders and professional societies), and principal investigators. Stakeholders commented on available recommendations and identified gaps, limitations and areas for expansion. The stakeholders reviewed, added to, and prioritized the list of future research recommendations. We subsequently reviewed the websites of 126 international agencies and institutes conducting health technology assessments for their guidance or standards for how they conduct and report decision and simulation modeling, with an emphasis on how systematic reviews incorporated modeling. We solicited input from senior researchers at EPCs and AHRQ with experience in the methods of conducting decision and simulation models.

Based on the gathered systematic review evidence on modeling recommendations and guidance, we developed a list of recommendations to serve as guidance for developing models in conjunction with systematic reviews. There are two major types of recommendations: (1) those that follow from principles and are not amenable to empirical testing, and (2) those that can be tested empirically or through simulation. We provide the rationale for each guidance recommendation, evidence that the recommendation should be preferred, or best judgment where adequate evidence is lacking. We have also categorized the recommendations as proposed by Sculpher (2000), Philips (2004), and Kuntz (2013), by whether they pertain to the model structure, model data, or consistency, and reporting.

**Terminology and definitions**

Table 1 defines terms used in the recommendations.

**Principles for best practice**

We begin by outlining general principles for the conduct and reporting of decision and simulation modeling studies found consistently in our systematic review. We believe that these principles represent generally accepted rules for sound practice and have used them to guide our more specific recommendations, which are presented in the next section.
Modelers should consider (1) the goals of the modeling exercise; (2) the nature of the phenomena-to-be-modeled; (3) their own abilities in math and computation; and (4) objective constraints in terms of available time, data, or resources. Further, when developing, implementing and ‘solving’ models, one makes many methodological decisions. One should report them explicitly, justify them, and subject them to appropriate stability and sensitivity analyses.

### Table 1. Definition of terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Explanation and elaboration</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>A simplified representation of reality.</td>
<td>We focus on models which represent reality by means of mathematical relationships.</td>
</tr>
<tr>
<td>Simulation</td>
<td>A typical process of “solving” the model, especially when analytical solutions are cumbersome or intractable.</td>
<td>Often even analytically tractable models are ‘solved’ with simulation methods.</td>
</tr>
<tr>
<td>Model component</td>
<td>An element of a model. Model components may include variables (parameters), health states, agents, processes, and so on.</td>
<td>The descriptor is on purpose generic, to encompass all model types.</td>
</tr>
<tr>
<td>Stochastic (aleatory) uncertainty</td>
<td>Statistical uncertainty around the estimates of model variables that are informed by empirical data.</td>
<td>See below.</td>
</tr>
<tr>
<td>Structural uncertainty</td>
<td>Uncertainty secondary to our incomplete understanding of the modeled phenomenon. Typically it pertains to functional forms of relationships between model variables. At a more fundamental level, structural uncertainty will always exist, because the ‘true’ relationship between variables in the real world cannot be uncovered from data.</td>
<td>In some cases, the distinction between stochastic and structural uncertainty is a matter of definition. Some structural uncertainty would become stochastic uncertainty, if appropriate data were available to the modeler.</td>
</tr>
<tr>
<td>Propagation of (stochastic) uncertainty</td>
<td>The process of obtaining the stochastic uncertainty in model outputs. This derives from the stochastic uncertainty in model inputs. Propagation of uncertainty can be done: • Analytically, exactly, or up to an approximation (e.g., first order delta method) • Numerically, e.g., with forward Monte-Carlo simulations or equivalently with Markov Chain Monte Carlo (MCMC) methods</td>
<td>It is customary to use the term “probabilistic sensitivity analysis” (PSA) to refer to numerical propagation of uncertainty by means of forward Monte-Carlo methods. We do not use this term (PSA) in this work to avoid confusion.</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>The process of varying model variables over all elements of a set of interesting values and examine impact on results.</td>
<td>Sensitivity analysis has a ‘continuous’ aspect.</td>
</tr>
<tr>
<td>Stability analysis</td>
<td>Performing discrete actions and evaluating their impact on results. Examples include • Changing the structure of the model, e.g., use alternative statistical distributions or different functional forms for relationships between variables • Systematically excluding input data (e.g., leave one-study-out in a meta-analysis</td>
<td>Stability analysis involves discrete actions.</td>
</tr>
</tbody>
</table>
Table 2. Principles for best practice in decision and simulation modeling

<table>
<thead>
<tr>
<th>Recommendation Areas</th>
<th>Description of what is encompassed</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualization and structure</td>
<td>Conceptualization pertains to the decision to use modeling, and the delineation of the perspective and scope. Structure pertains to variables, health states and other components of the model, and how they relate to each other (i.e., the mathematical scaffold of the model).</td>
<td>Consider a discrete-time Markov model. The disease states, variables informing transition probabilities, the mathematical relationships among the variables, the time horizon of the model and so on characterize the model's structure.</td>
</tr>
<tr>
<td>Data</td>
<td>Model inputs. May be obtained through empirical investigation, systematic elicitation of opinion, or best judgment/introspection.</td>
<td>Estimates for variables in the model, e.g., treatment effects, transition probabilities, costs, and utility weights.</td>
</tr>
<tr>
<td>Consistency</td>
<td>Whether the model achieves its stated goals, and the processes of assessing this.</td>
<td>Determination of whether the model has logical errors and whether the model output is consistent with expert opinion, observed data, or other models.</td>
</tr>
<tr>
<td>Reporting</td>
<td>Summarizing model output to achieve modeling (e.g., further one’s understanding the topic, inform decisionmaking).</td>
<td>Incremental cost effectiveness curves (to present the results of cost-effectiveness analyses), risk diagrams (to represent model-based risk assessments), and tornado diagrams (to summarize sensitivity analyses).</td>
</tr>
</tbody>
</table>

The following subsections provide detailed guidance on the conduct of decision and simulation modeling in the context of systematic reviews, organized by ‘conceptualization and structure’, ‘data’, ‘consistency’, and ‘reporting’. Briefly, structure and data are what constitute the model proper; consistency refers to an assessment of the model against its stated goals; and reporting considers issues related to results reporting and presentation. Table 3 provides operational definitions and examples for these areas of modeling.

Table 3. Conceptual definitions for the structure, data, consistency, reporting framework

<table>
<thead>
<tr>
<th>Recommendation Areas</th>
<th>Description of what is encompassed</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualization and structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This guidance is provided to facilitate the use of decision and simulation modeling in conjunction with systematic reviews, particularly as they are prepared within the AHRQ EPC program. The recommendations provide general guidance about conceptualizing, specifying, implementing, and evaluating decision and simulation models. It is not possible to provide detailed recommendations about which structures to use in which cases, or instructions about the (technical) implementation. Interested readers should consult some of the numerous books, technical reports available on this topic (several of which are cited below), including the vast literature on decision and simulation modeling in areas other than healthcare. Other general guidance documents, which were used as sources for the current set of recommendations, should also be consulted.7;27;31;33;34;36;40-48

**Recommendations for decision and simulation modeling in systematic reviews**

**Conceptualization and Structure**

The research question should be explicit. The decision to use modeling for addressing the research question should be described and justified.

As described earlier, modeling will be useful for many research questions, especially when the research questions are not directly answerable by existing empirical data. Defining the question at hand and the objective of the analysis may require using literature-based information, expert knowledge, and input from Key Informants and Technical Experts.30

The choice of perspective depends on the research question to be addressed and the relevant decisionmaker. There is no a priori preferred modeling perspective.

The modeling perspective determines the methods for choosing and handling consequences, values, and costs in the model; thus is should depend on the research question and the decisionmaker.49 For example, when modeling a specific clinical interaction where the decisionmaker is an identifiable patient, the appropriate perspective is that of the individual patient. In contrast, when the goal is to inform the decisionmaking of a public payer or a federal agency, one should prefer a payer or societal perspective.50

That said, the societal perspective (which considers impact on sectors beyond healthcare and includes time costs, opportunity costs, and community preferences) may allow for a more complete accounting of benefits and costs.51 For this reason, it has been recommended as the appropriate perspective in “base case” analyses.52 Obtaining appropriate data for modeling from a societal perspective is challenging and may be difficult to do well.51,53,54 Simplifications have been advocated, such as ignoring impact on outside non-healthcare sectors, to represent a ‘partial societal perspective’.

The model’s scope should be described and justified. The model’s scope should be consistent with the research question and the model’s perspective.

A model represents only some aspects of the phenomena under study. The research question defines how elaborate the modeling should be, and what aspects of reality it chooses to represent or omit (for parsimony). For example, many research questions in healthcare pertain to length of life, and mortality outcomes should be within the scope of models answering these questions.
More broadly, the scope of the model includes defining the condition or disease of interest, populations, risk factors, diagnostic or therapeutic interventions. For decision models one should also define the decision-relevant quantities, the decision (optimality) criteria, and the decisionmaking perspective. This is akin to defining a systematic reviews’ eligibility criteria (Population, Intervention, Comparator, Outcomes).

**The mathematical structure of the model and its implementation (computation) should correspond to the research question, the model's scope, and the decisionmaking perspective. The rationale for the choice of a mathematical structure should be provided, and structural assumptions should be explicated and justified.**

The preferred model structure depends on the research question and the model scope. The model structure should reflect the current understanding of the topic being modeled (e.g., disease prognosis and treatment effects, diagnostic test application, public health interventions) Disease states and transitions, or functional relationships should reflect understanding of the course of the disease. Detailed guidance on choosing among alternative mathematical structures (and computational implementations) is beyond the scope of this work. Readers are referred to the extensive technical literature available in healthcare and other fields. Of note, relatively simple models (e.g., decision trees, time-homogeneous Markov chain-based models) may be appropriate for use in the setting of most EPC evidence reports, particularly when the goal of modeling is to contextualize the evidence and extend review findings.

**The model should allow for comparisons among all interventions that are relevant to the research question, model scope, and decisionmaking context.**

In many cases the goal of modeling is to inform decisionmaking about the implementation of an intervention (e.g., a specific treatment or policy) or to assess the impact of modifying the levels of risk factor or an exposure (e.g., reducing cholesterol, or eradicating a disease agent from the environment). In such cases, the model should allow the inclusion of all relevant and feasible interventions (or exposures). In general, feasible options should not be excluded from the model; in the rare case that such exclusions are deemed necessary, they should be carefully justified.

**The time horizon should be long enough to allow all meaningful outcomes to be evaluated fully.**

When comparing alternative interventions, the time horizon should be long enough to allow the manifestation of differences in important outcomes. In some cases, a short time horizon may be adequate to compare interventions (e.g., when modeling the effectiveness of interventions for alleviating symptoms of the common cold); in many cases a life-time horizon is needed, particularly when modeling the effects of long-term treatment of chronic disease. Choice of long-term time-horizons has implications for the data used to populate models (e.g., life-time horizons almost always require the extrapolation of treatment effects well beyond the followup duration of available clinical trials).

**When deciding how to handle time, space, interagent interactions, and health states, one should consider the “nature” of the modeled phenomena, and the**
convenience of, and calculation errors associated with, alternative modeling choices.

For example, when deciding how to deal with time, we have three options: (1) do not model it explicitly (integrate it out, as for example in decision trees); (2) model it as a continuous quantity (as in differential-equation-based dynamic systems); (3) model it as a discrete quantity (as in discrete-time Markov processes). Whether time is modeled as continuous or discrete should be guided by the specifics of the system being modeled and the process for making decisions (e.g., whether decisions are made in a continuous fashion or only at specific timepoints). In some cases where discrete modeling may be appropriate (e.g., modeling the occurrence of an outcome when measurement is possible only at specific intervals), continuous time models may offer convenient mathematical approximations. The converse may be the case in problems of a continuous nature that can be approximated by more tractable discrete-time models (e.g., models describing the development of epidemics). For discrete-time models, the cycle length should match the speed of changes in the system being modeled (e.g., the natural history of the disease, or the anticipated temporal evolution of a system).

Analogous considerations pertain to modeling vs not modeling spatial location; inter-agent interactions (interactions between people); and modeling health states in various degrees of granularity (e.g., disease severity).

The targeted level of complexity (or parsimony) should be determined by the research question and the model’s scope.

Models should be complex enough to capture all pertinent aspects of the system being modeled but not more (‘rule of reason’). At the same time, models should be as simple as possible to facilitate timely development, error checking, and validation. Simple models are also generally more accessible to nontechnical stakeholders of the modeling process, and their results can be communicated more easily. The tradeoff between simplicity and complexity should be driven by considerations related to the research question and the context in which model results will be used.

Data

Methods for identifying, and analyzing data should be described. Data choices should be driven by the research question and the model’s scope and structure. All data sources should be reported clearly and appropriate references should be provided.

To enhance transparency and face validity, the source of each data element should be identified fully. Particularly for data that is not derived from systematic review and meta-analysis, the rationale for why the given value was chosen. This applies both to the base case data and, for each data element, the range of values to be tested in sensitivity analyses.

Estimates for influential variables should be obtained from systematic processes (systematic review).

For decision and simulation models prepared jointly with a systematic review of studies of interventions, the summary estimates of treatment effects and related parameters should provide the data to inform the relevant model parameters. In particular, model parameters likely to have a large influence on model results should be identified through a systematic and replicable process.
that aims to minimize bias.\textsuperscript{76-80} In most cases this will mean conducting systematic literature reviews to inform influential parameters. However, in many cases only part of the evidence retrieved by the systematic review will be appropriate for use in the model. The research question, decisional context, and goals of modeling should inform the choice of which studies to be included and the choice of synthesis methods.\textsuperscript{80-87}

Decision and simulation models typically require data on multiple parameters that are not collected in systematic reviews, such as prevalence, incidence, costs, and utilities. Appropriate sources of such data can include registries and other large observational studies, studies found through a nonsystematic approach, stakeholder panel opinions, and domain expert judgment. When retrieving and processing data, modelers make a large number of methodological decisions that can appreciably impact results. Thus, all such decisions should be reported and justified. Supplementary material describing detailed methods and data sources can be made available electronically.

**Obtaining estimates for model variables should follow epidemiological and statistical principles.**

All major assumptions and methodological choices should be reported and justified. Several excellent sources provide detailed guidance on data management and manipulation, exploratory data analysis, inference, estimation, and related computational techniques. When multiple studies contribute information on a parameter of interest (e.g., treatment effectiveness, prevalence of disease, accuracy of a diagnostic test) evidence should be synthesized across studies using appropriate meta-analytic methods.\textsuperscript{77,78} Detailed guidance on the conduct of quantitative synthesis for different types of data structures is beyond the scope of this document; interested readers should refer to the relevant EPC guidance, and the many sources on meta-analysis and evidence synthesis.\textsuperscript{82,83-113}

**A “best evidence approach” should be used when selecting data sources for model parameters.**

Data from randomized trials cannot be used to inform all model parameters because (1) some parameters are best estimated from alternative study designs (e.g., the prevalence of a risk factor is best estimated from a sampling survey of a representative population; the test performance of a diagnostic test is best estimated from a cohort study); (2) available randomized trials may not be sufficiently applicable to the population to be modeled (e.g., trials may enroll highly selected populations, provide inadequate information for subgroups of interest, or have short followup duration); and (3) trials may not be available. In all these cases, evidence from other study designs will have to be included in the model. Researchers contemplating decision and simulation modeling in the context of systematic reviews will have to select and appraise appropriate study types for each model parameter.\textsuperscript{114} General guidance on “best evidence” strategies in systematic reviews is provided by a recent EPC Methods Research report.\textsuperscript{115}

**The risk of bias of the available data should be assessed and accounted for when obtaining estimates for model parameters.**

One should avoid using unadjusted, incompletely adjusted, or inappropriately adjusted results that are potentially biased simply because no other information is available.\textsuperscript{116,117} Instead, modelers should consider using methods that allow the adjustment of study results to account for all sources of bias and related uncertainty (i.e., multiple bias adjustment, possibly in the setting of sensitivity analysis).\textsuperscript{118,119}
The factors that contribute to a study’s risk of bias depend on the specific modeling context and the study designs considered (e.g., sources of bias for surveys are distinct for those of randomized and non-randomized studies). These factors should be assessed for each study. For modeling, it is generally not adequate to assess the risk of bias of individual studies (or entire bodies of evidence). Because models typically are specified with respect to “true” parameters, it is desirable that model inputs be ‘corrected’ (adjusted) for biases.

The direction and magnitude of the bias associated with each item, the uncertainty around their effect, and the relationship between bias items (e.g., whether bias effects are additive or nonadditive effects), should be incorporated in the analyses. In most cases, the effect of bias items (also known as “bias parameters”) cannot be identified from study data; thus, one should use methods that incorporate external information (empirical or judgmental). Extensive literature exists on the assessment of specific risk of bias items for individual studies, as well as methods for multiple bias modeling (i.e., bias adjustment and risk analysis).116;118;120-132

Formal elicitation methods should be used to quantify expert opinion and the associated uncertainty.

When no empirical evidence is available for parameters of interest, modelers will have to rely on expert opinion. Current EPC processes for Key Informant and Technical Expert engagement (during the development, refinement and conduct of systematic reviews), or similar processes, can be leveraged to incorporate formal opinion elicitation methods. Modelers should be aware that elicitation methods (e.g., the framing of questions) can influence the information that is obtained, particularly when the subjects of the elicitation process have labile values for the parameters of interest.133 Interested readers should consult the extensive literature on elicitation methods for different types of parameters.134-143

Assumptions required for extrapolating from existing data should be reported and justified. They should also be subjected to stability and sensitivity analyses.

A particular challenge arises when there is need to extrapolate beyond the observed data (e.g., to longer followup periods, or to other populations). Such extrapolations are based on assumptions about unobserved data. These assumptions should be reported and justified; they should also be subjected to sensitivity analyses (e.g., assessing a range of values for the parameters of the chosen survival distributions) and stability analyses (e.g., using alternative survival distributions when extrapolating survival times).144

The assumptions required for transporting information across studies to a common (new) setting should be described, justified and subjected to stability and sensitivity analyses.

Decision and simulation models often use data obtained from diverse sources.15 In fact, modeling is often used with the explicit goal of synthesizing information from diverse domains (e.g., treatment effect estimates from trials of selected populations may be combined with natural history information from large observational cohorts). In such cases, the validity of modeling results depends on the validity of assumptions about the transportability of effects across domains. These assumptions should be identified explicitly and justified based on theoretical considerations and the understanding of the underlying mechanisms.145;146 Consideration should be given to formal (causal) methods for assessing the transportability of results across domains.147-152
Analyses should take into account heterogeneity (nonrandom variation) in all parameters.

As a general principle, decision and simulation models should account for clinical heterogeneity, defined as nonrandom (systematic) variation in parameters of interest.153;154 Attempts should be made to explain heterogeneity via appropriate statistical methods (e.g. subgroup or regression analyses) by incorporating information on determinants of variability. Because our current understanding of any topic is likely to be incomplete (e.g., important modifiers of effect may be unknown) and because data unavailability may limit our ability to explore heterogeneity (e.g., well-known modifiers may not be measured or reported in published studies), models should also allow for residual (unexplained) variation.

Unexplained heterogeneity arises very often in meta-analyses of treatment effects using published (group level) level data. In such cases efforts to explain heterogeneity rely primarily on meta-regression methods and residual heterogeneity is accounted for using random effects models.87;155-158 Modelers should be aware that random effects models can “average over” and obscure, important data patterns and – contrary to popular belief – are not always more conservative that fixed effect models.159;160 Person-level data can allow decision and simulation models to meaningfully incorporate heterogeneity;161-168 however, their use is very uncommon in systematic reviews prepared by EPCs or meta-analyses published in peer-reviewed journals.169

Models should propagate the uncertainty in inputs to outputs.

Data analysis should allow unbiased parameter estimation and appropriately account for parameter uncertainty from model inputs to model outputs.27;153;154;170-179 Sometimes this can be done analytically, either exactly, or approximating up to an order of error, such as with the delta method. In most cases it is computationally convenient to propagate uncertainty with numerical methods, typically with a forward-Monte-Carlo approach. It is customary to use the term “probabilistic sensitivity analysis” (PSA) to refer to numerical propagation of uncertainty by means of forward Monte-Carlo methods. We do not use this term in this work.

Detailed guidance on the conduct of probabilistic analyses is available elsewhere.153;154;173;180-186 Of note, probabilistic methods for incorporating and propagating uncertainty in models do not eliminate the need for stability and sensitivity analyses. For example, the use of a specific probability distribution to represent uncertainty around a model input does not eliminate the need to assess the impact of using alternative probability distributions (stability analysis) or the need to assess the impact of evaluating permutations of the distributional parameters (sensitivity analysis).

Depending on the goal of the model, in rare cases it may not be desirable, or necessary, to perform analyses that propagate uncertainty. For example, for decisional problems where optimality is judged with minimax or maximin criteria, an analysis of bounds (extreme values) may suffice. Furthermore, if substantial uncertainty exists about the appropriate distributional form for estimates of model inputs, it may be futile to insist on probabilistic analyses, and may be appropriate to set more modest and attainable goals for the modeling exercise (e.g., use models to gain insights or to communicate implications). When such cases arise, analysts should provide the rationale for not using probabilistic analyses.
Consistency: Anticipating and correcting errors, model verification

After implementing the model, attempts should be made to detect errors in its logic and implementation.

Errors are unavoidable in any nontrivial model. Mistakes in the research question formulation, the model structure, incorporation of data, or software implementation can become apparent during any phase of model development, and may require revising the structure, or collecting additional data. Errors in implementation can be challenging to detect, and can also have important consequences. The risk of mistakes in question formulation and model structure can be reduced by adhering to some of the principles outlined earlier in this document (e.g., consulting with topic experts, using a conceptual model to guide the mathematical model implementation), together with transparent reporting of methods and results and the use of teams with sufficient expertise. Several checking techniques have been advocated for healthcare-related models (e.g., sensitivity analysis, extreme value analysis). In addition, software production techniques such as unit testing, code review (review of one programmer’s work by another team member), paired programming (i.e., one programmer’s coding is monitored by another in real time) can be considered. Duplicate implementation of the same model by an independent team can also be used to identify errors in coding. Because these strategies can substantially increase the costs of model development, their use should be balanced against the modeling goals, model complexity, and anticipated frequency and impact of errors.

Consistency: Face validity, conceptual model validation

Topic experts should be invited to review the model structure and outputs and to judge whether they appear consistent with their expectations. Counterintuitive model results should be described and explained.

An evaluation of the model and its results by a group of topic experts can alert modelers to the presence of deficiencies in model structure or data. Counterintuitive model results ("paradoxical findings") may indicate errors. If an error has been ruled out, they should be described and explained, with reference to model structure, available data, and current understanding of the modeled phenomena.

Consistency: Operational model validation, confronting models with data

The consistency between model outputs and the data on which the model was based should be evaluated.

A combination of graphical and statistical methods should be used to compare model outputs with expected results. For parameters that are identifiable using available data, model validation is essentially an assessment of model fit. As such, comparisons of observed versus expected values (graphical or statistical) can be used to identify potential areas of improvement in model structure and assumptions.
Data should not be withheld from model development with the sole purpose of assessing model validity.

Generally, data should not be withheld during model development for the purpose of using the data for model validation. Using all available data during model development (for parameter estimation) is more efficient (because all information is incorporated), allows appropriate handling of correlated inputs, and permits the assessment of consistency across available sources of evidence.\textsuperscript{196} Resampling-style methods to compare the “fit” of the model to available data and the detection of outlying or influential observations can be used. Additional model validation methods are available under a Bayesian framework. Even when parameters are not identifiable by available data (e.g., parameters related to the rate of tumor growth in cancer microsimulation models) holding out data on identifiable parameters (or on functional combinations of identifiable and nonidentifiable parameters) is, in general, less efficient than joint modeling.

If multiple models addressing the same research question are available their results should be compared and any discrepancies explained.

Results from independently developed models addressing the same research question can be available by design (comparative modeling) or happenstance (e.g., multiple teams working on the same research question simultaneously).\textsuperscript{197} If such independent models are known about or identified through literature review, then outputs from different models should be compared as part of cross-model validation and any discrepancies need to be explained, with reference to the model structure and data inputs of each model.

Consistency: Predictive model validation

The appropriateness of using future observations to evaluate a model depends on the research question and the intended use of the model.

A comparison of model output with future empirical results (unavailable at the time model development) is not an appropriate method of evaluation for some models.\textsuperscript{33} In general, models used to guide decisionmaking or to contextualize and synthesize evidence at a specific point in time should generally not be evaluated with respect to their ability to predict the results of future empirical research.\textsuperscript{18,33,198} The majority of models developed in conjunction with systematic evidence reviews are likely to belong to this category. However, for models intended as predictive or forecasting tools, predictive validation is an important component of model assessment.

Models should be updated as understanding of disease improves, and as new interventions and empirical data become available.

Models should be updated as our understanding of disease mechanisms (causal agents, natural history), potential interventions (e.g., new treatments or variations of existing treatments) and their associated benefits and costs evolve. The model structure and its software implementation need to be flexible enough to accommodate this updating process.

Reporting and interpreting results

The model structure, data used to populate it, and results should be transparent.

Information about the model structure and data used to populate it should meet the standards of reproducible research.\textsuperscript{43,46,47,188,199} This is particularly important for models that are supported
by public funds (e.g., models that can be created in conjunction with EPC evidence reports) or 
models used to inform decisions that affect public policy. Transparent reporting will generally 
involve a detailed technical description of the model structure, an implementation of the model 
in computer code (or equivalent formats, such as spreadsheet files), a detailed tabular 
presentation of model inputs (e.g., probability distributions and their parameters) together with 
the data sources used to estimate these parameters. This level of transparency allows rigorous 
external peer review of the model, increases public trust in the modeling enterprise, and 
facilitates future research in the content area (e.g., extensions of the model to incorporate new 
data or to make it transferable to new settings) and in modeling methodology (e.g., cross-model 
type comparisons or technical extensions of the model).

**Reported results and their interpretation should convey uncertainty in model outputs.**

Results should be reported in a way that conveys uncertainty in model output. This may 
include the use of graphical and statistical summaries that convey the degree of uncertainty in 
model results (e.g., confidence bands, credible intervals, scatterplots of multiple model runs), 
together with summaries of sensitivity and stability analyses. Given the large number of 
methodological choices made at every step of model development, and the inherent subjectivity 
of drawing conclusions from complex research activities, we believe that general purpose 
algorithmic approaches cannot be developed or recommended for summarizing model results. 
Instead, we recommend complete reporting of model structure and data, coupled with 
transparency in presenting the modelers’ rationale for their decisions.

**Results should be reported in a way that addresses user needs.**

Because models have many different functions, reporting of results should be tailored to the 
goals of the modeling effort, while remaining faithful to the model structure and assumptions, 
and conveying all uncertainty in the results. Every effort should be made to report the model 
findings and analyses in a manner that will be most useful to the stakeholders who would be 
expected to use the report. It is impossible to provide specific guidance to address all possible 
model types and uses of modeling in this document. Interested readers are referred to the many 
available texts on healthcare modeling, the reporting of statistical and simulation analyses, and 
graphing quantitative information.

**All individuals who provided input to, developed and analyzed the model and interpreted its results should fully disclose any perceived conflicts of interest.**

As with any research, all investigators should provide full disclosures of any interests that 
can reasonably be perceived as a conflict. Both financial and nonfinancial conflicts of interest 
should be provided. For models produced for the AHRQ EPC program and many other 
HTA groups, it is necessary that conflicts of interest be avoided. Modelers should adhere to 
established guidance for managing conflicts of interest for EPC products (e.g., Institute of 
Medicine recommendations; existing EPC guidance).
References


Ref Type: Report


Ref Type: Generic


(13) Halpern EF, Weinstein MC, Hunink MG, Gazelle GS. Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. 2000;20:314-322


Ref Type: Report


(36) Sainfort F, Kuntz KM, Gregory S et al. Adding decision models to systematic reviews: informing a framework for deciding when and how to do so. *Value in Health* 2013;16:133-139


(50) Roy S, Madhavan SS. Making a case for employing a societal perspective in the evaluation of Medicaid prescription drug interventions. *PharmacoEconomics* 2008;26:281-296


(60) Caro JJ, Moller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? *Value in Health* 2010;13:1056-1060


(72) van Rosmalen J, Toy M, O'Mahony JF. A mathematical approach for evaluating Markov models in continuous time without discrete-event simulation. Medical Decision Making 2013;33:767-779


(79) Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. 2003;7:iii


(83) Graham PL, Moran JL. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *Journal of Clinical Epidemiology* 2012;65:503-510


(85) Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549


(115) Chopra A. 111In/125/131I-Labeled anti-mucin-1 murine, chimeric or humanized antibody hPAM4. 2004. PMID: 21834178


(118) Braithwaite RS, Roberts MS, Justice AC. Incorporating quality of evidence into decision analytic modeling. *Annals of internal medicine* 2007;146:133-141

(119) Goldhaber-Fiebert JD. Accounting for biases when linking empirical studies and simulation models. *Medical Decision Making* 2012;32:397-399


Ref Type: Report


(129) Molitor N-T, Best N, Jackson C, Richardson S. Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by products. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2009;172:615-637


(137) Frew EJ, Whynes DK, Wolstenholme JL. Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. *Medical Decision Making* 2003;23:150-159


(146) Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. 2005;21:165-171


(149) Causal transportability with limited experiments. 2013.

(150) Transportability of causal and statistical relations: A formal approach. IEEE, 2011.


Ref Type: Report

(152) Pressler TR, Kaizar EE. The use of propensity scores and observational data to estimate randomized controlled trial generalizability bias. *Statistics in Medicine* 2013;32:3552-3568


(156) Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* 2008;37:1158-1160


(160) Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999;150:469-475


(169) Kovalchik SA. Survey finds that most meta-analysts do not attempt to collect individual patient data. Journal of Clinical Epidemiology 2012;65:1296-1299


(183) Brisson M, Edmunds WJ. Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. *Medical Decision Making* 2006;26:434-446


(195) Stout NK, Knudsen AB, Kong CY, McMahon PM, Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. Pharmacoeconomics 2009;27:533-545


(199) Peng RD. Reproducible research in computational science. 2011;334:1226-1227


(201) Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL. Toward a peer review process for medical decision analysis models. Med Care 1994;32:JS52-JS64
Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 4--Analyzing the model and interpreting the results. *Medical Decision Making* 1997;17:142-151


Cleveland WS. The elements of graphing data. Monterey. *Calif, Wadsworth* 1985;


Barbieri M, Drummond MF. Conflict of interest in industry-sponsored economic evaluations: real or imagined? *Current Oncology Reports* 2001;3:410-413


Ref Type: Report