Methods Guide
for Comparative Effectiveness Reviews

Guidance for the Conduct and Reporting of Modeling and Simulation Studies in the Context of Health Technology Assessment
This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, 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.

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

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies, may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

Prepared by:
Tufts Evidence-based Practice Center
Boston, MA

Investigators:
Issa J. Dahabreh, M.D., M.S.\textsuperscript{1,2,3}
Thomas A. Trikalinos, M.D.\textsuperscript{1,2}
Ethan M. Balk, M.D., M.P.H.\textsuperscript{1,2}
John B. Wong, M.D.\textsuperscript{4}

\textsuperscript{1}Center for Evidence-based Medicine, School of Public Health, Brown University, Providence, RI
\textsuperscript{2}Department of Health Services, Policy & Practice, School of Public Health, Brown University, Providence, RI
\textsuperscript{3}Department of Epidemiology, School of Public Health, Brown University, Providence, RI
\textsuperscript{4}Division of Clinical Decision Making, Tufts Medical Center, Boston, MA
Editors’ Foreword
These guidelines encourage Evidence-based Practice Centers to quantify biases. The idea is not new—in fact, it was proposed and debated at length in the late 1980s and early 1990s. Most systematic reviews take a less transparent approach. In the example in Box 1, the investigators have pooled five studies and propose a “fair” rating for the overall quality of this evidence. Judging by the effect size, they seem to be saying that an intervention decreases mortality, but they also seem to be saying not to trust that estimate. When the investigators have a belief about the magnitude and direction of bias, the default approach (depicted in the box) is to put forward a numerical estimate they believe to be wrong and then qualify it—for example, “We think the estimate may be high because these are the first trials of this intervention, and early trials tend to have exaggerated effect sizes.” This approach makes it difficult to justify putting out a numerical estimate.

Box 1. Example

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pooled Effect Size</th>
<th>Overall Quality of Evidence</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>0.85 ± 0.13</td>
<td>Fair</td>
<td>5</td>
</tr>
</tbody>
</table>

Explicit adjustment for bias means that the pooled effect size estimate would take into account concerns about quality and be interpretable as the authors’ best estimate of the effect. While the explicit approach is appealing in many ways, most systematic reviewers fear that bias adjustments will introduce subjectivity and error rather than improve transparency. They note correctly that there is no reliable reproducible approach to making these estimates, and that the magnitude and direction of bias are often unpredictable.

Given this concern, it is important to note the following:
1. The basic recommendation is to use quantitative bias adjustments to integrate the reported effect sizes with the assessment of risk of bias or quality when meta-analysis is used alongside decision modeling or simulation.
2. Evidence-based Practice Centers are not required to use quantitative bias adjustments in systematic reviews and meta-analyses when decision or simulation modeling is not done.
3. If the investigators’ true belief is that no adjustment is needed (e.g., that the adjustment factor should be 1.0), it is important to convey that judgment. There is no requirement to use a factor different from the investigators’ true belief. What is important is to be transparent about our confidence in the estimate instead of using vague qualitative statements. If, in the example shown, the investigators are so unsure about the likely magnitude and direction of bias that they are unwilling to use an adjustment factor other than 1.0, this conveys important information to the reader about the emptiness of the “fair” rating.

Mark Helfand, M.D., M.P.H.
Stephanie Chang, M.D., M.P.H.

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.

Strong methodological approaches to systematic review improve the transparency, consistency, and scientific rigor of these reports. Through a collaborative effort of the Effective Health Care (EHC) Program, the Agency for Healthcare Research and Quality (AHRQ), the EHC Program Scientific Resource Center, and the AHRQ Evidence-based Practice Centers have developed a Methods Guide for Comparative Effectiveness Reviews. This Guide presents issues key to the development of Systematic Reviews and describes recommended approaches for addressing difficult, frequently encountered methodological issues.

The Methods Guide for Comparative Effectiveness Reviews is a living document, and will be updated as further empiric evidence develops and our understanding of better methods improves.

If you have comments on this Methods Guide paper, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Andrew Bindman, M.D.                Arlene S. Bierman, M.D., M.S.
Director                              Director
Agency for Healthcare Research and Quality Center for Evidence and Practice Improvement

Stephanie Chang, M.D., M.P.H.
Task Order Officer
Director
Evidence-based Practice Center Program Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality
Acknowledgments

The authors would like to thank the following EPC Directors, who offered comments on previous versions of this report, and the Associate Editor.

Eric B. Bass, M.D., M.P.H.
Director
Johns Hopkins Evidence-based Practice Center
Johns Hopkins University
Baltimore, MD

Roger Chou, M.D.
Director,
Pacific Northwest Evidence-based Practice Center
Professor of Medicine
Professor of Medical Informatics and Clinical Epidemiology
Oregon Health & Science University
Portland, OR

Mark Helfand, M.D., M.S., M.P.H.
Associate Editor
Evidence-based Practice Center Program
President
Society for Medical Decision Making
Professor of Medicine
Professor of Medical Informatics and Clinical Epidemiology
Oregon Health & Science University
Portland, OR

Robert L. Kane, M.D.
Director
Minnesota Evidence-based Practice Center
Professor and Minnesota Chair in Long-Term Care and Aging, Health Policy & Management
School of Public Health
University of Minnesota
Minneapolis, MN

Jennifer Lin, M.D., M.C.R.
Associate Director
Kaiser Permanente Research Affiliates Evidence-based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR
The authors acknowledge the contributions of Drs. William Lawrence, M.D., M.S., and Joanna Siegel, Ph.D., who served as Task Order Officers for some of the duration of this project.
Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who provided input to this report follows:

Eric B. Bass, M.D., M.P.H.
Director
Johns Hopkins Evidence-based Practice Center
John Hopkins University
Baltimore, MD

Andrew Briggs, D.Phil.
William R. Lindsay Chair in Health Economics
University of Glasgow
Glasgow, Scotland, UK

Robert M. Golub, M.D.
Deputy Editor
Journal of the American Medical Association
Professor of Medicine
Northwestern University Feinberg School of Medicine
Chicago, IL

Douglas Owens, M.D., M.S.
Henry J. Kaiser, Jr., Professor
Director
VA Postdoctoral Medical Informatics Fellowship
VA Palo Alto Health Care System
Director
Program on Clinical Decision Making and Guideline Development
Center for Primary Care and Outcomes Research
Stanford University
Palo Alto, CA
Prior to publication of the final version of this chapter, the EPC sought input from independent Peer Reviewers. However, the conclusions and synthesis of the scientific literature presented in this report do not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers follows.

Andrew Briggs, D.Phil.
William R. Lindsay Chair in Health Economics
University of Glasgow
Glasgow, Scotland, UK

Scott Cantor, Ph.D.
Professor
Department of Health Services Research
Division of Cancer Prevention and Population Sciences
The University of Texas MD Anderson Cancer Center
Houston, TX

Ashish Deshmukh, Ph.D.
Postdoctoral Student
Department of Health Services Research
Division of Cancer Prevention and Population Sciences
The University of Texas MD Anderson Cancer Center
Houston, TX
David B. Matchar, M.D.
Professor and Director
Program in Health Services and Systems Research
Duke-NUS Medical School
Singapore
Professor (Internal Medicine and Pathology)
Duke University Medical Center
Durham, NC

Steven Teutsch, M.D., M.P.H.
Senior Fellow
Leonard D. Schaeffer Center for Health Policy and Economics
University of Southern California
Senior Fellow
Public Health Institute
Adjunct Professor
UCLA Fielding School of Public Health
Los Angeles, CA
Guidance for the Conduct and Reporting of Modeling and Simulation Studies in the Context of Health Technology Assessment

Structured Abstract

Objectives. The U.S. Agency for Healthcare Research and Quality (AHRQ) solicited the development of guidance for modeling and simulation studies conducted in the context of health technology assessment and systematic comparative effectiveness reviews.

Guidance development process. We updated and expanded existing systematic reviews of recommendations for the conduct and reporting of modeling and simulation with input from a multidisciplinary team of clinical, policy, and decision analysis experts. The results of the systematic review were discussed in person with a panel of 28 stakeholders, including patient representatives, providers and purchasers of care, payers, policymakers, and principal investigators. Stakeholders commented on existing recommendations from various sources and identified gaps, limitations, and areas for elaboration. We subsequently reviewed the Web sites of 126 health technology assessment organizations providing guidance on the conduct and reporting of modeling and simulation. We also solicited input from senior researchers with experience in modeling and simulation from AHRQ and its Evidence-based Practice Centers, and from external reviewers.

Results. We developed principles and good-practice recommendations for modeling and simulation studies conducted to enhance and contextualize the findings of systematic reviews. The guidance applies to structural mathematical models, including declarative, functional, and spatial models. The recommendations address the identification, estimation, verification, and validation of such models, as well as the use of sensitivity, stability, and uncertainty analyses in model development and assessment. We organized recommendations by whether they pertain to the model conceptualization and structure, data, model assessment and consistency, or the interpretation and reporting of results. We provide the rationale for each recommendation, supportive evidence, or best judgment where adequate evidence was lacking.

Conclusions. We hope that this work will contribute to increased use and better conduct and reporting of modeling and simulation studies in health technology assessment.
Introduction

Understanding the effects of interventions and using evidence to inform decisions are difficult tasks. Systematic reviews generally do not fully address uncertainty, tradeoffs among alternative outcomes, and differences among individuals in their preferences (values) for the alternative outcomes. Uncertainty may remain when clinical studies provide evidence only for surrogate outcomes; have small sample sizes, limited follow-up durations, or deficiencies in their design and conduct; or provide insufficient information on relevant patient subgroups. Tradeoffs among patient-relevant outcomes are common; for example, effective treatments may be associated with adverse effects (e.g., drug reactions), and informative diagnostic tests may result in overdiagnosis and overtreatment. Patients’ preferences for different outcomes (along with those of their families and other caregivers) need to be considered when assessing the consequences of alternative actions.

Models and simulations are valuable tools for inference and decisionmaking in the presence of uncertainty, tradeoffs, and varying preferences. Models can also be used to structure investigators’ thinking, facilitate the communication of assumptions and results, synthesize data from disparate sources, make predictions, and examine and understand the impact of (possibly counterfactual) interventions. These goals are highly relevant to the evidence syntheses prepared by the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers (EPCs).

This document provides general guidance in the form of good-practice principles and recommendations for EPCs preparing modeling and simulation studies. Currently, EPCs have variable expertise and experience in modeling, but the growing complexity of the health care questions addressed in EPC reports suggests that use of modeling may increase in the future. AHRQ has recognized the need for an overview of good practices for modeling and simulation to guide these efforts within the EPC Program.2,3

The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Society for Medical Decision Making (SMDM) have recently published detailed modeling recommendations (including discussion of many applied examples reviewed in the context of those recommendations). The availability of this information makes the development and use of models in conjunction with systematic reviews prepared by EPCs more feasible.4-17 In view of this, we sought to produce general guidance for modeling and simulation in the context of health technology assessment* on the basis of a systematic review of existing guidance documents, discussions with technical experts, and extensive deliberation within the EPC Program. The guidance aims to (1) encourage the use of good modeling and reporting practices in conjunction with systematic reviews without being too prescriptive about how to develop specific models and (2) describe how systematic reviews can increase the transparency of the modeling process and contribute to the development of useful models. We believe that this guidance applies generally, but to maintain focus, we emphasize models that could accompany systematic reviews produced by the EPC Program. We aim to establish a baseline understanding of modeling and simulation for EPC reports. We also provide an extensive list of references that can be a source of detailed information (and numerous examples) about specific modeling and simulation methods. We

*Readers who wish to review the recommendations without going over this introductory material can turn to the section titled “Good-Practice Recommendations for Modeling and Simulation in the Context of Health Technology Assessment.”
deemphasize issues specific to economic modeling, because economic assessments are not a priority of the EPC Program.

**Model Types and Scope of the Guidance**

A model is a representation of select aspects of reality in a simplified way. There exist physical (e.g., a scaled-down airplane wing tested in a wind tunnel), analog (e.g., a DNA molecule is like a staircase), or theoretical (formal) models (e.g., a mathematical description of the flow of air around an airplane wing). Models that can be prepared in conjunction with systematic reviews are exclusively theoretical. The starting point for most theoretical models is a conceptual model, a simplified natural language or pictorial representation of reality. A sound conceptual model is a prerequisite for the development of mathematical (quantitative) models. The analytic frameworks that are used to guide the conduct of systematic reviews prepared by the EPCs can form the basis of conceptual models representing the underlying decision or care process; for example, analytic frameworks often resemble the gist of decision trees or influence diagrams. Background information on analytic frameworks is provided elsewhere in the Methods Guide for Comparative Effectiveness Reviews.

Mathematical models are a large and diverse group of formal models that use variables, together with mathematical symbols that represent relationships between the variables. Perhaps the most common mathematical models encountered in practice are multivariable regression models (e.g., ordinary least-squares regression, logistic regression). 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 (also referred to as “models of data”). They describe how the response varies over covariates, without necessarily referring to assumptions about the underlying mechanisms. The literature addressing these models is vast (e.g., in Statistics and Computer Science) and is not covered in this guidance document. Instead, we address mathematical models that attempt to capture “true” (structural) relationships among their components (also referred to as “models of phenomena”) and combine information from multiple sources. These models include declarative (e.g., Markov models), functional (e.g., compartmental models), and spatial (e.g., geographic information systems) models. In applied work, elements of these model subtypes are commonly combined (multimodels). Thus, the models covered by this guidance include those considered by the National Research Council (“replicable, objective sequences of computations used for generating estimates of quantities of concern”) and the 2003 ISPOR principles of good practice (“analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources”). Simulation studies operate ("run") fully specified models to understand the phenomenon or process of interest, to predict its behavior, or to obtain insight into how its course can be modified by an intervention.

**Goals of Modeling and Simulation in EPC Reports**

We briefly consider the potential goals of modeling when performed in conjunction with systematic reviews.

- **To structure investigators’ thinking and to facilitate the communication of data, assumptions, and results**: Modeling can help investigators organize knowledge about a

---

†In some cases behavioral models (e.g., regression) are used to estimate the parameters of structural models.
• **To synthesize data from disparate sources:** Evidence on a specific research question may be available from multiple sources, and a single source may contribute information to the estimation of more than one model parameter (or functional combinations of parameters). Modeling provides mathematical tools for evidence synthesis and the assessment of consistency among data sources. For example, models can be used to combine information from clinical trials on the effect of treatment on intermediate outcomes with information from cohort studies on 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”), the future (forecasts), or other populations or outcomes (sometimes referred to as “extrapolations”). They can also pertain to the prioritization and planning of future research. These predictions may be useful in themselves, even without reference to the anticipated effects of interventions.

• **To support causal explanations and infer the impact of interventions:** Modeling can be used to assess the effects of possible (hypothetical) interventions and to provide mechanistic explanations for observed phenomena. When used this way, models are taken to encode structural causal mechanisms or to approximate such mechanisms sufficiently well. This allows models to examine counterfactual scenarios (“what if” analyses).

• **To inform decisionmaking:** The decisions that can be informed by modeling, even in the relatively narrow context of systematic reviews, are extremely diverse. They include decisions about patient-level care (accounting for treatment heterogeneity and variation in patient preferences), drug or device licensing, health care policy, and the need to conduct additional research.

Communicating assumptions, synthesizing evidence, and informing decisions are probably the most common goals of the models and simulations that would be developed in conjunction with systematic reviews. That said, the listed goals are not mutually exclusive, and typically the same model is used to achieve multiple goals.

**When Is Modeling Worth the Effort?**

Because issues related to the appropriateness of modeling in EPC reports are addressed by existing guidance, this document does not provide detailed recommendations to help investigators decide whether modeling and simulation should be undertaken. However, we briefly describe the typical conditions under which modeling is worth the extra effort.

The development of models, especially models that can be used to understand complex phenomena and to inform difficult decisions, is a demanding process. 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 weaknesses. Although this document and the references cited herein provide information on methods for modeling and simulation, defining the circumstances under which modeling is worth the investment of time and resources beyond those required for a systematic review is challenging. In general, modeling is most useful
when the research question is complex, data sources have limitations (e.g., sparse or conflicting
evidence, high risk of bias, or followup durations shorter than the time horizon of interest),
outcomes involve complex tradeoffs, and choices are preference laden. Consideration should
also be given to whether modeling is likely to produce results that the review’s intended
audience will deem credible and useful. 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 health care policy should also be
considered when deciding whether modeling is worth the effort.

Modeling and Simulation Process
The specifics of model development and assessment vary across specific applications because
modeling and simulation studies are used to address diverse research questions. Nonetheless, the
key activities for the development of quantitative models, within the scope of this guidance, can
be identified:31,59-69

1. **Specifying the research question and setting the modeling goals:** Specifying an
answerable research question that addresses the needs of the relevant stakeholder(s)
2. **Conceptualizing the model and specifying its mathematical structure:** Determining
which components of a disease or process need to be represented in the model to address
the research question and describing their relationships
3. **Assembling information:** Identifying data sources, eliciting expert opinions, and
processing all information that will be used as an input to the model
4. **Implementing and running the model:** Running the model using mathematical or
numerical analysis methods
5. **Assessing the model:** Examining whether the model attains its stated goals; detecting
model shortcomings by examining the model, as well as by comparing its output with
prior beliefs, data, and other similar models
6. **Interpreting and reporting results:** Presenting the model findings in a way that
addresses the research question

Model development is an iterative and dynamic process.12,62 Multiple iterations are typically
needed between the phases outlined here 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
case conceptualized, additional data needs arise that require further data collection. Similarly, data
deficiencies that are detected during model assessment often require restructuring of the model,
supplemental data collection, or other modifications of the modeling strategy.

Guidance Development Process
This guidance document is the culmination of a multistep process of summarizing existing
recommendations and soliciting stakeholder input. Earlier steps of the process are described in a
companion report (“Modeling and Simulation in the Context of Health Technology Assessment:
Review of Existing Guidance, Future Research Needs, and Validity Assessment”; publication
details to be provided by OCKT). Briefly, with input from a multidisciplinary team of clinical,
policymaking, modeling, and decision analysis experts, we updated and expanded two systematic
reviews of recommendations for the conduct and reporting of modeling and simulation studies, as described in detail in the companion report. To assess the relevance of published methodological recommendations for modeling and simulation, we discussed the results of our systematic review in person with a panel of 28 stakeholders, including patient representatives, care providers, purchasers of care, payers, policymakers (including research funders and professional societies), and principal investigators. To examine worldwide health technology assessment procedures and practices, we reviewed the Web sites of 126 international health technology assessment agencies and institutes for their guidance or standards for whether to conduct modeling and simulation studies, and if so, how to perform and report such studies. We solicited feedback on a set of draft recommendations from senior researchers at EPCs and AHRQ with experience in modeling and simulation methods. The draft recommendations were presented for comment and affirmation or dissent at the 2014 Annual Meeting of SMDM and the 2015 Annual Meeting of ISPOR. In these oral presentations, participants in the audience were invited to express disagreement (by raising their hands) after the presenter read out each recommendation statement. These reactions and additional comments from the audience were incorporated in revised versions of the report; however, this should not be taken to imply endorsement of these recommendations by either SMDM or ISPOR. External peer reviewers and AHRQ personnel further vetted a draft version of this report, and we incorporated their input into the final version; however, this should not be taken to imply endorsement of the recommendations by the reviewers. Lastly, EPC Directors were invited to consider the final report for adoption as guidance for the EPC Program, and the majority voted in favor of including this report in the EPC Methods Guide for Comparative Effectiveness Reviews.

Based on the gathered systematic review evidence on modeling recommendations and the processes described here, we prepared the final version of good-practice principles and recommendations for developing models in conjunction with systematic reviews. We categorized the modeling recommendations by whether they pertain to the conceptualization and structure of the model, data, consistency, or interpretation and reporting of results, as was done in earlier related work.

Our approach in crafting the guidance was consistent with the framework for making methodological recommendations that we have described in an earlier publication. We made a concerted effort to review the relevant literature and consulted with recognized experts in the field. Although the final version of the guidance reflects the authors’ best judgment and is not the product of consensus among the stakeholders involved in the process, we have explained the rationale for each recommendation and, when available, have provided evidence that the recommendation should be preferred.
## Terminology and Definitions

Table 1 defines terms used in this document.

<table>
<thead>
<tr>
<th>Term</th>
<th>Explanation and Elaboration</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>A representation of select aspects of a phenomenon or system in simplified form.</td>
<td>Models are constructed following a process of abstraction and idealization. We focus on models that represent reality by means of mathematical relationships (using mathematical operations, symbols, and variables).</td>
</tr>
<tr>
<td>Decision model</td>
<td>A model of the choice between two or more alternative options (i.e., alternative actions or rules for making decisions over time).</td>
<td>These models are used to explore the consequences of different choices (e.g., the decision to use a particular treatment when two or more alternatives are available).</td>
</tr>
<tr>
<td>Simulation</td>
<td>The operation (“running”) of a model.</td>
<td>“Simulation” is sometimes defined more restrictively as using numerical methods (especially methods that use computer-generated pseudorandom numbers) when analytic solutions are cumbersome or intractable. We use a broader definition in this document.</td>
</tr>
<tr>
<td>Computer simulation</td>
<td>A simulation carried out using a computer.</td>
<td>Almost all simulations relevant to health technology assessment are computer simulations.</td>
</tr>
<tr>
<td>Model component</td>
<td>An element of a model (e.g., variables, health states, agents, and processes).</td>
<td>The term is purposely generic to encompass all model types.</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Lack of certainty; a state of limited knowledge.</td>
<td>Many typologies of uncertainty have been proposed in the context of modeling and simulation. Such typologies are useful to the extent that they contribute to the success of applied modeling work; a complete theoretical treatment of the concept of uncertainty is beyond the scope of this guidance but is central to modeling practice (and addressed by a vast literature to which we provide a few selective links). In applied work we find it useful to consider scenario, structural, and parametric uncertainty (defined in this table). In some cases it is also useful to consider residual (predictive) uncertainty (i.e., in nondeterministic models), to represent inherent stochasticity not captured by scenario, parametric, or structural sources.</td>
</tr>
<tr>
<td>Scenario uncertainty</td>
<td>Uncertainty in the scenarios for which the model is run.</td>
<td>Exogenous forces not under the control of the investigators can be a source of scenario uncertainty. For example, in a model of a hospital that uses data on the number of patients attending the emergency room, running the model under an extreme scenario (e.g., to simulate a doubling of the number of patients attending) is subject to scenario uncertainty.</td>
</tr>
<tr>
<td>Structural uncertainty</td>
<td>Uncertainty due to incomplete understanding of the modeled phenomenon. Typically, this pertains to functional forms of relationships between model variables. At a more fundamental level, structural uncertainty always exists because the “true” relationship between variables in the real world cannot be uncovered from data.</td>
<td>In some cases, the distinction between structural and parametric uncertainty is a matter of definition. Some structural uncertainty would become parametric uncertainty if relevant data were available to the modeler. Structural uncertainty is conditional on the modeled scenario.</td>
</tr>
<tr>
<td>Parametric uncertainty</td>
<td>Uncertainty about the values of model parameters.</td>
<td>Parametric uncertainty is sometimes referred to as “aleatory” or “stochastic” uncertainty. It is conditional on the model structure and the scenario being modeled.</td>
</tr>
<tr>
<td>Term</td>
<td>Explanation and Elaboration</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Uncertainty analysis</td>
<td>Analysis that addresses lack of knowledge regarding the model structure, parameter values, and scenarios, or any inherently nondeterministic aspect of the model.</td>
<td>The most common goal of uncertainty analyses is the propagation of uncertainty from model inputs to model outputs (e.g., accounting for lack of certainty when estimating treatment effects or event rates).</td>
</tr>
<tr>
<td>Propagation of uncertainty</td>
<td>The process of assessing uncertainty in model outputs by incorporating all sources of uncertainty in the model inputs. This derives from the uncertainty in model inputs. Uncertainty can be propagated analytically (exactly or up to an approximation) or numerically (e.g., with forward Monte Carlo simulations or 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 the term PSA in this work to avoid confusion. We draw a distinction between the propagation of uncertainty and sensitivity analysis. Propagation of uncertainty is important for obtaining valid results (especially in nonlinear models) and for correctly assessing the value of obtaining additional information. However, simply propagating uncertainty from inputs to outputs does not fulfill the goals of sensitivity analysis (assess the influence of inputs on outputs).</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>The process of varying model variables over a set of values that are of interest (e.g., because they are deemed plausible) and examining impact on results.</td>
<td>Sensitivity analysis can be used to evaluate the impact of different inputs on model outputs or to examine the implications of different values of unidentifiable model parameters on model results. Sensitivity analysis often has a “continuous” character (e.g., the magnitude of a treatment effect is varied over multiple finely spaced values within the plausible range).</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, such as using alternative specifications (e.g., different functional forms for relationships between variables) and systematically excluding input data (e.g., leave one study out in a meta-analysis).</td>
<td>Stability analysis involves discrete analytical decisions (e.g., the summary treatment effect can be obtained via a random effects or common effect meta-analysis model). Stability analyses assess the impact of alternative analytic approaches on model results (e.g., using a weekly or monthly cycle length in a discrete time Markov chain, or assessing robustness of model results to estimate treatment effects after excluding studies one at a time). Such analyses are often described as explorations of “methodological uncertainty;” however, the term “stability analyses” is more appropriate because their purpose is to examine robustness to alternative methodological choices, not to reflect uncertainty about the modeled phenomenon.</td>
</tr>
<tr>
<td>Model verification</td>
<td>The assessment of the correctness of the mathematical structure (e.g., absence of mistakes in the logic) and of the implementation of the computational model (e.g., absence of software bugs, suitability of numerical algorithms).</td>
<td>Model verification includes the identification and correction of mistakes in the model logic and software bugs, and an assessment of the suitability of numerical algorithms used in the model.</td>
</tr>
<tr>
<td>Model validation</td>
<td>Validation is the comparison of the model and its output with expert beliefs (conceptual validity and face validity), data (operational and predictive validity), and other models (cross-model validity).</td>
<td>Validation includes various checks of the face, operational, and external validity of the model. It is closely related to the concepts of representational fidelity (is the model a good representation of the actual system or process?) and behavioral fidelity (is the model output similar to the behavior of the actual system or process?). Because complete model validity cannot be established in the affirmative, a model can only be evaluated with respect to a specific purpose. The examination of model consistency (model assessment) includes attempts to verify and validate attributes of the model and establish its credibility.</td>
</tr>
<tr>
<td>Preferences (values)</td>
<td>We use the term in a broad sense, to denote how desirable a given outcome is for an individual or a group.</td>
<td>Preferences are used in the valuation of outcomes. Sometimes “utility” is defined as a “cardinal measure of the strength of one’s preference.”</td>
</tr>
</tbody>
</table>
Principles for Good Practice in Modeling and Simulation

We begin by outlining general principles for the conduct and reporting of modeling and simulation studies (Table 2). 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.

At the start of the modeling work, investigators should consider (1) the goals of the modeling application; (2) the nature of the modeled phenomena; (3) the available expertise; and (4) objective constraints in terms of available time, data, or other resources. Further, when developing, implementing, and running models, there are many methodological decisions to be made. These decisions should be recorded, justified, and subjected to stability analysis. This can be done most conveniently by specifying the modeling methods in a protocol (while the study is being planned) and by generating detailed documentation (while the study is conducted and after it is completed).

<table>
<thead>
<tr>
<th>Table 2. Principles for good practice in modeling and simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The research question, modeling goals, and the scope of the model should be clearly defined.</td>
</tr>
<tr>
<td>The model structure and assumptions should be explained and justified.</td>
</tr>
<tr>
<td>Model components and the relationships between them should be defined. The chosen relationships between model components should be justified.</td>
</tr>
<tr>
<td>The model should be informed by data. Data selection, analysis, and interpretation should be aligned with the research question and the model's scope; data sources should be described.</td>
</tr>
<tr>
<td>The model should reflect uncertainty in inputs.</td>
</tr>
<tr>
<td>Sensitivity analyses (to assess the influence of model inputs) and stability analyses (to evaluate the impact of modeling decisions) should be undertaken and reported.</td>
</tr>
<tr>
<td>Models should be assessed for their ability to address the research question within the stated scope.</td>
</tr>
<tr>
<td>Modeling methods should be transparent. Adequate details about the structure, data, and assessment methods should be reported so that the modeling process is replicable.</td>
</tr>
</tbody>
</table>

We provide good-practice recommendations for modeling and simulation in conjunction with systematic reviews, organized by conceptualization and structure, data, assessment and consistency, and interpretation and reporting. Briefly, structure and data comprise 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 of these areas of modeling.

This guidance is provided to facilitate the use of modeling and simulation 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 assessing models and simulations. In general, all recommendations should be viewed pragmatically when embarking on a specific project. (This is sometimes referred to as the “rule of reason.”) Systematic reviewers and modelers should exercise judgment when deciding whether specific recommendations are likely to have an appreciable impact on the review conclusions and should balance feasibility with the desire to conduct extensive analyses.
It is not possible to provide detailed recommendations about which structures to use in which cases or instructions about the implementation and manipulation of various model types. Interested readers should consult some of the numerous books, technical reports, and papers available on this topic (several of which are cited in the section on good-practice recommendations), including systematic overviews of existing guidance,2,3,70,74,103 the recent ISPOR-SMDM recommendations,4-17 other sources of guidance (including methodological appraisals),38,47,73,98,100,102,104-162 expository works with a focus on biomedical modeling,61,62,163-183 and the vast literature on modeling and simulation in other subject areas.184-210

Table 3. Operational definitions for the conceptualization and structure, data, model assessment and consistency, and interpretation and 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>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, as well as how they relate to each other (i.e., the mathematical scaffold of the model).</td>
<td>In a discrete-time Markov model the disease states, variables informing transition probabilities, mathematical relationships among the variables, and time horizon of the model 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>Model assessment and consistency</td>
<td>A model can only be evaluated with respect to the specific goals of modeling (the goals are determined by the research question). Model assessment examines the extent to which the model achieves the stated goals of representing the phenomenon of interest and the effects of alternative actions on pertinent outcomes. Model assessment activities occur throughout the process of model development. Model consistency includes attempts to verify and validate the model and establish its credibility.</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>Interpretation and reporting of results</td>
<td>Summarizing model output to achieve the goals of modeling (e.g., to further understanding the topic or to inform decisionmaking).</td>
<td>Risk diagrams (to represent model-based risk analyses) and tornado diagrams (to summarize sensitivity analyses using ordered bar charts).</td>
</tr>
</tbody>
</table>

Sensitivity, Stability, and Uncertainty Analyses

Many of the recommendations in this guidance emphasize the need to perform sensitivity and stability analyses. By sensitivity analysis we mean the process of varying model variables over a set of values that are of interest and examining impact on results. Such analyses can be used to evaluate the impact of different inputs on model outputs or to examine the implications of different values of unidentifiable model parameters on model results. Sensitivity analysis can be local (e.g., examining changes in output in response to infinitesimal perturbations of the inputs) or global (e.g., examining changes over a broader range of input values). Many methods for sensitivity analysis, both stochastic and deterministic, have been proposed; the choice among available methods should be dictated by the goals for the modeling effort.69,203,211-224
Stability analyses are assessments of the impact of alternative analytic approaches on model results (e.g., whether to use a monthly or weekly cycle length in a Markov model). Often such analyses are described as explorations of “methodological uncertainty;” however, the term “stability analyses” is more appropriate because their purpose is not to reflect uncertainty about the modeled phenomenon but simply to examine robustness to alternative methodological choices.101

Handling of uncertainty is related to, but distinct from, sensitivity analysis. We organize uncertainty into three types: scenario, structural, and parametric uncertainty. A fourth type, “residual” (predictive) uncertainty, may also be important to consider. (For definitions, see Table 1.) Structural uncertainty is perhaps the most challenging to address because empirical observations are always compatible with a large number of alternative model structures. Methods for handling structural uncertainty include stability analyses (i.e., building models with alternative structures),183 model expansion by “parameterizing” alternative structures, and formal model averaging.92,225-228 Proper handling of parametric uncertainty is necessary for valid inference using on models and simulations.90-92,225-227,229,230 Although uncertainty in model outputs should have no impact on decisionmaking under a Bayesian decision theoretic view,231,232 we believe that decisionmakers are often interested in the degree of certainty around model outputs and (heuristically) consider that information when making decisions. In addition, proper handling of uncertainty is critical for using models to determine the need for future research, prioritize specific research activities, and plan future studies.233

We draw a distinction between uncertainty analysis (i.e., the propagation of stochastic uncertainty from model inputs to outputs) and sensitivity analysis. Propagation of uncertainty (analytically or via various Monte Carlo methods) is important for obtaining valid results (especially in nonlinear models) and for correctly assessing the value of obtaining additional information. However, simply propagating uncertainty from inputs to outputs cannot fulfill the goals of sensitivity analysis (i.e., to assess the influence of model inputs). Confusingly, the term “probabilistic sensitivity analysis” is often used in the literature to describe uncertainty propagation that is not coupled with attempts to identify influential inputs. We propose that the term “uncertainty analysis” or “uncertainty propagation” should be used to describe such analyses, and that the term “sensitivity analysis” should be reserved for analyses that aim to assess the influence of model inputs.234-240 In many cases, sensitivity analysis will be conducted in models that also propagate uncertainty, but the two activities have different goals.

Model Assessment and Consistency

Model assessment is meaningful only with respect to the goals of modeling. Assessment activities occur throughout the life cycle of model development.65,241-243 The examination of model consistency includes attempts to verify and validate the model and establish its credibility; this is a critical issue in modeling and has been extensively considered in the literature (both in health care contexts and beyond).65,67,68,128,241,242,244-270 Verification (internal validity) is the assessment of the correctness of the mathematical structure (e.g., absence of mistakes in the logic) and of the implementation of the computational model (e.g., absence of software bugs, suitability of numerical algorithms). Validation is the comparison of the model and its output with expert beliefs, data, and other models. In practice, validation includes various “checks” of face, external, predictive, and cross-model validity. It is closely related to the concepts of representational fidelity (i.e., whether the model is a good representation of the modeled system or process) and behavioral fidelity (i.e., whether the model output is similar to the behavior of
the modeled system or process, sometimes referred to as dynamic fidelity), which arise in all types of modeling.\textsuperscript{271} Importantly, a model can be evaluated only with respect to a specific purpose; complete model validity cannot be established in the affirmative; in fact, it has been suggested that a model that successfully passes an evaluation should be considered corroborated, not validated.\textsuperscript{272,273}

**Good-Practice Recommendations for Modeling and Simulation in the Context of Health Technology Assessment**

**Conceptualization and Structure:**

Explicitly state the research question and the modeling goals. Describe and justify the decision to use modeling to address the research question. Use a conceptual model to guide the development of the mathematical model.

Choose a perspective depending on the research question and the relevant stakeholders. There is no a priori preferred modeling perspective.

Specify the model scope to be consistent with the research question and modeling perspective. Describe and justify the model scope.

Specify and implement the structure of the mathematical model to correspond to the research question, the model’s scope, and the modeling perspective. Provide the rationale for the chosen mathematical structure; explain and justify structural assumptions and computational approximations.

Allow for comparisons among all interventions that are relevant to the research question and the model’s scope.

Use a time horizon long enough to allow all relevant outcomes to be fully evaluated.

When deciding how to handle time, spatial location, interactions among agents, and health states, consider the nature of the modeled phenomenon and the convenience of (and approximation errors associated with) alternative choices.

Determine the targeted level of complexity (or parsimony) based on the research question and the model scope. It is often preferable to build a simpler model first and progressively increase the degree of complexity.

**Data:**

Describe the methods for identifying and analyzing data. Make data choices based on the research question and the model’s scope and structure. Report all data sources clearly and provide explicit references. Obtain values for model inputs following epidemiological and statistical principles. Use a “best evidence approach” when selecting data sources for model parameters. Obtain estimates for influential parameters using systematic review methods.
Assess the risk of bias of the available evidence and account for sources of bias when estimating values for model parameters.

Use formal elicitation methods to quantify expert opinion and its associated uncertainty. Use appropriate methods to quantify preferences for different outcomes.

Describe and justify the assumptions required for extrapolating beyond observed data and transporting information from various data sources to a common (target) setting. Subject these assumptions to stability and sensitivity analyses.

In statistical analyses, account for heterogeneity.

Use modeling methods that propagate uncertainty from inputs to outputs.

**Model Assessment and Consistency:**

Evaluate the model with respect to the specified modeling goals.

Anticipate, detect, and correct errors in the model’s logic and implementation.

 Invite topic experts to review the model’s structure and outputs and to judge whether these seem consistent with their expectations. Verify, describe, and explain counterintuitive model results.

Assess the consistency between model outputs and the data on which the model was based.

Do not withhold data from model development for the sole purpose of assessing model validity.

Decide whether using future observations to assess a model is appropriate based on the research question and the modeling goals.

Update the model as new data become available, new interventions are added, and the understanding of the investigated phenomenon improves.

If models addressing the same research question are available, compare their results to the new model and explain any discrepancies.

**Interpretation and Reporting of Results:**

Be transparent about the model structure, computational implementation, and data. Report results clearly and in a way that addresses users’ needs.

Interpret and report results in a way that communicates uncertainty in model outputs.

Fully disclose any potential conflicts of interest.
Explanation and Elaboration of Recommendations for Modeling and Simulation

Conceptualization and Structure

Explicitly state the research question and the modeling goals. Describe and justify the decision to use modeling to address the research question. Use a conceptual model to guide the development of the mathematical model. Modeling is useful for addressing many research questions, especially questions that are not directly answerable using existing empirical data. A well-specified and explicitly stated research question is critical for modeling and simulation.274,275 Models prepared in conjunction with systematic reviews should be based on a clear conceptual model.5,12,276-279 Defining the question and objective of the analysis may require using literature-based information, expert knowledge, and input from stakeholders (e.g., the Key Informants and Technical Experts who provide input during the preparation of EPC reviews).12 If a model is to be prepared in conjunction with a systematic review, issues related to the model should be considered during the planning of the review (e.g., possible model structures, anticipated data). Conversely, decisions related to the construction and use of models should be informed by the design of the systematic review (e.g., regarding the various populations, interventions, and outcomes that could be considered in the model).

Choose a perspective depending on the research question and the relevant stakeholders. There is no a priori preferred modeling perspective. The modeling perspective determines the methods for choosing and handling consequences, preferences, and, if examined, costs in the model; thus it should depend on the research question and the relevant stakeholders (e.g., decisionmakers).280 For example, when modeling aims to address the clinical options faced by an identifiable patient, the appropriate perspective is that of the individual patient. In contrast, when the goal is to inform health policy decisionmaking of a public payer or a Federal agency, one should prefer a payer or societal perspective.281 The societal perspective (which considers impact on sectors beyond health care and includes time costs, opportunity costs, and community preferences) may allow for a comprehensive accounting of benefits, harms, and costs and can serve as a “base case,” facilitating comparability of the results across health policy analyses.282 For this reason, it has been recommended as an appropriate “default” perspective.162 However, obtaining appropriate data for modeling from a societal perspective can be challenging (e.g., accommodating equity concerns).282-284

Specify the model scope to be consistent with the research question and modeling perspective. Describe and justify the model scope. The scope of a mathematical model includes the condition or disease of interest, populations, risk factors, and diagnostic or therapeutic interventions. For decision models, the scope also includes alternative strategies, decision-relevant outcome quantities (e.g., life-years gained, quality-adjusted life-years, disability-adjusted life-years), the decision (optimality) criteria, the time horizon, and the decisionmaking perspective. Determining the scope of the model is akin to defining a systematic review’s study selection criteria (e.g., population, intervention, comparator,
outcomes, timing, and setting). By necessity, a model represents only some aspects of the phenomenon or process under study. The research question defines how complex the model should be and what aspects of reality are represented or omitted (for parsimony). For example, many research questions in health care pertain to length of life; thus, mortality outcomes should be within the scope of models answering these questions.

Specify and implement the structure of the mathematical model to correspond to the research question, the model’s scope, and the modeling perspective. Provide the rationale for the chosen mathematical structure; explain and justify structural assumptions and computational approximations.

The preferred model structure depends on the research question and the model’s 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). Health states, transitions between health states, and functional relationships between parameters should reflect the understanding of the course of the disease. Detailed guidance on choosing among alternative mathematical structures and on implementing them in computational models is beyond the scope of this document. Readers are referred to the extensive technical literature in health care*4-17,38,46,61,165,167,181,285-307 and other fields.31,187,192,308-310 Of note, relatively simple models (e.g., decision trees, time-homogeneous Markov state transition models) may be appropriate for use in the setting of many EPC evidence reports, particularly when the goal of modeling is to contextualize the evidence and extend review findings.

Allow for comparisons among all interventions that are relevant to the research question and the model’s scope.

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 a 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 justified.

Use a time horizon long enough to allow all relevant outcomes to be fully evaluated.

When comparing alternative interventions, the time horizon should be long enough to allow the manifestation of differences in relevant 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 lifetime horizon is needed, particularly when modeling the effects of long-term treatment of chronic disease. The time horizon choice has implications for the data used to populate models; for example, lifetime horizons almost always require the extrapolation of treatment effects well beyond the followup duration of available clinical trials.
When deciding how to handle time, spatial location, interactions among agents, and health states, consider the nature of the modeled phenomenon and the convenience of (and approximation errors associated with) alternative choices. For example, when deciding how to deal with time, we have three options: (1) do not model it explicitly (as in simple 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 models). 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 spatial location, interagent interactions (e.g., interactions between modeled individuals), and health states in various degrees of granularity (e.g., disease severity).

Determine the targeted level of complexity (or parsimony) based on the research question and the model scope. It is often preferable to build a simpler model first and progressively increase the degree of complexity. Models should be as complex as needed 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 generally more accessible to nontechnical stakeholders, and results from such models 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. In general, it is preferable to first build a simpler model and progressively increase the degree of complexity in order to facilitate error checking and ultimately obtain a reliable model that satisfies the goals of the modeling effort.

Data

Describe the methods for identifying and analyzing data. Make data choices based on the research question and the model’s scope and structure. Report all data sources clearly and provide explicit references. Obtain values for model inputs following epidemiological and statistical principles. Use a “best evidence approach” when selecting data sources for model parameters. Obtain estimates for influential parameters using systematic review methods.

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

All major assumptions and methodological choices for determining model inputs should be reported and justified. Modelers must select, appraise, and synthesize appropriate study types for
A recent EPC Methods Research Report provides general guidance on “best evidence” strategies in systematic reviews. Data from randomized trials cannot be used to inform all model parameters because (1) some parameters are best estimated from other study designs (e.g., the prevalence of a risk factor is best estimated from a sampling survey of a representative population; the 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 at all. In all these cases, evidence from other study designs will have to be included in the model.

For modeling and simulation studies prepared jointly with a systematic review of studies of interventions, estimates of treatment effects and other inputs (together with corresponding measures of sampling variability) should be used to inform the relevant model parameters. In particular, model parameters likely to have a large influence on model results should be informed by a systematic and replicable process that aims to minimize bias. 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 include and the choice of synthesis methods.

Data on other model inputs (e.g., prevalence, incidence, resource use or costs, and utilities) may be obtained through processes other than systematic review. Appropriate sources of such data can include de novo analyses of registries and other large observational studies, completed studies found through a nonsystematic approach, stakeholder panel opinions, and domain expert judgments. When retrieving and processing data, modelers often make decisions that may appreciably impact results (e.g., use of operational selection criteria to determine the relevance of published studies or use of approximate calculations when extracting data from published studies). All such decisions should be recorded, justified, and reported in the model’s documentation. Supplementary material describing detailed methods and data sources can be made available electronically.

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 methods (meta-analysis, network meta-analysis, or generalized evidence synthesis). When data from multiple sources are combined to estimate model parameters, the examination of consistency among sources is an important task. For example, inconsistency in network meta-analyses can indicate the presence of effect measure modification or bias in the evidence base. This guidance does not provide detailed information on the conduct of quantitative synthesis for different types of data structures; both EPC guidance and many other sources can be consulted for detailed descriptions of meta-analysis and evidence synthesis methods.

Related to the idea of estimation of model parameters is the concept of model calibration, the tweaking of (typically unidentifiable or weakly identifiable) model parameters to improve the “closeness” of the model outputs with empirical data. We do not distinguish sharply between processes for calibration and estimation because the analytic goals and the methods to achieve them are similar. This becomes particularly clear when considering Bayesian simulation models. A more detailed discussion of model calibration is provided in a companion report (“Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment”).
Assess the risk of bias of the available evidence and account for sources of bias when estimating values for model parameters.

Models typically are specified with respect to “true” parameters, but empirical studies provide parameter estimates that are subject to bias. Consequently, model inputs should be adjusted (“corrected”) for biases. In general, modelers should avoid using unadjusted, incompletely adjusted, or inappropriately adjusted results simply because no other information is available.\textsuperscript{371,372} When the available evidence base is large, this may be possible by obtaining information from studies free of such problems. However, in many cases, studies free of bias are unavailable or represent a very small fraction of the available evidence. In such cases, modelers should adjust study results to account for bias and associated uncertainties (i.e., multiple bias modeling) and should undertake sensitivity analyses.\textsuperscript{373,374}

Because the factors that determine the direction and magnitude of bias depend on the modeling context and the design, conduct, and analysis of the studies under consideration, bias assessment has to be tailored on a case-by-case basis.\textsuperscript{375,376}

The direction and magnitude of bias introduced by different factors, uncertainty about bias, and the relationship between biasing factors should be incorporated into the analyses. In most cases, “bias parameters” cannot be identified from study data; thus, modelers have to use methods that incorporate external information (empirical and 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).\textsuperscript{371,373,377-388}

Use formal elicitation methods to quantify expert opinion and its associated uncertainty. Use appropriate methods to quantify preferences for different outcomes.

When no empirical evidence is available for parameters of interest, modelers have to rely on expert opinion (e.g., to estimate probabilities of event occurrence). Preferences for different outcomes can also be elicited using specialized methods. The literature on methods for eliciting expert opinions and for determining preferences is extensive and is not covered in this report; the measurement of preferences is a contentious topic.\textsuperscript{389-398}

Current technical expert and stakeholder engagement processes in systematic reviews can incorporate formal methods for eliciting expert opinions and quantifying preferences for different outcomes (e.g., by expanding the roles of Key Informants and Technical Experts involved in the development, refinement, and conduct of systematic reviews). 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 quantities of interest.\textsuperscript{399} When elicitation of preferences cannot be performed \textit{de novo}, the literature can be used as a source of information.

Describe and justify the assumptions required for extrapolating beyond observed data and transporting information from various data sources to a common (target) setting. Subject these assumptions to stability and sensitivity analyses.

A particular challenge arises when there is a need to extrapolate beyond the observed data (e.g., to longer followup periods or to other populations). Such extrapolations are based on untestable assumptions that 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).

Models often use data obtained from diverse sources. 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 mechanisms underlying the modeled phenomenon. Consideration should be given to formal (causal) methods for assessing the transportability of results across domains.

**In statistical analyses, account for heterogeneity.**

As a general principle, models and simulations should account for heterogeneity, defined as nonrandom (systematic) variation. Attempts should be made to explain heterogeneity by incorporating information on determinants of variability via appropriate statistical methods (e.g., subgroup or regression analyses). Because scientific understanding of any topic is likely to be incomplete (e.g., important modifiers of effect may be unknown) and because lack of data 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 is common in meta-analyses of treatment effects that use published (aggregate) level data. In such cases, efforts to explain heterogeneity rely primarily on metaregression methods, and residual heterogeneity is accounted for by using random-effects models. 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 than fixed-effect models. Person-level data can allow models and simulations to meaningfully incorporate heterogeneity, however, such data are rarely available in systematic reviews prepared by EPCs or in meta-analyses published in peer-reviewed journals.

**Use modeling methods that propagate uncertainty from inputs to outputs.**

Appropriate data analysis methods should be used to obtain valid parameter estimates and to propagate uncertainty from inputs to outputs. Sometimes this can be done analytically, either exactly or by approximating up to an order of error (e.g., with the delta method). In most cases, it is computationally convenient to propagate uncertainty with numerical methods, typically with a forward Monte Carlo approach; in the medical modeling literature, this is often, and somewhat inappropriately, termed “probabilistic sensitivity analysis.”

Detailed descriptions of methods for conducting probabilistic analyses are available elsewhere in the literature on modeling in health care and other fields. Of note, probabilistic methods for incorporating and propagating uncertainty in models do not eliminate the need for stability and sensitivity analyses. For example, the choice of the distribution is rarely unique. Thus, it may be important to assess the impact of using alternative probability distributions (stability analysis) or to assess the impact of varying the parameters determining the distribution (e.g., location, scale, as applicable) over a range (sensitivity analysis).
In rare cases, it may be unnecessary to perform analyses that propagate uncertainty, based on the goals of the model. 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 performing probabilistic analyses.

Model Assessment and Consistency

**Evaluate the model with respect to the specified modeling goals.**
A model can only be evaluated with respect to the specific goals of modeling (as determined by the research question). The preferred model assessment methods and criteria depend on the intended use of the model.

Model Verification

**Anticipate, detect, and correct errors in the model’s logic and implementation.**
Errors are unavoidable when developing any nontrivial model. Mistakes in research question formulation, 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 logic and implementation can be challenging to detect and can have important consequences. The risk of mistakes in question formulation and model structure can be reduced by adhering to some of the principles outlined previously in this document (e.g., consulting with topic experts, using a conceptual model to guide the implementation of the mathematical model), together with transparent reporting of methods and results and the use of teams with sufficient expertise. Several checking techniques have been advocated for health care–related models (e.g., sensitivity analysis, extreme value analysis, dimensional analysis). In addition, software production techniques, such as unit testing, code review (review of one programmer’s work by another team member), and paired programming (i.e., one programmer’s coding being monitored by another in real time), can be considered. Duplicate implementation of the same model by an independent team or implementation of the same model in a different software package can also be used to identify errors in coding. Because these strategies can substantially increase the time and resources required for model development, their use should be balanced against the modeling goals, model complexity, and anticipated frequency and impact of errors.
Model Validation

Face Validation

Invite topic experts to review the model’s structure and outputs and to judge whether these seem consistent with their expectations. Verify, describe, and explain counterintuitive model results.

An examination of the model and its results by a group of topic experts can alert modelers to the presence of deficiencies in the model’s structure or data. For example, a formal version of this examination involves providing model-generated output (e.g., incidence rates, mortality rates, distributions of patients across stages at diagnosis) and empirical data on the same quantities to users of the model. The experts are then asked to identify which results are “real” and which are model generated. This procedure can be used to assess the credibility of a given model (and relates to the Turing test in artificial intelligence research).

Counterintuitive model results (“paradoxical findings”) may indicate “bugs” or errors, so such results should be examined carefully. If an error has been ruled out, the results should be described and explained with reference to model structure, available data, and current understanding of the modeled phenomena.

External and Predictive Validation

Assess the consistency between model outputs and the data on which the model was based.

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 model-predicted values (graphical or statistical) can be used to identify potential areas of improvement in model structure, assumptions, and data.

Do not withhold data from model development for the sole purpose of assessing model validity.

Generally, data should not be withheld during model development for the purpose of using them for model validation. Using all of the available data during model development improves the efficiency of parameter estimation, facilitates the appropriate handling of correlated inputs, and allows an assessment of consistency across all available sources of evidence. For example, problems may exist when model predictions do not agree well with observations. Model validation, in terms of agreement of model predictions with the corresponding data, can be formalized with metrics of model fit. Resampling methods (cross-fold sampling, bootstrap) can be used to assess model fit and to detect outlying or influential observations that may guide further explorations. Additional model validation methods are available in a Bayesian framework (e.g., posterior predictive checks). Validation assessments that use ideas of model fit require careful application and interpretation in over-parameterized models that have parameters that cannot be fully identified from the data (e.g., parameters related to the unobservable rate of tumor cell growth in cancer microsimulation models). Even when such parameters (e.g., tumor growth) are not identifiable by available data, withholding data on identifiable parameters (or on
functional combinations of identifiable and nonidentifiable parameters) is, in general, less efficient than joint modeling.

**Decide whether using future observations to assess a model is appropriate based on the research question and the modeling goals.**

Predictive validation is an important component of the assessment of models intended as forecasting tools. However, a comparison of model output with empirical results unavailable at the time of model development is not an appropriate method of assessment for models intended to guide decisionmaking using the best available data at a specific point in time. Models developed in conjunction with systematic reviews are likely to belong in this category.

**Update the model as new data become available, new interventions are added, and the understanding of the investigated phenomenon improves.**

Models should be updated when new data about important parameters become available (e.g., updated systematic reviews with new or different effect estimates). In addition, as the understanding of disease mechanisms (causal agents, natural history) and interventions and their consequences evolve, model updating should be considered. The model structure and its software implementation must be flexible enough to accommodate this updating process.

**Cross-Model Validation**

If models addressing the same research question are available, compare their results to the new model and explain any discrepancies.

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). If such independent models are available (known to the modelers or identified through literature review), then their outputs should be compared as part of cross-model validation, and any discrepancies should be explained with reference to the structure and data inputs of each model.

**Interpretation and Reporting of Results**

Be transparent about the model structure, computational implementation, and data. Report results clearly and in a way that addresses users’ needs.

The implementation of the model structure and data used to populate it should meet the standards of reproducible research. This is particularly important for models that are supported by public funds (e.g., models created in conjunction with EPC evidence reports) or models used to inform decisions that affect health care 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), and 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). Using the best analytic approach might make complete reporting more challenging; however, accessibility should not be pursued at the expense of model performance (i.e., models should not be oversimplified in order to make their operation understandable to users). The recent ISPOR–SMDM good research practices report provides detailed guidance regarding appropriate elements for technical and nontechnical documentation for modeling studies.

Reporting of the results of modeling studies should be tailored to the goals of the relevant stakeholders while remaining faithful to the model structure and assumptions, and communicating uncertainty in the results. Every effort should be made to present the model findings and analyses in a manner that will be most useful to the stakeholders who would be expected to use them. For models prepared in conjunction with EPC reports, stakeholders (e.g., Key Informants and Technical Experts) can provide useful suggestions for presenting the results of modeling efforts. It is impossible to give specific guidance to address all model types and uses of modeling covered by this document. Interested readers are referred to the many available texts on health care modeling, the reporting of statistical and simulation analyses, and graphing quantitative information.

Interpret and report results in a way that communicates uncertainty in model outputs.

Results should be reported in a way that effectively communicates uncertainty in model outputs. This may include the use of graphical and statistical summaries that describe 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.

Fully disclose any potential conflicts of interest.

All persons who developed the model, conducted and analyzed simulations, or interpreted model results, and those who provided input during any stage of the modeling process should fully disclose any potential conflicts of interest. Both financial and nonfinancial conflicts of interest should be reported. For models produced for the AHRQ EPC Program and many other health technology assessment groups, it is necessary that conflicts of interest be avoided. Modelers should adhere to established guidance for avoiding and managing conflicts of interest for EPC products (e.g., Institute of Medicine recommendations and existing EPC guidance).

Concluding Remarks

This report provides guidance in the form of widely accepted principles and good-practice recommendations for the conduct and reporting of modeling and simulation studies in the context of health technology assessments. Development of the guidance was based on a systematic review; input from clinical, policy, and decision analysis experts; and stakeholder
discussions. Leadership within the EPC Program, AHRQ personnel, and external reviewers provided extensive feedback. The principles and recommendations are applicable to the class of structural mathematical models that can be developed and used in conjunction with systematic reviews. Because of this broad scope, the guidance does not prescribe specific modeling approaches. We hope that this work will contribute to increased use and better conduct and reporting of modeling and simulation studies in health technology assessment.
Bibliographic Note

Table 4 organizes the references cited throughout the report into categories by (1) modeling topics covered and (2) whether the exposition of the methods was primarily targeting applications in medical, epidemiological, or health services research versus other research fields. The categories are not exclusive, and some references are cited under more than one category. The list is by no means exhaustive of the vast literature on modeling and simulation; it is meant only as a starting point for readers who wish to further explore this literature. We obtained guidance documents on mathematical and simulation modeling in medical, epidemiological, and health services research through systematic review; we obtained all other references from our personal bibliographies or through recommendations by stakeholders, topic experts, or peer reviewers.

Table 4. References cited in the report

<table>
<thead>
<tr>
<th>Topic</th>
<th>Primarily Target Medical, Epidemiological, or Health Services Research</th>
<th>Primarily Target Other Research Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing methodological guidance</td>
<td>78</td>
<td>None</td>
</tr>
<tr>
<td>Structural modeling, representation; goals of modeling, modeling process; model complexity</td>
<td>19, 50, 56, 66</td>
<td>18, 32-36, 55, 57, 77-92, 271, 274, 275, 311, 313-317, 466</td>
</tr>
<tr>
<td>General modeling tutorials, overviews of modeling practices, expository papers</td>
<td>1.39, 42, 44, 46, 51, 53, 54, 61, 62, 73, 163-167, 179-182, 2575, 168, 170-176, 178, 183, 200, 296-303, 351</td>
<td>37, 52, 169, 177, 185-191, 193-196, 198, 199, 201, 203, 204, 211, 310, 312</td>
</tr>
<tr>
<td>Guidance for modeling (the decision to conduct modeling, the methods for conducting the modeling, and issues related to reporting); empirical assessments of published modeling studies</td>
<td>2.17, 38, 47-49, 58, 70, 74, 102-110, 111-119, 121-147, 400-401, 48-162, 170, 359-361, 445, 463-465</td>
<td>184, 191, 202, 207, 447, 459</td>
</tr>
<tr>
<td>Books on decision analysis, economic and mathematical modeling</td>
<td>93, 304-307, 352, 470, 471</td>
<td>31, 39, 60, 64, 192, 205, 206, 209, 210, 308, 309, 460</td>
</tr>
<tr>
<td>Analytic frameworks, influence diagrams, conceptual modeling; choice of model perspective</td>
<td>20, 28, 280-282, 284</td>
<td>276-279</td>
</tr>
<tr>
<td>Sources of information for obtaining values for model inputs; methods for systematic reviews and meta-analyses, evidence synthesis</td>
<td>45, 421-427, 93, 239, 318-336, 338-345, 347, 348, 350, 352-357, 413-420</td>
<td>337, 346, 349</td>
</tr>
<tr>
<td>Elicitation of probabilities and preferences</td>
<td>94-96, 389-392, 394, 395, 397, 398</td>
<td>393, 396, 399</td>
</tr>
<tr>
<td>Risk-of-bias assessment and bias adjustment</td>
<td>340-343, 371-381, 383-388</td>
<td>382</td>
</tr>
<tr>
<td>Statistical modeling (behavioral)</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>Visualization of quantitative information</td>
<td>None</td>
<td>450, 467-469</td>
</tr>
<tr>
<td>Heterogeneity in modeling</td>
<td>408-412</td>
<td></td>
</tr>
<tr>
<td>Transporting study and modeling results across settings; generalizing results</td>
<td>111, 121, 137, 153, 400, 407, 428</td>
<td>401-406</td>
</tr>
<tr>
<td>Conflict of interest; potential for bias in the modeling process</td>
<td>472-480</td>
<td></td>
</tr>
<tr>
<td>Reproducible research</td>
<td>458</td>
<td>457</td>
</tr>
</tbody>
</table>

Citations to the two abstracts presenting preliminary results from this report are not included in the table.
References


44. Hodges JS. Six (or so) things you can do with a bad model. Oper Res. 1991;39(3):355-65.


119. Drummond M, Sculpher M. Common methodological flaws in economic


136. Luce BR, Simpson K. Methods of cost-effectiveness analysis: areas of consensus


205. Tolk A. Engineering principles of combat modeling and distributed simulation. Wiley Online Library; 2012.


374. Goldhaber-Fiebert JD. Accounting for biases when linking empirical studies and


411. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to


439. Law AM. A tutorial on how to select simulation input probability distributions.


