Methods Research Report

Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment

Methods Research Report

Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2007-10055-I

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

Investigators:
Issa J. Dahabreh, M.D., M.S.
Jeffrey A. Chan, B.A.
Amy Earley, B.S.
Denish Moorthy, M.D.
Esther E. Avendano, B.A.
Thomas A. Trikalinos, M.D.
Ethan M. Balk, M.D., M.P.H.
John B. Wong, M.D.

AHRQ Publication No. 16(17)-EHC020-EF
January 2017

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 author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

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.

Suggested citation: Dahabreh IJ, Chan JA, Earley A, Moorthy D, Avendano EE, Trikalinos TA, Balk EM, Wong JB. Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment. Methods Research Report. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No.16(17)-EHC020-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2017. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

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

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

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

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Sharon B. Arnold, Ph.D.
Acting Director
Agency for Healthcare Research and Quality

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

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

Elisabeth Kato
Task Order Officer
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

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 follows.

Eric B. Bass, M.D., M.P.H.
Director
Evidence-based Practice Center (EPC)
Johns 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
JAMA
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

J. Sanford Schwartz, M.D.
Leon Hess Professor in Internal Medicine,
University of Pennsylvania
Philadelphia, PA

Milton C. Weinstein, Ph.D.
Henry J. Kaiser Professor of Health Policy and Management,
Harvard Medical School
Director
Program on the Economic Evaluation of Medical Technology,
Harvard T.H. Chan School of Public Health
Boston, MA

Peer Reviewers

Prior to publication of the final report, the EPC Program sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does 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 non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows.

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

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

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

Michael Pignone, M.D., M.P.H., F.A.C.P.
Division of General Internal Medicine
University of North Carolina- Chapel Hill
Chapel Hill, 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

Evelyn Whitlock, M.D., M.P.H.
Center for Health Research,
Kaiser Permanente Northwest,
Portland, OR

Timothy J. Wilt, M.D., M.P.H.
Minneapolis VA Health Care System
Minneapolis, MN

Acknowledgments

The authors would also like to 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.

Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment

Structured Abstract

Background. Despite rigorous systematic reviews of efficacy and effectiveness of health care interventions, patients, providers and policymakers may remain in doubt about what they should do because of uncertainty, tradeoffs among benefits and harms, and conflicting preferences. Modeling and simulation studies in health care can supplement systematic reviews to increase the usefulness of the evidence summary. The *aims* of this report are four-fold: (1) to summarize evidence- and consensus-based guidance on the conduct and reporting of health care modeling and simulation; (2) to summarize guidance from Health Technology Assessment (HTA) groups for modeling; (3) to prioritize future research needs to improve models; and (4) to provide an overview of methods for model calibration and validation.

Methods. With guidance from a Technical Expert Panel and a Clinical and Policy Advisory team with clinical, methodological research, and policymaking expertise, we completed the following projects: For *Aim 1*, we conducted a systematic review of articles that provided evidence- or consensus-based recommendations for the conduct and reporting of health care modeling and simulation studies. We classified recommendation statements into four domains: model structure, data, consistency, and communication of model results. To contextualize the findings of the systematic review, we organized a meeting with a group of 28 stakeholders, including modelers, users of models, and funders of research. For *Aim 2*, we searched the web sites of 126 international agencies and institutes conducting HTA for real-world practices regarding when to apply modeling and simulation methods, which we summarized. For *Aim 3*, from the systematic review and from the stakeholders in Aim 1 we identified and collected suggestions for future research needs. Stakeholders prioritized those needs based on importance, desirability of new research, feasibility, and potential impact. For *Aim 4*, we searched for articles that compared or applied alternative validation methods for modeling and simulation. We extracted and summarized descriptions and comparisons of any methods and reported results for face validity, internal validity, external validity, cross-model validation, and calibration.

Results. The systematic review of modeling recommendations (Aim 1) found 71 eligible articles. 90 percent of articles (n=64) provided recommendations regarding model structure. Almost all articles (n=68, 96%) also provided recommendations regarding obtaining appropriate data to populate models. Stakeholders highlighted the importance of establishing guiding principles for "good practice" but discouraged the use of "cookbook" checklists. Of the 71 articles, 38 (54%) provided suggestions for future research; stakeholders provided 28 additional suggestions. We found 21 HTA organizations provided guidance (Aim 2) through their web sites regarding the application of modeling and simulation in the context of conducting a HTA. The HTA organizations varied widely in what areas of modeling they provided guidance for and what specific recommendations they made. In general, HTA organizations favored incorporating

models into HTA, provided recommendations on how to model data and structure, and recommended inclusion of costs in cost-effectiveness models. Future research needs that were prioritized (Aim 3) included questions about model data, model structure, consistency, and reporting. Studies comparing validation methods (Aim 4) provided information on model validation (face validity and internal, external, and cross-model validation) and calibration (varying specifications of the calibration problem with the same and different algorithms and use of alternative algorithms for the same calibration problem).

Conclusion. Our systematic review and stakeholder meeting summary provides a comprehensive compendium of guidance documents for modeling and simulation studies, annotated with information on the domains covered by each document, and the methods used to arrive at specific recommendations. We also summarized modeling recommendations for HTA organizations. These processes enabled us to prioritize future research needs to form an empirical basis for and to improve recommendations for modeling. Our overview of model validation and calibration provides insights into the relative value and efficiency of different methods.

Contents

Tables

Figures

Appendixes

Preamble

Despite rigorous systematic reviews of health care interventions, patients, providers and policymakers may remain in doubt about which interventions to use because of uncertainty, tradeoffs between benefits and harms, and differences in preferences. First, even after a synthesis of best-available evidence, uncertainty may remain when evidence is available only for surrogate outcomes, or when clinical studies have small sample sizes, limited followup durations, deficiencies in their design and conduct, or do not provide sufficient information on relevant patient subgroups. Second, tradeoffs between benefits and harms occur and optimal decisionmaking for individuals and populations may depend on their values (preferences) for the different potential outcomes. Lastly, models allow consideration of the preferences of decisionmakers for different outcomes when choosing among health care interventions.

This project was designed to advance the credibility, transparency, and methodological rigor of modeling by 1) assembling a multidisciplinary modeling core methods group; 2) reviewing and evaluating current literature on best practices in prioritization, conduct, and dissemination of health care models; 3) identifying priority topics in methods research for modeling and simulation; and 4) reviewing evidence on comparisons of validation and calibration methods.

The report is written as a series of four stand-alone chapters, each in a manuscript format. Chapter 1 is a systematic review of recommendations for the conduct and reporting of health care modeling and simulation studies. Chapter 2 is a review of guidance provided by international Health Technology Assessment (HTA) web sites. Chapter 3 prioritizes future research needs to improve the conduct and reporting of modeling and simulation studies. Addressing one of the prioritized future research needs, in Chapter 4 we review reports of health care models that used multiple methods to either validate or calibrate their models.

Chapter 1. Systematic Review of Recommendations for the Conduct and Reporting of Health Care Modeling and Simulation Studies

Introduction

Rigorous systematic reviews of the literature have become the preferred way to identify, appraise, and synthesize studies on the comparative effectiveness of competing health care interventions. However, users of such reviews—including patients, providers, and policymakers—may remain in doubt about which interventions to use because of persistent uncertainties, tradeoffs between benefits and harms, differences in preferences, or insufficient evidence.[1] Even after a synthesis of best-available evidence, uncertainty may remain regarding the optimal choice among available interventions for important patient-relevant outcomes because studies may provide information mainly on surrogate outcomes, have short follow-up durations, or report inadequate subgroup data. Policy needs often require decision making under uncertainty and decision makers have become increasingly interested in complementing the results of systematic reviews of empirical evidence with information from modeling and simulation studies.

Modeling is especially useful when uncertainty exists about how specific evidence might be applied to a particular decisional context or a specific patient, or formulated into a recommendation. Modeling can provide a comprehensive and transparent integration of empirical evidence on benefits and harms, values (preferences), and resource utilization, while accounting for all relevant sources of uncertainty.[2] Model results can be used to guide decision making and to support the prioritization and planning of future clinical research activities.[3,4] Decision-making typically involves trade-offs in harms and benefits and thus may depend on individuals' values for a range of outcomes. Modeling offers a coherent approach for integrating clinical research evidence and patient values to optimize decision making. However, caution is warranted when models are based on insufficient, poor quality, or extrapolated evidence. Models whose structure and inputs are based on extensive background knowledge good quality evidence are likely to be more reliable.

Methods for the conduct of health care modeling and simulation have continued evolving to address the ever-increasing information needs of decision makers. The complexity and continued advances of the relevant methods have spurred the publication of recommendation statements on "best practices" for modeling and simulation. Two previous systematic reviews have assessed published recommendations and guidelines for decision analytic modeling for health technology assessment in an effort to identify methodological recommendations for decision-analytic modeling. Philips et al., in 2004, identified existing guidelines for best practices in health technology assessment and classified them into three domains—model structure, data, and consistency.[5] Kuntz et al., in 2013, reviewed recommendations published through 2009 for developing, validating and using decision-analytic models in the context of systematic reviews.[1] In addition to model structure, data, and consistency, they assessed a fourth domain pertaining to the communication of model results.

We sought to update and expand previous syntheses of evidence- and consensus-based guidance on health care modeling by performing a systematic review of published recommendation statements for modeling and simulation studies conducted on health care topics.

We sought to address health care models broadly, regardless of whether they accompany systematic reviews or of their intended audience. To help contextualize and interpret the systematic review findings, we convened a meeting of diverse stakeholders, including modelers, users of models, and funders of research. Here we summarize the results of the systematic review and stakeholder input, to provide an up-to-date compendium of recommendations for best practices in health care modeling and simulation.

Methods

The Tufts Evidence-based Practice Center, under contract with the Agency for Healthcare Research and Quality (AHRQ), conducted a systematic review of the published literature for evidence- and consensus-based guidance on the conduct of modeling and simulation studies. We convened a Clinical and Policy Advisory team (CaPA), and a Technical Expert Panel (TEP) to provide input in the design and conduct of the review (see **Appendix A**). The CaPA was formed to provide clinical and policymaking expertise from individuals who have used modeling input to inform decision-making. It included three members, with expertise and experience in applying decision analysis and simulation modeling for developing clinical guidelines, assessing public health risk management, and informing health care policy decisions. The TEP included four internationally-recognized experts in evidence synthesis and modeling and two policymakers and payers, who also have experience in modeling.

Systematic Review

Search Strategy

We searched four electronic databases (MEDLINE, Cochrane Methodology Register, Health Technology Assessment Database, NHS Economic Evaluation Database) for articles presenting best practices in prioritization, conduct, and reporting of modeling and simulation studies through October 30, 2012. We based our search strategy on that used by Philips et al. and Kuntz et al. but substantially expanded the search terms to include keywords related to modeling methods and guideline statements.[1,5] We confirmed that our search captured all articles reviewed by the two prior reviews, relevant articles from our personal bibliographies, and those suggested by CaPA, and TEP members. To make the size of the corpus that needed to be screened manageable (the search yielded 65,053 citations), we limited our screening to 37 journals identified based on the journals that published articles included by Philips et al. and Kuntz et al., supplemented with journals recommended by the CaPA and TEP as likely to publish guidance documents pertaining to decision analysis, simulation, economic analysis for health technology assessment, and health outcomes research. We also limited the search to articles published since 1990. The final search strategy, with the list of included journals, is presented in **Appendix B**.

Abstract Screening and Study Selection

Using the open-source abstract screening software Abstrackr (http://www.cebm.brown.edu/software),[6] five reviewers independently screened abstracts in duplicate and resolved disagreements by group consensus. Eligible studies had to provide guidance on the elements of a good model, address the key elements that constitute a good modeling or simulation study, or provide explicit criteria against which to assess the quality or validity of models or simulations. Furthermore, we required that the guidance had to be either evidence-based (i.e., from a systematic review) or consensus-based (e.g., from discussion or

collaboration of experts) and that articles provided a description of the process through which the authors arrived at their recommendations. Although this requirement excluded some older seminal articles,[7-10] we found that their recommendations had been incorporated into more recent eligible articles. Examples of acceptable processes included systematic reviews, a Delphi consensus process, or presentation at a conference with active feedback from meeting attendees. We excluded articles that appeared to be based only on the opinions of the authors.

Eligible recommendation statements had to provide general guidance for modeling and simulation (e.g., guidance applicable across multiple disease topics such as comparative effectiveness of treatments, screening strategies, infection control, telemedicine) or for making formulary decisions at a national or regional level. Because of the potential for limited generalizability, we excluded guidance for modeling of specific disease types (e.g., chronic diseases such as cancer), local decision makers (e.g., hospital formularies), single modeling methodologies (e.g., Markov model, microsimulation model), or specific aspects of the modeling process (e.g., uncertainty propagation in models, reporting or dissemination of modeling guidance or results). Instead we focus on works that provide more broad guidance and make recommendations for a range of methods for and components of models and simulations. Such an example is the recent ISPOR-SMDM guidance, which covers all key specific aspects of modeling.[11]

We further excluded articles that provided recommendations for decision aids or for statistical analyses for estimating effect size, inference, or prediction; evaluated only costs; were related to clinical chemistry and laboratory medicine, occupational health, or vaccines; or dealt with multi-criteria decision analysis and analytic hierarchy process methods.

Data Extraction

We created a data extraction form based on the three-component framework originally proposed by Sculpher et al. in 2000[8] and employed in a previous systematic review of methodological recommendations:[5] model structure, model data, and model consistency. In addition, we added a component pertaining to the "communication of model results" as suggested by Kuntz et al. which we also found to be a useful addition to the original framework.[1] Subcategories within the four thematic components were identified and operationally defined following Philips et al. (also see **Figure 2** and **Table 1** in this chapter).[5] In each article, we identified specific recommendations as statements of policy or procedure that used modal verbs (e.g., "should," "must"), that were placed in the context of other text as a suggestion for action, that were explicitly noted as a recommendation, or that were included in a checklist of items used to critically review a model. We then mapped each recommendation statement to the appropriate component—structure, data, consistency, presentation of results—and subcategories therein. For each publication, we also extracted the process for generating recommendations, the purpose of the recommendations or article, the target audience, and the intended scope of the recommendations.

Stakeholder Meeting

We invited workshop participants who would represent expertise in modeling and simulation, systematic review and evidence-based medicine, epidemiology, and biostatistics, and perspectives from six stakeholder groups: patient representatives, providers of care, purchasers of care, payers, policy makers (including research funders and professional societies), and

principal investigators[12] In total 43 individuals were invited to participate and 28 attended the 1-day meeting in-person or remotely.

The goals of the workshop were to review and expand the list of recommendations and research needs identified by the systematic review of methodological recommendations for modeling and simulation and to develop a list of priority research areas aiming to improve the usefulness and credibility of models used to inform decision making. Results regarding future research needs will be reported separately.

One of the authors (JBW) began the meeting with a presentation about modeling to ensure common background knowledge and vocabulary and to set the ground rules for the ensuing discussions. We then presented the categories and topics of recommendation statements identified by the systematic review, followed by a group discussion involving all participants. After assignment into three smaller groups (each comprising 7 to 12 participants), stakeholders reviewed and discussed the modeling recommendations in-depth over two breakout sessions. One investigator facilitated each session (JBW, EMB, and IJD) and a second investigator kept detailed notes (DM, TAT, or JAC) to supplement the tape-recording of the discussions. Facilitators reviewed the goals for each session and used a list of topics to guide the unstructured discussions. Stakeholders were encouraged to comment on available recommendations and identified gaps, limitations and areas for expansion within each component of interest. Key points from each of the small group discussions were presented to the whole group for further discussion. Following the meeting, we circulated summaries from all discussions to participants and solicited additional comments via email.

Results

Systematic Review of Published Recommendations

The search **(Appendix B)** yielded 6825 citations (**Figure 1**); the TEP and CaPA suggested 20 additional articles. We reviewed 358 articles in full text, of which 71 met our eligibility criteria and reported recommendations for conducting modeling and simulation studies. [5,7,10,13-80] Of these, 28 provided recommendations based on expert panel deliberations, 14 on systematic literature reviews, 14 on nonsystematic literature reviews, 7 on conference discussions, and 8 on a combination of these methods.

The complete list of extracted information is available electronically on the Systematic Review Data Repository (project title: *Recommendations for modeling and simulation studies*).

Figure 1. Flow diagram for papers presenting recommendations for modeling and simulation

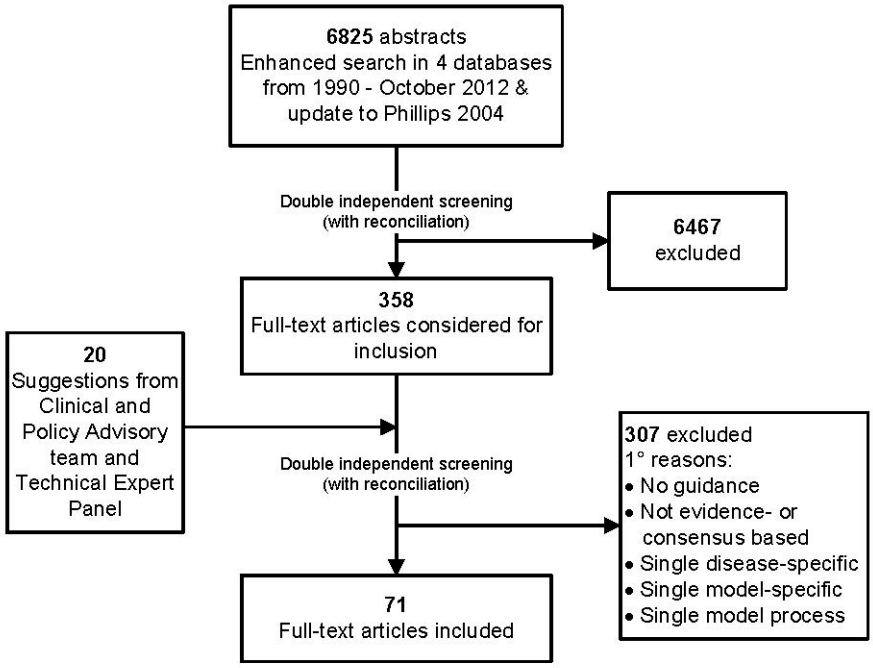
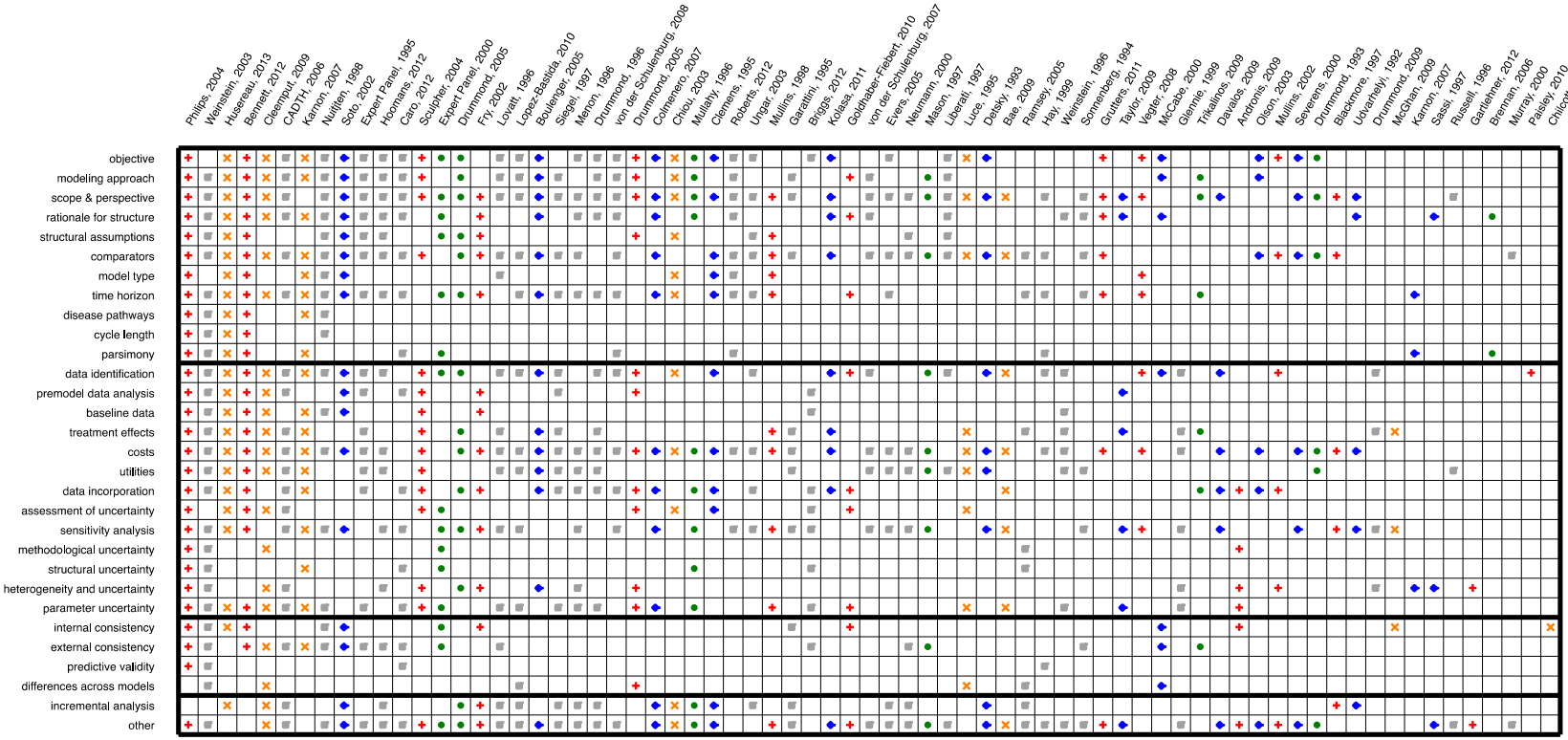


Figure 2 presents a summary of the recommendation types extracted from each article, along with information on the process used to generate the recommendations. When examining a single row of the matrix, a preponderance of filled cells indicates that most articles have provided recommendations addressing a specific area. For example, defining the scope and perspective of the analysis is the most common area of recommendations across the articles we reviewed. When examining a single column of the matrix, a preponderance of filled cells indicates that a given paper has provided recommendations addressing many different areas. For example, the paper by Philips et al. (2004) addressed all but two of the areas we examined.[5]

The majority of articles (64/71; 90%) provided one or more recommendations regarding model structure, including modeling objectives (40/71; 56%); model scope and perspective (52/71; 73%), and choice and justification of comparators used in modeling (46/71; 65%). Nearly all (68/71; 96%) made recommendations about obtaining appropriate data to populate models (main and sensitivity analyses), including obtaining cost data (49/71; 69%), methods for data identification and synthesis (40/71; 56%), and the conduct of sensitivity analyses (40/71; 56%). A minority of articles provided recommendations regarding the assessment of model consistency (27/71; 38%). Recommendations about model validity mainly pertained to internal (14/71; 20%) and external validity (19/71; 27%), with only a small number addressing the predictive validity of models (4/71; 6%). **Table 1** gives the count of recommendation types for all 71 articles and stratified by process used to generate the recommendations. Overall, no clear pattern of association between process and specific recommendation types was apparent.

Figure 2. Map of recommendation statements from the systematic literature review



Note: The figure presents a visual summary of the recommendation statements extracted from the systematic review. Columns represent the individual eligible papers (sorted left to right by number of recommendation subcategories). Rows represent recommendation topics. Thick black lines group subcategories belonging to the same component (from top to bottom: structure, data, consistency, and miscellaneous). Red crosses indicate recommendations based on systematic reviews, grey squares indicate expert panels, green circles indicate conferences or meeting, blue diamonds indicate "nonsystematic" reviews, and orange boxes indicate combinations of methods.

.

Table 1. Recommendation types, by process of generation

	Reporting Items, by category	All documents N=71	SR N=14	Expert panel N=28	Conf/Mtg N=7	Non-SR N=14	Combo N=8
Structure	objective	40 (56)	7 (10)	15 (21)	4 (6)	9 (13)	5 (7)
	modeling approach	32 (45)	5 (7)	15 (21)	4 (6)	4 (6)	4 (6)
	scope & perspective	52 (73)	9 (13)	22 (31)	6 (8)	10 (14)	5 (7)
	rationale for structure	33 (46)	5 (7)	14 (20)	3 (4)	8 (11)	3 (4)
	structural assumptions	17 (24)	5 (7)	7 (10)	2 (3)	1 (1)	2 (3)
	comparators	46 (65)	8 (11)	22 (31)	3 (4)	8 (11)	5 (7)
	model type	12 (17)	4 (6)	3 (4)	0 (0)	2 (3)	3 (4)
	time horizon	36 (51)	7 (10)	17 (24)	3 (4)	5 (7)	4 (6)
	disease pathways	6 (8)	2 (3)	2 (3)	0 (0)	0 (0)	2 (3)
	cycle length	5 (7)	2 (3)	2 (3)	0 (0)	0 (0)	1 (1)
	parsimony	12 (17)	2 (3)	5 (7)	2 (3)	1 (1)	2 (3)
Data	data identification	40 (56)	8 (11)	17 (24)	3 (4)	7 (10)	5 (7)
	premodel data analysis	15 (21)	5 (7)	6 (8)	0 (0)	2 (3)	2 (3)
	baseline data	12 (17)	4 (6)	4 (6)	0 (0)	1 (1)	3 (4)
	treatment effects	25 (35)	4 (6)	11 (15)	2 (3)	3 (4)	5 (7)
	costs	49 (69)	9 (13)	20 (28)	4 (6)	10 (14)	6 (8)
	utilities	28 (39)	3 (4)	17 (24)	2 (3)	2 (3)	4 (6)
	data incorporation	30 (42)	8 (11)	10 (14)	3 (4)	6 (8)	3 (4)
	assessment of uncertainty	13 (18)	5 (7)	2 (3)	1 (1)	1 (1)	4 (6)
	sensitivity analysis	40 (56)	6 (8)	19 (27)	4 (6)	7 (10)	4 (6)
	methodological uncertainty	6 (8)	2 (3)	2 (3)	1 (1)	0 (0)	1 (1)
	structural uncertainty	8 (11)	1 (1)	4 (6)	2 (3)	0 (0)	1 (1)
	heterogeneity and uncertainty	18 (25)	7 (10)	6 (8)	1 (1)	3 (4)	1 (1)
	parameter uncertainty	29 (41)	7 (10)	13 (18)	2 (3)	2 (3)	5 (7)
Consistency	internal validity	14 (20)	5 (7)	3 (4)	1 (1)	2 (3)	3 (4)
	external validity	19 (27)	2 (3)	10 (14)	3 (4)	2 (3)	2 (3)
	predictive validity	4 (6)	1 (1)	3 (4)	0 (0)	0 (0)	0 (0)
	differences across models	7 (10)	1 (1)	3 (4)	0 (0)	1 (1)	2 (3)
Other	incremental analysis	24 (34)	2 (3)	12 (17)	2 (3)	5 (7)	3 (4)
	other, including reporting	52 (73)	9 (13)	24 (34)	5 (7)	11 (15)	3 (4)

All results are expressed as number of papers (percentage of papers, out of 71).
Combo = combination of methods, Conf/Mtg = conference or meeting, SR = systematic review.

Stakeholder Panel Discussions

The stakeholder group, together with the CaPA, TEP, and the investigators, as well as representatives from AHRQ, met for a day-long series of sessions on 27 February 2013 at AHRQ headquarters (Rockville, MD), to comment and deliberate on the findings of the systematic review (see **Appendix A**). After reviewing the list of recommendations, stakeholders uniformly agreed on the need to increase transparency of various modeling processes. Initial discussions centered on clarifying individual recommendations and, occasionally, their classification into the four components (structure-data-consistency-communication of results). Stakeholders highlighted the importance of recommendations that apply broadly, across multiple types of models. Questions were raised on the appropriateness of using the Philips et al. report as a framework and whether individual recommendations should be bundled/re-bundled or stand alone.[5] While recognizing that the classification of recommendations into the four components involved judgments, we believe that it will be helpful to modelers seeking to identify guidance related to a specific aspect of their work.

In small group discussions, stakeholders questioned the value of yet another "checklist of recommendations". Many felt it would be more helpful to identify "best practices" or "principles" on how to integrate modeling and the systematic review processes. They stressed the importance of understanding the needs and perspectives of the different stakeholders who might use model results, with particular consideration of patient preferences. Participants strongly indicated that the presentation and contextualization of modeling results to different stakeholders (e.g., policy-makers) remains an understudied area that is as important as developing methodological guidance for the conduct of modeling and simulation studies.

Conclusions

This systematic review updates and organizes recommendations for health care modeling and simulation studies into four domains. It classifies recommendations by domain and by the methods used to develop them. A synthesis of these recommendations, together with additional sources of information on modeling and simulation, has been attempted in the accompanying report titled *Guidance for the Conduct and Reporting of Modeling and Simulation Studies in the Context of Health Technology Assessment*.

Over 90 percent of recommendation documents provided one or more recommendations pertaining to model structure and data appropriateness, but only a minority provided recommendations regarding model validity. One might conjecture that the recommendation topics that were most frequently addressed in the papers we reviewed were either the most important (as perceived by those making the recommendations) or the easiest to address (on the basis of available theoretical or empirical evidence). Given the critical importance of assessing model validity,[11] the absence of recommendations on methods for model validation more likely reflects uncertainty about the most appropriate methodological approaches. We review issues related to model validation and calibration in Chapter 4 of this report.

We discussed the findings of the systematic review with a diverse stakeholder panel that included patient representatives, providers of care, purchasers of care, payers, policy makers (from a range of national and international agencies and professional societies), and principal investigators, many with expertise in modeling and simulation, systematic review methods, epidemiology, and biostatistics (See **Appendix A** for the list of stakeholders). These discussions highlighted the importance of transparency in modeling methods in decision analysis

applications and the need to identify best practices, rather than use "cookbook" checklists. When applied to complex methodological decisions (such as the design and conduct of modeling and simulation studies) checklists tend to promote mechanistic approaches that do not adequately address the practical challenges faced by modelers. Alternatively, determining the theoretical considerations relevant to each research problem and working from "first principles" may result in more fundamentally sound models. Nonetheless, checklists can increase consistency in the reporting of the methods and findings of modeling and simulation studies, and in the peer review of modeling research.[31,81]

Our work has limitations that need to be considered when interpreting our results. The majority of the recommendation statements we reviewed were derived from the medical literature, leading to a preponderance of statements regarding economic analyses (using modeling and simulation methods). Nonetheless, outside of cost-related considerations, many of the recommended methodological principles apply across multiple model types, as presented in our annotated bibliography of recommendations. More generally, systematic reviews of methodological topics are usually less likely to be comprehensive than reviews of empirical research studies (e.g., clinical trials) because of the unavailability of standardized indexing terms and the large number of sources that need to be searched. Further, we only classify guideline statements into four broad domains (components) and sub-components, without attempting to reconcile conflicting statements or to provide a synthesis across statements. Such a synthesis would be an appropriate task for a recommendation-making panel, such as the group updating the recommendations of the Second Panel on Cost Effectiveness in Health and Medicine (personal communication, Prof. Peter Neumann).[82]

In conclusion, this systematic review and stakeholder meeting summary provides a comprehensive compendium of guidance documents for modeling and simulation studies, annotated with information on the domains covered by each document, and the methods used to arrive at specific recommendations. Our annotated bibliography will be useful to modelers and others looking for sources of evidence- or consensus-based guidance. We have also incorporated this review of the existing guidance into a separate document: *Guidance for the Conduct and Reporting of Modeling and Simulation Studies in the Context of Health Technology Assessment*.

Chapter 2. Review of Guidance from Health Technology Assessment Organizations

Introduction

Health technology assessment (HTA) is a method of evidence synthesis that utilizes systematic review methodology to study the medical, social, ethical, and economic implications of development, diffusion, and use of health technology.[83] HTA is evolving with the need for continued decisionmaker support when evaluating new technologies.[84] Systematic review, utilizing meta-analytic techniques, provides best estimates of average effects of interventions and rates of outcomes (including benefits and harms). Given sufficient evidence, systematic review may also provide an indication about how effects of interventions may differ in different settings, among different people, and across variations of the interventions (e.g., different doses or diagnostic thresholds). However, systematic reviews alone often provide a poor basis to balance benefits and harms in different scenarios or for different patients, to balance benefits and the resources required to achieve them, or to incorporate individual values (preferences) into decisionmaking. Modeling can provide a comprehensive, transparent, and interpretable integration of empirical evidence on benefits and harms, preferences, and resource utilization, while accounting for all relevant sources of uncertainty. However, it is hard to determine when a model may be of sufficient value to justify the time and resources required to incorporate it alongside systematic reviews.[85]

The goal of this review is to identify and summarize guidance from international HTA organizations, agencies and institutes that assess new health technologies or economic evaluations regarding when and how to incorporate modeling alongside systematic reviews, specifically searching for descriptions regarding whether to integrate modeling as part of HTA, the timing of modeling in relation to the conduct of systematic review, modeling methodological recommendations, the potential budget impact of introducing the technology, and the effect of including modeling on the HTA project budget. We then categorized all available recommendations across HTA organizations to examine both differences and commonalities in guidance.

Methods

Mathes et al. originally identified the HTA organizations through member lists of the International Network of Agencies for Health Technology Assessment (INAHTA), Health Technology Assessment International (HTAi) not-for-profit organizations, and the European Network for Health Technology Assessment (EUnetHTA).[86] Building on this review we searched the web sites of 126 international agencies and institutes conducting HTA (HTA organizations) for guidance regarding when and how to apply modeling methodology. We used the list of “HTA organizations” in Mathes et al. without making further determination of whether each is truly an HTA organization. For our review, two researchers independently reviewed identified HTA organization web sites in January 2014 (**Appendix C**). Standardized screening procedures included use of web site navigation (e.g., hyperlinks utilizing key terms such as "publications", "recommendation" and "methods"), web site search engines, and sitemaps to identify appropriate Web content and documents.

Relevant data were collected verbatim directly from web site text and linked documents available through the web site such as handbooks or guidelines on HTA methods. Non-English language web sites and documents were excluded from extraction. From each source, we extracted language from statements that could be read as recommendations, guidance, or standards regarding when and how to model including integration of modeling as part of HTA, modeling alongside systematic review of literature, timing of modeling with respect to the systematic review (i.e., concurrent or sequential), use of pre-existing versus established models, how to model, and budgetary considerations. The choice of guidance topics was determined prior to review of any of the relevant HTA organization documents. We did not search for methods pertaining to how HTA organizations determined their recommendations. Relevant language was extracted verbatim to standardized spreadsheets. Reviewers further categorized extracted language on when to model into 11 categories, as listed in **Table 2**. All extracted language and categorizations were reviewed by the entire project team for consensus.

Table 2. Questions for categorization of HTA guidance

Category	Question
Integration of modeling in HTA	Is modeling always a part of HTA?
Systematic review alongside modeling	Is modeling always informed by a systematic review?
Timing of modeling	Should modeling be performed concurrent with or after a systematic review once the parameters are set?
Use of pre-existing or established models	Should one use pre-existing or established models? Include the process for this decision (e.g., Is a systematic review of models conducted?; How is a model deemed adequate for use?)
Modeling recommendations	What are the methods for modeling?
How systematic review incorporated into the model	How should one incorporate the systematic review into the model?
Who conducts the model	Who should perform the modeling? What is the composition of the modeling team?
Inclusion of quality of life	Should the model include quality of life?
Inclusion of costs	Should the model include costs?
Budget analysis done	Should one conduct a budget analysis (e.g., a national budget impact analysis)?
Impact on project budget	Should one conduct an analysis of the impact of modeling on a project budget?

Results

Of the 126 web sites, 21 (17%) provided relevant text regarding the application of modeling in the context of conducting a HTA.[87-107] The remaining 106 web sites (84%) did not provide any guidance or were written in non-English languages and could not be translated (42 web sites, 33%). The 21 web sites with relevant, extractable data included HTA organizations from four continents including two HTA organizations in Asia (Taiwan and Thailand), three in Australia and New Zealand, 11 in Europe, and five in North America (3 U.S. and 2 Canada). A summary of available guidance on modeling across the 21 Web sites is presented in **Table 3** and all recommendation made by the HTA organizations are tabulated in **Appendix D**. Details specific to each of the 11 subcategories are described below.

Table 3. Summary of available guidance on modeling from international HTA agencies

Web site	Integration of modeling in HTA	SR alongside modeling	Timing of modeling	Use of pre-existing or established models	Modeling recommendations	How SR incorporated into the model	Who conducts the model	Inclusion of quality of life	Inclusion of costs	Budget analysis done	Impact on project budget
Agency for Health Technology Assessment in Poland (AHTApol/Poland)	X				X			X	X		
Canadian Agency for Drugs and Technologies in Health (CADTH/Canada)					X				X		
Danish Centre for Health Technology Assessment (DACEHTA/Denmark)	X				X			X	X		
Health Information and Quality Authority (HIQA/ Ireland)	X	X	X		X			X	X	X	
National Authority of Medicines and Health Products (INFARMED/Portugal)			X		X						
Institute for Quality and Efficiency in Health Care (IQWIG/Germany)	X								X		
Belgian Federal Health Care Knowledge Centre (KCE/Belgium)	X							X	X		

Web site	Integration of modeling in HTA	SR alongside modeling	Timing of modeling	Use of pre-existing or established models	Modeling recommendations	How SR incorporated into the model	Who conducts the model	Inclusion of quality of life	Inclusion of costs	Budget analysis done	Impact on project budget
MAS (Medical Advisory Secretariat, within the Ontario Ministry of Health and Long-Term Care Health Strategies Division/Canada)					X				X		
Medical Services Advisory Committee (MASC/Australia)					X				X		
National Institute for Clinical Excellence (NICE/UK)	X	X	X	X	X	X		X	X	X	
Pharmaceutical Benefits Advisor Committee (PBAC, Australia)	X				X						
Pharmaceutical Management Agency of New Zealand (PHARMAC/New Zealand)	X										
AAZ (Agency for Quality and Accreditation in Health Care/Croatia)					X				X		
HITAP (Health Intervention and Technology Assessment Program/Thailand)	X				X				X	X	
ICER (Institute for Clinical and Economic Review/US)	X	X		X	X	X					

Web site	Integration of modeling in HTA	SR alongside modeling	Timing of modeling	Use of pre-existing or established models	Modeling recommendations	How SR incorporated into the model	Who conducts the model	Inclusion of quality of life	Inclusion of costs	Budget analysis done	Impact on project budget
LBI (Ludwig Boltzmann Institute for Health Technology Assessment/Austria)	X				X						
MHRA (Medicines and Healthcare Products Regulatory Agency/UK)	X				X						
NLM (National Library of Medicine/US)	X				X			X	X		
AHRQ (US Agency for Healthcare Research and Quality/US)	X		X	X	X	X	X	X	X		X
CAST (Centre for Applied Health Services Research and Technology Assessment, University of Southern Denmark)	X	X	X		X				X		
CDE (Center for Drug Evaluation/Taiwan)		X								X	

Abbreviation: SR, systematic review

Despite the relatively large number of international HTA organization web sites, only 17 percent (21 agencies) provided guidance regarding the application of modeling in the context of conducting an HTA. On average, each web site provided guidance on roughly 4 of the 11 subcategories regarding when and how to incorporate modeling. NICE and AHRQ provided the most guidance among the 21 web sites, addressing 9 out of 11 subcategories, followed by the Health Information and Quality Authority (HIQA/Ireland) providing guidance on 7 of the subcategories. All but one HTA organization (Pharmaceutical Management Agency of New Zealand) provided guidance on multiple subcategories. The most frequently reported guidance across HTA organization web sites addressed recommendations on how to model (17 agencies) and integration of modeling as part of HTA (15 agencies); only one HTA organization (AHRQ) provided guidance on who develops and implements the model.

Integration of Modeling in HTA

Fifteen HTA organizations addressed incorporation of modeling in the context of conducting an HTA. Of these 15, six agencies required a model for their HTAs; three specifically as part of HTA development process and three as part of an economic evaluation process. Three other agencies recommended, but did not require, the incorporation of a model as part of HTA under certain conditions. For example, two recommended including models when conducting an economic evaluation. The remaining six agencies neither required nor recommended modeling as part of HTA. Rather they remained neutral on the topic while acknowledging that the technique has been used in prior agency reports.

Systematic Review Alongside Modeling

Five HTA organizations addressed the conduct of systematic reviews to inform modeling. One mentioned modeling but did not specify whether it should always be conducted alongside systematic review or alongside HTA in general. Of the remaining four HTA organizations, only one web site stated that a model must always be prepared jointly with a systematic review when conducting HTA. The other three organizations did not always require a systematic review.

Timing of Modeling

Five HTA organizations provided guidance on the timing of modeling in the context of conducting a HTA. Two recommended that the modeling be done concurrently with the systematic review while the other three recommended the modeling be done after the review once parameters have been estimated with one, however, acknowledging that it may not be feasible or timely to conduct the systematic review and model in sequence rather than in parallel.

Use of Pre-existing or Established Models

Only three of the 15 HTA organizations addressed the use of pre-existing versus established models in the context of conducting a HTA. One did not favor the use of pre-existing or established models in HTA and recommended the development of a *de novo* model to accompany each systematic review. The remaining two HTA organizations acknowledged that it was acceptable to use pre-existing models under certain conditions; one required that the model be conducted by manufacturers and sponsors of the HTA and the other cautioned against using established models that are not flexible enough to represent the consequences of all interventions of interest.

Modeling Recommendations

Seventeen of the 21 HTA organizations (81%) provided guidance on how to model. We divided these recommendations into four categories addressing whether or not the advice featured a statement on data, structure, validity and assumptions. Twelve of the 17 HTA organizations included a statement on model data, 14 addressed model structure, 6 addressed model validation, and 7 addressed model assumptions.

How Systematic Reviews Were Incorporated Into the Modeling Process

Whether a model is prepared concurrently with or after the completion of a systematic reviews, the question remains how to incorporate the review results into the model. Three HTA organizations gave guidance on this issue with two stating that the model outcome estimates should be based on the systematic reviews. The remaining HTA organizations suggested that the systematic reviews be used to produce parameter estimates for use in sensitivity analyses.

Who Conducts the Model

AHRQ was the only HTA organization addressing the question of who specifically should perform the modeling. They note that it is not always feasible for the systematic review team to also conduct a model since different expertise is needed. When separate teams are used, they should collaborate closely and should, ideally, reside in the same location.

Inclusion of Quality of Life

All seven HTA organizations that comment on quality of life suggest that models should take into account differences in quality of life among health states. Four of the seven state that a model should always include quality of life. The other three state that quality of life should be incorporated when appropriate (e.g., when final utility results are needed or when there is adequate evidence about quality of life to include in the model).

Inclusion of Costs

Fourteen HTA organizations provided advice on whether costs should be included in a model. Only four HTA organizations stated outright that costs should be incorporated into the model. Ten HTA organizations suggest using costs only for the conduct of a cost-effectiveness analysis.

Budget Analysis Done

Four HTA organizations provided guidance on conducting a budget analysis and all recommended a budget impact analysis as part of HTA.

Impact on Project Budget

AHRQ was the only HTA organization that addressed the impact of conducting a modeling project on the systematic review budget and suggested that “modeling efforts could easily consume 20 to 40 percent of the budget for a systematic review” if it were included as part of the review.[108]

Conclusions

Mathes (2013),[86] upon which our search was based, summarized recommendations on methods for the preparation of economic evaluation by international HTA organizations. We extended their review by summarizing the HTA organizations' recommendations within the framework laid out by Sculpher (2000),[8] Philips (2004),[5] and Kuntz (2013),[109] as elaborated on by our systematic review of recommendations for modeling and simulation (*Chapter 1*). This structured framework, which allowed us to categorize recommendations into the 11 domains in **Table 2**, highlighted differences across HTA organizations with respect to the breadth and detail of guidance provided to modelers. The recommendations made by HTA organizations pertain to models conducted for the purposes of these organizations, primarily evaluation of new or costly technologies. The models used by the HTA organizations are often used by national agencies to determine coverage or clinical recommendations.

Despite the relatively large number of international HTA organization web site, only 17 percent (21 agencies) provided guidance regarding the application of modeling in the context of conducting an HTA. The majority of these HTA organizations are from Europe, Australia/New Zealand, and Canada. AHRQ provided the most guidance among the 21 web site followed by NICE and HIQA/Ireland. HTA organization mostly addressed how to model and how to integrate modeling into HTA.

The HTA organizations varied widely in what areas of modeling they provided guidance for and what specific recommendations they made. This is consistent with the heterogeneity across researchers and organizations in the recommendations on the conduct and reporting of modeling and simulation studies, described in *Chapter 1*. For example, although 17 HTA organizations provided guidance on how to model, guidance did not consistently address the same themes (i.e., data, structure, consistency) regarding how to model. Moreover, while 15 HTA organizations commented on the integration of modeling as part of HTA, not all HTA organizations required a model as part of HTA development. Finally, variation in the frequency of reporting across the 11 subcategories may reflect the relative importance of these aspects of modeling to specific agencies and/or countries.

A number of practical constraints may have limited our review of HTA organization web site. We relied on a previous review of HTA organization web site,[86] which provided an apparently comprehensive listing of international HTA organization web site, but agencies not present on these lists may have provided additional modeling guidance. In addition, HTA organization guidance not provided on a web site were not included. We considered using Google Site Search (https://www.google.com/cse/sitesearch/create) to more comprehensively screen individual web site to identify modeling guidance however, the often large number of potentially relevant items identified by this search tool made using this methodology infeasible. In addition, about one-third of the web sites accessed (38 agencies) lacked an English language translation and could not be reviewed to identify additional relevant modeling guidance. Not all HTA organizations web site included "guidance," per se, and it required a collective judgment call to categorize their guidance statements. Other readers may have categorized statements differently. We were also unable to obtain descriptions of methods used by the HTA organizations to determine their recommendations.

In summary, only a small number of HTA organizations (21 of 126) provide guidance on their web sites for incorporating modeling alongside systematic reviews. Most HTA organization that provided guidance made recommendations about whether to incorporate models into HTA, generally favoring including models. Most also provided recommendations on how to model,

focusing primarily on model data and structure, with fewer recommendations on model assumptions or validation. Similarly, most addressed inclusion of costs in cost-effectiveness models. Few HTA organizations provided guidance related to other aspects of modeling in the context of HTA, including whether modeling should be conducted alongside systematic review, when modeling should be done related to the HTA, whether pre-existing models could be used, how systematic reviews should be incorporated into the modeling process, whether models should incorporate quality of life, whether a budget impact analysis should be done as part of the HTA, or who should conduct the modeling or the project budget impact of adding modeling to an HTA. Variability in recommendations probably reflects the heterogeneous needs addressed by HTA agencies operating in different jurisdictions. Harmonizing guidance across HTA agencies, and adopting a common set of best practices whenever possible, would allow for more efficient transport of modeling results (or specific model implementations) across agencies.

Chapter 3. Future Research Needs for Health Care Modeling and Simulation Studies

Introduction

Trade-offs between benefits and harms are common in most clinical contexts and should be weighed against each other in decisionmaking. Providers, patients, and policymakers are increasingly interested in complementing evidence of benefits and harms with information from modeling and simulation studies to explicitly answer pressing health care policy needs. Modeling of health care conditions and management options, ideally based on evidence from systematic reviews, can provide a single, comprehensive, explicit and interpretable analysis of uncertainty,[2] values and resource utilization to guide decisionmaking and to support the prioritization of future clinical research activities.[3,4] Modeling and simulation studies offer a coherent approach for integrating clinical research evidence and patient values to optimize choices and maximize expected utility for individual and for population health. However, as highlighted in Chapter 1, gaps remain regarding how best to conduct and report modeling and simulation studies.

The systematic review and panel discussion described in Chapter 1 underscored the need for further research on ways to improve upon the performance of and uses of modeling and simulation studies in the context of systematic reviews. Here, we describe the stakeholders' prioritization of future research needs topics to advance modeling and simulation.

Methods

We conducted a systematic review of the published literature for evidence- and consensus-based guidance on the conduct of modeling and simulation studies. From the systematic review articles, we extracted suggestions for future research (both future research needs and research gaps) made by their authors. We then categorized the finding from the systematic review and convened an expert and stakeholder panel to discuss our findings and to prioritize future research needs in that context.

As described in Chapter 1, we formed a stakeholder panel that included experts in modeling and simulation, systematic review and evidence-based medicine, and epidemiology and biostatistics (See **Appendix A** for the list of stakeholders). Perspectives from six stakeholder categories were represented in the workshop[12]—patients and the public; providers of care; purchasers of care; payers; policy makers; and principal investigators, researchers and research funders. Several stakeholders represented more than one perspective. Policymakers, principal investigators, and providers were the most prevalent perspectives. Workshop participants included policy makers (from a range of national and international agencies and professional societies), AHRQ Evidence-based Practice Centers, guideline developers, CISNET (Cancer Intervention and Surveillance Modeling Network), modelers, epidemiologists, statisticians, professional societies, payers, and patient advocates. **Appendix A** lists the 28 stakeholders who participated in the panel. The panel met in-person or remotely at a workshop at AHRQ on 27 February 2013. The goals of the workshop were to review and expand the list of recommendations and research needs identified by the systematic review of methodology recommendations for modeling and simulation, and to prioritize research to improve the usefulness and credibility of models to inform decisionmaking.

Stakeholders were encouraged to comment on individual recommendations, identifying gaps, limitations and areas for expansion. They then discussed future research needs topics. The groups reviewed and discussed the list of future research needs gathered from our systematic review. The stakeholders were then asked to prioritize the list of future research needs derived from the systematic review. In addition, during the meeting, we solicited additional topics from participants, and following the meeting, the stakeholders prioritized these also. We engaged stakeholders in three groups, which met separately, as well as jointly. We refer to these three groups as "policymakers", "principal investigators", and "providers" to refer to the majority of the participants' perspective in each group. Stakeholders representing the perspectives of patients and the public, researchers and purchasers were dispersed in the three groups.

To direct discussions about future research needs and deliberations about their prioritization, we asked stakeholders to consider four dimensions of need—Importance, Desirability of new research, Feasibility, and Potential impact—described more fully in **Table 4**. Due to methodological restrictions to comply with Federal policies, we used multiple methods with our stakeholders to assess their priorities. At the meeting, stakeholders were provided with a form listing the future research needs derived from our systematic review and were asked to rank order each on a scale of 1 ("Not desirable") to 10 ("Essential"), with an option for no opinion. After the meeting, stakeholders were sent the list of additional future research needs proposed by the stakeholders themselves and were asked for their ratings but, no specific method for rating was suggested. Most stakeholders who responded used a 1-10 scale but alternative methods used included categorizing into high, moderate, and low priority; 1 to 7 stars, and others. We therefore normalized all scales to a 10 point scale. Because of the different systems used to prioritize future research needs from the systematic review and the stakeholders, these two priority lists were kept separately. We selected the approximately five highest priority future research needs from each stage, using a natural breakpoint between higher and lower priority topics rather than a strict threshold of five topics.

Table 4. Dimensions of need (approach to prioritization)*

Dimension	Definition
Importance	• Represents a critical uncertainty for decision makers • Advances credibility, transparency, and methodological rigor of modeling • Represents important variation or controversy in modeling practice • Represents high burden in time, effort, or resources to modelers
Desirability of New Research/Duplication	• Would not replicate ongoing or prior research, or established knowledge
Feasibility	• Ability to perform research
Potential Impact	• Potential for significant health or economic impact with clear implications for resolving important dilemmas in health and health care decisions or inequities or vulnerable population • Frequency (high frequency implies greater potential impact, and vice versa)

* Based on Standardized Selection Criteria of the Agency for Healthcare Research and Quality's Effective Health Care Program.[110]

Results

We reviewed 71 papers reporting recommendations for conducting modeling and simulation studies. The future research needs and research gaps presented in these papers are summarized in **Table 5**. Suggestions for future research were made in 38 of the 71 papers (54%).[5,10,15-17,19-22,24-27,29,30,32,36,37,39-41,43,44,48,50,53,54,57,59,60,62-64,66,67,70,72,75] Initial discussion among stakeholders focused

on the purpose of prioritizing future research needs with reasons including using models to guide grant funding and the distribution of research and funding across diseases, intervention types, and methods, and choosing appropriate topics for evidence review. Stakeholders felt that future research needs should span various model types. Aspects of model evaluation, including model validation and calibration, emerged as important targets for methodological research. Methodological advances in model evaluation were considered important for ensuring the validity and enhancing the credibility and acceptability of modeling results. Stakeholders presented multiple views across various future research needs. An example includes conflicting views expressed on the level of prioritization to place on using non-randomized trial data for assessing treatment effectiveness based on the existence of ongoing research on this topic.

One group of stakeholders discussed the use of multilevel modeling of primary or secondary (aggregate) data, with a particular focus on using data from multicenter/multiregional studies as an important field for future research. This group also noted that the topics identified by the systematic review as targets for future research were somewhat "Euro-centric", possibly reflecting the origin of a large number of the publications included in the review. The panel suggested that effort should be directed toward identifying additional research priorities with an emphasis on the United States health care system and its needs. Stakeholders uniformly emphasized the importance of performing research on "widening the audience for using models," including research on how to communicate the results of modeling studies to different audiences. Discussions concluded with the identification of 30 additional future research needs to be included in prioritization exercise (**Table 6)**.

Based on stakeholder feedback, the 10 future research needs that were considered high priority by the stakeholder panel as a whole are highlighted in **Tables 5 and 6**. The highest priority future research need about model structure is a review of the standards for best practices in fields outside of medicine, such as engineering and environmental modeling. The goal would be to ensure that modeling of medical topics uses the most up-to-date practices and to determine how medical models could be improved by using, testing and adapting methodological advances from outside health care. This future research need was of particular interest to principal investigators, those most likely to develop models and conduct simulations.

Another priority for future research pertained to incorporating surrogate outcomes in models, and to evaluating the assumptions invoked when using surrogate outcomes. Three future research needs were prioritized regarding where model data come from. Stakeholders wanted to better understand how nonrandomized trial data should be appropriately used, and if and how these data need to be assessed; how the bias inherent to many studies, regardless of design, ought to be handled; and what is the validity of, the indications for, and the best practices for conducting multiparameter synthesis.

Table 5. List of future research needs derived from the systematically reviewed articles

Future Research Needs	Prioritized Future Research Needs			
	All stakeholders	Policy-makers	Principal investigators	Providers
MODEL STRUCTURE				
Incorporation of surrogate outcomes, often done naively, and in a 1-to-1 relationship between surrogate and clinical outcomes	X		X	
Methods for assessing transferability/generalizability of economic analyses				
Develop and standardize techniques and processes for structuring complex models in the setting of HTA that are accessible to decision-makers				
MODEL DATA				
Use of non-randomized trial data for assessing treatment effectiveness, including bias assessment, bias modeling/corrections, and selection modeling	X	X	X	X
Multiparameter evidence synthesis, particularly for parameters not related to treatment effectiveness (NB: this includes indirect/network meta-analysis)	X	X	X	X
Parameterizing models for probabilistic sensitivity analysis, particularly when data for a parameter are sparse				
Formalizing methods for interpreting non-probabilistic sensitivity analysis				
Conceptualizing the search process for parameters other than treatment effectiveness; developing practical methods for searching, including standardized searches				
Use of multi-level modeling of primary or secondary (aggregate) data, with a particular focus on using data from multi-center/multi-region studies				
Assessing willingness to pay and developing conversion factors (for health outcomes)				
MODEL CONSISTENCY				
Determine optimal methods for model validation and calibration	X	X	X	X
Assessment of structural uncertainty	X		X	
Error research (research on methods for preventing and identifying errors in the modeling process)		X		
Methods for automated model checking (for structural errors as well as logical/numerical errors)				
Methods for assessing indirect costs (e.g. for individuals outside the labor force)				
RESULTS REPORTING AND INTERPRETATION				
Implementation research on the application, use, and impact of modeling (are decisions/outcomes improved)	X	X	X	X
Work on developing flexible and comprehensive systems for evaluating completed economic analyses				

Future Research Needs	Prioritized Future Research Needs			
	All stakeholders	Policy-makers	Principal investigators	Providers
Relationship between uncertainty, inference, and decision making?				X
Determining threshold values for the incremental cost-effectiveness ratio				

Future research needs prioritized by all stakeholders are in bold; those prioritized by only policymakers, principal investigators, or providers are in italics. Within categories, future research needs are listed in the order of prioritization by all stakeholders.

Table 6. List of future research needs derived from stakeholders

Future Research Needs	Prioritized Future Research Needs			
	All stakeholders	Policymakers	Principal investigators	Providers
MODEL STRUCTURE				
Review of standards for best practice in fields outside medicine (e.g., engineering, operations research, environmental modeling, etc.)	X		X	
Research on using models in decision aids for shared decision-making				X
Review of standards for best practices in the development of decision analysis and simulation models for patient-centered comparative effectiveness questions		X		
Research on the use of duplicate modeling (building independent models with common inputs) to explore structural uncertainty				
Research on individualizing models – predictive value for individual patients				X
How to build decision models to illustrate trade-offs in patient-centered outcomes				
Developing multi-purpose/multi-disease models; research on "reusable models", models that can be repurposed to be used for different decision problems that fit under the same model structure		X		
Identifying cases where relatively simple models are "good enough" for guiding decisions – for example, exploring cases where simple and more elaborate models agree or disagree, to identify patterns				
Research on the choice of "appropriate" modeling approaches for different decisionmakers (considering policy vs. patient-level decisions and the trade-off between complexity and transparency)				X
Review of standard for best practice within more specific topics (e.g., specific model types)				

Future Research Needs	Prioritized Future Research Needs			
	All stakeholders	**Policymakers**	**Principal investigators**	**Providers**
Computational complexity for some modeling is too high – and there is computer science or applied math approaches that could be explored			X	
Research on multi-level modeling for populating models, with particular focus on using multi-level models to reflect variability in patterns of care				
Framework for deducing the sufficient complexity of a model for a given question				
When to model, what to model, how to model				
Framework for pushing the use of conceptual modeling as the first step -- to help understand what's important and what is not				
MODEL DATA				
Modelers often "take data as they are" and plug them in; however we can break down the variability in the data into sampling error, bias, and heterogeneity; we need better understanding of the role of bias, and of how to handle it	X		X	X
Emphasis on "not is it cost-effective" – but for whom is it cost-effective? In most cases this will involve the use of individual participant data			X	
How to account for distributional justice/equity/utility tradeoffs (benefits for some are not accrued by others)				
Research on appropriate measures of economic value (i.e., without focusing exclusively on willingness-to-pay)				
What is a structural sensitivity analysis in one modeling approach is a parameter in another – understanding this duality is important as it is easier to handle the latter				
Methods for multidimensional utility assessment (e.g., a joint utility for treatment and outcome sequences, a joint utility for combinations of morbidities versus combining separate utilities for single morbidities)				
MODEL CONSISTENCY				
Methods for accounting for heterogeneity including baseline risk and benefit, health status, and individual patient preferences	X	X		
Development of methods to use simulation models to address questions on heterogeneity of treatment effect		X		
Assess quality and applicability of models				
Performing cross model comparison and selecting a model				

Future Research Needs	Prioritized Future Research Needs			
	All stakeholders	**Policymakers**	**Principal investigators**	**Providers**
RESULTS REPORTING AND INTERPRETATION				
Optimal methods of communication of models to end-users; additional education needs for communication; how models are used and communicated; Widening audience for using models – research on how to communicate results to different audiences	X	X	X	X
OTHER				
Methods to engage and tailor methods and objectives to end users				X
Methods for using value of information to choose among a broad range of alternative study designs for different interventions				

Future research needs prioritized by all stakeholders are in bold; those prioritized by only policymakers, principal investigators, or providers are in italics. Within categories, future research needs are listed in the order of prioritization by all stakeholders.

Stakeholders prioritized three future research needs regarding model consistency. The highest priority among these is to determine which methods for model validation and calibration[a] are most appropriate, most improve the validity and applicability of models, and which methods are most likely to be feasible for use. In addition, which methods should be used to examine the impact of alternative model structures and when this should be done.

Stakeholders also prioritized research into methods for accounting for heterogeneity within models, including baseline risk and benefit (or treatment heterogeneity), health status, and individual patient preferences. The stakeholders' logic, particularly policymakers, was that, in general, the most useful models are those that can be individualized for particular patients.

Implementation research on the application, use, and impact of modeling was identified as a future research priority for results reporting and. The goal of this research would be to determine how models can be framed and presented to maximize their value for real-world decisionmaking. Stakeholders also prioritized research into optimal methods to communicate models to end-users, including education of end-users, how models are presented, and how to fulfill the needs of different audiences.

It is worth noting that policymakers and providers, in particular, prioritized several different future research needs than the stakeholder panel as a whole, in line with their particular perspectives. Policymakers prioritized future research needs that address ensuring that models are accurate, patient-centered, re-usable, and address heterogeneity. Specifically, these included research on methods for preventing and identifying errors in the modeling process, best practices for developing models for patient-centered comparative effectiveness questions, developing multi-purpose and multi-disease models that can be repurposed, and methods to use simulation models to address questions on heterogeneity of treatment effects.

Providers' highest priority future research needs generally revolved around how to individualize models for patients and how to make them most useful in clinical practice. Specifically, these included research into the relationship between model results and decision making – on the relationship between uncertainty, inference, and decision making; how to use models in decision aids that are used for shared decision-making; how to individualize models and to provide predictive value for individual patients; and how to engage and tailor the model methods and objectives to end-users.

Principal investigators, in contrast, tended to prioritize future research needs regarding the mechanics of developing models. In addition, to the future research needs described above on assessment of structural uncertainty and best practice standards from outside of medicine, principal investigators also prioritized research into how to handle models for which the computational complexity is too high to develop and how to incorporate individual participant data into models to determine for which people are interventions cost-effective.

Conclusions

Based on a systematic review of 71 publications providing recommendations for the conduct and reporting of modeling and simulation studies, we summarized a list of future research needs and

[a] Validation and calibration are methods to test how the models comport with reality as measured by empirical data. As such they can provide information that enhances model credibility and acceptability as well as provide insights into the potential use of modeling decisions in practical settings.[111]

presented these statements to a diverse stakeholder panel with expertise in modeling and simulation, systematic review and evidence-based medicine, and epidemiology and biostatistics. In addition, we solicited additional future research needs proposed by these stakeholders, who prioritized both lists of future research needs. The stakeholders prioritized future research needs with the goals of improving the methodology for the conduct of modeling, the validity of the models, and the communication of their findings. The future research needs were also prioritized primarily from the stakeholder perspectives it was believed most commonly directly use health care models, namely, policymakers, principal investigators and physicians. Other stakeholders (e.g., patients) were considered, but were thought to have a lesser direct impact from research into improving modeling methodology. The broad range of stakeholders included in this exercise highlighted a difference in priorities between policymakers and providers. Policymakers prioritized future research needs that address issues related to expanding the applicability of the models, in particular ensuring that models are accurate, patient-centered, re-usable, and address heterogeneity. Providers' highest priority future research needs generally revolved around interactions between providers and their patients, namely how to individualize models for patients and how to make them most useful in clinical practice.

Overall, the stakeholders prioritized a number of issues.

First, perhaps in recognition of the imminent "big data" revolution in health care, our stakeholders acknowledged the need to assess the appropriate use of non-randomized trial data for determining treatment effectiveness, including the potential for bias assessment, bias modeling/corrections, and selection modeling.[112] For example, the Food and Drug Administration's Sentinel Initiative has led to its Mini-Sentinel Initiative and Observational Medical Outcomes Partnership which have now blossomed into the Reagan-Udall Foundation (RUF),[113] an independent nonprofit public-private collaboration to generate post-marketing evidence from huge heterogeneous data sources of the use of FDA-regulated drugs, devices and procedures in the real-world so that the health care community can reliably identify harms and opportunities to improve patient care. Moreover, across health technology assessments agencies,[86,114] recent reviews concerning the use of observational versus randomized trial data have identified conflicting recommendations with one agency preferring observational data.

Second, much work is needed to better understand how data from different sources—including randomized trials, other trials, database analyses, observational studies, epidemiological data—should be used and assessed in a given model and, possibly, adjusted for risk of bias. Current guidance on how to handle data from multiple sources relies on transparency, researcher judgment, and assessment of uncertainty.[5] However, modelers would benefit from better evidence on how and whether to use, assess, and adjust for potentially biased data from multiple sources. The other related future research need considered multiparameter evidence synthesis, particularly for parameters not related to treatment effectiveness. Of note, multiparameter evidence synthesis includes indirect/network meta-analysis, an analysis tool for which there is a need for guidance about its use in models. Using Bayesian methods and Markov Chain Monte Carlo software, multiparameter evidence synthesis synthesizes a broad range of alternative evidence sources, but can also examine the consistency of the evidence provided by these multiple information sources.[115] Stakeholders discussed the need for "chains of evidence" reasoning to piece together disparate pieces of evidence, such as evidence on intermediate and terminal outcomes, or evidence from different study designs. Having quantified the uncertainty in the underlying evidence base, such types of analyses can also be used to prioritize future research by examining the impact of reducing uncertainty. Although examples of multi-

parameter evidence synthesis exist, interest has grown as demonstrated by a seven-part tutorial in Medical Decision Making.[116-122]

Third, regarding model consistency—specifically validation and verification methods—to improve model acceptability, models need to be validated and calibrated to ensure the credibility of modeling results.[11,123] Multiple methods are available for both validation and calibration, but there is limited evidence comparing specific methods.[4] Although CISNET colon cancer models have been systematically compared,[109] studies are needed specifically to address which validation or calibration methods and approaches are most appropriate for alternative types of models for different diseases. Calibration methods vary in their time and resource requirements; thus, the most appropriate method may not be the "best" method in all circumstances. Lastly, increasing scientific journals have called for reproducible research as a foundation for scientific evidence.[11,124,125] In modeling this would consist of cross model (between model) validation where independently produced models yield similar results (convergent validity).[126]

High profile journals such as Science,[127] have called for shining light into computational science, i.e., black boxes. As articulated by Weinstein, "Decision makers will not readily accept results and cost-effectiveness. unless they can understand them intuitively and explain them to others in relatively simple terms."[126] Consistent with these trends, the stakeholder identified optimal methods of communication of models to end-users; additional education needs for communication; how models are used and communicated; widening audience for using models – research on how to communicate results to different audiences as a future research need.

Lastly, in an upcoming era of value-based payment for outcome, stakeholder prioritized the need to explore implementation research on the application, use, and impact of modeling (are decisions/outcomes improved). Models are commonly accepted in decision-making in such fields as environmental protection, weather prediction, and defense strategy, but less so in health care.[128] Better use of up-to-date methods used in these and other fields could only improve medical models.

In conclusion, this systematic review and expert panel provides a comprehensive collation of methodological guidance developed through various methodological processes and identifies ways to improve upon and standardize the use of modeling and simulation in the context of systematic reviews. This review updates previous syntheses of evidence- and consensus-based guidance on the conduct of modeling and simulation studies and the stakeholder panel prioritizes future research topics needed to advance the current state-of-the-art in modeling and simulation.

Chapter 4. A Review of Validation and Calibration Methods for Health Care Modeling and Simulation

Introduction

In practice, models of at least moderate complexity will be 'solved' with computer-based numerical analysis and simulation. Because these computer models are used to inform predictions or decisions in the real world, assessing their credibility (trustworthiness) is paramount. A recent ISPOR-SMDM Good Research Practices Task Force identified model validation as one of the two determinants of confidence in models (the other being transparency).[11] Further, because some aspects of reality are unmeasured or unknowable, models will often require inputs for which no or only indirect data exist. In such cases, model calibration can be used to select input values that lead to model outputs "as close as possible" to available empirical data. Because model validation and calibration entail a "confrontation of models with data" they can inform judgments about the credibility of models, and can guide the use of modeling results in practical settings.

Our systematic review of evidence- and consensus-based guidance on the conduct and reporting of modeling and simulation studies (Chapter 1) identified model calibration and validation as a major methodological research area. This was triangulated by a panel of multiple stakeholders, including developers and users of health care models, who also identified aspects of model evaluation, and model validation and calibration in particular, as important targets for future methodological research. There is limited previous work surveying calibration and validation methods used for health care models and most existing reviews have either focused on a limited topic area (e.g., treatment of cardiovascular disease, cancer natural history) or modeling methodology (e.g., micro-simulation models).

Based on the above, we conducted a project aiming to provide a unifying overview of validation and calibration methods, and a survey of studies comparing validation and calibration methods used in health care modeling and simulation studies.

Methods

Sources of Information and Review Methods

Issues pertaining to model evaluation and assessment arise in many methodological disciplines (e.g., mathematics and statistics, economics, theory of simulation, operations research), as well as many topic areas (e.g., health care, disease modeling, biology, environmental science, mechanical engineering, material science). The relevant literature is vast, poorly categorized in standard literature databases (i.e., specific search terms are lacking), and published as journal papers, conference proceedings, books, and technical reports that are not always easy to identify, obtain, and comprehend. Thus, a comprehensive systematic review of all relevant methodological papers was deemed infeasible. Instead, we relied on a mixed approach that combined consultation with expert methodologists; hand-searching the reference lists of related papers, technical reports, and books; review of our personal reference collections; and a systematic review of studies comparing validation and calibration methods for disease modeling-related or health care-related models.

The systematic component of our literature review covered four electronic databases (MEDLINE, Cochrane Methodology Register, Health Technology Assessment Database, NHS Economic Evaluation Database), through June 3, 2013, for articles presenting validation and calibration methods in reports of modeling and simulation studies. We also rescreened the citations retrieved by the search strategy of our recently completed systematic review of recommendations for the conduct and reporting of modeling and simulation studies. The final search strategy, with the list of 37 included journals, is presented in **Appendix B**.

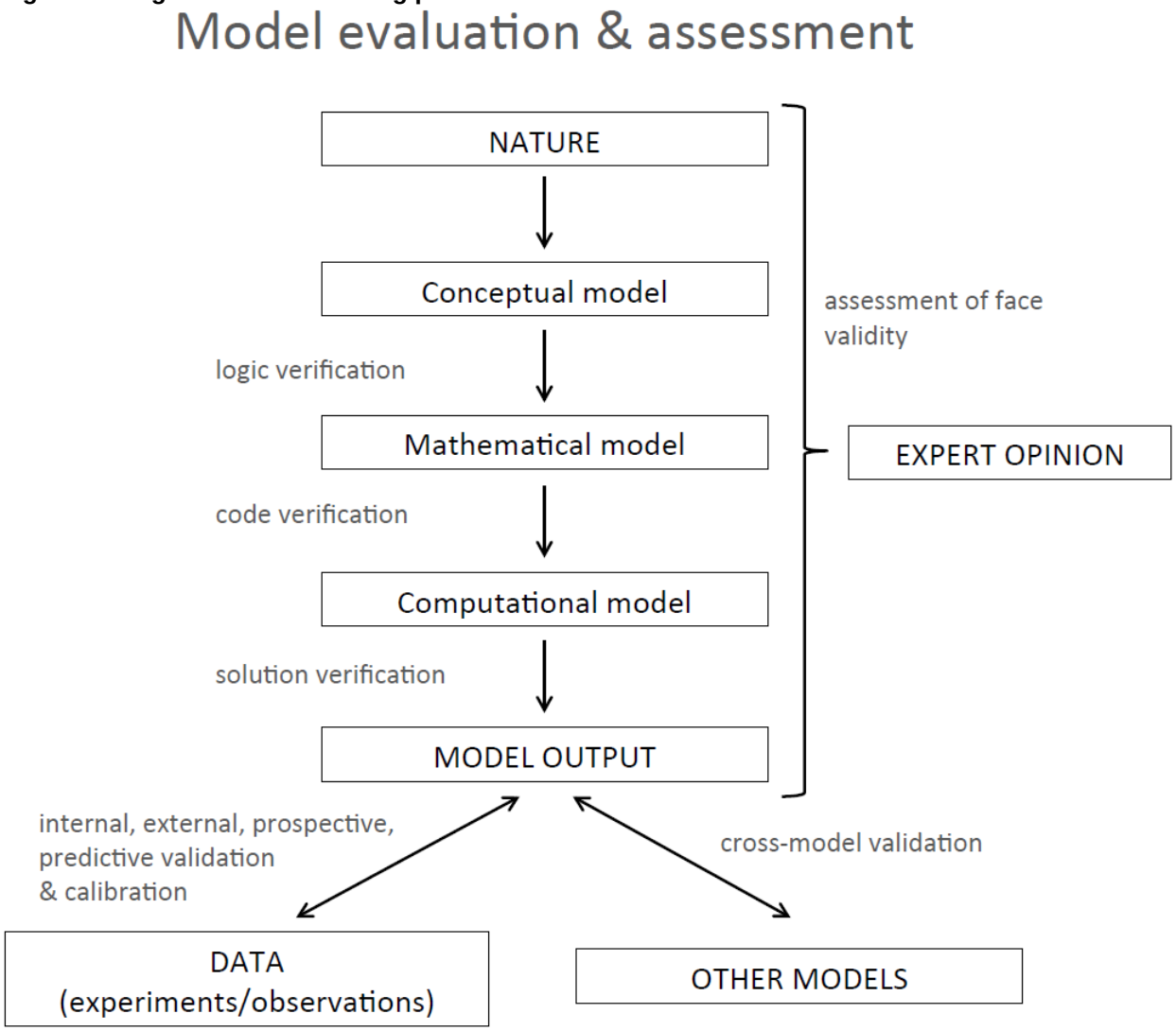
Six reviewers independently screened 6825 abstracts in duplicate and resolved disagreements by group consensus. Eligible studies had to compare or apply at least two methods related to model validation, model calibration, or goodness of fit, in the context of modeling or simulation. We excluded studies that applied only a single method of validation or calibration. Three reviewers extracted descriptive information from included articles. Completed data extraction forms were verified by a second reviewer and were discussed during group meetings. Extracted data included population characteristics, outcomes, basic model description, and methods and results related to model validation and calibration.

Definitions and Preliminaries

By 'modeling' we mean a multistep iterative process to conceptualize, by abstraction and idealization, a representation of salient aspects of reality (conceptual model), specify it mathematically (mathematical model) and implement it in computer code (computational model) so that it can be 'solved'. **Figure 3** outlines the modeling process. Because the modeled natural phenomena can be complex and the model implementation process is often intricate, it is important to perform checks throughout. Terminology about these checks varies across fields and topic areas, but the underlying concepts are similar.

In this work we use the terms 'verification,' 'assessment of face-validity,' 'validation,' and 'calibration' to describe various checks. ***Verification*** refers to assessing 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). ***Face-validity*** refers to whether the model is deemed a satisfactory representation of the salient aspects of reality and whether the model results appear to be plausible. We use the term ***calibration*** to describe the process of determining the values of parameters so that model outputs match (i.e., "fit") observed empirical data. We define ***validation*** to be the comparison of model outputs with expert judgment, observed data, or other models, without any attempt to modify model parameters to improve fit. A more detailed description of types of model validation is provided in the next section.

Figure 3. Diagram of the modeling process



Overview of Validation and Calibration Methods

Overview of Model Validation

Validation is the assessment of the "congruence" between model predictions and actual observed data, or the results of other models addressing the same (or similar) research question, or expert predictions of what the results should be. The literature identifies several aspects of model validation, including face validity, verification and internal validation, cross-model comparisons, external validation, and prospective and predictive validation.[11,50,129,130] To a large extent, the descriptions provided below follow those adopted by the recent ISPOR-SMDM Modeling Good Research Practices Task Force.[11]

- *Face validity* ("first order" validation) refers to the determination, by a suitable group of experts, that the model reflects the current understanding of the science and available evidence. Expert review should cover all aspects of the modeling process, including the

question formulation, model structure, model data, and the model output. Evaluation of aspects of modeling other than the model output, are best performed "blinded" to the model results to reduce the possibility of biased assessment.

- *Verification and internal validation* ("second order" validation) refers to an assessment of whether the model has been implemented correctly and behaves as expected. Verification can pertain to the computer code ("code verification") and the solutions that it produces in well-understood problems with known solutions ("solution verification").[129] Some consider the comparison of model outputs with the data used to populate the model to be a component of internal validity. However, when using a study to estimate model parameters only a small part of the study results are used; in which case other data can be used for validation. For example, if a randomized trial is used to inform a model regarding the relationship between treatment (e.g., statin vs. no statin) and a surrogate outcome (e.g., LDL cholesterol at 6 months), the trial data on other outcomes (e.g., mortality, morbidity) can be used for model validation.[11]
- *External validation* ("third order" validation): In external validation, the model outputs are compared with empirical observations that were not used in model development. As noted above, external validation is sometimes taken to mean the comparison of model outputs against observations in datasets that are disjoint from those used in model development. In general, external validation using disjoint data provides a more stringent assessment of model performance.
- *Prospective and predictive validation* ("fourth order" validation) assess the model's ability to reproduce ("predict") empirical results that were not available and were not used during its development.[130] Prospective validation refers to the use of data that accrues over additional followup in studies that were used for model development. Predictive validation refers to the use of data from independent studies that were unavailable at the time of model development.
- *Cross-model validation* involves a comparison of results among different models for the same (or sufficiently similar) analyses.[11] Such comparisons can increase the credibility of the models and provide methodological insights. Cross-model comparisons have been used extensively in cancer simulation models supported by the Cancer Intervention and Surveillance Modeling Network (CISNET).[131]

Two general types of methods are in common use for internal, external, and prospective or predictive validation: 'informal' methods using graphical and tabular presentations of model results (e.g., time series and scatter plots, cumulative frequency distributions) and 'formal' (statistical) using a distance function or goodness-of fit metric.[132] The value of graphical and tabular data displays cannot be overemphasized. However, these methods may not always have adequate sensitivity for detecting poor fit to the data and are operator dependent. For this reason, graphical methods are usually combined with statistical methods. The latter rely on assessments of goodness-of-fit that quantify the discrepancy between observed data and model outputs;[133,134] in many ways these quantitative assessments are similar to standard model fit criteria used in statistics.[135] When a Bayesian approach is adopted, posterior predictive checks (i.e., comparisons between the observed data and the model's posterior predictive distribution) can be used to assess model fit.[136,137] For example, Gardner (2011) proposed and studied alternative model fit statistics for individual-level infectious disease models, based on posterior predictive checks.[138]

It is not possible to identify a universally preferred method for statistical model validation.[129] However, some general principles to guide the choice are identified in the literature: first,

statistical validation should be applied to quantities of interest that are relevant to the model scope and the perspective of the analysis. Second, statistical criteria for model fit need to be appropriate for the mathematical structure of the model.[129] For example, if there is dependence (clustering) in the statistical model (e.g., repeat measurements within individuals or groups) the statistical criteria should account for that dependence. Third, because model fit is generally improved by increasing the number of model parameters (which may lead to over-fitting the data used for model development and limit generalizability), criteria that "account for" the number of model parameters may be preferable. Fourth, any statistical method for model validation should take into account uncertainty in both the empirical data and the model outputs.

More generally, it is not possible to establish model validity in the affirmative; i.e., there is no criterion that, if met, establishes a model as generally valid. In fact, some experts suggest that it is only possible to demonstrate model invalidity in a specific setting and for a specific purpose (e.g., by showing that model predictions do not fit a set of observations that is relevant to the anticipated model use).[11]

Empirical Studies of Model Validation

Our searches for studies using alternative model assessment methods identified several studies that applied various approaches to assess model fit.[130,138-145] This list is obviously not exhaustive, in that it is limited to health care applications and is based on a systematic, but non-comprehensive process for identifying relevant studies. We found that validation methods were specific to the research question that was addressed by each study and the investigators' choice of methods for model implementation. For these reasons, we did not use them to draw general conclusions.

Overview of Model Calibration

Calibration involves the optimization of a subset of model parameters to improve the fit of model predictions to empirical data. Traditionally, calibration is distinguished from other estimation tasks by its use to obtain estimates for parameters for which only indirect data is available. For example, in microsimulation models of cancer in humans, parameters for which direct empirical data are unavailable (e.g., growth rates of preclinical cancers) are determined by selecting input rate values to produce model outputs (e.g., cancer incidence rates, which are functionally dependent on the parameters for which direct data are unavailable) that "are as close as possible" to empirical data.

Calibration efforts are tailored to the specific needs of a particular model. Calibration is fundamentally an optimization (estimation) problem. To specify a *calibration problem,* one has to define the following components:

- *Calibration parameters* are the typically weakly identifiable (e.g., supported by indirect evidence) model parameters that are subjected to calibration.[43] A most important issue is whether the *feasible domain of the calibration parameters is convex* or not. This is because convex calibration problems (problems where the parameter domain and the objective functions are convex) can be solved easier than non-convex ones.
- *Calibration targets* are the data against which the model output is compared; calibration aims to select parameter values that produce model outputs that are "close" to the calibration targets (while "close" may be assessed graphically or visually, it is preferably encoded quantitatively by an objective function). The choice of calibration targets

depends on the model quantities of interest, the availability of "high quality" data, and the goals of modeling. For example, when calibrating a model, the calibration targets should be derived from data relevant to the decisional context, and obtained from well-designed and conducted studies of populations "similar" to those who will be affected by the decision.

- *Objective functions* are typically scalar functions of the calibration parameters used to assess "closeness" quantitatively. Common choices include a *distance* of the calibration target data from model outputs, a convexity-preserving transformation of a distance, or a *likelihood* or *pseudo-likelihood*.[133,134,146-148] Example distances are the sum of absolute or squared differences between model outputs and calibration targets (L1- and L2-norms, respectively). Examples of convexity-preserving transformations of distances are various chi-squared statistics. Distances are convex objective functions, and for many problems the likelihood and pseudo-likelihood is convex as well. As mentioned above, convexity of the objective function is an important property, because convex problems are much easier to solve than non-convex ones.

The goal is to solve (optimize) the calibration problem, that is, identify the feasible values of the calibration parameters that optimize the objective function. Solving the calibration problem entails defining the following:

- *Algorithm for optimizing* the objective function. These algorithms search for values of calibration parameters in the feasible domain that optimize the objective function. Principled searching uses mathematical programming to obtain values. Descriptions of *ad hoc* approaches, such as 'manual'-tuning, however, also occur in the health care literature.

- *Acceptance criteria* are used to determine whether an algorithm has converged to a solution. Typically, this means that further iterations do not change the value of the objective function and the estimated values of the calibration parameters beyond prespecified tolerances (strict tolerances can be of the order of machine precision).

- The *stopping rule* is the criterion for terminating the calibration process. Usually, calibration is stopped when the acceptance criteria are satisfied, the search space is exhausted (e.g., all points in a grid search have been evaluated), or a predetermined maximum number of iterations has been reached.

As mentioned already, the solution of the calibration problem (identifying the global optimum of the objective function within the feasible domain of the calibration parameters) depends greatly on the objective function and the shape of the feasible domain of the calibration targets. When the problem is convex (both the objective function and the feasible domain are convex) or can be restated to be convex, a single optimum exists and the mathematical programming methods to find it are very robust, representing a readily-usable technology.[149] Problems that are not convex—or cannot be recast as convex—have local optima, and demand global optimization approaches. For such problems, exact solutions often become computationally expensive, and only approximate solutions may be practical.

Based on the above, we make the following general observations, which are important for interpreting the empirical research found in the health care literature:

- Judging which specification of the calibration problem (in particular, which objective function) is most appropriate is not answerable from data alone. This is a general statement for optimization problems. The choice of the objective function should reflect the decisionmakers' perspective, and the nature of the problem.[150] Thus comparisons of solutions to alternative objective functions are difficult to interpret.
- From a theoretical basis, solutions to different specifications of the calibration problem clearly need not be identical. We interpret empirical demonstrations of this phenomenon in specific applications as stability analyses.
- Once a calibration problem has been specified, it is straightforward to rank the different optimization algorithms according to their performance, by ordering (within machine precision) the value that the (scalar) objective function has with each algorithm's solution. To learn from such comparisons one must (a) be confident that the algorithms were implemented correctly and efficiently; and (b) be able to characterize salient mathematical attributes of the calibration problem, e.g., whether common regularity conditions were met.[149,151,152] Because this requires a very deep understanding of the problem at hand and of the mechanics of the various algorithms, it is generally not possible to draw generalizable conclusions.

Empirical Studies of Model Calibration

Our search for studies comparing alternative calibration methods for health care models identified four relevant studies (**Table 7**). They include comparisons between different specifications of the calibration problem solved with the same algorithm, or with different algorithms; and comparisons between alternative algorithms for the same calibration problem, that highlighted various mathematical aspects. Further, each empirical study was limited to a single modeled process (and the same model was examined in 2 of the 4 studies). Many reported results (e.g., running time, performance, etc.) are expected to be dependent on the specific computer implementation. Thus, we deem that it is not possible to draw general conclusions from these studies.

Methodological Appraisals of Validation and Calibration Methods

Several systematic methodological appraisals of health care models and simulations provide information on the validation and calibration methods used in practice. These studies have found that more than half of all modeling studies do not report any use of calibration or validation methods.[43,153-164] When some aspect of model validation or calibration is mentioned, reporting is often incomplete. Below, we review the results of methodological appraisals that provided a more in depth assessment of validation and calibration methods.

Cancer

Stout (2009) reviewed 131 studies of cancer microsimulation models that could have used calibration methods to determine input values for unobservable parameters.[153] Approximately 50 percent of the studies (n=66) referred to "calibration" or "model fitting" and an additional 16 percent (n=21) provided references to methodological publications on model calibration. Nearly all studies (95%, n=83 of 87) identified the calibration targets they used, 54 percent (n=47) reported information on the goodness-of-fit metric used. Information on the search algorithms used was not well described. The authors used the results of this investigation to derive a 7-item

"*Calibration Reporting Checklist*," which may be useful as a reminder of issues to consider when evaluating calibration methods in modeling studies.

Table 7. Studies comparing alternative calibration methods applied to the same problem

Author, year	Area of application	Model structure	Methods compared	Study findings
Kong, 2009[165,166]	Lung cancer development, progression, detection, treatment, and survival (Lung Cancer Policy Model)	Agent-based; state transition model; 1 month cycle length	• Search algorithms (simulated annealing vs. genetic algorithm)	• Both search algorithms attained study-determined threshold GOF scores within 1000 search iterations • Simulated annealing outperformed the genetic algorithm • The model predictions after calibrations matched other mathematical models of cancer development
Taylor, 2010[167,168]	Cervical cancer epidemiology, natural history, and effectiveness of vaccination	Cohort-based; 6-state Markov model; 6-month cycle length; lifetime horizon; implemented in Excel with Visual Basic for Applications.	• Search algorithms ('manual' calibration vs. random search of parameter domain vs. Nelder-Mead)	• The Nelder-Mead algorithm and manual calibration achieved the best fit (weighted mean percent deviations of 7% and 10%, respectively); random search performed poorly (weighted mean percent deviation of 39%) • Use of the Nelder-Mead algorithm required less analyst time but was more computationally demanding, compared with manual calibration.
Karnon, 2011[169,170]	Choice of adjuvant chemotherapy for early breast cancer	Cohort-based; Markov model; 1 year cycle length; 50-year time horizon; implemented in Excel with an add-on component (Microsoft Excel Solver; Frontline Systems).	• GOF metrics (chi-square vs. likelihood) • Search algorithms (random vs. gradient-based guided search) • Alternative convergence criteria (narrow vs. broad)	• The chi-square GOF metric "differentiated between the accuracy of different parameter sets" to a greater than the log-likelihood statistics • The guided search strategy produced results of higher accuracy and greater precision than random search • Broader convergence criteria produced less accurate results that were closer to the non-calibrated results
Taylor, 2012[167,171]	Cervical cancer epidemiology, natural history, and effectiveness of vaccination	Cohort-based; 6-state Markov model; 6-month cycle length; lifetime horizon	• Alternative starting values for the Nelder-Mead search algorithm (5, randomly chosen) • GOF metrics [weighted MPD with weights for the cancer incidence and mortality parameters that were 6- and 3-fold larger that those of corresponding carcinoma in situ endpoints (1-6-3 weights) vs. MPD with 1-3-3 weights vs. MSPD with 1-3-3 weights vs. MSPD with 1-3-6 weights vs. ML]	• The sensitivity/stability analyses to the choice of initial values and alternative weighting schemes revealed a substantial amount of uncertainty in the model output – far greater than that revealed by forward propagation of uncertainty

GOF = goodness-of-fit; ML = maximum likelihood; MPD = mean percentage deviation; MSPD = mean squared percentage deviation.

Cardiovascular Disease

Haji Ali Afzali (2013) reviewed 81 model-based studies (including cohort and agent-based models) for cardiovascular disease.[160] They found that 73 percent (59 studies) reported some element of model evaluation, but only 6 percent (5 studies) reported a calibration process. Usually multiple calibration targets were employed in each study but only a single study provided information on the goodness-of-fit metric and no studies reported information on the acceptance criteria. Search algorithms were generally not well documented.

Unal (2006) reviewed the methodology of 42 coronary heart disease models (reported in 75 publications).[163] In general validation and calibration methods were not used systematically and were not reported in detail. Six of the 42 models were considered "principal coronary heart disease models" – of these, two reported some calibration procedure and only one reported the performance of model validation.

Neurological Disease

Siebert (2004) reviewed 8 studies using mathematical models to evaluate treatments for Parkinson's disease.[164] None of the eight studies reported any internal or external validation of their models. Dams (2011) surveyed 11 cost-effectiveness studies for Parkinson's disease including therapeutic and diagnostic evaluations.[161] They found that only four models reported performing some form of model validation and none provided adequate details of their validation methods and results.

Respiratory Disease & Smoking Cessation

Ferdinands (2008) reviewed 13 disease simulation models of asthma or chronic obstructive pulmonary disease (11 state-transition models and 2 dynamic population models).[162] Only two studies provided information on code and solution verification; seven studies reported comparisons of model outputs with data used to develop the model; seven studies reported results of external validation; and no studies reported performing predictive validation or plans to undertake such efforts.

Bolin (2012) assessed 78 economic evaluations of smoking cessation therapies,[157] 30 of which were considered "highly relevant" (defined as studies applying "intertemporal modeling with a time horizon" of at least 20 years). They found that "several studies"[b] used simulation models – that were not described as previously validated – without performing any model validation.

Calibration as Estimation

As outlined above, calibration is very similar to statistical estimation.[172,173] Both processes have the same goal, namely to find input values that lead to the best possible model fit. For example, if the objective function of the calibration is a likelihood function, calibration is—by most any definition—equivalent to maximum likelihood estimation. The conceptualization of calibration as estimation is helpful when assessing the identifiability of the parameters of the mathematical model and examining the consistency between data inputs.

[b] An exact count was not provided in the main text of the paper and the supplementary appendix was not downloadable from the journal Web site.

As an example, we consider simulation models that use meta-analysis to inform some of their inputs. First, note that the empirical data inputs for the model comprise two potential approaches and two types of data (1) *meta-analysis-estimation* of input data to estimate some model parameters, and (2) *calibration-estimation* in which calibration targets are used to estimate the remaining model parameters. Modelers have two options: do the meta-analysis-estimation and the calibration-estimation separately (as two steps; most common practice) or jointly (in one step; least common). We make the following observations about one-step versus two-step procedures:

- Compared with one-step estimation, *two-step estimation is generally less efficient (in the statistical sense) and does not guarantee that the best-fitting values for parameters will be identified*. It can also hinder the complete characterization of parameter uncertainty and the representation of correlations between data sources or dependencies among model parameters. The one-step method is consistent with the scientific maxim of using all available evidence when making decisions. Further, it may help avoid under-assessments of uncertainty. The one-step approach is closely related to methods for synthesizing evidence from diverse sources, including multi-parameter and generalized evidence synthesis,[115,174] the confidence profile method,[175-178] cross-design synthesis,[179-181] and teleo-analysis.[182]
- *One-step estimation allows for formal tests of consistency of parameter estimates obtained by different sources of evidence.* One-step approaches enable an assessment of whether the various data sources 'square up'. If the data are inconsistent (do not 'square up'), a serious problem exists that requires resolution (discussion of possible methods for resolving inconsistencies is beyond the scope of the current work).[85,183] If the data are consistent, the one-step approach maximizes use of all of the available information.
- *One-step estimation allows one to use well-established quantitative methods for comparing differences between model outputs and empirical data while using all available data*.[184] Examples of such methods include posterior predictive checks, posterior mean deviance statistics, and various model cross-validation approaches.
- Under some circumstances, which can be formalized, the one-step approach and the two-step approach (as described above) are mathematically equivalent.[c]

Parameter Identifiability

The ability of Bayesian methods to incorporate external information or subjective beliefs, in the form of informative prior distributions, is particularly appealing when some model parameters are unidentifiable. For example, Rutter (2009) used a Bayesian approach to calibrate a microsimulation model of colorectal cancer natural history.[185] Briefly, a model of colorectal cancer natural history was programmed and prior distributions were specified for all model parameters. Markov Chain Monte Carlo (MCMC) methods were used to estimate model parameters using data from multiple sources. For parameters that were unidentifiable using available data, informative prior distributions were specified; these distributions appropriately accounted for parameter uncertainty (as opposed to fixing the parameters to arbitrary values).

[c] For example, this is true when the objective function is differentiable and the gradient of the objective function with respect to the calibration parameters is not a function of the remaining (other) parameters in the mathematical model, and the gradient of the objective function with respect to the other parameters is not a function of the calibration parameters.

The finite sample size performance of the proposed methodology was assessed in a simulation study, which demonstrated that the proposed method was an unbiased estimator for parameters for which data were available.

Nonetheless, jointly performing calibration and estimation of model parameters does not eliminate problems of identifiability: model parameters for which there is only limited (e.g., indirect) or no information are effectively unidentifiable.[186] Their posterior distribution is determined by the prior distribution chosen for them. In addition, in complex models, identifiability is hard to assess by just examining the model equations or inspecting the posterior distributions it produces. Instead, quantitative assessment is necessary. In the above-mentioned colorectal cancer microsimulation study,[185] informative prior distributions were specified for unidentifiable model parameters and the model diagnostics proposed by Garrett & Zeger (2000) were used to assess identifiability via overlap statistics.[187] The utility of this approach was also demonstrated in the aforementioned simulation study.[185]

Examples of Calibration as Estimation

In addition to Rutter (2009),[185] other examples of using Bayesian methods for model calibration, validation, and parameter estimation exist, both for health care and non-health care models and simulations. These studies vary in their complexity, the number of data sources and the amount of information available for model development and evaluation.[188,189] Jackson (2013) and Whyte (2011) provide tutorials on using Bayesian evidence synthesis methods and provide code and data to reproduce the analyses.[188,189]

Conclusions

This chapter provides an overview of the state-of-the science on model validation and calibration for health care models. It appears that in health care, methodological research on the calibration and validation of models has been limited to case-studies applying a small number of alternative approaches to a small number of models. Because such case studies produce results that are applicable to these particular models, and address only a small part of the complex and multifaceted methodological decisions that modelers make, we believe that there is need for further research on validation and verification methods.

Based on our review of the literature and discussions with the stakeholders (described in *Chapter 1*), we have identified the following broad areas for future research, with a focus on areas that may be of interest to the Effective Health Care Program:

- Consideration should be given to the *development of default ("reference") models* to facilitate the use of validated models as adjuncts to systematic reviews.[190,191] Because model validation and calibration are time consuming activities and because systematic reviews need to be prepared in a timely fashion, the use of modeling in systematic reviews could be facilitated by developing and validating reference models for high-impact conditions (e.g., as has been done in CISNET).[192] Such conditions could be selected among AHRQ's priority areas, by taking into consideration the potential value of using models to supplement reviews of published evidence in each area.
- Further research is needed for the *development, validation, and calibration of complex models that incorporate evidence from multiple sources*. Systematic reviews (e.g., comparative effectiveness reviews prepared by Evidence-based Practice Centers) often retrieve evidence that is *flawed* (as indicated by risk of bias assessments), *indirect* (e.g., addressing laboratory surrogates instead of clinical outcomes), *incomplete* (e.g., with

missing data), and *conflicting* (clinically and methodologically heterogeneous). Under these conditions "global subjective assessments" of the evidence are prone to error (and bias).[178] Modeling can address these problems by synthesizing evidence in a statistically valid way and allowing a formal assessment of consistency, while making all assumptions explicit.

- Research is needed to determine "best practices" for validating and calibrating models that are intended for use across different settings and patient populations.[67,193] Such methods would rely on developing criteria for *formalizing judgments on the adequacy of the validation process* (especially external, prospective, and predictive validation).
- Given the importance of cross-model validation (especially in the absence of relevant empirical data) and the increasing availability of models addressing the similar research questions further research in needed to *explore how discrepancies among models relate to the models' potential for being prospectively and externally validated (against data).*
- Methodological work is also needed to *identify optimal methods for communicating (e.g., visualizing) the validation and calibration methods used in complex models*. Such research is necessary for presenting complex models to applied modelers and – more importantly – lay "consumers" of modeling and simulation results.

In summary, model validation and calibration are fundamental processes for establishing the credibility of models and simulations. "Confronting models with data" is an important component of establishing their validity and correct parameterization.[111] Ongoing progress in statistical, operational, and computational methods can provide modelers with an expanding toolkit for validating and calibrating models. However, current empirical research is limited to methodological appraisals or case-studies of alternative methods. Future research should advance our understanding of the theoretical basis of model evaluation and use comprehensive simulation methods to compare alternative approaches.

References

1. Kuntz K, Sainfort F, Butler M, et al. Decision and Simulation Modeling in Systematic Reviews. Methods Research Report. (Prepared by the University of Minnesota Evidence-based Practice Center under Contract No. 290-2007-10064-I.) AHRQ Publication No. 11(13)-EHC037-EF. 2013. PMID: 23534078.
2. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Econ 2006;15(4):373-81. PMID: 16389628.
3. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999 Jun;18(3):341-64. PMID: 10537899.
4. Claxton K, Ginnelly L, Sculpher M, et al. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. Health Technol Assess 2004;8:1-103. PMID: 15248937.
5. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess 2004 Sep;8(36):iii-158. PMID: 15361314.
6. Wallace BC, Trikalinos TA, Lau J, et al. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010;11:55. PMID: 20102628.
7. Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 9;276(14):1172-77. PMID: 8827972.
8. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. Pharmacoeconomics 2000;17:461-77. PMID: 10977388.
9. Siegel JE, Weinstein MC, Russell LB, et al. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276:1339-41. PMID: 8861994.
10. Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA 1996 Oct 16;276(15):1253-58. PMID: 8849754.
11. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making 2012 Sep;32(5):733-43. PMID: 22990088.
12. Concannon TW, Meissner P, Grunbaum JA, et al. A new taxonomy for stakeholder engagement in patient-centered outcomes research. J Gen Intern Med 2012 Aug;27(8):985-91. PMID: 22528615.
13. Economic analysis of health care technology. A report on principles. Task Force on Principles for Economic Analysis of Health Care Technology. Ann Intern Med 1995 Jul 1;123(1):61-70. PMID: PM:7762918.
14. Decision analytic modelling in the economic evaluation of health technologies. A consensus statement. Consensus Conference on Guidelines on Economic Modelling in Health Technology Assessment. Pharmacoeconomics 2000 May;17(5):443-44. PMID: PM:10977386.
15. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health Technol Assess 2009 Jun;13(29):iii, ix-61. PMID: 19500484.
16. Bae EY, Lee EK. Pharmacoeconomic guidelines and their implementation in the positive list system in South Korea. Value Health 2009 Nov;12 Suppl 3:S36-S41. PMID: 20586979.
17. Boulenger S, Nixon J, Drummond M, et al. Can economic evaluations be made more transferable? Eur J Health Econ 2005 Dec;6(4):334-46. PMID: 16249933.

18. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ 2006 Dec;15(12):1295-310. PMID: 16941543.

19. Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value Health 2012 Sep;15(6):796-803. PMID: 22999128.

20. Chilcott J, Tappenden P, Rawdin A, et al. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. Health Technol Assess 2010 May;14(25):iii-107. PMID: 20501062.

21. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care 2003 Jan;41(1):32-44. PMID: 12544542.

22. Cleemput I, van WP, Huybrechts M, et al. Belgian methodological guidelines for pharmacoeconomic evaluations: toward standardization of drug reimbursement requests. Value Health 2009 Jun;12(4):441-49. PMID: 19900251.

23. Clemens K, Townsend R, Luscombe F, et al. Methodological and conduct principles for pharmacoeconomic research. Pharmaceutical Research and Manufacturers of America. Pharmacoeconomics 1995 Aug;8(2):169-74. PMID: 10155611.

24. Colmenero F, Sullivan SD, Palmer JA, et al. Quality of clinical and economic evidence in dossier formulary submissions. Am J Manag Care 2007 Jul;13(7):401-07. PMID: 17620035.

25. Davalos ME, French MT, Burdick AE, et al. Economic evaluation of telemedicine: review of the literature and research guidelines for benefit-cost analysis. Telemed J E Health 2009 Dec;15(10):933-48. PMID: 19954346.

26. Detsky AS. Guidelines for economic analysis of pharmaceutical products: a draft document for Ontario and Canada. Pharmacoeconomics 1993 May;3(5):354-61. PMID: 10146886.

27. Drummond M, Brandt A, Luce B, et al. Standardizing methodologies for economic evaluation in health care. Practice, problems, and potential. Int J Technol Assess Health Care 1993;9(1):26-36. PMID: 8423113.

28. Drummond M, Sculpher M. Common methodological flaws in economic evaluations. Med Care 2005 Jul;43(7 Suppl):5-14. PMID: 16056003.

29. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. Int J Technol Assess Health Care 2005;21(2):165-71. PMID: 15921055.

30. Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value Health 2009 Jun;12(4):409-18. PMID: 19900249.

31. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996 Aug 3;313(7052):275-83. PMID: 8704542.

32. Evers S, Goossens M, de VH, et al. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess Health Care 2005;21(2):240-45. PMID: 15921065.

33. Fry RN, Avey SG, Sullivan SD. The Academy of Managed Care Pharmacy Format for Formulary Submissions: an evolving standard--a Foundation for Managed Care Pharmacy Task Force report. Value Health 2003 Sep;6(5):505-21. PMID: 14627057.

34. Garattini L, Grilli R, Scopelliti D, et al. A proposal for Italian guidelines in pharmacoeconomics The Mario Negri Institute Centre for Health Economics. Pharmacoeconomics 1995 Jan;7(1):1-6. PMID: 10155289.

35. Gartlehner G, West SL, Mansfield AJ, et al. Clinical heterogeneity in systematic reviews and health technology assessments: synthesis of guidance documents and the literature. Int J Technol Assess Health Care 2012 Jan;28(1):36-43. PMID: 22217016.

36. Glennie JL, Torrance GW, Baladi JF, et al. The revised Canadian Guidelines for the Economic Evaluation of Pharmaceuticals. Pharmacoeconomics 1999 May;15(5):459-68. PMID: 10537963.

37. Goldhaber-Fiebert JD, Stout NK, Goldie SJ. Empirically evaluating decision-analytic models. Value Health 2010 Aug;13(5):667-74. PMID: 20230547.

38. Graf von der Schulenburg JM, Greiner W, Jost F, et al. German recommendations on health economic evaluation: third and updated version of the Hanover Consensus. Value Health 2008 Jul;11(4):539-44. PMID: 18194408.

39. Grutters JP, Seferina SC, Tjan-Heijnen VC, et al. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. Value Health 2011 Jul;14(5):777-84. PMID: 21839418.

40. Hay J, Jackson J. Panel 2: methodological issues in conducting pharmacoeconomic evaluations--modeling studies. Value Health 1999 Mar;2(2):78-81. PMID: 16674337.

41. Hoomans T, Severens JL, van der RN, et al. Methodological quality of economic evaluations of new pharmaceuticals in The Netherlands. Pharmacoeconomics 2012 Mar;30(3):219-27. PMID: 22074610.

42. Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Med Decis Making 2007 Jul;27(4):491-99. PMID: 17761961.

43. Karnon J, Goyder E, Tappenden P, et al. A review and critique of modelling in prioritising and designing screening programmes. Health Technol Assess 2007 Dec;11(52):iii-xi, 1. PMID: 18031651.

44. Kolasa K, Dziomdziora M, Fajutrao L. What aspects of the health technology assessment process recommended by international health technology assessment agencies received the most attention in Poland in 2008? Int J Technol Assess Health Care 2011 Jan;27(1):84-94. PMID: 21262087.

45. Liberati A, Sheldon TA, Banta HD. EUR-ASSESS Project Subgroup report on Methodology. Methodological guidance for the conduct of health technology assessment. Int J Technol Assess Health Care 1997;13(2):186-219. PMID: 9194352.

46. Lopez-Bastida J, Oliva J, Antonanzas F, et al. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ 2010 Oct;11(5):513-20. PMID: 20405159.

47. Lovatt B. The United Kingdom guidelines for the economic evaluation of medicines. Med Care 1996 Dec;34(12 Suppl):DS179-DS181. PMID: 8969324.

48. Luce BR, Simpson K. Methods of cost-effectiveness analysis: areas of consensus and debate. Clin Ther 1995 Jan;17(1):109-25. PMID: 7758053.

49. Mason J. The generalisability of pharmacoeconomic studies. Pharmacoeconomics 1997 Jun;11(6):503-14. PMID: 10168092.

50. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. Pharmacoeconomics 2000 May;17(5):501-13. PMID: 10977390.

51. McGhan WF, Al M, Doshi JA, et al. The ISPOR Good Practices for Quality Improvement of Cost-Effectiveness Research Task Force Report. Value Health 2009 Nov;12(8):1086-99. PMID: 19744291.

52. Briggs A, Weinstein MC, Fenwick E, et al. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value Health 2012;15:range. PMID: 22999133.

53. Menon D, Schubert F, Torrance GW. Canada's new guidelines for the economic evaluation of pharmaceuticals. Med Care 1996 Dec;34(12 Suppl):DS77-DS86. PMID: 8969316.

54. Mullahy J. What you don't know can't hurt you? Statistical issues and standards for medical technology evaluation. Med Care 1996 Dec;34(12 Suppl):DS124-DS135. PMID: 8969321.

55. Mullins CD, Ogilvie S. Emerging standardization in pharmacoeconomics. Clin Ther 1998 Nov;20(6):1194-202. PMID: 9916612.

56. Mullins CD, Wang J. Pharmacy benefit management: enhancing the applicability of pharmacoeconomics for optimal decision making. Pharmacoeconomics 2002;20(1):9-21. PMID: 11817989.

57. Murray CJ, Evans DB, Acharya A, et al. Development of WHO guidelines on generalized cost-effectiveness analysis. Health Econ 2000 Apr;9(3):235-51. PMID: 10790702.

58. Blackmore CC, Magid DJ. Methodologic evaluation of the radiology cost-effectiveness literature. Radiology 1997 Apr;203(1):87-91. PMID: 9122421.

59. Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals (Brief record). 2013PMID: <[31] UNIQUE ID (DOI)>.

60. Neumann PJ, Stone PW, Chapman RH, et al. The quality of reporting in published cost-utility analyses, 1976-1997. Ann Intern Med 2000 Jun 20;132(12):964-72. PMID: 10858180.

61. Nuijten MJ, Pronk MH, Brorens MJ, et al. Reporting format for economic evaluation. Part II: Focus on modelling studies. Pharmacoeconomics 1998 Sep;14(3):259-68. PMID: 10186465.

62. Olson BM, Armstrong EP, Grizzle AJ, et al. Industry's perception of presenting pharmacoeconomic models to managed care organizations. J Manag Care Pharm 2003 Mar;9(2):159-67. PMID: 14613345.

63. Paisley S. Classification of evidence in decision-analytic models of cost-effectiveness: a content analysis of published reports. Int J Technol Assess Health Care 2010 Oct;26(4):458-62. PMID: 20923588.

64. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. Value Health 2005 Sep;8(5):521-33. PMID: 16176491.

65. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2. Value Health 2012 Sep;15(6):804-11. PMID: 22999129.

66. Sassi F, McKee M, Roberts JA. Economic evaluation of diagnostic technology. Methodological challenges and viable solutions. Int J Technol Assess Health Care 1997;13(4):613-30. PMID: 9489253.

67. Sculpher MJ, Pang FS, Manca A, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess 2004 Dec;8(49):iii-192. PMID: 15544708.

68. Severens JL, van der Wilt GJ. Economic evaluation of diagnostic tests. A review of published studies. Int J Technol Assess Health Care 1999;15(3):480-96. PMID: 10874376.

69. Siegel JE, Torrance GW, Russell LB, et al. Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on cost Effectiveness in Health and Medicine. Pharmacoeconomics 1997 Feb;11(2):159-68. PMID: 10172935.

70. Sonnenberg FA, Roberts MS, Tsevat J, et al. Toward a peer review process for medical decision analysis models. Med Care 1994 Jul;32(7 Suppl):JS52-JS64. PMID: 8028413.

71. Soto J. Health economic evaluations using decision analytic modeling. Principles and practices--utilization of a checklist to their development and appraisal. Int J Technol Assess Health Care 2002;18(1):94-111. PMID: 11987445.

72. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technol Assess 2009 Jan;13(8):iii, ix-50. PMID: 19203465.

73. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. Med Decis Making 2009 Sep;29(5):E22-E29. PMID: 19734441.

74. Udvarhelyi IS, Colditz GA, Rai A, et al. Cost-effectiveness and cost-benefit analyses in the medical literature. Are the methods being used correctly? Ann Intern Med 1992 Feb 1;116(3):238-44. PMID: 1530808.

75. Ungar WJ, Santos MT. The Pediatric Quality Appraisal Questionnaire: an instrument for evaluation of the pediatric health economics literature. Value Health 2003 Sep;6(5):584-94. PMID: 14627065.

76. Vegter S, Boersma C, Rozenbaum M, et al. Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: a systematic review on content and adherence to guidelines. Pharmacoeconomics 2008;26(7):569-87. PMID: 18563949.

77. von der SJ, Vauth C, Mittendorf T, et al. Methods for determining cost-benefit ratios for pharmaceuticals in Germany. Eur J Health Econ 2007 Sep;8 Suppl 1:S5-31. PMID: PM:17582539.

78. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health 2003 Jan;6(1):9-17. PMID: 12535234.

79. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ 2013;346:f1049.

80. Bennett C, Manuel DG. Reporting guidelines for modelling studies. BMC Medical Research Methodology 2012;12:168. PMID: 23134698.

81. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ 2013;346:f1049. PMID: 23529982.

82. Gold M. Panel on cost-effectiveness in health and medicine. Medical Care 1996;34(12):DS197-DS199.

83. International Network of Agencies for Health Technology Assessment. Health Technology Assessment Resources. 2014.

84. Luce BR, Drummond M, Jonsson B, et al. EBM, HTA, and CER: clearing the confusion. Milbank Q 2010 Jun;88(2):256-76. PMID: 20579285.

85. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making 2002 Jul;22(4):359-71. PMID: 12150601.

86. Mathes T, Jacobs E, Morfeld JC, et al. Methods of international health technology assessment agencies for economic evaluations--a comparative analysis. BMC Health Serv Res 2013;13:371. PMID: 24079858.

87. Agency for Health Technology Assessment in Poland (AHTApol/Poland). 2014.

88. Canadian Agency for Drugs and Technologies in Health (CADTH/Canada). 2014.

89. Danish Centre for Health Technology Assessment (DACEHTA/Denmark). 2014.

90. Health Information and Quality Authority (HIQA/Ireland). 2014.

91. National Authority of Medicines and Health Products (INFARMED/Portugal). 2014.

92. Institute for Quality and Efficiency in Health Care (IQWIG/Germany). 2014.

93. Belgian Federal Health Care Knowledge Centre (KCE/Belgium). 2014.

94. MAS (Medical Advisory Secretariat, within the Ontario Ministry of Health and Long-Term Care Health Strategies Division). 2014.

95. Medical Services Advisory Committee (MASC/Australia). 2014.

96. National Institute for Clinical Excellence (NICE/UK). 2014.

97. Pharmaceutical Benefits Advisor Committee (PBAC, Australia). 2014.

98. Pharmaceutical Management Agency of New Zealand (PHARMAC/New Zealand). 2014.

99. AAZ (Agency for Quality and Accreditation in Health Care, Croatia). 2014.

100. National Institute for Clinical Excellence (NICE/UK). 2014.

101. ICER (Institute for Clinical and Economic Review). 2014.

102. LBI (Ludwig Boltzmann Institute for Health Technology Assessment). 2014.

103. MHRA (Medicines and Healthcare Products Regulatory Agency). 2014.

104. NLM (National Library of Medicine). 2014.

105. AHRQ (US Agency for Healthcare Research and Quality). 2014.

106. CAST (Centre for Applied Health Services Research and Technology Assessment, University of Southern Denmark). 2014.

107. CDE (Center for Drug Evaluation). 2014.

108. Agency for Healthcare Research and Quality. Guidelines and Recommendations. 2012.

109. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. Med Decis Making 2011 Jul;31(4):530-39. PMID: 21673186.

110. Agency for Healthcare Research and Quality. The Effective Health Care Program Stakeholder Guide, Appendix C. 2011.

111. Cooper BS. Confronting models with data. J Hosp Infect 2007 Jun;65 Suppl 2:88-92.

112. Groves, P, Kayyali, B, Knott, D, et al. The 'big data' revolution in healthcare. Center for US Health System Reform Business Techology Office. McKinsey & Company; 2013.

113. Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel System--a national resource for evidence development. N Engl J Med 2011 Feb 10;364(6):498-99.

114. Zechmeister-Koss I, Schnell-Inderst P, Zauner G. Appropriate Evidence Sources for Populating Decision Analytic Models within Health Technology Assessment (HTA): A Systematic Review of HTA Manuals and Health Economic Guidelines. Med Decis Making 2013 Oct 17

115. Ades AE, Welton NJ, Caldwell D, et al. Multiparameter evidence synthesis in epidemiology and medical decision-making. J Health Serv Res Policy 2008 Oct;13 Suppl 3:12-22. PMID: 18806188.

116. Ades AE, Caldwell DM, Reken S, et al. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making 2013 Jul;33(5):679-91. PMID: 23804511.

117. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 1: introduction. Med Decis Making 2013 Jul;33(5):597-606. PMID: 23804506.

118. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Med Decis Making 2013 Jul;33(5):671-78. PMID: 23084510.

119. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making 2013 Jul;33(5):657-70. PMID: 23804509.

120. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making 2013 Jul;33(5):641-56. PMID: 23804508.

121. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making 2013 Jul;33(5):618-40. PMID: 23804507.

122. Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013 Jul;33(5):607-17. PMID: 23104435.

123. ISPOR. Value in Health Guide for Authors. 2012;2012

124. Laine C, Goodman SN, Griswold ME, et al. Reproducible research: moving toward research the public can really trust. Ann Intern Med 2007;146:450-53. PMID: 17339612.

125. Peng RD. Reproducible research in computational science. Science 2011;334:1226-27. PMID: 22144613.

126. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics 2006;24(11):1043-53. PMID: 17067190.

127. Morin A, Urban J, Adams PD, et al. Research priorities. Shining light into black boxes. Science 2012;336:159-60. PMID: 22499926.

128. Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions: uses, roles, and validity. Value Health 2001;4(5):348-61. PMID: 11705125.

129. Committee on Mathematical and Statistical Foundations of Verification VaUQ. Assessing the Realiability of Complex Models: Mathematical and Statistical Foundations of Verification,Validation,and Uncertainty Qualification. Washington, D.C.: National Academies Press; 2012.

130. Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. Health Econ 2010 Jan;19(1):43-55. PMID: 19206080.

131. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005;353(17):1784-92. PMID: 16251534.

132. Moriasi D, Wilson B, Douglas-Mankin K, et al. Hydrologic and water quality models: use, calibration, and validation. Transactions of the ASABE 2012;55(4):1241-47.

133. Legates DR, McCabe GJ. Evaluating the use of "goodness of fit measures in hydrologic and hydroclimatic model validation.". Water resources research 1999;35(1):233-41.

134. Moriasi DN, Arnold JG, Van Liew MW, et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 2007;50(3):885-900.

135. Rao C, Wu Y. On model selection. Institute of Mathematical Statistics Lecture Notes-Monograph Series. 2001;38:1-57. PMID: <[31] UNIQUE ID (DOI)>.

136. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica. 1996;6:733-807.

137. Gelman A. A Bayesian Formulation of Exploratory Data Analysis and Goodness–of–fit Testing. International Statistical Review 2003;71(2):369-82.

138. Gardner A, Deardon R, Darlington G. Goodness-of-fit measures for individual-level models of infectious disease in a Bayesian framework. Spat Spatiotemporal Epidemiol 2011 Dec;2(4):273-81. PMID: 22748225.

139. Deuchert E, Brody S. Plausible and implausible parameters for mathematical modeling of nominal heterosexual HIV transmission. Ann Epidemiol 2007 Mar;17(3):237-44. PMID: 17320790.

140. Hur C, Hayeck TJ, Yeh JM, et al. Development, calibration, and validation of a U.S. white male population-based simulation model of esophageal adenocarcinoma. PLoS One 2010;5(3):e9483. PMID: 20208996.

141. Ishida H, Wong JB, Hino K, et al. Validating a Markov model of treatment for hepatitis C virus-related hepatocellular carcinoma. Methods Inf Med 2008;47(6):529-40. PMID: 19020689.

142. Nijhuis RL, Stijnen T, Peeters A, et al. Apparent and internal validity of a Monte Carlo-Markov model for cardiovascular disease in a cohort follow-up study. Med Decis Making 2006 Mar;26(2):134-44. PMID: 16525167.

143. Perreault S, Levinton C, Laurier C, et al. Validation of a decision model for preventive pharmacological strategies in postmenopausal women. Eur J Epidemiol 2005;20(1):89-101. PMID: 15756909.

144. Sendi PP, Craig BA, Pfluger D, et al. Systematic validation of disease models for pharmacoeconomic evaluations. Swiss HIV Cohort Study. J Eval Clin Pract 1999 Aug;5(3):283-95. PMID: 10461580.

145. Willis M, Asseburg C, He J. Validation of economic and health outcomes simulation model of type 2 diabetes mellitus (ECHO-T2DM). J Med Econ 2013 Aug;16(8):1007-21. PMID: 23718682.

146. Vanni T, Karnon J, Madan J, et al. Calibrating models in economic evaluation: a seven-step approach. Pharmacoeconomics 2011 Jan;29(1):35-49. PMID: 21142277.

147. Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: Theory. Econometrica 1984;52(3):681-700.

148. Gong G, Samaniego FJ. Pseudo maximum likelihood estimation: theory and applications. Ann Statist 1981:861-69.

149. Boyd SP, Vandenberghe L. Convex optimization. Cambridge university press; 2004.

150. Bertsimas D, Farias VF, Trichakis N. On the efficiency-fairness trade-off. Management Science 2012;58(12):2234-50.

151. Kuhn H, Tucker A. Nonlinear Programming. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, California: University of California Press; 2008. p. 481-92.

152. Vapnik V. Statistical learning theory. 1998.

153. Stout NK, Knudsen AB, Kong CY, et al. Calibration methods used in cancer simulation models and suggested reporting guidelines. Pharmacoeconomics 2009;27(7):533-45. PMID: 19663525.

154. Earnshaw SR, Wilson M, Mauskopf J, et al. Model-based cost-effectiveness analyses for the treatment of acute stroke events: a review and summary of challenges. Value Health 2009 Jun;12(4):507-20. PMID: 19900253.

155. Rochau U, Schwarzer R, Jahn B, et al. Systematic assessment of decision-analytic models for chronic myeloid leukemia. Appl Health Econ Health Policy 2014 Apr;12(2):103-15. PMID: 24385259.

156. Abuelezam NN, Rough K, Seage GR, III. Individual-based simulation models of HIV transmission: reporting quality and recommendations. PLoS One 2013;8(9):e75624. PMID: 24098707.

157. Bolin K. Economic evaluation of smoking-cessation therapies: a critical and systematic review of simulation models. PharmacoEconomics 2012 Jul 1;30(7):551-64. PMID: 22591112.

158. Punyacharoensin N, Edmunds WJ, De AD, et al. Mathematical models for the study of HIV spread and control amongst men who have sex with men. Eur J Epidemiol 2011 Sep;26(9):695-709. PMID: 21932033.

159. Goehler A, Geisler BP, Manne JM, et al. Decision-analytic models to simulate health outcomes and costs in heart failure: a systematic review. PharmacoEconomics 2011 Sep;29(9):753-69. PMID: 21557632.

160. Haji Ali AH, Gray J, Karnon J. Model performance evaluation (validation and calibration) in model-based studies of therapeutic interventions for cardiovascular diseases : a review and suggested reporting framework. 2013 Apr;11(2):85-93.

161. Dams J, Bornschein B, Reese JP, et al. Modelling the cost effectiveness of treatments for Parkinson's disease: a methodological review. Pharmaeconomics 2011 Dec;29(12):1025-49. PMID: 22077577.

162. Ferdinands JM, Mannino DM. Obstructive lung disease models: what is valid? COPD 2008 Dec;5(6):382-93. PMID: 19353353.

163. Unal B, Capewell S, Critchley JA. Coronary heart disease policy models: a systematic review. BMC Public Health 2006;6:213. PMID: 16919155.

164. Siebert U, Bornschein B, Walbert T, et al. Systematic assessment of decision models in Parkinson's disease. Value Health 2004 Sep;7(5):610-26. PMID: 15367256.

165. McMahon PM, Kong CY, Johnson BE, et al. Chapter 9: The MGH-HMS lung cancer policy model: tobacco control versus screening. Risk Anal 2012 Jul;32 Suppl 1:S117-S124. PMID: 22882882.

166. Kong CY, McMahon PM, Gazelle GS. Calibration of disease simulation model using an engineering approach. Value Health 2009 Jun;12(4):521-29. PMID: 19900254.

167. Kohli M, Ferko N, Martin A, et al. Estimating the long-term impact of a prophylactic human papillomavirus 16/18 vaccine on the burden of cervical cancer in the UK. Br J Cancer 2007 Jan 15;96(1):143-50. PMID: 17146475.

168. Taylor DC, Pawar V, Kruzikas D, Gilmore KE, Pandya A, Iskandar R, Weinstein MC. Methods of model calibration: observations from a mathematical model of cervical cancer. Pharmacoeconomics. 2010;28(11):995-1000. doi:

10.2165/11538660-000000000-00000. PubMed PMID: 20936883.

169. Karnon J, Vanni T. Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. Pharmacoeconomics 2011 Jan;29(1):51-62. PMID: 21142278.

170. Karnon J, Delea T, Barghout V. Cost utility analysis of early adjuvant letrozole or anastrozole versus tamoxifen in postmenopausal women with early invasive breast cancer: the UK perspective. Eur J Health Econ 2008 May;9(2):171-83. PMID: 17602251.

171. Taylor DC, Pawar V, Kruzikas DT, et al. Incorporating calibrated model parameters into sensitivity analyses: deterministic and probabilistic approaches. PharmacoEconomics 2012 Feb 1;30(2):119-26. PMID: 22149631.

172. Hansen LP, Heckman JJ. The empirical foundations of calibration. J Econ Perspect 1996:87-104.

173. Dawkins C, Srinivasan TN, Whalley J. Calibration. Handbook of econometrics 2001;5:3653-703.

174. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med 2003 Dec 15;22(23):3687-709. PMID: 14652869.

175. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. Oper Res 1989 Mar;37(2):210-28. PMID: 10292450.

176. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence. The Confidence Profile Method. International Journal of Technology Assessment in Health Care 1990;6(1):31-55.

177. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: The confidence profile method. Med Decis Making 1990 Jan;10(1):15-23. PMID: 2182960.

178. Eddy D, Shachter R. Meta-Analysis by the Confidence Profile Method. Academic Press; 1992.

179. Droitcour J, Silberman G, Chelimsky E. A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. Int J Technol Assess Health Care 1993;9(3):440-49. PMID: 8340208.

180. Kaizar EE. Estimating treatment effect via simple cross design synthesis. Stat Med 2011 Nov 10;30(25):2986-3009. PMID: 21898521.

181. Nixon RM, Duffy SW. Cross-issue synthesis: potential application to breast cancer, tamoxifen and genetic susceptibility. J Cancer Epidemiol Prev 2002;7(4):205-12. PMID: 12846491.

182. Wald NJ, Morris JK. Teleoanalysis: combining data from different types of study. BMJ 2003 Sep 13;327(7415):616-18. PMID: 12969936.

183. Welton NJ, Ades AE. Estimation of markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Med Decis Making 2005 Nov;25(6):633-45. PMID: 16282214.

184. Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. Statistics in Medicine 2003 May 30;22(10):1649-60. PMID: 12720302.

185. Rutter CM, Miglioretti DL, Savarino JE. Bayesian Calibration of Microsimulation Models. J Am Stat Assoc 2009 Dec 1;104(488):1338-50. PMID: 20076767.

186. Basu A. Identifiability. In: Kotz S, Johnson N (eds). Encyclopedia of Statistical Sciences. Wiley Interscience: 2006. p. 2-6.

187. Garrett ES, Zeger SL. Latent class model diagnosis. Biometrics 2000 Dec;56(4):1055-67. PMID: 11129461.

188. Jackson CH, Jit M, Sharples LD, et al. Calibration of Complex Models through Bayesian Evidence Synthesis: A Demonstration and Tutorial. 2013 Jul 25

189. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. Med Decis Making 2011 Jul;31(4):625-41. PMID: 21127321.

190. Afzali HH, Karnon J, Merlin T. Improving the accuracy and comparability of model-based economic evaluations of health technologies for reimbursement decisions: a methodological framework for the development of reference models. Med Decis Making 2013 Apr;33(3):325-32. PMID: 22961101.

191. Haji Ali AH, Karnon J. Addressing the challenge for well informed and consistent reimbursement decisions: the case for reference models. PharmacoEconomics 2011 Oct;29(10):823-25. PMID: 21770483.

192. Cancer Intervention and Surveillance Modeling Network. 2014.

193. Mason JM, Mason AR. The generalisability of pharmacoeconomic studies: issues and challenges ahead. PharmacoEconomics 2006;24(10):937-45. PMID: 17002476.

Appendix A. Acknowledgments

Clinical and Policy Advisory team (CaPA): Joshua Cohen (Tufts Medical Center), Peter Neumann (Tufts Medical Center), Stephen Pauker (Tufts Medical Center).

Technical Expert Panel (TEP): Eric Bass (Johns Hopkins University), Andrew Briggs (University of Glasgow), Robert Golub (Northwestern University), Douglas Owens (Stanford University), J. Sanford Schwartz (University of Pennsylvania), Milton Weinstein (Harvard University).

Stakeholder meeting attendees (stakeholders): Andrew Briggs (University of Glasgow), Martin Brown (Patient-Centered Outcomes Research Institute), Paul Coates (National Institutes of Health), Joshua Cohen (Tufts Medical Center), Mark Ebell (University of Georgia), Eric Feuer (National Institutes of Health), Rachael Fleurence (Patient-Centered Outcomes Research Institute), Scott Grosse (Center for Disease Control), Mark Helfand (Oregon Health & Science University), Mark Hltaky (Stanford University), Don Husereau (Institute of Health Economics), Karen Kuntz (University of Minnesota), Timothy Lash (Wake Forest University), Angela Lasher (U.S. Food and Drug Administration), Jeanne Mandelblatt (Georgetown University), Robert McDonough (Aetna), Peter Neumann (Tufts Medical Center), Deborah Olster (National Institutes of Health), Douglas Owens (Stanford University), Stephen Pauker (Tufts Medical Center), Diane Petitti (Arizona State University), Mark Roberts (University of Pittsburgh), James Rollins (Centers for Medicare and Medicaid Services), James Scanlon (U.S. Department of Health and Human Services), J. Sanford Schwartz (University of Pennsylvania), Uwe Siebert (University of Health Sciences, Medical Informatics, and Technology in Austria), Pamela Wescott (Foundation for Informed Medical Decision Making), Ann Zauber (Memorial Sloan-Kettering Cancer Center).

This manuscript reflects solely the opinions of the authors and does not reflect the opinions of AHRQ, the Department of Health and Human Services, meeting participants, or the institutions with which they are affiliated.

Appendix B. Systematic Review Search Strategy

Summary	Search strings	
Recreated Philips et al. 2004 search{ Philips, 2004 198 /id}:		
"Recommendation-related" terms	1	(checklist? or check list? or standards or standardi?ation or peer review$ or rules or critiquing or criteria or good or bad or correct$ or bias or fundamentals or recommend$ or best or strength$ or weakness$ or quality or qualities or validity or guideline? or validation or checkpoint?).ti.
	2	(properly or critically appraise or problems or limitations or rating scale? or framework$ or protocol? or audit or principles or methodology$).ti.
	3	(validate or validation or evaluating or properties or guidance or integrity or evaluation or pros or cons).ti.
	4	or/1-3
"Modeling-related" terms	5	(decision adj (tree or triage or data or analytic or analysis)).ti.
	6	exp models, economic/ or exp models, econometric/
	7	(exp decision support techniques/ or exp data interpretation, statistical/ or exp decision theory/ or exp models, statistical/ or exp likelihood functions/ or exp linear models/ or exp logistic models/ or exp proportional hazards models/) and exp costs/ and cost analysis/
	8	((economic? or pharmacoeconomic? or decision? or cost? or costing?) and model$).ti.
	9	(markov or crystal ball).ti.
	10	exp markov chain/
	11	or/4-9
Critical appraisal of models	12	((markov model$ or economic model$ or mathematical model$ or cost$ model$ or pharmacoeconomic model$ or decision model$) adj2 (checklist? or check list? or standards or standardi?ation or peer review$ or rules or critiquing or criteria or good or bad or correct$ or bias or fundamentals or recommend$ or best or strength$ or weakness$ or quality or qualities or validity or guideline? or validation or checkpoint?)).ab.
	13	((markov model$ or economic model$ or mathematical model$ or cost$ model$ or pharmacoeconomic model$ or decision model$) adj2 (properly or critically appraise or problems or limitations or rating scale$ or good practice$ or framework$ or protocol$ or audit or principles or ethodology$)).ab.
	14	((markov model$ or economic model$ or mathematical model$ or cost$ model$ or pharmacoeconomic model$ or decision model$) adj2 (validate or validation or evaluating or properties or guidance or integrity or avoiding bias or evaluation or pros or cons)).ab.
	15	((decision tree or decision triage or decision data or decision analytic$ or decision analysis or crystal ball) adj2 (checklist? or check list? or standard$ or peer review$ or rules or critiquing or criteria or good or bad or correct$ or bias or fundamentals or recommend$ or best or strength$ or weakness$ or quality or qualities or validity or guideline? or validation or checkpoint?)).ab.
	16	((decision tree or decision triage or decision data or decision analytic$ or decision analysis or crystal ball) adj2 (properly or critically appraise or problems or limitations or rating scale$ or framework$ or protocol$ or audit or principles or methodology$)).ab.
	17	((decision tree or decision triage or decision data or decision analytic$ or decision analysis or crystal ball) adj2 (validate or validation or evaluating or properties or guidance or integrity or evaluation or pros or cons)).ab.

Summary	Search strings	
	18	((economic evaluation? or economic analysis or economic stud$ or economic submission?) and guideline$).ti.
	19	or/12-18
Total Philips search	20	or/4, 11, 19
Additional terms added for this systematic review, restricted to targeted journals:		
Targeted journals	21	Value in health.jn.
	22	(Health technology assessment or "Health Technology Assessment (Winchester, England)").jn.
	23	Pharmacoeconomics.jn.
	24	Journal of Medical Economics.jn.
	25	Annals of Internal Medicine.jn.
	26	Medical Decision Making.jn.
	27	BMC Health Services Research.jn.
	28	Clinical Therapeutics.jn.
	29	European Journal of Health Economics.jn.
	30	(The Journal of the American Medical Association or jama).jn.
	31	(British Medical Journal or BMJ).jn.
	32	Current Medical Research & Opinion.jn.
	33	Health Economics.jn.
	34	Journal of Health Economics.jn.
	35	Medical care.jn.
	36	International Journal of Technology Assessment in Health Care.jn.
	37	BMC Medical Research Methodology.jn.
	38	Journal of Clinical Epidemiology.jn.
	39	Journal of general internal medicine.jn.
	40	American journal of managed care.jn.
	41	Journal of managed care pharmacy.jn.
	42	The European journal of health economics.jn.
	43	Journal of evaluation in clinical practice.jn.
	44	The Journal of the american board of family practice.jn.
	45	Statistics in medicine.jn.
	46	Archives of Internal Medicine.jn.
	47	Clinical Therapeutics.jn.
	48	Current Medical Research & Opinion.jn.
	49	The New England Journal of Medicine.jn.
	50	Lancet.jn.
	51	PLOS medicine.jn.
	52	Annual review of genomics.jn.
	53	Human genetics.jn.
	54	Population health metrics.jn.
	55	Radiology jn.
	56	Journal of the national cancer institute.jn.
	57	Health care management science.jn.
	58	(Canadian medical association journal or cmaj).jn.
	59	or/21-58
Recommendation-related terms, restricted to journals	60	(consensus or standard$ or framework$ or principle$ or committee$).mp. [mp=ti, ab, ot, nm, hw, ps, rs, ui, tx]
	61	59 and 60

Summary	Search strings	
	62	(cost$ and (effectiv$ or utility or benefit) and (analy$ or model$)).mp. [mp=ti, ab, ot, nm, hw, ps, rs, ui, tx]
	63	exp cost-benefit/
	64	(decision and (analy$ or model$ or analy$)).mp. [mp=ti, ab, ot, nm, hw, ps, rs, ui, tx]
	65	exp decision support techniques/
	66	((model and (microsimulation or dynamic or discrete event or simulation or state transition or agent based or infectious disease transmission or transmission or seir)) or computer simulation).mp. [mp=ti, ab, ot, nm, hw, kf, ps, rs, ui, tx]
Modeling-related terms, restricted to journals	67	or/62-66
	68	59 and 67
Additional terms, total	69	61 or 68
Total (combined Philips search and additional terms)	70	20 or 69
Restriction by publication date	71	limit 70 to yr = 1990 -Current

Appendix C. Health Technology Assessment Organizations

Health Technology Assessment (HTA) Organization	Web site
AAZ (Agency for Quality and Accreditation in Health Care, Croatia)	http://www.aaz.hr/
AETMIS (Agence d´Évaluation des Technologies et des Modes d´Intervention en Santé)	http://www.aetmis.gouv.qc.ca/site/accueil.phtml
AETS ICS III (Agencia de Evaluación de Tecnologias Sanitarias)	http://www.isciii.es/htdocs/en/investigacion/Agencia_quees.jsp
AETSA (Andalusian Agency for Health Technology Assessment)	http://www.juntadeandalucia.es/salud/servicios/aetsa/
Age.Na:s (Agenzia Nazionale per I Servizi Sanitari Regionali)	http://www.agenas.it/
Agency for Health Technology Assessment in Poland (AHTApol/Poland)	http://www.aotm.gov.pl/index.php?id=397
AHRQ (US Agency for Healthcare Research and Quality)	http://www.effectivehealthcare.ahrq.gov/tools-and-resources/researcher-resources/
AHTA (Adelaide Health Technology Assessment)	http://www.adelaide.edu.au/ahta/
AIFA (Agenzia Italiana Del Farmaco)	http://www.agenziafarmaco.it/en
ARESS (Agenzia Regionale per i Servizi Sanitari)	http://www.aress.piemonte.it/Links.aspx
ARSENÁL (Veneto's Research Centre for e-Health Innovation)	http://www.consorzioarsenal.it/en/web/guest/home
ASERNIP-S (Australian Safety and Efficacy Register of New Interventional Procedures –Surgical)	http://www.surgeons.org/racs/research-and-audit/asernip-s/asernip-s-publications
ASSR (Regione Emilia Romagna, Agenzia Sanitaria e Sociale Regione Emilia Romagna)	http://asr.regione.emilia-romagna.it/
AVALIA-T (Galician Agency for Health Technology Assessment)	http://www.sergas.es/MostrarContidos_Portais.aspx?IdPaxina=60538
BAG (Bundesamt für Gesundheit) / FOPH (Federal Office of Public Health)	http://www.bag.admin.ch/index.html?lang=de
BCBS (Blue Cross BlueShield Association)	http://www.bcbs.com/
Belgian Federal Health Care Knowledge Centre (KCE/Belgium)	https://kce.fgov.be/
BS-CA (Blue Shield of California Foundation)	http://www.blueshieldcafoundation.org/
CAHIAQ (Catalan Agency for Health Information, Assessment and Quality) (formerly CAHTA)	http://www.gencat.cat/salut/depsan/units/aatrm/html/en/dir394/index.html
Canadian Agency for Drugs and Technologies in Health (CADTH/Canada)	http://www.cadth.ca/en
CAST (Centre for Applied Health Services Research and Technology Assessment, University of Southern Denmark)	http://www.sdu.dk/Om_SDU/Institutter_centre/CAST?sc_lang=en
CDE (Center for Drug Evaluation)	http://www.cde.org.tw/English/Pages/e-default.aspx
CEDIT (Comité d´Evaluation et de Diffusion des Innovations Technologiques)	http://cedit.aphp.fr/-Pays-.html?rubrique&lang=en&dir=ltr
CEM (Cellule d'expertise médicale)	http://www.ms.public.lu/fr/actualites/2011/04/02-offre-d-emploi/index.html
CENETEC (Centro Nacional de Excelencia Tecnológica en Salud)	http://www.cenetec.salud.gob.mx/

Health Technology Assessment (HTA) Organization	Web site
CMeRC - HTA Unit	Not available
CMTP (Center for Medical Technology Policy)	http://www.cmtpnet.org/
CNHTA (Committee for New Health Technology Assessment)	http://www.cha.ac.kr/
CRD (Centre for Reviews and Dissemination)	http://www.york.ac.uk/inst/crd/
CVZ (College voor Zorgverzekeringen, Dutch health care insurance board)	http://www.cvz.nl/en/home
DAHTA@DIMDI (Deutsche Agentur für Health Technology Assessment - Bewertung gesundheitsrelevanter Verfahren – Deutsches Institut für medizinische Dokumentation und Information)	http://www.dimdi.de/static/de/index.html
Danish Centre for Health Technology Assessment (DACEHTA/Denmark)	http://www.sst.dk/English/DACEHTA.aspx
DECIT-CGATS - Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia	http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=25516
DSI (Danish Institute for Health Services Research)	http://dsi.dk/english/
EMKI (Institute for Healthcare Quality Improvement and Hospital Engineering)	http://www.emki.hu/site/index.php
ESKI (National Institute for Strategic Health Research)	http://www.eski.hu/index_en.php
ETESA (Department of Quality and Patient Safety of the Ministry Health of Chile)	http://www.redsalud.gov.cl/portal/url/page/minsalcl/g_home/home.html
FEGAS (School of Health Administration)	http://www.sergas.es/MostrarContidos_Portais.aspx?IdPaxina=50200
FIMEA (Finnish Medicines Agency)	http://www.fimea.fi/frontpage
FinOHTA (Finnish Office for Health Technology Assessment)	http://finohta.stakes.fi/EN/index.htm
G-BA (Gemeinsamer Bundesausschuss)	http://www.g-ba.de/
GÖG/BIQG (Gesundheit Österreich GmbH)	http://www.goeg.at/
GR (Gezondheidsraad)	http://www.gezondheidsraad.nl/
GYEMSZI (National Institute for Quality- and Organizational Development in Healthcare and Medicines)	http://www.ogyi.hu/gyemszi/
HA (Hospital authority Hong Kong)	http://www.ha.org.hk/visitor/ha_index.asp
HAS (Haute Autorité de Santé)	http://www.has-sante.fr/portail/jcms/c_5443/english?cid=c_5443
Health Information and Quality Authority (HIQA/ Ireland)	http://www.hiqa.ie/
HIS (Health Care Improvement Scotland)	http://www.healthcareimprovementscotland.org/home.aspx
HITAP (Health Intervention and Technology Assessment Program)	http://www.hitap.net/en/splash
HSAC (Health Services Assessment Collaboration)	http://www.healthsac.net/
HTA-HSR/DHTA (HTA & Health Services Research)	http://www.centerforfolkesundhed.dk/om+centret/in+english

Health Technology Assessment (HTA) Organization	Web site
HVB, Hauptverband der Österreichischen Sozialversicherungsträger	http://www.sozialversicherung.at/portal27/portal/esvportal/start/startWindow?action=2&p_menuid=2&p_tabid=1
ICER (Institute for Clinical and Economic Review)	http://www.icer-review.org/
ICTAHC (Israel Center for Technology Assessment in Health Care)	http://www.health.gov.il/subjects/
IECS (Institute for Clinical Effectiveness and Health Policy)	http://www.iecs.org.ar/
IER (Institute for Economic Research)	http://www.ier.si/index.php
IHE (Institute of Health Economics)	http://www.ihe.ca/
INESSS - Institut national d'excellence en santé et en services	http://www.inesss.qc.ca/index.php?id=50&L=1
Institute for Quality and Efficiency in Health Care (IQWIG/Germany)	https://www.iqwig.de/en/home.2724.html
IPP (Institut für Public Health und Pflegeforschung, Universität Bremen)	http://www.ipp.uni-bremen.de/index.php
IRF (Institute for Rational Pharmacotherapy)	http://www.irf.dk/en/home.htm
JAZMP (Agency for Medicinal Products and Medical Devices)	http://www.jazmp.si/index.php?id=105
Kaiser Permanente	https://www.kaiserpermanente.org/
KDTD (Turkish Evidence-Based Medicine Association)	http://www.kanitadayalitip.org/index_eng.html
Kela (The Social Insurance Institution of Finland)	http://www.kela.fi/in/internet/english.nsf
Laziosanità (Agenzia di Sanità Pubblica, Regione Lazio)	http://www.regione.lazio.it/web2/contents/sanita.php
LBI (Ludwig Boltzmann Institute for Health Technology Assessment)	http://hta.lbg.ac.at/page/homepage
MaHTAS (Health Technology Assessment Section, Ministry of Health Malaysia)	http://www.moh.gov.my/health_assesments
MAS (Medical Advisory Secretariat, within the Ontario Ministry of Health and Long-Term Care Health Strategies Division)	http://www.health.gov.on.ca/english/providers/program/mas/tech/tech_mn.html
Medical Services Advisory Committee (MASC/Australia)	http://www.msac.gov.au/
MHRA (Medicines and Healthcare Products Regulatory Agency)	http://www.mhra.gov.uk/index.htm/
MOH Indonesia (Ministry of Health – Republic of Indonesia)	http://www.depkes.go.id/en/
MOH RS (Ministry of Health – Serbia)	http://www.zdravlje.gov.rs/index.php?
MOH Singapore (Ministry of Health – Singapore)	http://www.moh.gov.sg/content/moh_web/home.html
MOH Spain (Ministry of Health – Spain)	http://www.msc.es/
MOH-CZ (Ministry of Health - Czech Republic)	http://www.mzcr.cz/En/
MTAA (Medical Technologies Association of Australia)	http://www.mtaa.org.au/pages/index.asp
MTU-SFOPH (Medical Technology Unit - Swiss Federal Office of Public Health)	http://www.bag.admin.ch/
National Authority of Medicines and Health Products (INFARMED/Portugal)	http://www.infarmed.pt/portal/page/portal/INFARMED/ENGLISH

Health Technology Assessment (HTA) Organization	Web site
National Institute for Clinical Excellence (NICE/UK)	http://www.nice.org.uk/
NBoH (National Board of Health)	http://www.sst.dk/
NCPE (National Centre for Pharmacoeconomics, St. James's Hospital)	http://www.stjames.ie/Departments/DepartmentsA-Z/N/NationalCentreforPharmacoeconomics/DepartmentOverview/
NCPHP (National Centre of Public Health Protection)	http://ncphp.government.bg/
NECA - National Evidence-based healthcare Collaboration Agency	http://www.neca.re.kr/eng/
NETSCC, HTA - NIHR (Coordinating Centre for Health Technology Assessment)	http://www.hta.ac.uk/
Newcastle University	http://www.ncl.ac.uk/
NHG (National Healthcare Group)	http://www.nhg.com.sg/
NHMRC	http://www.nhmrc.gov.au/
NHS QIS (Quality Improvement Scotland)	http://www.nhshealthquality.org/nhsqis/CCC_FirstPage.jsp
NHSC (National Horizon Scanning Centre)	http://www.haps.bham.ac.uk/publichealth/horizon/
NIPH-RS (National Institute of Public Health of the Republic of Slovenia)	http://www.ivz.si/
NLM (National Library of Medicine)	http://www.nlm.nih.gov/
NOKC (Norwegian Knowledge Centre for the Health Services)	http://www.kunnskapssenteret.no/Home?language=english
NSPH (National School of Public Health)	http://www.nsph.gr/default.aspx?page=home
OSTEBA (Basque Office for Health Technology Assessment)	http://www.osanet.euskadi.net/r85-osteba/es/contenidos/informacion/osteba/es_osteba/osteba.html
PATH (Programs for Assessment of Technology in Health Research Institute)	http://www.path-hta.ca/Home.aspx
PenTAG (Peninsula Technology Assessment Group)	Not available
Pharmaceutical Benefits Advisor Committee (PBAC, Australia)	http://www.health.gov.au/internet/main/publishing.nsf/Content/pbac-outcomes-info
Pharmaceutical Management Agency of New Zealand (PHARMAC/New Zealand)	http://www.pharmac.govt.nz/
QPACT (Queensland Policy and Advisory Committee for New Technology)	http://www.health.qld.gov.au/newtech/html/QPACT.asp
Regione Veneto (Regione Veneto, Direzione Piani e Programmi Socio Sanitari)	http://www.regione.veneto.it/channels
Reglom-DGSAN (Regione Lombardia Direzione Generale Sanita)	http://www.sanita.regione.lombardia.it/cs/Satellite?c=Page&childpagename=DG_Sanita/DGHomeLayout&cid=1213277054618&pagename=DG_SANWrapper
RIZIV (Rijksinstituut voor ziekte- en invaliditeitsverzekering)	http://www.riziv.fgov.be/presentation/nl/index.htm
santésuisse (Branchenverband der schweizerischen Krankenversicherer)	http://www.santesuisse.ch/de/dyn_output.html?content.vname=portal
SBU (Swedish Council on Technology Assessment in Health Care)	http://www.sbu.se/en/
ScHARR (Technology Assessment Group, University of Sheffield)	http://www.shef.ac.uk/scharr/sections/heds/collaborations/tag

Health Technology Assessment (HTA) Organization	Web site
SIDC (State Institute for Drug Control)	http://www.sukl.sk/en/about-us
SingHealth (Singapore Health Service)	http://www.singhealth.com.sg/Pages/Home.aspx
SLOVATHA (Slovak Agency for Health Technology Assessment)	
SNHTA (Swiss Network for HTA)	http://www.snhta.ch/
SNSPMS (National School of Public Health, Management d Professional Development)	http://www.snspms.ro/
SPC on Standardization and HTA	Not available
SSD/MSOC (Ministry for Social Policy, Strategy and Sustainability Division)	Not available
Sundhed.dk (Centre for Public Health, Central Denmark Region, department HTA & Health Services Research)	http://www.cfk.rm.dk/om+os/in+english/health+technology+assessment+and+health+services+research
TLV (Dental and Pharmaceutical Benefits Agency)	http://www.tlv.se/in-english-old/in-english/
UCEETS - The National Coordination Unit of Health Technology Assessment and Implementation	http://www.msal.gov.ar/pngcam/
UETS (Unidad de Evaluación de Tecnologías Sanitarias)	http://www.madrid.org/cs/Satellite?cid=1142494649964&language=es&pagename=PortalSalud/Page/PTSA_pintarContenidoFinal&vest=1142494649964
UFI-SALUD (Unidad de Financiamiento Internacional de Salud)	http://www.ufisalud.gov.ar/
UMIT (Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik)	http://www.umit.at/page.cfm?vpath=index
University Hospital A. Gemelli	http://www.rm.unicatt.it/
UTA (University of Tartu , Department of Public Health)	http://www.ut.ee/en
UVT (HTA Unit in A. Gemelli Teaching Hospital)	http://www.policlinicogemelli.it/area/?s=206
VASPVT (State Health Care Accreditation Agency under the Ministry of Health of the Republic of Lithuania)	http://www.vaspvt.gov.lt/index.php?2719160486
VATAP (VA Technology Assessment Program)	Not available
VEC (Centre of Health Economics)	http://www.vec.gov.lv/english/default.html
ZonMw (The Medical and Health Research Council of The Netherlands)	http://www.zonmw.nl/

Appendix D. HTA Organization Data Extraction

Question	Agency for Health Technology Assessment in Poland (AHTApol/Poland)
Integration of modeling	The situations in which modeling is recommended include: - the need to evaluate the results in real practice when only the results of experimental tests are available and the results obtained in one country can be transposed into another one, - indirect comparative synthesis if relevant direct trials are missing, - providing estimates if direct measurements are missing, - preliminary assessment and scheduling of trials, - early stage of development of a new technology if comprehensive trials are missing. - the need to extrapolate the results beyond the time horizon of the clinical trials included in the clinical analysis, - the need to transpose the experimental effectiveness measured (i.e. indirect results expressed on a disease-specific scale) to final utility results (e.g. life years gained, gained QALY),
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	If modeling is necessary, the model structure should be presented. Assumptions of the model should be clear, well justified and tested in a sensitivity analysis. If data in the model are extrapolated over time horizon of the primary trials, the following scenarios should be analyzed: optimistic, pessimistic and neutral. The analytical task consists in taking into account Polish data concerning the use of resources and costs.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	The situations in which modeling is recommended include: - the need to transpose the experimental effectiveness measured (i.e. indirect results expressed on a disease-specific scale) to final utility results (e.g. life years gained, gained QALY),
Inclusion of costs	The analytical task consists in taking into account Polish data concerning the use of resources and costs.
Budget analysis done	nd
Impact on project budget	nd

Question	Canadian Agency for Drugs and Technologies in Health (CADTH/Canada)
Integration of modeling	nd
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	Economic evaluations of health care technologies typically involve building and then using models to synthesize evidence and assumptions from multiple sources to estimate the long-term incremental costs and outcomes of new therapies. Because the outputs (results) depend on the model structure, the data, and the assumptions used, the model should be as transparent as possible. As a result, decision makers should be critical when reviewing the results of a model-based evaluation.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	Economic evaluations of health care technologies typically involve building and then using models to synthesize evidence and assumptions from multiple sources to estimate the long-term incremental costs and outcomes of new therapies.
Budget analysis done	Nd
Impact on project budget	Nd

Question	Danish Centre for Health Technology Assessment (DACEHTA/Denmark)
Integration of modeling	Modeling is used frequently in connection with HTA since it is here attempted to take existing literature as the basis. There is often evidence for the effect of a technology in the form of clinical data, survival data and/or data concerning health-related quality of life, and one will then, where appropriate, content oneself with collecting cost data and comparing these with the effects in a model … In some cases, modeling will need to be used in the economic analysis – whether completely or only partially. There are a number of reasons for this (Buxton et al.)
Modeling alongside SR	Nd
Timing of modeling	Nd
Use of pre-existing vs. established models	Nd
Model recommendations	Extrapolation of short-term clinical data for the purpose of predicting these data in the longer term, e.g. survival probabilities, or linkage of intermediate endpoints to final endpoints, can lead to modeling in the economic analysis. The performance of the clinical study in a controlled and randomised design which ensures a high degree of internal validity often conversely means that the study has a low degree of external validity. Here, it can be necessary to model the economic analysis in order to be able to generalise about daily practice or between regions in the country. As mentioned previously, it may also happen to be placebo that the new technology is compared with in the clinical study. Here, it may be necessary to use models in the economic analysis to investigate the cost-effectiveness of the new technology in relation to daily practice. Lastly, there may be insufficient economic and clinical data, particularly early in the development/life cycle of a health technology. The economic analysis can, in such a situation, be modeled entirely on the basis of the best available evidence and the expectations that one may have. Regardless of whether modeling is necessary, or the economic analysis can be based directly on the clinical study, it may be a good idea, purely in order to gain a comprehensive view, to draw up a decision tree for the possible patient streams as referred to above.
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	There is often evidence for the effect of a technology in the form of clinical data, survival data and/or data concerning health-related quality of life, and one will then, where appropriate, content oneself with collecting cost data and comparing these with the effects in a model …
Inclusion of costs	As mentioned previously, it may also happen to be placebo that the new technology is compared with in the clinical study. Here, it may be necessary to use models in the economic analysis to investigate the cost-effectiveness of the new technology in relation to daily practice.
Budget analysis done	Nd
Impact on project budget	Nd

Question	Health Information and Quality Authority (HIQA/Ireland)
Integration of modeling	The use of modeling is typically required as part of an economic evaluation to make clinical and cost-effectiveness estimates relevant to the time frame under review.
Modeling alongside SR	In the reference case, evidence on outcomes should be obtained by means of a systematic review with all data sources clearly described.(15) Evidence generated from this phase is necessary to inform decision making, but may also be used to populate economic decision-analytic models. These models can be used to project the potential health and economic consequences of using different technologies over an adequate time frame.
Timing of modeling	Economic evaluations may be run alongside a clinical trial, where the patient outcomes and associated costs generated in the trial are used to populate the economic model, rather than data from multiple trials or gathered in a systematic review. In such cases there are a number of risks of bias (e.g., protocol-driven costs, lack of longer-term follow-up data, inappropriate outcomes) that can impact on the results. Adequate steps must be taken to show that the data are appropriate and generalisable to the relevant population in Ireland (e.g., it may be reasonable to make the trial data available for independent assessment). Models will frequently require numerous additional parameters which may be directly or indirectly related to the effectiveness of a technology (e.g., uptake rate, disease severity). The values for these sorts of parameters will often be informed by local data on disease prevalence, service utilisation and expert opinion. As they are not typically derived from systematic review, care must be taken to adequately address potential bias in the parameter estimates and to take into account the uncertainty or lack of precision in the estimates. As such, a sensitivity analysis should also include these parameters. Where expert opinion is used, it should be elicited in a manner which minimises bias and the process should be documented in sufficient detail.
Use of pre-existing vs. established models	Nd
Model recommendations	modeling (See section 2.12) There is no one optimal modeling technique, rather the choice of model should depend on the research question to be addressed. Models used to synthesise and extrapolate available evidence should be developed in accordance with good modeling practice guidelines. The model should be clearly described, with the assumptions and inputs documented and justified. The methods for the quality assurance of the model should be detailed and the model validation results documented. The model and its key inputs should be subjected to comprehensive sensitivity analysis. Uncertainty (Section 2.15) The effects of model uncertainty (i.e., structure, methods and assumptions) and parameter uncertainty on the outcome of the economic evaluation must be systematically evaluated using sensitivity analysis and scenario analyses for the range of plausible scenarios. The range of values provided for each parameter must be clearly stated and justified. Justification for the omission of any model input from the sensitivity analysis should be included. For the reference case, a one-way sensitivity analysis should be conducted to identify the key model inputs/assumptions contributing most to uncertainty. Multivariate analysis should be used for key model inputs. Probabilistic sensitivity analysis (PSA), in the form of a Monte Carlo simulation, should be used to assess parameter uncertainty. The expected value of perfect information (EVPI) should also be evaluated.

	The use of extrapolation modeling is typically required when adopting a lifetime horizon as long-term primary data on the safety and effectiveness of a new technology will only be available after the product has been in routine clinical use for some time. When extrapolating data beyond the duration of the clinical trials, inherent assumptions regarding future treatment effects and disease progression should be clearly outlined and tested as part of the sensitivity analysis (see also Section 2.15). Models will frequently require numerous additional parameters which may be directly or indirectly related to the effectiveness of a technology (e.g., uptake rate, disease severity). The values for these sorts of parameters will often be informed by local data on disease prevalence, service utilisation and expert opinion. As they are not typically derived from systematic review, care must be taken to adequately address potential bias in the parameter estimates and to take into account the uncertainty or lack of precision in the estimates. As such, a sensitivity analysis should also include these parameters. Where expert opinion is used, it should be elicited in a manner which minimises bias and the process should be documented in sufficient detail. Currently, there are no agreed Irish cost models available. As a result, the generation of valid Irish cost data is challenging and time consuming. Until a valid Irish cost model is established, there is a need for flexibility regarding cost valuation. To maximise reproducibility and transferability, all assumptions and cost estimates must be clearly reported and subjected to one-way and probabilistic sensitivity analysis (see also Section 2.15). In particular, where costs are applied from other countries, the assumptions necessary to transfer this data must be explicitly reported, with all costs converted to their Irish equivalent in euro using Purchasing Power Parity indices.(21) An example of how to transfer costs is included in Appendix 2. The evidence supporting the biological or clinical plausibility of the subgroup effect should be fully documented, including details of statistical analyses. Since the goal of the health system is to maximise the potential for health gain from its finite resources, a stratified analysis that allows cost-effectiveness to be modeled separately for each subgroup, may contribute important information to the final advice. Probabilistic sensitivity analysis (PSA) is the preferred approach for exploring uncertainty arising from parameter imprecision (e.g. uncertainty around the true mean values of cost and efficacy inputs) in decision-analytic modeling. With this approach, probability distributions are applied using specified plausible ranges for the key parameters rather than the use of varied point estimates for each parameter. IPD from a single or small number of trials may also be used as a basis for developing a micro-simulation model. Patient characteristics are used to populate the model and simulate the impact of introducing a treatment in terms of endpoints and costs. Such an exercise should not be considered as either evidence synthesis or meta-analysis, but rather a form of subgroup analysis. The use of IPD for micro-simulation is beyond the scope of these Guidelines.
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	The preferred evaluation type for the reference case is a cost-utility analysis (CUA) with the outcomes expressed in terms of quality-adjusted life-years (QALYs).

	Valuing Outcomes (See section 2.11) For the reference case, health effects should be valued in QALYs.
Inclusion of costs	The use of modeling is typically required as part of an economic evaluation to make clinical and cost-effectiveness estimates relevant to the time frame under review. Currently, there are no agreed Irish cost models available. As a result, the generation of valid Irish cost data is challenging and time consuming. Until a valid Irish cost model is established, there is a need for flexibility regarding cost valuation. To maximise reproducibility and transferability, all assumptions and cost estimates must be clearly reported and subjected to one-way and probabilistic sensitivity analysis (see also Section 2.15). In particular, where costs are applied from other countries, the assumptions necessary to transfer this data must be explicitly reported, with all costs converted to their Irish equivalent in euro using Purchasing Power Parity indices.(21) An example of how to transfer costs is included in Appendix 2. The evidence supporting the biological or clinical plausibility of the subgroup effect should be fully documented, including details of statistical analyses. Since the goal of the health system is to maximise the potential for health gain from its finite resources, a stratified analysis that allows cost-effectiveness to be modeled separately for each subgroup, may contribute important information to the final advice.
Budget analysis done	Entire report recently released on Budget Impact Analysis: http://www.hiqa.ie/system/files/Budget-Impact-Analysis-Guidelines-2014.pdf
Impact on project budget	Nd

Question	National Authority of Medicines and Health Products (INFARMED/Portugal)
Integration of modeling	Nd
Modeling alongside SR	Nd
Timing of modeling	An important problem that pharmaco-economic studies have to face is that only efficacy data are available when a new product is launched. Any studies carried out at this stage will inevitably have to extrapolate the effectiveness of the treatment on the basis of its estimated efficacy in the clinical trials. modeling is normally used to do this.
Use of pre-existing vs. established models	Nd
Model recommendations	If no data on effectiveness are available from clinical trials…efficacy data obtained in appropriate clinical trials can be used after being corrected by modeling.
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	Nd
Inclusion of costs	Nd
Budget analysis done	Nd
Impact on project budget	Nd

Question	Institute for Quality and Efficiency in Health Care (IQWIG/Germany)
Integration of modeling	Economic data are not regularly collected in clinical trials. If this is done, how ever, these data alone are often not sufficient for a full and substantiated depiction of the costs of a health technology. Clinical trials seldom provide information on the long-term economic consequences associated with the introduction of a new technology. In addition, they do not always adequately and comprehensively reflect all cost aspects relevant to the German health care setting. Moreover, protocol-induced resource consumption in clinical trials may bias cost estimation. For these reasons, the modeling of the economic effects of a health technology is an essential component in health economic evaluation.
Modeling alongside SR	Nd
Timing of modeling	nd
Use of pre-existing vs. established models	Nd
Model recommendations	Nd
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	Economic data are not regularly collected in clinical trials. If this is done, how ever, these data alone are often not sufficient for a full and substantiated depiction of the costs of a health technology. Clinical trials seldom provide information on the long-term economic consequences associated with the introduction of a new technology. In addition, they do not always adequately and comprehensively reflect all cost aspects relevant to the German health care setting. Moreover, protocol-induced resource consumption in clinical trials may bias cost estimation. For these reasons, the modeling of the economic effects of a health technology is an essential component in health economic evaluation.
Inclusion of costs	Nd
Budget analysis done	Nd
Impact on project budget	Nd

Question	Belgian Federal Health Care Knowledge Centre (KCE/Belgium)
Integration of modeling	Modeling should be applied if the available data are insufficient to allow a full assessment of the cost-effectiveness or cost-utility of a product.
Modeling alongside SR	Nd
Timing of modeling	Nd
Use of pre-existing vs. established models	Nd
Model recommendations	Nd
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	Modeling should be applied if the available data are insufficient to allow a full assessment of the cost-effectiveness or cost-utility of a product.
Inclusion of costs	Modeling should be applied if the available data are insufficient to allow a full assessment of the cost-effectiveness or cost-utility of a product.
Budget analysis done	Nd
Impact on project budget	Nd

Question	MAS (Medical Advisory Secretariat, within the Ontario Ministry of Health and Long-Term Care Health Strategies Division/Canada)
Integration of modeling	Nd
Modeling alongside SR	Nd
Timing of modeling	Nd
Use of pre-existing vs. established models	Nd
Model recommendations	The time horizon chosen for an economic evaluation is important and can dramatically affect the size of the incremental cost-effectiveness ratio. However, the data on which efficacy is based usually is derived from randomized trials or non-experimental studies that follow patients for a relatively short period of time. modeling techniques must be used to project lifetime costs and effects if such a time frame is appropriate. Unfortunately, however, the data on which to project lifetime costs and clinical effects must almost certainly be much more speculative than those with a short time frame. Submissions should clearly state the time horizon chosen. The analysis should delineate the time horizon on which estimates can be based from currently available high quality empirical data (e.g., randomized trials that follow patients for months to a few years) or from modeled data based on extrapolations.
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	Nd
Inclusion of costs	The time horizon chosen for an economic evaluation is important and can dramatically affect the size of the incremental cost-effectiveness ratio. However, the data on which efficacy is based usually is derived from randomized trials or non-experimental studies that follow patients for a relatively short period of time. modeling techniques must be used to project lifetime costs and effects if such a time frame is appropriate. Unfortunately, however, the data on which to project lifetime costs and clinical effects must almost certainly be much more speculative than those with a short time frame. Submissions should clearly state the time horizon chosen. The analysis should delineate the time horizon on which estimates can be based from currently available high quality empirical data (e.g., randomized trials that follow patients for months to a few years) or from modeled data based on extrapolations.
Budget analysis done	Nd
Impact on project budget	Nd

Question	Medical Services Advisory Committee (MASC/Australia)
Integration of modeling	Nd
Modeling alongside SR	Nd
Timing of modeling	Nd
Use of pre-existing vs. established models	Nd
Model recommendations	The aim of the economic evaluation is to use the clinical studies ... to determine the economic cost of substituting the proposed service for the main comparator in the setting for the requested listing (the base-case economic evaluation). MSAC requires a full and transparent description of the variables used in the economic evaluation. Generally, two steps are involved: 1. a study-based economic evaluation (effectively, a cost-consequences analysis), which is based on the study variables (eg population, setting, time horizon) 2. a modeled economic evaluation, in which study-based variables are modified using modeling techniques ('translated') to take account of differences between the study variables and the target variables for the proposed service.
How SR incorporated into the model	Nd
Who conducts the model?	Nd
Inclusion of quality of life	Nd
Inclusion of costs	The aim of the economic evaluation is to use the clinical studies ... to determine the economic cost of substituting the proposed service for the main comparator in the setting for the requested listing (the base-case economic evaluation).
Budget analysis done	Nd
Impact on project budget	Nd

Question	National Institute for Clinical Excellence (NICE/UK)
Integration of modeling	Modeling provides an important framework for synthesising available evidence and generating estimates of clinical and cost effectiveness in a format relevant to the Appraisal Committee's decision-making process. Models are required for most technology appraisals. Situations when modeling is likely to be required include those where: -all the relevant evidence is not contained in a single trial -patients participating in trials do not match the typical patients likely to use the technology within the NHS -intermediate outcomes measures are used rather than effect on HRQL and survival -relevant comparators have not been used or trials do not include evidence on relevant subgroups -the long-term costs and benefits of the technologies extend beyond trial follow-up In the multiple technology assessment (MTA) process, the Assessment Group prepares the assessment report, which is an independent synthesis of the evidence from published information and the submissions from manufacturers and sponsors about the clinical and cost effectiveness of the technology/technologies. The report provides a systematic review of the literature and a review of manufacturer and sponsor economic models submitted to the Institute. It usually includes a new assessment of cost effectiveness based on an economic model.
Modeling alongside SR	In the multiple technology assessment (MTA) process, the Assessment Group prepares the assessment report, which is an independent synthesis of the evidence from published information and the submissions from manufacturers and sponsors about the clinical and cost effectiveness of the technology/technologies. The report provides a systematic review of the literature and a review of manufacturer and sponsor economic models submitted to the Institute. It usually includes a new assessment of cost effectiveness based on an economic model.
Timing of modeling	nd
Use of pre-existing vs. established models	The report provides a systematic review of the literature and a review of manufacturer and sponsor economic models submitted to the Institute. It usually includes a new assessment of cost effectiveness based on an economic model.
Model recommendations	Economic models should also: be replicable have face validity (that is, be plausible) be open to external scrutiny. The models used to synthesise available evidence to generate estimates of clinical and cost effectiveness for the Institute's needs should follow accepted guidelines...Providing an all-embracing definition of what constitutes a high-quality model is not possible, but some guidelines are available….(see page 42). It is essential that clinical and cost effectiveness is considered over an appropriate time horizon to reflect UK practice and patients, and to compare treatment options that represent routine care and/or current best practice for the relevant patient groups. Therefore, it will be necessary to construct an analytical framework within which to synthesise the available evidence so that estimates of clinical and cost effectiveness can be made that are relevant to the clinical decision-making context. This framework will usually require the development of a model using aggregated or individual patient data to estimate parameters. It is essential that clinical and cost effectiveness is considered over an appropriate time horizon to reflect UK practice and patients, and to compare treatment options that

	represent routine care and/or current best practice for the relevant patient groups. Therefore, it will be necessary to construct an analytical framework within which to synthesise the available evidence so that estimates of clinical and cost effectiveness can be made that are relevant to the clinical decision-making context. This framework will usually require the development of a model using aggregated or individual patient data to estimate parameters. Trial data may not be sufficient to quantify baseline risk of some health outcomes or events for the population of interest. Quantifying the baseline risk of health outcomes and how the disease would naturally progress with the comparator intervention can be a useful step when estimating absolute health outcomes in the economic analysis. Relative treatment effects observed in randomised trials may then be applied to data on the baseline risk of health outcomes for the populations or subgroups of interest. The methods used to identify and critically appraise sources of data for these estimates should be stated and justified. Table 5.1 Summary of the reference case.
How SR incorporated into the model	Synthesis of evidence on outcomes [always] based on a systematic review
Who conducts the model?	Nd
Inclusion of quality of life	Multiple sections, as follows: 2.2.6, Table 5.1, 5.12, 5.2.11, 5.2.12, 5.4.1, 5.4.2,5.4.9,5.4.10,5.9.2,5.9.3,6.2.26
Inclusion of costs	It is essential that clinical and cost effectiveness is considered over an appropriate time horizon to reflect UK practice and patients, and to compare treatment options that represent routine care and/or current best practice for the relevant patient groups. Therefore, it will be necessary to construct an analytical framework within which to synthesise the available evidence so that estimates of clinical and cost effectiveness can be made that are relevant to the clinical decision-making context. This framework will usually require the development of a model using aggregated or individual patient data to estimate parameters. Further details of modeling methods are provided in section 5.7. The models used to synthesise available evidence to generate estimates of clinical and cost effectiveness for the Institute's needs should follow accepted guidelines...Providing an all-embracing definition of what constitutes a high-quality model is not possible, but some guidelines are available….(see page 42). If the use of the technology is conditional on the outcome of a diagnostic test, the accuracy of the test and associated costs should be incorporated into the assessments of clinical and cost effectiveness.
Budget analysis done	Multiple sections as follows: 5.2.12, 5.5.9, 5.13.6, 5.13.8, 6.2.14
Impact on project budget	

Question	Pharmaceutical Benefits Advisor Committee (PBAC/Australia)
Integration of modeling	The primary purpose of submission section C is to guide the presentation of analyses conducted to translate the systematic overview of the results of direct randomised trial evidence to the listing requested, and thus to the framework of the economic evaluation (submission section D--NEED TO FURTHER EXTRACT SECTION). This is particularly important when one or more variables incorporated into the economic evaluation are derived from, but not directly based on, the clinical evaluation presented in submission section B....The need for premodeling studies arises because the study protocols for the trials used for the clinical evaluation might differ from the proposed clinical practice setting for the main indication
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	This is particularly important when one or more variables incorporated into the economic evaluation are derived from, but not directly based on, the clinical evaluation presented in submission section B. These variables may be derived using a number of analyses that modify the results of the clinical evaluation to help construct a modeled economic evaluation.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	nd
Budget analysis done	nd
Impact on project budget	nd

Question	Pharmaceutical Management Agency of New Zealand (PHARMAC/New Zealand)
Integration of modeling	Decisions have to be made regardless of data availability. modeling in economic analysis is necessary in order to inform decision making at a particular point in time.
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	nd
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	nd
Budget analysis done	nd
Impact on project budget	nd

Question	AAZ (Agency for Quality and Accreditation in Health Care/Croatia)
Integration of modeling	nd
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	It is important to identify potential selection bias in the inputs to the model and for the model to quantify the decision uncertainty associated with a technology (that is, the probability that a different decision would be reached if the true cost effectiveness of each technology could be ascertained before making the decision).The models used to synthesize available evidence to generate estimates of clinical and cost-effectiveness for the Agency's needs should follow accepted guidelines. Full documentation and justification of structural assumptions and data inputs should be provided. When there are alternative plausible assumptions and inputs, sensitivity analyses of their effects on model outputs should be undertaken. It is important to identify potential selection bias in the inputs to the model and for the model to quantify the decision uncertainty associated with a technology (that is, the probability that a different decision would be reached if the true cost effectiveness of each technology could be ascertained before making the decision).
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	The models used to synthesize available evidence to generate estimates of clinical and cost-effectiveness for the Agency's needs should follow accepted guidelines.
Budget analysis done	nd
Impact on project budget	nd

Question	HITAP (Health Intervention and Technology Assessment Program/Thailand)
Integration of modeling	Time frame for economic evaluation: A full report may include an evidence review, an economic model and a budget impact analysis. If the evidence reviews systematic, the time frame will be longer than if the review is non-systematic.
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	Provides table of required economic modeling protocol components
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	May include economic evaluation
Budget analysis done	May include budget analysis
Impact on project budget	nd

Question	ICER (Institute for Clinical and Economic Review/US)
Integration of modeling	ICER's appraisal process includes the development of a de novo decision-analytic model to accompany the systematic review.
Modeling alongside SR	ICER's appraisal process includes the development of a de novo decision-analytic model to accompany the systematic review.
Timing of modeling	nd
Use of pre-existing vs. established models	ICER's appraisal process includes the development of a de novo decision-analytic model to accompany the systematic review.
Model recommendations	These models are aligned closely with the parameters of the systematic review to ensure that model outputs are generalizable to the appropriate patient populations and treatment settings. Sensitivity analyses of companion decision-analytic models
How SR incorporated into the model	To produce parameter estimates for use in sensitivity analyses of companion decision-analytic models, as a means of exploring the potential for these estimates to affect how the value of multiple interventions compares.
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	nd
Budget analysis done	nd
Impact on project budget	nd

Question	LBI (Ludwig Boltzmann Institute for Health Technology Assessment/Austria)
Integration of modeling	Follow good modeling practices when constructing the model used to conduct the evaluation. Analysts are encouraged to consult good modeling practice guidelines as required.
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	Modeling considerations: - Follow good modeling practices whenconstructing the model used to conduct the evaluation. Analysts are encouraged to consult good modeling practice guidelines as required. - Describe the model, including its scope, structure, and assumptions. Provide justification for assumptions and choices. - Use a model structure that is appropriate for addressing the study question. Build the model in such a way to permit updating of results as more data become available. - Explain and justify any causal relationships and extrapolation techniques used in the model. Base the extrapolation of data on valid techniques that reflect reasonable scientific evidence, and test through sensitivity analysis. - Formally validate the model, and state how this was done.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	nd
Budget analysis done	nd
Impact on project budget	nd

Question	MHRA (Medicines and Healthcare Products Regulatory Agency/UK)
Integration of modeling	Models, which are typically detailed and complex formulations of the consequences of drug treatments, are required to consider disease evolution and treatment outcomes
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	Models, which are typically detailed and complex formulations of the consequences of drug treatments, are required to consider disease evolution and treatment outcomes
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	nd
Budget analysis done	nd
Impact on project budget	nd

Question	NLM (National Library of Medicine/US)
Integration of modeling	Decision models are also used to set priorities for HTA (Sassi 2003).
Modeling alongside SR	nd
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	The basic steps of decision analysis are: 1. Develop a model (e.g., a decision tree) that depicts the set of important choices (or decisions) and potential outcomes of these choices. For treatment choices, the outcomes may be health outcomes (health states); for diagnostic choices, the outcomes may be test results (e.g., positive or negative). 2. Assign estimates (based on available literature) of the probabilities (or magnitudes) of each potential outcome given its antecedent choices. 3. Assign estimates of the value of each outcome to reflect its utility or desirability (e.g., using a HRQL measure or QALYs). 4. Calculate the expected value of the outcomes associated with the particular choice(s) leading to those outcomes. This is typically done by multiplying the set of outcome probabilities by the value of each outcome. 5. Identify the choice(s) associated with the greatest expected value. Based on the assumptions of the decision model, this is the most desirable choice, as it provides the highest expected value given the probability and value of its outcomes. 6. Conduct a sensitivity analysis of the model to determine if plausible variations in the estimates of probabilities of outcomes or utilities change the relative desirability of the choices. (Sensitivity analysis is used because the estimates of key variables in the model may be based on limited data or simply expert conjecture.) The assumptions and estimates of variables used in models should be validated against actual data as it becomes available, and the models should be modified accordingly. Modeling should incorporate sensitivity analyses to quantify the conditional relationships between model inputs and outputs. The assumptions and estimates of variables used in models should be validated against actual data as it becomes available, and the models should be modified accordingly. Modeling should incorporate sensitivity analyses to quantify the conditional relationships between model inputs and outputs.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	Assign estimates of the value of each outcome to reflect its utility or desirability (e.g., using a HRQL measure or QALYs). Conduct a sensitivity analysis of the model to determine if plausible variations in the estimates of probabilities of outcomes or utilities change the relative desirability of the choices.
Inclusion of costs	Models and their results are only aids to decision-making, not statements of scientific, clinical, or economic fact. The report of any modeling study should carefully explain and document the assumptions, data sources, techniques, and software. Modelers should make clear that the findings of a model are conditional upon these components. The use of decision modeling in cost-effectiveness analysis in particular has advanced in recent years,

	with development of checklists and standards for these applications (Gold 1996; Soto 2002; Weinstein 2003).
Budget analysis done	nd
Impact on project budget	nd

Question	AHRQ (US Agency for Healthcare Research and Quality/US)
Integration of modeling	Implicitly not always: p 26 "Out of 193 evidence reports, 10 reports and 1 supplement to a technology assessment were identified through the search process."
Modeling alongside SR	nd
Timing of modeling	p ES5 "The timing of a modeling project in connection with a systematic review is important. One approach would be to have the report from the modeling study coincide with that of the systematic review. However, the results from the systematic review typically will be required to conduct the final modeling analysis. Thus, the addition of a decision model could delay the overall project. Another concern is the ability to determine the opportunity or need for a model before the project has started or before the question refinement phase has been completed. The proposal process could be augmented to include a more collaborative question refinement prior to proposal submissions, which would involve a relatively quick review of the literature to determine if there were aspects of the disease and interventions that were suitable for modeling." p 14 "The first step in the process should be to engage the stakeholder in discussions about the goals of decision modeling and how it could potentially add value to the topic being addressed (though there may be timing issues discussed below). This will likely require that the stakeholder be educated on what a decision model is, how they have been used in practice, and what their value is in this context." p 42 "When to conduct a modeling project in connection with a systematic review is a concern. Ideally, one would complete the systematic review first and then develop/refine a decision model that is designed to optimize the use of the evidence results. For example, the final results from a systematic review could inform modeling decisions about ways to categorize a disease that maximizes the use of the evidence. Or the results may indicate several options for categorizing a disease that would allow the modelers to build in different structural assumptions that could be evaluated in sensitivity analyses. This ideal situation, however, is unlikely to happen in practice and the modeling work will likely need to be completed at the same time, or close to the same time, as the systematic review. This is not an insurmountable problem and it is reasonable to assume that, with adequate interactions between the systematic review team and the modeling team, the modeling work could be done concurrently with the systematic review, with interim model parameter estimates used prior to completion of the reviews. Figure 1 illustrates this framework." p 55 "The issues surrounding the timing of when a decision analysis is conducted alongside a review pose several challenges. Ideally, a decision analysis would not be done unless it was deemed to add substantial value to the questions being addressed by the systematic review. This may not become clear until after the systematic review has begun. However, it typically takes about the same time to develop and analyze a decision model as it does to conduct a systematic review, and the final decision analysis results should incorporate the results from the review. Thus the addition of a simulation model alongside a systematic review may add time to the overall project in some cases."
Use of pre-existing vs. established models	p 54 "[pre-existing or established models] may not fit the question precisely and it does not allow for input from the stakeholders" p 55 "[Don't use pre-existing models in] cases where the structure of existing models is not flexible enough to simulate the interventions of interest."
Model recommendations	Table 21. Assessing the quality of decision and simulation models p 75 "key issues to be addressed: the scientific and technical quality of the model, the interaction between the model and the decisionmaker(s) the model is intended to inform"
How SR incorporated into the model	p 6 "Decision models provide a way to synthesize multiple pieces of direct evidence in cases where only indirect evidence exists on the relationship between an intervention and the health outcomes of interest. Decision models can be used to structure the linkages between the intervention and the key health outcomes, where direct evidence can be used to inform each link. Thus, even though both systematic reviews and decision models are

	used to combine data, we view systematic reviews as an interpolation of the evidence with a goal of enhancing our knowledge, and decision modeling as an extrapolation of the evidence with the goal of decisionmaking."
Who conducts the model?	p 43 "Because decision modeling requires a different skill set, it is not always feasible to have the modeling work done by systematic review research teams, such as EPCs. Modeling is a multidisciplinary field that requires several disciplinary experts in order to conduct a credible modeling analysis on a wide variety of topics on timelines typical of a systematic review. It is beneficial for those conducting the modeling to have frequent interactions with researchers conducting the systematic review to ensure that the model is developed in such a way to incorporate the synthesized data, and that all relevant data are collected and synthesized to inform the model structure. In the ideal circumstance, the systematic review team and the decision analysis team would reside in the same place in order to facilitate a close working relationship."
Inclusion of quality of life	p ES3"Models can be used to: . . . (3) incorporate data from multiple sources (e.g., clinical and health-related quality-of-life endpoints), "
Inclusion of costs	p 1 "One type of decision analysis is a cost-effectiveness analysis, which incorporates both the benefits and the costs of competing alternatives and explicitly considers a limited budget. Our report is focused on modeling more broadly and not on economic evaluations that use modeling to project costs and health benefits. Our framework would, in general, allow for inclusion of costs as an outcome."
Budget analysis done	nd
Impact on project budget	p 37 "An essential issue is the resource intensiveness of models and modeling efforts. Most interviewees with experience with models in EPC reports responded that modeling efforts could easily consume 20–40 percent of the budget for a systematic review, and thus could not be accomplished without either inclusion in the budget at project inception, or an increased budget and timeline after the question refinement phase."

Question	CAST (Centre for Applied Health Services Research and Technology Assessment, University of Southern Denmark)
Integration of modeling	The state of the art of economic evaluations carried out as part of health technology assessments do not differ remarkably from that of economic evaluations in general. A notable exception is in the design, where the majority of the HTAs completed an economic evaluation retrospectively using secondary data in the form of a literature review or a meta-analysis. These data were often put together in a decision analytical model. This picture is not seen to this extent in economic evaluation in general, and is probably due to the nature of a health technology assessment as a synthesis of clinical and other evidence gathered from a systematic literature review.
Modeling alongside SR	The majority of the HTAs completed an economic evaluation retrospectively using secondary data in the form of a literature review or a meta-analysis. These data were often put together in a decision analytical model. This picture is not seen to this extent in economic evaluation in general, and is probably due to the nature of a health technology assessment as a synthesis of clinical and other evidence gathered from a systematic literature review.
Timing of modeling	The majority of the HTAs completed an economic evaluation retrospectively using secondary data in the form of a literature review or a meta-analysis. These data were often put together in a decision analytical model.
Use of pre-existing vs. established models	nd
Model recommendations	A model is an excellent way to combine this information; usually, a decision tree or a Markov model is applied in modeling studies.
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	Economic evaluations often seek to estimate lifetime costs and consequences
Budget analysis done	nd
Impact on project budget	nd

Question	CDE (Center for Drug Evaluation/Taiwan)
Integration of modeling	nd
Modeling alongside SR	HTA team members retrieve and summarize the key issues stated in the health technology assessment or appraisal reports from the world leading HTA agencies, follows with analyzing the possible product adoptability in Taiwan, eventually conduct systematic review of published literatures before recommendations sent to BNHI.
Timing of modeling	nd
Use of pre-existing vs. established models	nd
Model recommendations	nd
How SR incorporated into the model	nd
Who conducts the model?	nd
Inclusion of quality of life	nd
Inclusion of costs	Budget Impact analysis for all parties involved was constantly conducted.
Budget analysis done	nd
Impact on project budget	nd