Available online at www.sciencedirect.com
SciVerse ScienceDirect
journal homepage: www.elsevier.com/locate/jval

ISPOR TASK FORCE REPORTS

Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1

J. Jaime Caro, MDCM, FRCPC, FACP[1,2,], Andrew H. Briggs, Dphil[3], Uwe Siebert, MD, MPH, MSc, ScD[4,5], Karen M. Kuntz, ScD[6], on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force*

[1]Faculty of Medicine, McGill University, QC, Montreal, Canada; [2]United BioSource Corporation, Lexington, MA, USA; [3]Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow, Glasgow, UK; [4]UMIT – University for Health Sciences, Medical Informatics and Technology, Hall i.T., and Oncotyrol Center for Personalized Cancer Medicine, Innsbruck, Austria; [5]School of Public Health and Medical School, Harvard University, Boston, MA, USA; [6]School of Public Health, University of Minnesota, Minneapolis, MN, USA

ABSTRACT

Models—mathematical frameworks that facilitate estimation of the consequences of health care decisions—have become essential tools for health technology assessment. Evolution of the methods since the first ISPOR Modeling Task Force reported in 2003 has led to a new Task Force, jointly convened with the Society for Medical Decision Making, and this series of seven articles presents the updated recommendations for best practices in conceptualizing models; implementing state-transition approaches, discrete event simulations, or dynamic transmission models; and dealing with uncertainty and validating and reporting models transparently. This overview article introduces the work of the Task Force, provides all the recommendations, and discusses some quandaries that require further elucidation. The audience for these articles includes those who build models, stakeholders who utilize their results, and, indeed, anyone concerned with the use of models to support decision making.

Keywords: best practices, guidelines, methods, modeling.

Introduction

The use of models to support scientific endeavor is ubiquitous. Models are essentially communication tools that allow the complexity of a given system to be reduced to its essential elements. As such, models represent a simplification of reality and modeling is necessarily a reductionist methodology. This series of articles [1–6] relates to the application of modeling techniques to the area of health care decision making. This can include not only clinical decision models, designed to assist individual clinicians and their patients with decisions regarding their care, but also policy decision models, designed to more broadly evaluate whether particular health care technologies should be provided within the context of an organized health care system. These latter types of models are characterized by the need to explicitly include a budget constraint and therefore necessarily include both resource consequences and health outcomes in a health economic evaluation framework. Therefore, while these articles focus on modeling, drawing broadly on general methods, and apply beyond health economic assessment, the series necessarily touches on many aspects pertaining to economic evaluation.

Although the use of models to inform policy decision about the use of health technologies has been increasing [7], there remain strong concerns with their credibility [8,9]—a concern that is not unique to our field [10–12]. To help allay these concerns, several guidelines for good practices in modeling have been issued [13]. In 2000, the International Society for Pharmacoeconomics and Outcomes Research Task Force on Good Research Practices in Modeling Studies was created and after an extensive process of consultation, it issued its report in 2003. This report defined a model and its purpose, laid out the approach to evaluating a model, and described the Task Force's consensus regarding the attributes that characterize a good model, in terms of structure, data, and validation [14].

In the intervening years, the range of modeling techniques for medical and economic decision modeling has advanced substantially [15,16], as modelers in our discipline have become acquainted with more sophisticated modeling techniques. The relative simplicity of cohort-based models is still an attraction for many modelers

Conflict of Interest Statement: All members of the Task Force include modeling activities among their professional practices and in many cases these are funded by organizations with interest in the results. Some members of the Task Force own intellectual property relating to models. Many of the points made in these articles are supported by citations to articles authored by Task Force members. The Task Force took pains to avoid commercial considerations influencing the deliberations or the content of these articles.

* *Address correspondence to:* J. Jaime Caro, United BioSource Corporation, 430 Bedford Street, Lexington, MA 02420.
E-mail: jaime.caro@mcgill.ca.

http://dx.doi.org/10.1016/j.jval.2012.06.012

and decision makers. Nevertheless, there are situations when the decision problem demands taking extensive history into account and individual-level microsimulation methods are required [17]. One approach to implementing individual-level simulations is an adaptation of methods borrowed from engineering and operations research, which frames the problem in terms of the states individuals can be in and the events that can happen to them and their consequences [18,19]. These individual-level and stochastic techniques present additional challenges and require somewhat different approaches to modeling. Infectious disease modeling is a further approach that can handle interaction between individuals, and this dynamic form of modeling has also developed its own set of challenges and techniques [20]. The methods for simultaneously handling multiple parameters of a model and addressing uncertainty have also progressed significantly, and the approach to validation of models has received increasing attention.

Background to the Task Force

To ensure that the guidelines for good practices in modeling remain current, effective, and helpful, ISPOR judged it necessary to update them to accord with the newer methods being used in practice. As a result, a new Good Research Practices in Modeling Task Force was constituted to build on the excellent work done by the initial one from 2000 to 2003. To bring to bear the broadest expertise in this area, the Society for Medical Decision Making (SMDM) was invited to join the effort. The Task Force was asked to provide guidelines for designing the approach, selecting a technique, implementing and validating the model, parameterizing the inputs and assessing uncertainty, and using the resulting tool to inform decision making.

Early in 2010, the ISPOR and SMDM boards appointed the cochairs and consented to the proposed members of the Task Force. The Task Force convened expert developers and experienced users of models from academia, industry, and government, with representation from many countries. Given the breadth of the field at this point, a decision was made to divide the topic into six components and leads were appointed for each working group. Three of these topics covered the aspects felt to be general to all models in our field: conceptualization of a model, estimation of model parameters and handling of uncertainty, and validation of models and concerns for transparency. The other three dealt with specific techniques in common use: state-transition modeling, discrete event simulation, and dynamic transmission models. While there are undoubtedly topics of interest that are not addressed in these six articles, it was felt that these reports would cover the major areas that are at a stage of development appropriate for issuing guidelines.

The Task Force held its first meeting via teleconference on May 7, 2010, and hosted information sessions during 2010 at the ISPOR 15th Annual International Meeting in Atlanta, GA, at the 32nd Annual Meeting of the Society for Medical Decision Making in Toronto, ON, and at the ISPOR 13th Annual European Congress in Prague. Over numerous teleconferences, and occasional in-person meetings, the working groups produced draft reports for each section. Although the groups referred to the literature frequently, there was no systematic attempt to review it. Although substantiated as much as possible, the recommendations that emerged represent the opinions of the experts in the Task Force. These were not forced to consensus, and had substantial differences of opinion remained, they would have been documented as such. The draft recommendations were discussed by the Task Force as a whole in a meeting held in Boston in March 2011 and subsequently edited and circulated to the Task Force members in the form of a survey where each one was asked to agree or disagree with a recommendation, and if the latter, to provide the reason(s). Each group received the results of the survey and endeavored to address all rejections. In the end, there were no dissenting positions. The final drafts of the articles were posted on the ISPOR and SMDM Web sites for comment by the general membership of the societies.

A second group of experts—again, with broad representation of modelers and users of models—was invited to formally review the articles. Their comments were addressed by each working group, and revised drafts of each article were circulated to the Task Force as a whole. After receiving any additional comments and considering any further revisions, the final version of each article was prepared. (A copy of the original draft of this article, as well as the reviewer comments and author responses, is available at the ISPOR Web site: http://www.ispor.org/workpaper/Modeling-Good-Research-Practices-Overview.asp.) A summary of these articles was presented at a plenary session at the ISPOR 16th Annual International Meeting in Baltimore, MD, in May 2011, and again at the 33rd Annual Meeting of the Society for Medical Decision Making in Chicago, IL, in October 2011. These articles are jointly published in the Societies' respective journals, *Value in Health* and *Medical Decision Making*.

The audience for this set of articles encompasses both the researchers who develop models and those who use models to inform decisions. Investigators charged with reviewing others' models should find the guidelines helpful in their assessments. Even those affected by the decisions informed by models and those who report on the results of modeling analyses should find these recommendations useful.

It is important to note, however, that these articles are not intended as primers on their subjects. General textbooks and tutorial articles covering these techniques exist [21–25], and specific publications that address the methods are cited throughout. By the same token, these articles are not methodological treatises that address every aspect of a particular topic. Instead, they propose a set of best practices for modeling. They focus on the types of models and approaches taken today, not on nascent ones or even on those whose use is currently being debated (e.g., model averaging [26]). Further development of the methods will require that these guidelines be updated in due course.

Although it may not be possible to follow the entire set of recommendations in every modeling exercise, these do represent what the Task Force felt to be the best practices for modeling today and each recommendation should be given serious consideration. Nevertheless, the guidelines are not intended for use as a checklist to be followed unthinkingly. We encourage modelers who believe that they should not, or cannot, follow a particular recommendation to document this divergence, its rationale and likely consequences for their model, and its results and the inferences that will guide decision makers.

This overview article presents the process and methods of the Task Force and gives the reader an orientation to the contents of each of the detailed articles. It also provides all the recommendations of the Task Force, but without their detailed rationales and caveats. General quandaries and gaps in knowledge not covered in the other articles are addressed in the final section, along with some thoughts on developments in this area.

Orientation to the Series

A reader wishing to have a comprehensive view of the recommendations should approach the articles in order: first the one dealing with conceptualizing a model, then the three articles addressing specific techniques, followed by the article on parameter estimation and uncertainty, and concluding with the one on transparency and validation. If the detailed explanations are not required, then this overview article can be consulted for the list of recommendations.

The articles generally follow the same structure. After an introduction, the key concepts and definitions pertinent to each topic are laid out, followed by presentation of the recommendations and their rationale. Each group was given leeway, however, to approach its subject in the way it felt was best. The conceptualization article begins by defining the term "model." In the sense used by the Task Force, a *model* is a mathematical framework representing some aspects of reality at a sufficient level of detail to inform a clinical or policy decision [1]. The article then outlines the modeling process that begins with carefully examining the decision problem and laying out the elements required. Along the way, the designers will make many decisions. They will select what aspects of reality to include and to what extent they are detailed, driven by who their audience is (the "perspective") and the imperative to sufficiently reflect reality. They will choose the time span the model is to cover ("time horizon"), its target population(s), which interventions to consider, how to structure the model, which outcomes to report, and many other features. The article underscores that conceptualizing the model should precede examination of the available data to avoid designing a model that lacks key components and, thus, poorly represents the decision problem. It also emphasizes the importance of understanding the policy context and broad consultation with experts. This article concludes with guidance on choosing a technique for implementing the model.

The three articles on modeling techniques begin with definition of the technique and details of the key elements that characterize it. Indications on when to use the technique are given, and structuring of the model using that technique is described. All three conclude with suggestions on how to communicate that type of model. The state-transition article also deals briefly with decision trees, a simpler technique that is sometimes adequate for the decision problem at hand. The article on discrete event simulation spends somewhat more effort orienting readers to the technique, given that today it is less commonly used in medical contexts. The article on dynamic transition models devotes more space to the parameters involved because these are a crucial and prominent component of this type of model.

In the article regarding parameter estimation and uncertainty, significant effort is made to bring order to the unruly terminology pertaining to uncertainty (though, undoubtedly, practitioners will still want to cling to their favorite terms). The connection between estimation and subsequent uncertainty analyses is emphasized and the choices of distributions are described. Calibration methods and structural uncertainty are briefly addressed, and the article concludes with extensive guidance on the reporting of uncertainty.

The final article in the series deals with the twin aspects of transparency and validation. It begins with the thorny topic of trusting the results of a model enough to allow them to guide a decision and how this can be achieved. It discusses both technical and nontechnical documentation of a model, designed to achieve the required transparency, emphasizing that it is inappropriate to require that a model be understandable at full depth by someone without the necessary technical know-how. The types of validation, their necessity and sufficiency, and interpretation are then addressed in detail.

Best Practices

II. Conceptualizing the model

II-1 The modeling team should consult widely with subject experts and stakeholders to assure that the model represents disease processes appropriately and adequately addresses the decision problem.

II-2 A clear, written statement of the decision problem, modeling objective, and scope of the model should be developed. This should include: the spectrum of disease considered, perspective of the analysis, target population, alternative interventions, health and other outcomes, and time horizon.

II-2a The scope and structure of the model should be consistent with, and adequate to address, the decision problem/objective and the policy context.

II-2b The perspective of the analysis should be stated and defined. Outcomes modeled in the analysis should be consistent with the stated perspective. Analyses which take a perspective narrower than the societal perspective should report which outcomes are included and which are excluded.

II-2c The target population should be defined in terms of geography, patient characteristics (including co-morbid conditions), and disease stage, each of which should be appropriate to the decision problem.

II-2d Health outcomes modeled in the analysis, which may be measured as events, cases of disease, deaths, quality-adjusted life-years, disability-adjusted life-years, or other measures important to decision makers and stakeholders, should be directly relevant to the question being asked.

II-2e Interventions or strategies modeled in the analysis should be clearly defined in terms of frequency, component services (including services that may have preceded the intervention and that would affect its course), dose or intensity, duration, and any variations required for target subgroups.

II-3 Although data are an essential component of a model, the conceptual structure of a model should be driven by the decision problem or research question and not determined by data availability.

II-3a The choice of strategies/comparators crucially affects results, and should be determined by the decision problem and not by data availability or quality. All feasible and practical strategies should be considered. Constraining the range of strategies should be justified.

II-3b The time horizon of the model should be long enough to capture relevant differences in outcomes across strategies. A lifetime horizon may be required.

II-4 The conceptual representation of the decision problem should be used to identify key uncertainties in model structure where sensitivity analyses could inform the impact of structural choices. For example, where a lifetime horizon is used, the impact of alternative methods of extrapolating beyond the observed data should be explored.

II-5 The policy context of the model should be clearly stated. This includes who funded the model, who developed the model, whether the model was developed for a single application or multiple potential applications, and who the policy audience for the modeling work is.

II-6 An explicit process (expert consultations, influence diagrams, concept mapping, or similar method) should be used to convert the conceptualization of the problem into an appropriate model structure to ensure that the model reflects current theory of disease or the process being modeled.

II-7 There are often several types of models that are suitable for the decision problem, and versions of each of the three modeling types in the series can be used for the same decision problem. Some problems are more naturally represented in certain modeling types than in others.

II-7a For relatively simple models, or decision problems with special characteristics (e.g. very short time horizons, complex value structures) a decision tree may be appropriate.

II-7b If the conceptualization involves representing the disease or treatment process as a series of health states, state-transition models are often appropriate as they may be simple to develop, debug, communicate, analyze and readily accommodate the evaluation of parameter uncertainty. Their primary disadvantage, the Markovian assumption that transition probabilities do not depend on past history, can be addressed by increasing the number of states. Individual-based state-transition models (termed “microsimulations”), which do not require this assumption, are an alternative when the number of states grows too large.

II-7c When the disease or treatment process includes interactions between individuals, the modeling methods should be able to represent and evaluate the effects of those interactions (dynamic transmission models, discrete event simulations, agent-based models).

II-7d When the decision problem involves resource constraints, the modeling method should be able to represent and evaluate the effects of those constraints (discrete event simulation, agent-based models).

II-7e For some decision problems, combinations of model types, hybrid models, and other modeling methodologies are appropriate.

II-8 Model simplicity is desirable for transparency, ease of validation and description. However, the model should be sufficiently complex to answer the question at a level of detail consistent with the problem being modeled, and to preserve face validity to clinical experts. Greater complexity may be necessary in policy models that are intended to be used for many decision problems.

III. State-transition models

III-1 If the decision problem can be represented with a manageable number of health states that incorporate all characteristics relevant to the decision problem, including the relevant history, a cohort simulation should be chosen because of its transparency, efficiency, ease of debugging, and ability to conduct specific value of information analyses. If, however, a valid representation of any aspect of the decision problem would lead to an unmanageable number of states, then an individual-level state-transition model is recommended. Validity should not be sacrificed for simplicity.

III-2 The strategies being evaluated should be clearly defined. In particular, sequential decisions should not be modeled within the Markov cycle tree but rather be part of the specification of the alternative intervention strategies that precede the Markov tree.

III-3 The starting cohort should be defined by the demographic and clinical characteristics that affect the transition probabilities or state values (e.g., quality of life and cost).

III-4 Specification of states and transitions should generally reflect the biological/theoretical understanding of the disease or condition being modeled.

III-5 States should adequately capture the type of intervention (i.e., prevention, screening, diagnostics, treatment) as well as the intervention’s benefits and harms.

III-6 States need to be homogeneous with respect to both the observed and unobserved (i.e., not known by the decision maker) characteristics that affect transition probabilities.

III-7 The time horizon for the model should be sufficiently large to capture all health effects and costs relevant to the decision problem.

III-8 Cycle length should be short enough to represent the frequency of clinical events and interventions.

III-9 Components of state-transition models that reflect similar clinical courses should not be recreated but rather should be incorporated once and linked to that structure throughout the model.

III-10 Transition probabilities and intervention effects should be derived from the most representative data sources for the decision problem.

III-11 All methods and assumptions used to derive transition probabilities and intervention effects should be described.

III-12 Parameters relating to the effectiveness of interventions derived from observational studies should be correctly controlled for confounding. Time-varying confounding is of particular concern in estimating intervention effects.

III-13 The valuation of intermediate outcomes/states should be justified.

III-14 A half-cycle correction factor should be applied to both costs and effectiveness and should be applied in the first cycle. A half-cycle correction should also be applied in the final cycle for analyses that do not use a lifetime horizon.

III-15 For certain decision problems, it may be important to report not only the expected value but also the distribution of the outcomes of interest.

III-16 The number of individuals modeled in an individual-based simulation should be large enough to generate stable estimates of the expected value of interest.

III-17 The report should use nontechnical language and clear figures and tables that enhance the understanding of the model to communicate its key structural elements, assumptions, and parameters.

III-18 In addition to final outcomes, intermediate outcomes that enhance the understanding and transparency of the model results should also be presented.

IV. Discrete event simulation (DES)

IV-1 Discrete event simulation (DES) models should be used when the problem under study involves constrained or limited resources. DES is also an attractive option in nonconstrained models when there are interactions between individuals; populations and/or their environment, when time-to-event is best described stochastically rather than with fixed time intervals and time dependencies are important; when individual pathways through the model are influenced by multiple characteristics of the entity; and when recording individual entity experience is desirable.

IV-2 Constrained resource models should consider health-related outcomes, and not focus solely on measures of throughput.

IV-3 The effects of constrained resources should be modeled if:

- evaluated technologies result in differing levels of access (e.g. different referral rates), and

- time to access referred events can have significant effects on costs and/or outcomes (e.g. surgery).

IV-4 If downstream decisions can have significant effects on the differences in costs and/or outcomes of the primary object of an evaluation, the model should be structured to facilitate analyses of alternative downstream decisions.

IV-5 Where there are competing risks of events, parameterization approaches that represent correlation between the likelihood of competing events are preferred to the specification of separate time to event curves for each event.

IV-6 Where possible, progression of continuous disease parameters and the likelihood of related events should be defined jointly to maintain the discrete event nature of DES, for example, sample the level of the continuous measure (e.g. HbA1c score) at which an event occurs, and then sample the time at which the level is reached.

IV-7 To simplify debugging and updating, sub-models should be used to structure the model. When comparing two or more strategies within the same system (e.g. for the same condition in a health technology assessment model), sub-models common to all strategies (e.g. progression following disease recurrence) should be defined once, and called from each strategy (i.e. all patients experiencing a recurrence pass through the common disease recurrence module).

IV-8 For structural sensitivity analyses, alternative structures should be implemented within a single DES.

IV-9 Analysts should ensure that the mechanism for applying ongoing risk(s) over multiple events remains active over the relevant time horizon.

IV-10 Model implementation should account for the outputs required for the validation and the final analyses of the model. Where individual level data is required, relevant outputs should be stored as attributes, otherwise aggregated values should be collected from each model run to reduce the simulation burden.

IV-11 The choice between using general programming or dedicated DES (“off-the-shelf”) software should be informed by the relative importance of flexibility and execution speed (general programming languages) vs. modeling efficiency, automated structure and transparency (dedicated DES software). Spreadsheet software is generally inappropriate for implementing DES and should not be used without justification.

IV-12 When run times for probabilistic sensitivity analysis are constrained, the optimal combination of run size (per input parameter set) and numbers of alternative input parameter sets tested should be estimated empirically to optimize the precision of the outputs of interest.

IV-13 If the number of strategies to compare is large or there are many structural assumptions to test, then “factorial design” and optimum seeking approaches should be used.

IV-14 When computing time precludes adequate representation of uncertainty, meta-modeling (statistical representation of the model input-output relationship) should be used.

IV-15 If the system to be modeled is not empty at the start of the time horizon to be evaluated, a warm-up period should be used build the system up to the starting point if:

- it can be reasonably assumed that the key parameters have remained constant over time, or
- the history of the key parameters can be incorporated into the warm-up period (e.g. the introduction of new health technologies can be described).

IV-16 Animated representation of DES that displays the experience of events by individuals is recommended as a means of engaging with users, as well to helping to debug the model through the identification of illogical movements.

IV-17 Both general and detailed representations of a DES model’s structure and logic should be reported to cover the needs of alternative users of the model. Detailed event documentation figures are also of benefit to the analyst, as a point of referral when returning to a model after a period of absence.

V. Dynamic transmission models

V-1 A dynamic model is needed when a modeler is trying to evaluate an intervention against an infectious disease that 1) has an impact on disease transmission in the population of interest, and/or 2) alters the frequency distribution of strains (e.g., genotypes or serotypes).

V-2 The appropriate type of dynamic transmission model should be used for the analysis in question, based in part on the complexity of the interactions as well as the size of the population of interest and the role of chance effects. This model could be deterministic or stochastic, and population or individual-based. Justification for the model structure should be given.

V-3 Conduct sensitivity analysis on the time horizon and discount rate.

V-4 Conduct uncertainty analyses on known key structural assumptions that may have an impact on the conclusions, or justify the omission of such analyses.

V-5 When conducting sensitivity analyses, consideration of important epidemic thresholds is helpful when there is a possibility of the model exhibiting alternate behaviors.

V-6 For differential equation-based models, adaptive time step methods for numerical integration, that allow the degree of error tolerance to be specified in advance, are preferred to those that use a fixed time step of indeterminate accuracy.

V-7 If using a differential equations model, provide the model equations. Tabulate all initial values and parameters, including the mixing matrix and supply details of the type of mixing considered.

V-8 If using an agent-based model, thoroughly describe the rules governing the agents, the input parameter values, initial conditions and all sub-models.

V-9 Show the transmission dynamics over time (e.g., incidence and prevalence of infection and disease). When applicable, report changes in other infection-specific outcomes such as strain replacement and the emergence of resistance to antimicrobial drugs.

VI. Parameter estimation and uncertainty

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

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

VI-3 Terminology to describe concepts relating to parameter estimation and representation of uncertainty varies within the medical decision modeling field and in comparison to related fields. Authors should be aware of this and seek to carefully define their use of terminology to avoid potential confusion.

VI-4 All decision models will have parameters that need to be estimated. In populating models with parameter estimates, analysts should conform to the broad principles of evidence-based medicine. For example, analysts should: seek to identify and incorporate all relevant evidence, rather than cherry picking the best single source of evidence for that parameter; use best practice methods to avoid potential biases in parameter estimates that might arise (for example, when estimating treatment effectiveness from observational sources); and employ formal evidence syntheses techniques (meta analysis and network meta analysis) as appropriate.

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

VI-6 While completely arbitrary analyses, such as the presentation of the effect on model outputs of varying each input parameter by +/− 50%, can be used as a measure of sensitivity, such analyses should not be used to represent uncertainty.

VI-7 Analysts should give consideration to using commonly adopted standards from statistics for point estimate and interval estimation for input parameters, such as 95% confidence intervals, or distributions based on agreed statistical methods for a given estimation problem. Where departures from these standards are deemed necessary (or where no such standard exists for a given estimation problem), these should be justified.

VI-8 When there is very little information on a parameter, analysts should adopt a conservative approach such that the absence of evidence is reflected in a very broad range of possible estimates. On no account should parameters be excluded from a sensitivity analysis on the grounds that 'there is not enough information from which to estimate uncertainty'.

VI-9 In choosing distributional forms for parameters in a probabilistic sensitivity analysis, favor should be given to continuous distributions that provide a realistic portrayal of uncertainty over the theoretical range of the parameter of interest. Hence careful consideration should be given to whether distributions like the triangular should have any role in a probabilistic sensitivity analysis.

VI-10 Correlation among parameters should be considered. Jointly estimated parameters, such as those from a regression analysis, will have direct evidence on correlation which should be reflected in the analysis. Independently estimated parameters will have no such evidence, but this should not necessarily lead to an assumption of independence. Possible approaches are 1) to include a correlation coefficient as a parameter to the model where concern exists that an unknown correlation between parameters could be important, or 2) to reparameterize the model so that that the uncertain parameters can be reasonably assumed to be independent.

VI-11 Where uncertainties in structural assumptions were identified in the process of conceptualizing and building a model, those assumptions should be tested in a sensitivity analysis. Consideration should be given to opportunities to parameterize these uncertainties for ease of testing. Where it is not possible to perform structural sensitivity analysis it is nevertheless important that analysts be aware of the potential for this form of uncertainty to be at least as important as parameter uncertainty for the decision maker. (Linked to conceptual modeling recommendations)

VI-12 Uncertainty analyses can be either deterministic or probabilistic, and often it is appropriate to report aspects of both types within a single evaluation. Tornado diagrams, threshold plots, or simple statements of threshold parameter values, are all appropriate ways of reporting results from deterministic sensitivity analyses.

VI-13 When additional assumptions or parameter values are introduced for purposes of uncertainty analyses, such as distributional parameters for probabilistic sensitivity analyses, or parameter ranges for deterministic sensitivity analyses, these values should be disclosed and justified. Technical appendices are often appropriate for this purpose.

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

VI-15 When the purpose of a probabilistic sensitivity analysis is to guide decisions about acquisition of information to reduce uncertainty, results should be presented in terms of expected value of information.

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

VII. Transparency and validation

VII-1 Every model should have non-technical documentation that is freely accessible to any interested reader. At a minimum it should describe in non-technical terms the type of model and intended applications; funding sources; structure of the model; inputs, outputs, other components that determine the model's function, and their relationships; data sources; validation methods and results; and limitations.

VII-2 Every model should have technical documentation, written in sufficient detail to enable a reader with the necessary expertise to evaluate the model and potentially reproduce it. The technical documentation should be made available openly or under agreements that protect intellectual property, at the discretion of the modelers.

VII-3 Validation of a model should include an evaluation of face validity of the structure, evidence, problem formulation, and results of the model. A description of the process used to evaluate face validity should be made available on request. Evaluation of face validity should be made by people who have expertise in the problem area, but are impartial to the results of an analysis. If face validation raises questions about a model, these issues should be discussed by the modelers in their report of an analysis.

VII-4 Models should be subjected to rigorous verification. The verification methods should be described in the non-technical documentation of the model. The pertinent results of verification should be made available on request.

VII-5 Modelers should search for previously published modeling analyses of the same or similar problems and discuss insights gained from similarities and differences in results.

VII-6 Builders of models should have a formal process for conducting external validation that includes:

- Systematic identification of suitable data sources; justification of the selection; specification of whether a data source is dependent, partially dependent, or independent; and description of which parts of the model are evaluated by each data source.
- Simulation of each data source and comparison of results
- Quantitative measures of how well the model's results match the outcomes observed in the data source

VII-7 Comparison of results should include descriptions of:

- Data source
- Set up of the simulation
- Discrepancies between the data source and simulation setup, and implications of the discrepancies
- Comparisons of simulation results with observed results
- Discussion of discrepancies between simulation results and observed results
- Sensitivity analyses

VII-8 Modelers should make available on request a description of the external validation process and results

VII-9 Modelers should identify parts of a model that cannot be validated because of lack of suitable data sources, and describe how uncertainty about those parts is addressed.

VII-10 For multi-application models, modelers should describe criteria for determining when validations should be repeated and/or expanded.

VII-11 When feasible with respect to the decision being addressed and the availability of a future data source, a model should be tested for its prediction of future events. Builders of multiple-use models should seek opportunities to conduct predictive validations as part of their overall validation process.

Quandaries

A major quandary in modeling is the choice of technique that will be used to structure and analyze the model. Many techniques and variations are available and, with sufficient effort and ingenuity, most problems can be structured in any of the techniques [1]. This does not mean that the techniques are interchangeable and that the choice should be made casually. With advances in computing that render massive calculations quite feasible, it is expected that there will be increasing use of individual-based simulations because the computational challenges of simultaneously addressing stochastic uncertainty and parameter uncertainty in probabilistic sensitivity analysis diminish. Individual-based simulations are less subject to limitations imposed by the cohort simulation approach, particularly with regard to patient history based on events experienced within the model. Whether to frame them in terms of states or events will be of lesser concern. Indeed, there is no reason to treat these as mutually exclusive alternatives: hybrid models with some components represented as states and others as events are readily constructed and can be a very flexible and accurate approach with no restrictions in terms of how time is handled. At the same time, overly complex models should be avoided if a simpler one will accurately reflect all aspects of the decision problem.

Another quandary is whether to allow the design of a model to be driven by the data at hand. The idea that this should not be the case—detailed in the conceptualization article [1]—may rankle some on the grounds that there is not much point in designing a detailed model that can then not be populated by existing data. While a model lacking information on many inputs is not very useful, the reasons for designing first and looking at data after are that this produces more appropriate, relevant designs and often leads to looking for, and finding, data that might otherwise have been overlooked. The choice of data and their processing to yield suitable inputs for the model is a vast topic covered in other fields such as epidemiology. While this series does not address this in any detail, it is emphasized that good practice of evidence-based medicine should be followed. Whatever choices are made, the model parameters should reflect the uncertainty over the data gaps [5], which will ensure that a value of information analysis will provide the necessary impetus to launch studies to obtain the necessary data.

The converse situation is also of note: just because detailed data exist is not a sufficient reason to build a very complex model. The art of building models rests on the principle of parsimony, or Occam's Razor. We do not suggest that finding the balance between simplicity of modeling and avoidance of oversimplification is easy, but it is perhaps the most important skill a modeler can learn if a model is to truly fulfill its potential as a communication tool. Excessive detail and complexity reduce transparency and can lead to distrust in models and in the modeling community among those we seek to inform.

Throughout the series of articles, the subject of structural uncertainty keeps cropping up. This is a particularly difficult quandary [27]. There is no doubt that the choices made in structuring the model can significantly affect the results, and thus inferences made from them. In many cases, those choices are made on the basis of expert opinion, or influenced by concerns for simplicity, feasibility of implementation, and so on. This process leaves much room for uncertainty, but it is very difficult to quantify and analyze this uncertainty. This hurdle is augmented by using software specific for simulation not designed for modeling—many of those that are used include features that facilitate structural sensitivity analysis. In principle, all the alternatives considered could be modeled and the impact on the results examined. The effort involved tends to be perceived as prohibitive though, and even if the investment were made, the universe of possibilities is vast and extends beyond what the individual modeler might consider. Clearly, structural uncertainty is a topic ripe for intensive research. It is hoped that the next edition of the guidelines will be able to provide firm recommendations in this regard.

Another quandary arises in the article on uncertainty, in which it is stated that arbitrary ranges should not be used when examining the impact of uncertainty on the results. The reason is that such an analysis reveals how sensitive the model is to changes in that input but does not address uncertainty since the range of values is not a reflection of the latter. This poses a practical dilemma for modelers: how to address uncertainty when it is clear that an input is not exact but it is not clear to what degree. One could argue that the solution is to collect data on that input and use that process to quantify the uncertainty. This will usually not be feasible in a timely way, however. Thus, an option would be to employ Bayesian methods to create a probability distribution around the estimate and use this to quantify the relevant uncertainty. Given the practical and methodological difficulties, this area should be a focus of research.

Often, when reporting the results of a model analysis, the term "robust" is used. This term may be misinterpreted to mean that the results are unaffected by changes to the inputs, whereas it should indicate that within the uncertainty of the inputs, the conclusions (i.e., suggested decisions) were not altered. It would be quite worrisome if a model did not react to changes in its inputs; were this to be the case, one would conclude that there must be a major problem with the model as a proper one should respond to changes in inputs. Also, this is not a property of a model but rather of a particular analysis and set of results. Robustness is not per se a desirable feature. Instead, what investigators should examine are the conditions that alter the implications for the decision at issue and their credibility.

A particularly difficult quandary tackled by the Task Force is the conflict between the scientific desirability of making all methodological and technical details of a model available to peer reviewers and to other researchers and the need to protect intellectual property generated by substantial investments in the development of a model. As rejecting the latter would significantly reduce the incentive to devote major efforts to creating models—particularly those intended for multiple uses—the Task Force agreed, reluctantly, to recommend that intellectual property not be ignored. Instead, the proposal is to ask that modelers make full technical documentation available within whatever agreements they feel are necessary to grant them adequate protection. This should allow for detailed review of any model by other scientists, provided they are willing to abide by the confidentiality restrictions.

A final quandary is that models are often created in our field to address decisions regarding the use of limited resources but the modelers typically ignore the actual short-term resource constraints. Despite the availability of methods to simulate those constraints, most models regularly assume that any resource that is needed is immediately available and consumed, regardless of actual supply (or likely demand). Thus, the vexing health care queues common in many countries are not incorporated, nor are changes in waiting times as a consequence of new interventions. The potential for helping health systems adapt to changing practices in the short term has been overlooked. Incorporating this aspect would, undoubtedly, add another layer of complexity to models and a further demand for data that might be difficult to obtain, but it is a gap in current practice. Perhaps by the time the next guidelines are developed, our field will have advanced to assist decision makers not only with the challenge of which interventions to adopt but also with that of handling the implementation of system changes more efficiently.

Conclusion

The recommendations for best practices provided in this article and detailed in the accompanying six articles are intended, in the first instance, for practitioners who build models. Nevertheless, they should be of use to the decision makers who are the audience for the models' results, as well as those who commission models, granting agencies that fund them, and even those who report on the results and their implications.

Acknowledgments

The members of the task Force thank the two societies (ISPOR and SMDM) for coming together to support this endeavor. Danielle Mroz from ISPOR was instrumental in keeping the Task Force moving forward. We are also very appreciative of the efforts of the many people who agreed to review one or more of the articles in the series and who contributed greatly with their thoughtful comments.

Source of financial support: This Task Force was supported by ISPOR.

REFERENCES

[1] Roberts M, Russel L, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force-2. Value Health 2012;15:804–811.
[2] Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. Value Health 2012;15:812–820.
[3] Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task force-4. Value Health 2012;15:821–827.
[4] Pitman R, Fisman D, Zaric GS, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5. Value Health 2012;15:828–834.
[5] Briggs AH, 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;15:835–842.
[6] Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. Value Health 2012;15:843–850.
[7] Petrou S, Grey A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ 2011;342:1766–72.
[8] Caro JJ, Fox J, Schaecher KL, et al. Is Health Outcomes research valuable to real decision makers: A candid discussion with medical and pharmacy directors. [Online]; [cited 2012 January 21]. Available from: http://www.ispor.org/meetings/atlanta0510/releasedpresentations.asp. [Accessed June 22, 22].
[9] Buxton M, Drummond M, Van Hout B, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ 1997;6:217–27.
[10] Steele MJ. Dimensions of Credibility in Models and Simulations. [Online]; 2008 [cited 2012 January 21]. Available from: http://aegistg.com/Technical_Papers/08E-SIW-076%20Dimensions%20of%20Credibility%20in%20Models%20and%20Simulations%20Presentation_Slides.pdf. [Accessed June 22, 22].
[11] Lorscheid I, Heine BO, Meyer M. Opening the 'Black Box' of Simulations: Increased Transparency of Simulation Models and Effective Results Reports through the Systematic Design of Experiments. [Online].; 2011 [cited 2012 January 21]. Available from: http://ssrn.com/abstract=1860904. [Accessed June 22, 22].
[12] Kijowski DJ. Responsible Conduct Of Research With Computational Models. [Online]; 2010 [cited 2012 January 21]. Available from: http://hdl.handle.net/2142/18349. [Accessed June 22, 22].
[13] Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics 2006;24:355–71.
[14] 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;6:9–17.
[15] Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics 2006;24:1043–53.
[16] Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ 2006;15:1295–310.
[17] Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? Eur J Health Econ 2003;4:143–50.
[18] Caro JJ. Pharmacoeconomic analyses using discrete event simulation. Pharmacoeconomics 2005;23:323–32.
[19] Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. Pharmacoeconomics 2008;26:131–48.
[20] Borshchev A, Filippov A. From system dynamics and discrete event to practical agent based modeling: Reasons, techniques, tools. [Online]; [cited 2012 January 21]. Available from: http://www.systemdynamics.org/conferences/2004/SDS_2004/PAPERS/381BORSH.pdf. [Accessed June 22, 22].
[21] Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press, 2006.
[22] Macal CM, North MJ. Tutorial on agent-based modelling and simulation. J Simul 2010; 4:151–62.
[23] Jacobson SH, Hall SN, Swisher JR. Discrete-event simulation of health care systems. In: Hall RW, ed., Patient Flow: Reducing Delay in Healthcare Delivery. New York: Springer, 2006.
[24] Law AM, Kelton WD. Simulation Modeling and Analysis 4th ed.). New York: McGraw-Hill, 2007.
[25] Hunink MGM, Glasziou P, Siegel JE, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. United Kingdom: Cambridge University Press, 2001.
[26] Price M, Welton NJ, Briggs AH, Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. Value Health 2011;14:205–18.
[27] Bilcke J, Beutels P, Brisson M, Jit M. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Med Decis Making 2011;31:675–92.